AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Nonlinearity, Breaks, and Long-Range Dependence in Time-Series Models

Eric Hillebrand and Marcelo C. Medeiros

CREATES Research Paper 2012-30

Department of Economics and Business
Aarhus University
Bartholins Allé 10
DK-8000 Aarhus C
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

NONLINEARITY, BREAKS, AND LONG-RANGE DEPENDENCE IN TIME-SERIES MODELS

ERIC HILLEBRAND AND MARCELO C. MEDEIROS

ABSTRACT. We study the simultaneous occurrence of long memory and nonlinear effects, such as parameter changes and threshold effects, in ARMA time series models and apply our modeling framework to daily realized volatility. Asymptotic theory for parameter estimation is developed and two model building procedures are proposed. The methodology is applied to stocks of the Dow Jones Industrial Average during the period 2000 to 2009. We find strong evidence of nonlinear effects.

KEYWORDS: Smooth transitions, long memory, forecasting, realized volatility.

JEL CODES: C22.

ACKNOWLEDGMENTS: We thank 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 in December 2006, at the Midwest Econometrics Meetings in St. Louis in October 2007, 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, and while the second author was visiting CREATES, Aarhus University, Denmark. Their hospitality is gratefully appreciated. EH acknowledges support from the Danish National Research Foundation. MCM's research is partially supported by the CNPq/Brazil. Mihaela Craioveanu and Pierre Xu provided excellent research assistance. Special thanks goes to Eduardo Mendes for insightful discussions about consistent model selection. An earlier draft of the paper was circulated under the title "Asymmetries, Breaks, and Long-Range Dependence in Realized Volatility: A Simultaneous Equations Approach."

1. INTRODUCTION

Long memory and nonlinear effects such as parameter changes in time or in dependence on a state variable have long been recognized as confounding (Lobato and Savin 1998, Diebold and Inoue 2001, Granger and Hyung 2004, Mikosch and Starica 2004, Hillebrand 2005). We propose a modeling framework for time series that allows for the joint estimation and thus possible disentanglement of these effects. The method is applied to a set of realized stock volatility time series, but the methodology is general and can be applied to any time series that displays these effects such as, for example, unemployment rates (van Dijk et al. 2002), exchange rates (Baillie and Kapetanios 2008), river flows (Franses and Ooms 2001, Elck and Makus 2004), sea surface temperatures (Lewis and Ray 1997), and lung mechanics (Zhang et al. 1999), among many others.

Our proposal can be seen as a varying-coefficient model where the parameters of the conditional mean change according to a nonlinear function. We study the asymptotic behavior of the nonlinear least-squares estimator. In addition, we show under which conditions information criteria (IC) can be used to specify the model structure and also propose a sequence of tests that is robust to heteroskedasticity and non-normality of the error. The class of nonlinear functions considered in this paper is quite general: only local stationarity and finite fourth moments of the nonlinear function and its first and second derivatives are required. We also allow for a framework where time is the driving force for parameter changes. In this case, asymptotic theory cannot be achieved in the standard way, because as the sample size tends to infinity, the proportion of finite regimes converges to zero. Our solution is to scale the transition variable so that the location of the transition is a constant fraction of the sample rather than a fixed point (Andrews and McDermott 1995, Saikkonen and Choi 2004).

Simulations show that our strategy is successful in correctly determining the structure of models in a variety of situations. Furthermore, the long-memory parameter is precisely estimated (as zero) even in nonlinear short-memory models, where the risk of detecting spurious long-memory is high. Applying our model and testing framework to the 30 stocks of the Dow Jones Industrial Average from 2000 to 2009, we find evidence of structural breaks in the individual realized volatility time series. Dependence of volatility on the level of lagged returns is a robust finding across all stocks and in different model specifications, indicating pronounced asymmetry effects (Black 1976). We conclude that both long memory and non-linear effects coexist in the data. Accounting for non-linear terms in the volatility specification yields forecast gains in a multi-step out-of-sample comparison with the HAR-RV model of Corsi (2009) as well as the non-linear HAR-RV specification of (Corsi and Renò 2012).

The paper is organized as follows. Section 2 presents the model and the asymptotic theory. Model building is introduced in Section 3. Simulations are presented in Section 4. Empirical results are shown in Section 5. Section 6 concludes. All proofs are presented in the appendix. Additional results are provided in a supplement.

2. The Model

2.1. **Model Specification.** Let y_t be a zero-mean time series that possibly displays long memory and nonlinear behavior, such as structural breaks and/or threshold effects. For example, let $y_t := \log(RV_t) - \mu$, where RV_t is any consistent estimator of daily integrated variance and $\mu = \mathbb{E}[\log(RV_t)] < \infty$.[1] Consider the following model with time-varying coefficients:

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

or $\Phi(\boldsymbol{s}_t;\boldsymbol{\xi})v_t = \Theta(L)u_t$, where $\Phi(\boldsymbol{s}_t;\boldsymbol{\xi}) = 1 - \phi_1(\boldsymbol{s}_t;\boldsymbol{\xi}_1) - \ldots - \phi_p(\boldsymbol{s}_t;\boldsymbol{\xi}_p)$. The autoregressive (AR) coefficients $\phi_i(\boldsymbol{s}_t;\boldsymbol{\xi}_i)$, $i = 1,\ldots,p$ are nonlinear functions to be specified. They are indexed by the vector of parameters $\boldsymbol{\xi}_i \in \mathbb{R}^{k_{\xi_i}}$ and a vector of state variables $\boldsymbol{s}_t \in \mathbb{R}^{k_s}$. The

[1] Here the model is specified for realized variance (observed) and not for integrated variance (unobserved).

fractional differencing operator with parameter $d \in (-1/2, 1/2)$ is defined as usual $(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}$, with $\Gamma(\cdot)$ denoting the Gamma function. $\Theta(L) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)$ is a moving average (MA) lag polynomial and the error process u_t has zero mean.

2.2. **Interpretation.** The choice of the function $\phi_i(\cdot)$, $i = 1, \ldots, p$, is flexible and allows for different specifications, such as polynomials, logistic functions, exponential functions, splines, or others. The following examples list some possibilities.

EXAMPLE 1 (Linear ARFIMA). *Set* $\phi_i(\boldsymbol{s}_t; \boldsymbol{\xi}_i) = \phi_i$, $i = 1, \ldots, p$. *In this case,* $\Phi(L)(1-L)^d y_t = \Theta(L)u_t$, *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.*

EXAMPLE 2 (ARFIMA with smoothly changing parameters). *Set* $\boldsymbol{s}_t = t$. *Let* $\phi_i(\cdot)$, $i = 1, \ldots, p$: $\phi_i(\boldsymbol{s}_t; \boldsymbol{\xi}_i) = \phi_{i0} + \phi_{i1} f[\gamma(t-c)]$, *where* $f(y) = (1 + e^{-y})^{-1}$ *is the logistic function. Equation (1) becomes*

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

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* $t = c$.

EXAMPLE 3 (General Nonlinear ARFIMA). *A general alternative is to leave the type of non-linearity very general. Write*

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

where $f(\cdot)$ *is the logistic function,* $\gamma_m > 0$, *and* $\|\boldsymbol{\omega}_m\| = 1$, *with* $\omega_{m1} = \sqrt{1 - \sum_{j=2}^{q} \omega_{mj}^2}$.

Martens et al. (2009) 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 (1). The models put forward in Baillie and Kapetanios (2007, 2008) are also nested in our specification.

2.3. **Parameter Estimation.** In this section, we denote the parameter vector of the entire model as $\boldsymbol{\zeta} = (d, \boldsymbol{\xi}', \boldsymbol{\theta}', \sigma_u^2)' \in \mathbb{R}^{k_\zeta}$. Here, $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_0', \boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_p'\right)' \in \mathbb{R}^{k_\xi}$ denotes the vector of the parameters of the coefficient functions. The parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)' \in \mathbb{R}^q$ indexes the MA polynomial. Sometimes it is convenient to consider the parameter vector of the model excluding the error variance σ_u^2, which we denote $\boldsymbol{\psi} = (d, \boldsymbol{\xi}', \boldsymbol{\theta}')'$.

2.3.1. *Time Transformation.* Let T_0 be the size of a given data sample. For any sequence $\{x_t\}$, $t = 1, \ldots, T$, define $x_{tT} := (T_0/T)x_t$. Then, (1) is embedded in a sequence of models: $\Phi_{tT}(L)v_{tT}(d) = \Theta(L)u_t$, where $\Phi_{tT}(L) = 1 - \phi_1(\boldsymbol{s}_{tT}; \boldsymbol{\xi}_1)L - \ldots - \phi_p(\boldsymbol{s}_{tT}; \boldsymbol{\xi}_p)L^p$. 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 (Andrews and McDermott 1995).

2.3.2. *Assumptions.* We denote the true parameter as $\boldsymbol{\zeta}_* = \left(d_*, \boldsymbol{\xi}_*', \boldsymbol{\theta}_*', \sigma_{u,*}^2\right)' = \left(\boldsymbol{\psi}_*', \sigma_{u,*}^2\right)'$, where $\boldsymbol{\psi}_* = (d_*, \boldsymbol{\xi}_*', \boldsymbol{\theta}_*')'$, $\boldsymbol{\xi}_* = (\boldsymbol{\xi}_{1,*}', \ldots, \boldsymbol{\xi}_{p,*}')'$, $\boldsymbol{\theta}_* = (\theta_{1,*}, \ldots, \theta_{q,*})'$, and $\sigma_{u,*}^2$ is the error variance. Define $u_t(\boldsymbol{\psi}) = \Theta^{-1}(L)\,[\Phi_{tT}(L)v_{tT}]$. We use the shorthand notation $u_{t,*} := u_t(\boldsymbol{\psi}_*)$ and $v_{tT,*} := v_{tT}(d_*)$ and u_t and v_{tT} for $u_t(\boldsymbol{\psi})$ and $v_{tT}(d)$, respectively.

ASSUMPTION 1 (Parameter Space). *The parameter vector $\boldsymbol{\zeta}_* \in \mathbb{R}^{k_\zeta}$ is an interior point of $\boldsymbol{Z} \subset \mathbb{R}^{k_\zeta}$, a compact parameter space.*

ASSUMPTION 2 (Errors).

(1) *The sequence $\{u_{t,*}\}_{t=1}^{T}$ is drawn from an absolutely continuous distribution (with respect to the Lebesgue measure) that has positive density on the entire real line. $\mathbb{E}(u_{t,*}) = \mathbb{E}(u_{t,*}|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}(u_{t,*}^2) = \sigma_{u,*}^2 < \infty$, and $\mathbb{E}(u_{t,*}^2|\mathcal{F}_{t-1}) = \sigma_{t,*}^2$ such that $0 < \sigma_{t,*} < \infty$ for all t. Furthermore, $\lim\limits_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sigma_{t,*}^2 = \mathbb{E}(u_{t,*}^2)$. $\mathcal{F}_t$ is the σ-algebra formed by the information available at time t.*

(2) $\mathbb{E}|u_{t,*}|^n < \infty$ *for* $n = 1, \ldots, 4$.

ASSUMPTION 3 (Stationarity and Moments).

(1) $\mathbb{E}|\boldsymbol{z}_{tT}|^n < \infty$, $n = 1, \ldots, 4$, *where* $\boldsymbol{z}_{tT} = (v_{tT}, \boldsymbol{s}_{tT}')'$.

(2) $d_* \in (-1/2, 1/2)$.

(3) $\Theta_*(L)$ *is invertible.*

ASSUMPTION 4 (Autoregressive Nonlinear Function).

(1) *The transition functions are parameterized such that they are well defined.*

(2) *For all $\boldsymbol{s}_t$, $\boldsymbol{\xi}$, the roots of $\Phi_{tT,*}(\boldsymbol{s}_t;\boldsymbol{\xi})$ are outside the unit circle.*

(3) $\mathbb{E}\left|\frac{\partial}{\partial\boldsymbol{\xi}}\Phi_{tT}(L)v_{tT}\right|^n < \infty$, $n = 1, \ldots, 4$.

(4) $\mathbb{E}\left|\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'}\Phi_{tT}(L)v_{tT}\right|^n < \infty$, $n = 1, \ldots, 4$.

EXAMPLE 4 (for Assumption 4 (1): Logistic Transition). *If there are $M+1$ different regimes of volatility depending on a state variable s_t, with transitions governed by logistic functions, then the transition parameters c_m and γ_m, $m = 1, \ldots, M$, are such that: (1) $-\infty < c_1 < \ldots < c_M < \infty$; (2) $\gamma_m > 0$ for all m; and (3) $f[\gamma_1(s_t - c_1)] \geq f[\gamma_2(s_t - c_2)] \geq \ldots \geq f[\gamma_M(s_t - c_M)]$.*

Assumption 1 and 3 are standard for nonlinear models. In comparison with the extant literature, Assumption 2 is relatively weak. Baillie and Kapetanios (2008), for example, consider maximum likelihood estimation of long memory and nonlinear autoregressive models and derive their results under the assumption of i.i.d. errors and without considering time as possible non-linear state variable. Assumption 4 allows for a large number of functions.

2.3.3. *Least-Squares Estimation.* We estimate the parameters by nonlinear least squares (NLS), which in this case is equivalent to quasi-maximum likelihood estimation (QMLE):

$$\widehat{\boldsymbol{\psi}} = \arg\min_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathcal{Q}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T} q_t(\boldsymbol{\psi}) = \frac{1}{2T}\sum_{t=1}^{T} u_t^2(\boldsymbol{\psi}).$$

Let $\widehat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} u_t^2(\widehat{\boldsymbol{\psi}})$.

THEOREM 1 (Consistency). *Under Assumptions 1 – 4,* $\widehat{\boldsymbol{\psi}} \xrightarrow{p} \boldsymbol{\psi}_*$.

THEOREM 2 (Asymptotic Normality). *Under Assumptions 1 – 4,* $\widehat{\boldsymbol{\psi}}$ *is asymptotically normally distributed:* $\sqrt{T}\left(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_*\right) \xrightarrow{d} \mathcal{N}\left[0,\ \boldsymbol{A}(\boldsymbol{\psi}_*)^{-1}\boldsymbol{B}(\boldsymbol{\psi}_*)\boldsymbol{A}(\boldsymbol{\psi}_*)^{-1}\right]$*, where*

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

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

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

$$\boldsymbol{A}_T(\boldsymbol{\psi}) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 q_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'} \text{ and } \boldsymbol{B}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial q_t}{\partial\boldsymbol{\psi}}\frac{\partial q_t}{\partial\boldsymbol{\psi}'}.$$

3. MODEL SPECIFICATION

We propose a model-building procedure that selects the AR order and the number of nonlinear terms (M). M can be determined in two ways, either through an information criterion (IC) or by a sequence of tests for remaining nonlinearity.

3.1. Number of Nonlinear Terms.

3.1.1. *Information Criteria.* In this section we apply the results in Mendes (2012a,b) to show the consistency of an IC to determine the number of nonlinear terms in our model (M). Collect the data points in a vector $\boldsymbol{y} \in \mathbb{R}^T$. Define $M \in \{0, 2, \dots, \overline{M}\}$ and consider a class of models $\mathcal{M}(\boldsymbol{y}; \boldsymbol{\psi}_M)$ indexed by M. The parameter vector $\boldsymbol{\psi}_M \in \mathbb{R}^{k_{\psi,M}}$ is defined as in Section 2.3 and is also indexed by M. Denote M^* as the true value of M. Our goal is to estimate M by minimizing the following IC:

$$\mathsf{IC}(M) = \mathcal{Q}(\boldsymbol{\psi}, M) + \lambda_T(M), \tag{3}$$

where $\mathcal{Q}(\boldsymbol{\psi}, M) = \sum_{t=1}^{T} u_t^2(\boldsymbol{\psi}_M)$ and $\lambda_T(M)$ is a penalty term to be defined later.

Define the sets

$$\boldsymbol{\Psi}_M^* = \{\boldsymbol{\psi} \in \boldsymbol{\Psi} : \boldsymbol{\psi} = \arg\min \mathbb{E}[\mathcal{Q}(\boldsymbol{\psi}, M)]\} \text{ and } \boldsymbol{\Psi}_{T,M} = \{\boldsymbol{\psi} \in \boldsymbol{\Psi} : \boldsymbol{\psi} = \arg\min \mathcal{Q}(\boldsymbol{\psi}, M)\}.$$

ASSUMPTION 5 (Class of Models and Penalty Function). *Assume that:*

(1) $\mathcal{M}(\boldsymbol{y}; \boldsymbol{\psi}_1) \subseteq \mathcal{M}(\boldsymbol{y}; \boldsymbol{\psi}_2) \subseteq \cdots \subseteq \mathcal{M}(\boldsymbol{y}; \boldsymbol{\psi}_{\overline{M}})$;

(2) $\lambda_T(M)$ *is a positive and increasing function of* M;

(3) $\frac{1}{T}\lambda_T(M) \to 0$ *and* $\lambda_T(M) \to \infty$, *as* $T \to \infty$ *for every* $M \in \{0, 2, \dots, \overline{M}\}$;

(4) $\overline{M}$ *is such that*

$$\sum_{M=M^*+1}^{\overline{M}} \left[\frac{k_{\psi,M}}{\lambda_T(M) - \lambda_T(M^*)}\right] \to 0, \textit{ as } T \to \infty;$$

(5) *for every* $M = 0, 1, 2, \ldots, \overline{M}$ *and some positive constant* c

$$\max_{\widehat{\boldsymbol{\psi}} \in \boldsymbol{\Psi}_{T,M}} \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}^*_M} \mathbb{E} \left| \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) - \mathcal{Q}(\boldsymbol{\psi}, M) \right| \leq c k^2_{\psi,M} \text{ with probability } 1 \text{ as } T \to \infty.$$

Assumptions 5(1)-(4) define how the penalty term can be specified as well as the order of increase in the number of candidate models. Assumption 5(5) requires that the elements in $\boldsymbol{\Psi}_{T,M}$ and $\boldsymbol{\Psi}^*_M$ are arbitrarily close.

THEOREM 3. *Under Assumptions 1–3 and 5,* $P(\widehat{M} \neq M^*) \to 0$*, as* $T \to \infty$*.*

In contradistinction to Baillie and Kapetanios (2007), note that we consider the number of non-linear terms and not just the number of AR lags. Furthermore, Theorem (3) is general enough and can be used to jointly determine the order of the AR terms and the number of nonlinear terms.

3.1.2. *Sequence of Tests for Nonlinearity.* The testing procedure is inspired by van Dijk et al. (2002), Medeiros and Veiga (2005), and Strikholm and Teräsvirta (2006). To simplify the exposition we consider the case where there is no moving average term ($q = 0$) and the transition variable is scalar, $s_t \in \mathbb{R}$. However, it is not difficult to extend our results. Let $\boldsymbol{V}_{t-1} \equiv \boldsymbol{V}_{t-1}(d) = [v_{t-1}(d), \ldots, v_{t-p}(d)]'$ and consider the following model:

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

We wish to test $M = \tilde{M}$ against $M > \tilde{M}$. The appropriate null hypothesis is

$$\mathbb{H}_0 : \gamma_{\tilde{M}+1} = \gamma_{\tilde{M}+2} = \cdots = \gamma_M = 0. \quad (5)$$

Model (4) is identified only under the alternative, which means that standard asymptotic inference is not available. This problem is circumvented, as in Teräsvirta (1994), by expanding $f\left[\gamma_m \left(s_t - c_m\right)\right]$, $m = \tilde{M} + 1, \ldots, M$, into a Taylor series around the null hypothesis. 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:

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

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 $\mathbb{H}_0:\ \boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \boldsymbol{\rho}_3 = \boldsymbol{0}$. Under $\mathbb{H}_0$, $R_3(\boldsymbol{z}_t; \boldsymbol{\xi}) = 0$. We can use (6) to test for absence of remaining nonlinearity. Write $\widehat{\boldsymbol{h}}_t = \left(\widehat{\boldsymbol{h}}_{0,t}', \widehat{\boldsymbol{h}}_{a,t}'\right)'$, where $\widehat{\boldsymbol{h}}_{0,t} = -\frac{\partial u_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\Big|_{\mathbb{H}_0}$ and $\widehat{\boldsymbol{h}}_{a,t} = \left(\widehat{\boldsymbol{V}}_{t-1}' s_t, \widehat{\boldsymbol{V}}_{t-1}' s_t^2, \widehat{\boldsymbol{V}}_{t-1}' s_t^3\right)'$. Define $\boldsymbol{\iota} = (1, 1, \ldots, 1)' \in \mathbb{R}^T$ and $\widehat{\boldsymbol{H}} = \left(\widehat{\boldsymbol{h}}_1', \ldots, \widehat{\boldsymbol{h}}_T'\right)'$.

Following Wooldridge (1990, 1991) and under the additional assumption $\mathbb{E}|\boldsymbol{V}_{t-1}\boldsymbol{V}_{t-1}' s_t|^\delta < \infty$, for some $\delta > 6$, the test can be carried out in steps as follows:

(1) Estimate the parameters under $\mathbb{H}_0$ and compute the residuals $\widehat{u}_t$. If the sample size is small, estimation is difficult, such that $\widehat{\boldsymbol{h}}_{0,t} = \boldsymbol{0}$ is not met. This has an adverse effect on the empirical size of the test. To solve this problem, we regress the $\widehat{u}_t$ on $\widehat{\boldsymbol{h}}_{0,t}$. We compute a new sequence of residuals from this regression and use them to compute $\widetilde{\boldsymbol{H}}$.

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

(3) Compute the χ^2 statistic $LM_\chi = T - SSR$.

The test proposed above is robust against departures from normality as well as conditionally heteroskedastic errors. This is important in financial applications, where the errors are rarely normal and homoskedastic.

We now combine the procedure above into a sequence of tests. Start testing a linear model against a model with one or more nonlinear terms at a α_1-level of significance. In case $\mathbb{H}_0$ is rejected, one nonlinear term is added, the new model is re-estimated, and then tested against an alternative with one more nonlinear term. The procedure continues testing J nonlinear terms against alternative models with $\tilde{J} \geq J + 1$ terms at significance level $\alpha_J = \alpha_1 C^{J-1}$ for some 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, it is possible to control the overall level of significance. This procedure ensures that such a sequence of tests is consistent, and that $\alpha^* = \sum_{J=1}^{\bar{J}} \alpha_J$ acts as an upper bound on the overall level of significance. As for the determination of C, it is good practice to perform the sequence of tests with different values of C to avoid selecting models that are too parsimonious. On the other hand, one can fix the initial significance level and choose the value of C which gives a pre-specified upper bound.

3.2. **Autoregressive Order.** To determine of the AR order of the model, we follow Rech et al. (2001) and use a polynomial approximation of equation (4). By the Stone-Weierstrass theorem, the approximation can be made arbitrarily accurate under mild conditions. We select the number of lags in $\boldsymbol{V}_{t-1}$ in the approximate model in order to minimize a given IC.

4. Monte-Carlo Evidence

We generate 1000 replications of the models below with 500 and 1000 observations. We consider three distinct processes for u_t. In the simplest case, $u_t \sim \mathsf{NID}(0, 0.25)$. The second case is a GARCH specification where $\sigma_t^2 = 0.0001 + 0.95\sigma_{t-1}^2 + 0.049u_{t-1}^2$, $u_t = \sigma_t\varepsilon_t$, and $\varepsilon_t \sim \mathsf{NID}(0, 1)$. This implies that u_t has infinite fourth moment. The third error process is formed by a sequence of independent and t-distributed random variables with five degrees of freedom. Let the process $r_t = \exp(y_t)e_t$, where y_t is defined as below and $e_t \sim \mathsf{NID}(0, 1)$.

The shot-memory (SM) DGPs are given as:

$$y_t = 0.04 + 0.55y_{t-1} + 0.34y_{t-2} + \sigma_t\varepsilon_t. \tag{7}$$

$$y_t = 0.55y_{t-1} + 0.34y_{t-2} - (0.4y_{t-1} + 0.2y_{t-2})\, f\left[12\left(r_{t-1} + 0.5\right)\right] \tag{8}$$
$$+ (0.4y_{t-1} + 0.2y_{t-2})\, f\left[4\left(r_{t-1} - 1\right)\right] + \sigma_t\varepsilon_t.$$

$$y_t = 0.55y_{t-1} + 0.34y_{t-2} - (0.4y_{t-1} + 0.2y_{t-2})\, f\left[200\left(t/T + 0.25\right)\right] \tag{9}$$
$$+ (0.4y_{t-1} + 0.2y_{t-2})\, f\left[100\left(t/T - 0.75\right)\right] + \sigma_t\varepsilon_t.$$

Define $v_t \equiv (1-L)^{0.4}y_t$ and write the long-memory models as below.

$$v_t = 0.55v_{t-1} + 0.34v_{t-2} + \sigma_t\varepsilon_t. \tag{10}$$

$$v_t = 0.55v_{t-1} + 0.34v_{t-2} - (0.4v_{t-1} + 0.2v_{t-2})\, f\left[12\left(r_{t-1} + 0.5\right)\right] \tag{11}$$
$$+ (0.4v_{t-1} + 0.2v_{t-2})\, f\left[4\left(r_{t-1} - 1\right)\right] + \sigma_t\varepsilon_t.$$

$$v_t = 0.55v_{t-1} + 0.34v_{t-2} - (0.4v_{t-1} + 0.2v_{t-2})\, f\left[200\left(t/T + 0.25\right)\right] \tag{12}$$
$$+ (0.4v_{t-1} + 0.2v_{t-2})\, f\left[100\left(t/T - 0.75\right)\right] + \sigma_t\varepsilon_t.$$

The first class of models is a simple short-memory AR specification. The next two classes of DGPs are nonlinear short-memory processes, while the remaining specifications are all long-memory models. The specification and estimation results are reported in Tables 1 and 2.

Tables 1 shows, for $T = 1000$, the average bias and the mean-squared error (MSE) of the parameter estimates under the assumption of correct specification, i.e., correct number of regimes (M) and AR order (p). Results for $T = 500$ can be found in the supplemental material. The results show that the estimation is reliable. Note that the estimation of γ is known to be noisy.

In order to evaluate the performance of the modeling strategy, we also check the frequency of correct specification when the regime structure is unknown. The number of regimes is determined by the sequence of robust LM tests, while the AR order is determined by the BIC. Alternatively, we consider selection of both M and p by the BIC. The results are reported in Table 2. In order to evaluate the effects of different values of the significance-level adjusting parameter $C \in (0, 1)$ on the frequency of correct specification, we also run the sequence of robust LM tests considering $C \in \{1/3,\, 1/2,\, 2/3\}$. The results are reported in the supplemental material. There are no differences in the results between $C = 1/2$ and $C = 1/3$ and the sequence of tests tends to underestimate the number of regimes. On the other hand, when $C = 2/3$, the frequency of correct specification is a bit higher than expected.

The following conclusions emerge from Table 2. Firstly, in the linear case both methodologies work very well. Secondly, the selection of p is very precise in almost all the cases. The number of regimes is underestimated by both methods, but the performance improves as the sample size increases. As expected, the sequence of robust LM tests seems to work better when the errors are not normal. For the nonlinear short-memory models, the sequence of LM tests works better than the BIC. On the other hand, for nonlinear long-memory models, the LM tests are superior than the BIC only when the second nonlinear model is considered (breaks). For the first nonlinear long-memory model, the BIC delivers better results.

TABLE 1. PARAMETER ESTIMATES UNDER CORRECT SPECIFICATION.

The table reports average bias and mean-squared error (MSE) of parameter estimates from 1000 simulations of the models listed in Section 4 with $T = 1000$. The regime and lag structure M and p are assumed to be known in the estimation.

Short-Memory Models

	Linear						Nonlinear I						Nonlinear II					
	Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ϕ_{10}	0.00	0.00	0.01	0.01	0.05	0.02	0.00	0.01	0.06	0.01	0.03	0.02	0.01	0.01	0.01	0.03	0.02	0.02
ϕ_{11}	–	–	–	–	–	–	0.00	0.01	-0.03	0.01	-0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
ϕ_{12}	–	–	–	–	–	–	0.02	0.01	0.04	0.06	0.04	0.02	0.01	0.01	0.04	0.02	0.12	0.03
ϕ_{20}	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	-0.03	0.01	0.04	0.02	0.01	0.01	0.02	0.02	0.05	0.03
ϕ_{21}	–	–	–	–	–	–	0.04	0.02	0.05	0.01	0.06	0.04	0.01	0.01	0.08	0.02	0.05	0.05
ϕ_{22}	–	–	–	–	–	–	0.06	0.08	-0.01	0.01	0.04	0.02	0.02	0.01	0.04	0.08	0.02	0.02
γ_1	–	–	–	–	–	–	1.14	304.12	-3.10	15.46	-3.14	45.74	-33.42	16.74	-50.11	14.93	-47.24	54.64
γ_2	–	–	–	–	–	–	-13.10	403.13	2.15	19.31	-9.12	14.44	-45.13	23.16	-45.42	24.92	-116.73	46.56
c_1	–	–	–	–	–	–	0.01	0.02	0.03	0.01	-0.02	0.01	0.01	0.01	0.02	0.01	0.05	0.01
c_2	–	–	–	–	–	–	0.01	0.01	0.05	0.02	0.09	0.02	0.01	0.01	0.09	0.06	0.02	0.08
d	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.01	0.02	0.01	0.05	0.08	0.00	0.00	0.05	0.08	0.01	0.02
σ_u	0.00	0.00	0.00	0.01	0.10	0.12	0.00	0.01	0.02	0.14	0.04	0.04	0.00	0.00	0.01	0.00	0.09	0.14

Long-Memory Models

	Linear						Nonlinear I						Nonlinear II					
	Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ϕ_{10}	0.01	0.01	0.01	0.01	0.03	0.01	0.00	0.01	0.05	0.02	0.04	0.01	0.00	0.01	0.04	0.03	0.02	0.02
ϕ_{11}	–	–	–	–	–	–	0.00	0.01	0.06	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.04
ϕ_{12}	–	–	–	–	–	–	0.02	0.01	0.10	0.07	0.02	0.01	0.00	0.01	0.02	0.01	0.03	0.03
ϕ_{20}	0.03	0.01	0.01	0.01	0.01	0.01	0.03	0.00	0.03	0.02	0.02	0.01	0.00	0.02	0.01	0.01	0.03	0.02
ϕ_{21}	–	–	–	–	–	–	0.01	0.01	0.08	0.02	0.05	0.02	0.03	0.01	0.04	0.01	0.01	0.03
ϕ_{22}	–	–	–	–	–	–	0.03	0.05	0.01	0.01	0.08	0.04	0.01	0.01	0.06	0.03	0.03	0.01
γ_1	–	–	–	–	–	–	-3.21	99.24	4.38	94.63	18.93	19.84	18.12	38.14	25.27	15.54	14.84	47.18
γ_2	–	–	–	–	–	–	-16.14	83.84	2.85	84.37	9.47	39.36	15.28	47.34	36.46	42.45	37.18	94.18
c_1	–	–	–	–	–	–	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.09	0.07
c_2	–	–	–	–	–	–	0.01	0.01	0.04	0.02	0.13	0.08	0.01	0.01	0.04	0.04	0.09	0.13
d	0.01	0.01	0.00	0.01	0.01	0.01	0.08	0.02	0.02	0.00	0.09	0.09	0.08	0.03	0.06	0.07	0.01	0.02
σ_u	0.00	0.00	0.00	0.00	0.08	0.07	0.00	0.00	0.05	0.08	0.12	0.05	0.00	0.00	0.01	0.01	0.19	0.23

Table 2. Frequency of Correct Specification: LM Tests and BIC.

The table reports the proportion of correctly determined numbers of regimes and lag structures in 1000 simulations of the models listed in Section 4. The selection method is either the sequence of LM tests or the BIC (values in parentheses). The order p is always selected by BIC using a third-order approximation to the nonlinear function. We simulate the cases $T = 500$ and $T = 1000$.

Short-Memory Models: 500 observations

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
p	0.98 (0.98)	0.98 (0.98)	0.94 (0.94)	0.86 (0.86)	0.98 (0.98)	0.92 (0.92)	0.66 (0.66)	0.64 (0.64)	0.48 (0.48)
M	1.00 (1.00)	1.00 (1.00)	1.00 (0.98)	0.16 (0.08)	0.20 (0.06)	0.06 (0.10)	0.42 (0.18)	0.50 (0.12)	0.52 (0.12)
M and p	0.98 (0.98)	0.98 (0.98)	0.94 (0.92)	0.14 (0.06)	0.20 (0.06)	0.06 (0.10)	0.28 (0.14)	0.38 (0.08)	0.30 (0.08)

Short-Memory Models: 1000 observations

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
p	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.98 (0.98)	0.95 (0.95)	0.90 (0.90)
M	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.37 (0.26)	0.42 (0.18)	0.21 (0.30)	0.60 (0.40)	0.73 (0.32)	0.69 (0.32)
M and p	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.37 (0.26)	0.42 (0.18)	0.21 (0.30)	0.60 (0.40)	0.73 (0.32)	0.69 (0.32)

Long-Memory Models: 500 observations

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
p	0.96 (0.96)	0.98 (0.98)	0.98 (0.98)	0.94 (0.94)	0.90 (0.90)	0.86 (0.86)	0.96 (0.96)	0.72 (0.72)	0.70 (0.70)
M	1.00 (1.00)	0.96 (1.00)	0.98 (1.00)	0.08 (0.06)	0.10 (0.06)	0.10 (0.06)	0.56 (0.04)	0.50 (0.14)	0.44 (0.12)
M and p	0.96 (0.96)	0.94 (0.98)	0.96 (0.98)	0.08 (0.06)	0.10 (0.06)	0.10 (0.06)	0.54 (0.04)	0.38 (0.10)	0.38 (0.10)

Long-Memory Models: 1000 observations

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
p	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.96 (0.96)	0.95 (0.95)
M	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.19 (0.27)	0.32 (0.26)	0.30 (0.17)	0.97 (0.15)	0.90 (0.44)	0.88 (0.32)
M and p	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.19 (0.27)	0.32 (0.26)	0.30 (0.17)	0.97 (0.15)	0.90 (0.44)	0.88 (0.32)

5. Empirical Application

5.1. **Long Memory and Nonlinearity in Volatility.** Andersen et al. (2001,2003) and Barndorff-Nielsen and Shephard (2002) have pioneered the use of intraday data to construct measures of realized volatility (RV). Several studies have proposed extensions of the basic RV estimator that are robust to microstructure noise (Zhang et al. 2005, Barndorff-Nielsen et al. 2008) and the presence of jumps (Andersen et al. 2009). The daily dynamics of RV exhibit high persistence. Andersen et al. (2003) use an ARFIMA specification to model this long-range dependence. An alternative to ARFIMA are models that approximate long memory by aggregation. In this case, volatility is modeled as a sum of different processes, each with low persistence. This is used in the HAR-RV model proposed by Corsi (2009). On the other hand, there is evidence of nonlinearity in volatility, such as multiple regimes (Black 1976). Regime changes can take the form of switches in the parameters, for instance governed by a Markov chain (Hamilton and Susmel 1994), hard thresholds (Liu et al. 1997), or smooth transitions as in Medeiros and Veiga (2009).

McAleer and Medeiros (2008) consider a nonlinear heterogeneous autoregressive (HAR) model to describe both long range dependence and nonlinear dynamics. In their approach, long memory is approximated by aggregation and is not explicitly modeled by fractional differencing as it is here; see also Scharth and Medeiros (2009) and Chen et al. (2010).

5.2. **Data.** We use tick-by-tick trade data from the 30 stocks that comprise the Dow Jones Industrial Average in September of 2010: Alcoa Inc. (AA), Altria Group (MO), American Express Inc. (AXP), AT&T (T), Bank of America (BAC), Boeing Co. (BA), Caterpillar Inc. (CAT), Chevron Corp. (CVX), Cisco Systems (CSCO), Coca Cola (KO), Du Pont De Nemours (DD), Exxon Mobil (XOM), General Electric (GE), Hewlett Packard (HPQ), The Home Depot Inc. (HD), Intel Co. (INTC), International Business Machines Corp. (IBM), Johnson and Johnson (JNJ), JPMorgan Chase (JPM), Kraft Foods (KFT), McDonald's (MCD), Merck Co.

(MRK), Microsoft Corp. (MSFT), Pfizer Inc. (PFE), Procter and Gamble (PG), United Technologies Corp. (UTX), Verizon Communications (VZ), Wal-Mart Stores (WMT), Walt Disney Co. (DIS), and 3M Company (MMM). The data are obtained from the NYSE TAQ (Trade and Quote) database. The sample period starts in January 3, 2000, and ends in December 31,2009.

5.2.1. *Realized Volatility Estimation.* In calculating daily RV, we employ the realized kernel estimator with modified generalized Tukey-Hanning weights of order two according to Barndorff-Nielsen et al. (2008), hereafter BHLS. We also used the MedRV proposed in Andersen et al. (2009) and the results are very similar. The BHLS estimator is robust to jumps in estimating quadratic variation and the MedRV is robust to jumps in estimating integrated variance. We discard transactions outside trading hours, considering transactions between 9.30am through 4.00pm. Following Barndorff-Nielsen et al. (2008) we use 60-second activity-fixed tick time sampling schemes, such that we obtain the same number of observations each day. Changes between consecutive trades of more than five standard deviations of intra-day returns for any given day are discarded. For our data set of widely traded stocks and in this sample period, this removes most of the obvious recording errors but no meaningful price changes.

5.3. **Model Specification and Estimation.** We report specification results for the 30 series described above for the full sample period. Table 3 shows the number of regimes determined by the sequence of robust LM tests as well as BIC. The numbers in parentheses are the p-values for the remaining nonlinearity test when the sequence of LM tests is used. We consider past returns as well as time as transition variables. Several interesting results emerge. Firstly, for most of the stocks, the p-values are quite high, indicating that the choice of the initial significance level as well as the constant C does not influence the final number of regimes. Secondly, both criteria select a small number of regimes, which indicates that there is no overfitting. Thirdly, the LM tests tend to select a smaller number of regimes than the BIC. However, there are a few cases where both criteria agree on the number of regimes.

We discuss estimation and in-sample fit for one stock (KFT) that is fairly representative for the set of 30 stocks. The sample period for KFT is Jun 13, 2001, through Dec 31, 2009. Figure 1, Panel (a), shows the MedRV estimator for the logarithm of realized variance (solid) and the BHLS estimator (dots) on the left ordinate, and the estimated transition function with respect to time on the right ordinate. Panel (b) displays a Gaussian kernel estimate for the log returns of KFT in percent on the left ordinate and the estimated transition function with respect to lagged returns on the right ordinate.

For example, the estimated specification for the MedRV estimator and time transitions is as follows (standard errors in parentheses).

$$
\begin{aligned}
y_t &= \log(RV_t) - 0.2218,\ v_t = (1-L)^{\underset{(0.0036)}{0.4416}} y_t, \\
v_t &= \Big[\underset{(0.0055)}{-0.1230} + \underset{(0.0166)}{0.2295} f(\gamma, c)\Big] v_{t-1} - \Big[\underset{(0.0032)}{0.0851} - \underset{(0.0151)}{0.1375} f(\gamma, c)\Big] v_{t-2} + u_t, \qquad (13) \\
f(\gamma, c) &= \{1 + \exp[\gamma(t/T - c)]\}^{-1} = \left\{1 + \exp\left[\underset{(42973)}{69.30}\left(t/T - \underset{(1.26e-4)}{0.7661}\right)\right]\right\}^{-1}.
\end{aligned}
$$

The time transition captures the change in volatility dynamics during the subprime crisis, as can be seen in Panel (a) of Figure 1. Long-range dependence is captured by the fractional differencing parameter $d = 0.4416$, and the autoregressive parameters reflect changes in the short-run dynamics from anti-persistence at a scale of $1/(1 + 0.2082) \approx 0.8$ days to persistence at a scale of $1/(1+0.2082-0.367) = 1/(1-0.1588) \approx 1.2$ days. This corresponds to the short decorrelation scale in stock volatility reported, for example, in Hillebrand (2006).

The estimated specification for the BHLS estimator and asymmetry effects is as follows:

$$
\begin{aligned}
y_t &= \log(RV_t) - 0.1356,\ v_t = (1-L)^{\underset{(0.0033)}{0.4900}} y_t, \\
v_t &= \left[\underset{(0.0215)}{-0.0678} - \underset{(0.0193)}{0.0873} f(\gamma, c)\right] v_{t-1} + \left[\underset{(0.0468)}{0.2804} - \underset{(0.0473)}{0.3834} f(\gamma, c)\right] v_{t-2} \\
&\quad - \left[\underset{(0.0330)}{0.1543} - \underset{(0.0345)}{0.0837} f(\gamma, c)\right] v_{t-3} + u_t, \qquad (14) \\
f(\gamma, c) &= \left(1 + \exp\left\{\gamma\left[100\left(\log\frac{S_{t-1}}{S_{t-2}}\right) - c\right]\right\}\right)^{-1} \\
&= \left(1 + \exp\left\{\underset{(436.69)}{7.5875}\left[100\left(\log\frac{S_{t-1}}{S_{t-2}}\right) + \underset{(9.9e-5)}{2.2422}\right]\right\}\right)^{-1}.
\end{aligned}
$$

The estimation identifies an asymmetry effect at a threshold of -2.2422 percent return, as can be seen in Panel (b) of Figure 1. Above this threshold, there are essentially no short-term dynamics. The sum of the AR parameters in the regime above -2.2422 percent return is 0.0583. In the regime below this threshold, the sum of the AR parameters is -0.3286, indicating anti-persistence on the scale of three quarters of a day. The long-term dynamics are captured by an estimated d of 0.49, close to the non-stationary region, and the anti-persistent short scale will partly offset the long-range dependence. This corresponds to earlier findings that large negative returns influence volatility on shorter scales (Medeiros and Veiga 2009).

5.4. **Forecasting Exercise.** For each stock, we consider a rolling window of 1,500 observations for model specification and parameter estimation. Then, we use the model for one-, five-, and ten-days-ahead forecasting of log RV. For each forecast horizon we compare the models against a benchmark specification using the unconditional Giacomini and White (GW) test for equal predictive accuracy (Giacomini and White 2006).

We consider the following alternative specifications: (1) linear ARFIMA ("ARFIMA"); (2) nonlinear ARFIMA with time as transition variable ("STARFIMA I"); and (3) nonlinear ARFIMA with past daily return as transition variable ("STARFIMA II"). The benchmark models are: (1)

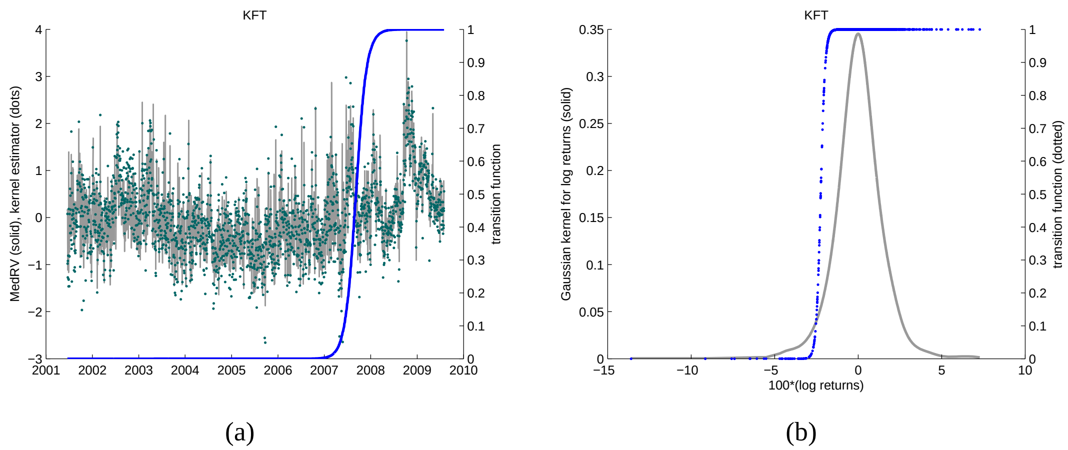

FIGURE 1. (a) KFT logarithm of realized variance (solid: MedRV, dots: kernel estimator) and transition function with respect to time, (b) KFT Gaussian kernel for log returns and transition function with respect to lagged returns (estimated from kernel estimator).

the linear HAR-RV model of Corsi (2009) ("Ratio I") and (2) the nonlinear HAR-RV of Corsi and Renò (2012) ("Ratio II"). Table 4 reports the results for one-step-ahead while Table 5 shows the results for five- and ten-steps ahead.

The reporting format are ratios of root mean-square errors (RMSE) of the proposed model to the RMSE of the benchmark model. The p-values of the Giacomini and White (2006) test are displayed in the column "GW I" for comparison with the linear HAR-RV model and in the column "GW II" for comparison with the nonlinear HAR-RV model. The null hypothesis of this test is that both models have equal predictive ability; p-values smaller than common significance levels indicate that one model outperformed the other. The more powerful model is indicated by the RMSE ratio; we are looking for ratios smaller than one. The models are re-specified for each time window. The sequence of LM tests is used to determine the number of regimes.

The main conclusion from the out-of-sample results is that the linear and nonlinear ARFIMA models outperform the benchmark for five- and ten-steps ahead. For one-step-ahead the benchmarks are superior. More specifically, for five-days-ahead, the ARFIMA is statistically superior

TABLE 3. FULL SAMPLE RESULTS: NUMBER OF REGIMES.

The table displays the estimated number of regimes for nonlinear ARFIMA models with lagged returns or time as transition variables, respectively. Three different criteria for the selection of regimes are considered. LM is the sequence of robust LM tests. BIC is the Bayes information criterion. The numbers in parentheses are p-values of the test for remaining nonlinearity.

	Transition Variable: Past Return		Transition Variable: Time	
	LM	BIC	LM	BIC
AA	2 (0.102)	1	2 (0.903)	1
AXP	1 (0.212)	2	1 (0.053)	1
BA	2 (0.193)	1	1 (0.405)	1
BAC	1 (0.140)	1	2 (0.883)	2
CAT	1 (0.229)	1	1 (0.051)	1
CSCO	2 (0.226)	2	1 (0.256)	1
CVX	1 (0.282)	2	2 (0.444)	1
DD	1 (0.345)	1	2 (0.329)	1
DIS	1 (0.379)	1	2 (0.047)	2
GE	1 (0.139)	2	2 (0.196)	2
HD	2 (0.156)	1	1 (0.244)	1
HPQ	1 (0.087)	1	2 (0.321)	1
IBM	2 (0.241)	1	2 (0.090)	1
INTC	1 (0.119)	1	1 (0.074)	1
JNJ	2 (0.722)	2	2 (0.815)	1
JPM	1 (0.096)	2	2 (0.579)	2
KFT	2 (0.158)	1	2 (0.088)	1
KO	1 (0.200)	1	1 (0.476)	1
MCD	1 (0.670)	1	1 (0.077)	1
MMM	2 (0.269)	2	2 (0.836)	1
MO	1 (0.224)	1	2 (0.087)	1
MRK	1 (0.228)	1	1 (0.200)	1
MSFT	1 (0.339)	1	2 (0.113)	1
PFE	1 (0.474)	1	1 (0.493)	1
PG	1 (0.199)	1	1 (0.221)	1
T	1 (0.357)	1	1 (0.147)	1
UTX	1 (0.055)	1	2 (0.650)	1
VZ	2 (0.082)	1	1 (0.204)	1
WMT	2 (0.436)	1	1 (0.218)	1
XOM	1 (0.087)	2	2 (0.214)	2

than the benchmarks in 27 cases. For ten-days-ahead, the ARFIMA model outperforms the HAR-RV and nonlinear HAR-RV in 21 and 23 cases, respectively. For five days-ahead, the nonlinear ARFIMA with breaks outperforms the HAR-RV and the nonlinear HAR-RV in 24 and 27 cases, respectively. When ten-days-ahead forecasts are considered, the nonlinear ARFIMA outperforms the benchmarks in 20 (HAR-RV) and 21 (nonlinear HAR-RV) cases. The performance of the asymmetric ARFIMA is similar. For five-days-ahead, it is superior than the benchmarks in 21 (HAR-RV) and 19 (nonlinear HAR-RV) cases, while for ten-days-ahead it outperforms the benchmarks in 19 cases. Finally, the benchmarks do not deliver any superior forecast for horizons larger than one. On the other hand, for one-step ahead, the nonlinear HAR-RV model statistically outperforms the competing models in about 50% of the cases.

In summary, with regard to financial volatility, we recommend the model proposed in this paper for forecast horizons longer than a single day. The model also has the advantage of identifying economically interpretable nonlinear effects in-sample, and it separates the long and short decorrelation scales found in financial volatility.

6. Conclusion

Usually, 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 asymmetry in the presence of long memory.

We show consistency and asymptotic normality of the nonlinear least-squares estimator. Asymptotic theory requires a time transformation that ensures that regimes of finite length remain identified as the sample size grows to infinity. We also propose two different model building procedures to determine the structure of the model.

Using stocks in the Dow Jones Industrial Average between 2000 and 2009, we find strong evidence for nonlinear effects driven by time and lagged returns in financial volatility time series. A forecast competition indicates that a specification with long memory and asymmetry can outperform the standard HAR-RV model and linear ARFIMA specifications, in particular at long forecast horizons.

APPENDIX A. PROOF OF CONSISTENCY

Proof of Theorem 1. By Theorem 4.1.1 of Amemiya (1985), $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_*$ if the conditions below hold: (1) $\boldsymbol{\Psi}$ is a compact parameter set; (2) $\mathcal{Q}_T(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi}$ and measurable in u_t; (3) As $T \to \infty$, $\mathcal{Q}_T(\boldsymbol{\psi})$ converges in probability to a deterministic function $\mathcal{Q}(\boldsymbol{\psi}) = \mathbb{E}\left[\mathcal{Q}_T(\boldsymbol{\psi})\right] < \infty$ uniformly on $\boldsymbol{\Psi}$; and (4) $\mathcal{Q}(\boldsymbol{\psi})$ attains a unique global maximum at $\boldsymbol{\psi}_0$.

Item (1) is given by assumption. Item (2) holds by definition of $\mathcal{Q}_T(\boldsymbol{\psi})$ and u_t. To prove item (3) we first notice that Assumptions 3 and 4 imply that $\mathbb{E}\left[q_t(\boldsymbol{\psi})\right] < \infty, \forall\, t$. Hence, $\mathbb{E}\left[\mathcal{Q}_T(\boldsymbol{\psi})\right] < \infty$. Now set $g_t(\boldsymbol{\psi}) = q_t(\boldsymbol{\psi}) - \mathbb{E}\left[q_t(\boldsymbol{\psi})\right]$. Also, $\mathbb{E}\left[\sup|q_t(\boldsymbol{\psi})|\right] < \infty$ by Assumptions 3 and 4(2) and (3). Application of Theorem 3.1 in Ling and McAleer (2003) proves item (3).

Consider Item (4). Rewrite the maximization problem as $\max\limits_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left[q_t(\boldsymbol{\psi}) - q_t(\boldsymbol{\psi}_*)\right]$. Now, $\mathbb{E}\left[q_t(\boldsymbol{\psi}) - q_t(\boldsymbol{\psi}_*)\right] = \frac{1}{2}\mathbb{E}\left(u_t^2 - u_{t,*}^2\right)$. 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

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

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.

□

APPENDIX B. PROOF OF ASYMPTOTIC NORMALITY

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. By the Cauchy-Schwarz inequality, $\mathbb{E}\left(XY\right) \leq \left[\mathbb{E}\left(X^2\right)\right]^{\frac{1}{2}} \left[\mathbb{E}\left(Y^2\right)\right]^{\frac{1}{2}}$, 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. By the same token, if both X and Y have finite second moments,

$$\mathbb{E}\left[(X+Y)^2\right] \leq \mathbb{E}\left(X^2\right) + \mathbb{E}\left(Y^2\right) + 2\left[\mathbb{E}\left(X^2\right)\right]^{\frac{1}{2}}\left[\mathbb{E}\left(Y^2\right)\right]^{\frac{1}{2}},$$
$$\leq K\left[\mathbb{E}\left(X^2\right) + \mathbb{E}\left(Y^2\right)\right] \text{ for some } K < \infty.$$

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

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

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

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

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

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

since $\frac{\partial}{\partial \boldsymbol{\theta}}\Theta^{-1}(L)\left[\Phi_{tT}(L)v_{tT}\right]$ does not contain u_t. □

LEMMA 2. *Under Assumptions 2-4,* $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\partial q_t}{\partial \boldsymbol{\psi}}\right| < \infty$ *and* $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\partial q_t}{\partial \boldsymbol{\psi}}\frac{\partial q_t}{\partial \boldsymbol{\psi}'}\right| < \infty$.

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:

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

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

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

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

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

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

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

by Assumptions 4(2), 2(2), and 3(2). By Assumption 4, all other elements are bounded:

$$\mathbb{E}\left|\frac{\partial q_t}{\partial\boldsymbol{\xi}}\right| = \mathbb{E}\left|u_t\Theta^{-1}(L)\frac{\partial}{\partial\boldsymbol{\xi}}\Phi_{tT}(L)v_{tT}\right| \leq \left(\mathbb{E}\left|u_t\right|^n\right)^{\frac{1}{n}}\left\{\mathbb{E}\left|\Theta^{-1}(L)\frac{\partial}{\partial\boldsymbol{\xi}}\Phi_{tT}(L)v_{tT}\right|^n\right\}^{\frac{1}{n}} < \infty.$$

By Assumptions 1, 2(2), 3(2), and 4,

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

This shows the first statement of Lemma 2. The second statement 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 $n = 4$. We omit the details of the second statement for the sake of brevity. □

LEMMA 3. *The function* $h_t(\boldsymbol{\psi}) := -\frac{\partial^2 q_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'} - \boldsymbol{A}(\boldsymbol{\psi})$*, where* $\boldsymbol{A}(\boldsymbol{\psi}) = -\mathbb{E}\left(\frac{\partial^2 q_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right)$*, is absolutely uniformly integrable:* $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |h_t(\boldsymbol{\psi})|\right] < \infty$*; it is continuous in* $\boldsymbol{\psi}$ *and* $\mathbb{E}\left[h_t(\boldsymbol{\psi})\right] = 0$*.*

Proof. By the triangular inequality, showing absolute uniform integrability is equivalent to show that $\mathbb{E}\left(\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\frac{\partial^2 q_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|\right) < \infty$. We will consider the second derivative of q_t with respect to d. There are 21 distinct second derivatives in $\boldsymbol{A}(\cdot)$; proving finiteness of the expected value of the supremum consists of applications of the Lebesgue Dominated Convergence Theorem.

First, note that

$$\frac{\partial^2}{\partial d^2}(1-L)^d = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left[\left(\sum_{i=0}^{j-1}\frac{1}{d-i}\right)^2 - \sum_{i=0}^{j-1}\left(\frac{1}{d-i}\right)^2\right]\prod_{i=0}^{j-1}(d-i)L^j, \tag{15}$$

$$= \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left[\sum_{\substack{i,k=0\\ i\neq k}}^{j-1}\frac{1}{(d-i)(d-k)}\right]\prod_{i=0}^{j-1}(d-i)L^j.$$

Then, we have

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

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

$$|R_1| = \left|\left[\Theta^{-1}(L)\Phi_{tT}(L)\frac{\partial}{\partial d}(1-L)^d y_{tT}\right]^2\right| \text{ and } |R_2| = \left|u_t\Theta^{-1}(L)\left[\Phi_{tT}(L)\frac{\partial^2}{\partial d^2}(1-L)^d y_{tT}\right]\right|.$$

The expected values of the terms on the right-hand sides are finite by arguments similar to those in the proof of Lemma 2. Therefore, the suprema of the left-hand sides are dominated by the right-hand sides and $\mathbb{E}\left[\sup|R_i|\right] < \infty$, $i = 1, 2$, by the Lebesgue Dominated Convergence Theorem. Thus, $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|h_t(\boldsymbol{\psi})|\right] < \infty$. □

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

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

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) $\left\{ \left.\frac{\partial q_t}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_*}, \mathcal{F}_t \right\}$ 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 $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_*$,

$$\boldsymbol{A}_T(\widehat{\boldsymbol{\psi}}_T) = -\frac{1}{T}\sum_{t=1}^{T} \left.\frac{\partial^2 q_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right|_{\widehat{\boldsymbol{\psi}}_T} \quad \text{and} \quad \boldsymbol{A}(\boldsymbol{\psi}_*) = -\mathbb{E}\left(\left.\frac{\partial^2 q_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_*} \right)$$

is non-singular. Conditions for this convergence can be found 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. $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|\boldsymbol{A}_T(\boldsymbol{\psi}) - \boldsymbol{A}(\boldsymbol{\psi})| \xrightarrow{p} 0$. Ling and McAleer (2003, Theorem 3.1) 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 $h_t(\boldsymbol{\psi}) = -\frac{\partial^2 q_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \boldsymbol{A}(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi}$, $\mathbb{E}h_t(\boldsymbol{\psi}) = 0$, and is absolutely uniformly integrable $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|h_t(\boldsymbol{\psi})|\right] < \infty$. This was shown in Lemma 3. Thus, we have established all conditions of 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 $m_t(\boldsymbol{\psi}) := \frac{\partial q_t}{\partial \boldsymbol{\psi}}\frac{\partial q_t}{\partial \boldsymbol{\psi}'} - \boldsymbol{B}(\boldsymbol{\psi})$.

As we did for $\boldsymbol{A}$ in Lemma 3, we have to show that m_t is absolutely uniformly integrable, continuous in $\boldsymbol{\psi}$, and $\mathbb{E}\left[m_t(\boldsymbol{\psi})\right] = 0$. By the triangular inequality, showing absolute uniform integrability reduces to showing that $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \frac{\partial q_t}{\partial \boldsymbol{\psi}}\frac{\partial q_t}{\partial \boldsymbol{\psi}'}\right] < \infty$. 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. m_t is continuous in $\boldsymbol{\psi}$ by the Continuous Mapping Theorem and has zero-mean by construction. □

APPENDIX C. PROOF OF MODEL SELECTION CONSISTENCY

Proof of Theorem 3. Write the event

$$\begin{aligned}
\{\widehat{M} \neq M^*\} &\Leftrightarrow \{\widehat{M} \neq M^* \cap M > M^*\} \cup \{\widehat{M} \neq M^* \cap M < M^*\} \\
&\Rightarrow \{\widehat{M} > M^* \cap M > M^*\} \cup \{\widehat{M} < M^* \cap M < M^*\} \\
&\Leftrightarrow \{\mathsf{IC}(\widehat{M}) < \mathsf{IC}(M^*) \cap M > M^*\} \cup \{\mathsf{IC}(\widehat{M}) < \mathsf{IC}(M^*) \cap M < M^*\} \\
&\Rightarrow \left[\bigcup_{M=M*+1}^{\overline{M}} \{\mathsf{IC}(M) < \mathsf{IC}(M^*)\}\right] \cup \left[\bigcup_{M=1}^{M^*-1} \{\mathsf{IC}(M) < \mathsf{IC}(M^*)\}\right] =: A \cup B.
\end{aligned}$$

We show that $\mathbb{P}(A \cup B) \to 0$ as $T \to \infty$. It is clear that

$$\{\mathsf{IC}(M) < \mathsf{IC}(M^*)\} \Leftrightarrow \{\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) > \lambda_T(M) - \lambda_T(M^*)\}.$$

Applying the Markov inequality to the right hand side of the above equation, we have, for $M > M^*$,

$$P\left(\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) > \lambda_T(M) - \lambda_T(M^*)\right) \leq \frac{\mathbb{E}\left|\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M)\right|}{\lambda_T(M) - \lambda_T(M^*)}.$$

Using the triangular inequality, we obtain

$$
\begin{aligned}
\mathbb{E}\left|\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M)\right| \leq & \max_{\widehat{\boldsymbol{\psi}} \in \boldsymbol{\Psi}_{T,M^*}} \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}_M^*} \mathbb{E}\left|\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\boldsymbol{\psi}, M^*)\right| \\
& + \max_{\widehat{\boldsymbol{\psi}} \in \boldsymbol{\Psi}_{T,M}} \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}_M^*} \mathbb{E}\left|\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) - \mathcal{Q}(\boldsymbol{\psi}, M)\right| \\
& + \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}_M^*} \mathbb{E}\left|\mathcal{Q}(\boldsymbol{\psi}, M) - \mathcal{Q}(\boldsymbol{\psi}, M^*)\right| = A_1 + A_2 + A_3.
\end{aligned}
$$

Since $M > M^*$, Assumptions 5(1) and 5(5) guarantee that $A_1 + A_2 \leq 2ck_{\psi,M}^2$. $A_3 = 0$ as, by definition, $\mathbb{E}\left[\mathcal{Q}(\cdot, M^*)\right]$ is minimum over all values of M.

Using the union bound on A we have

$$
P(A) \leq 2c \sum_{M=M^*+1}^{\overline{M}} \frac{k_{\psi,M}^2}{\lambda_T(M) - \lambda_T(M^*)} \to 0
$$

as $T \to \infty$.

Now assume that $M < M^*$ and write for $\boldsymbol{\psi} \in \boldsymbol{\Psi}_M^*$

$$
\begin{aligned}
\frac{1}{T}\left[\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M)\right] = & \frac{1}{T}\left\{\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathbb{E}\left[\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*)\right]\right\} - \frac{1}{T}\left\{\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) - \mathbb{E}\left[\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M)\right]\right\} \\
& + \frac{1}{T}\mathbb{E}\left[\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\boldsymbol{\psi}, M^*)\right] - \frac{1}{T}\mathbb{E}\left[\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) - \mathcal{Q}(\boldsymbol{\psi}, M)\right] \\
& - \frac{1}{T}\mathbb{E}\left[\mathcal{Q}(\boldsymbol{\psi}, M) - \mathcal{Q}(\boldsymbol{\psi}, M^*)\right] \\
= & \, o_p(1) - K_M,
\end{aligned}
$$

where $K_M = \frac{1}{T}\mathbb{E}\left[\mathcal{Q}(\boldsymbol{\psi}, M) - \mathcal{Q}(\boldsymbol{\psi}, M^*)\right]$.

The first line of the above expression is $o_p(1)$ by Assumption 3 and the law of large numbers, the second line is $o(1)$ by Assumption 5(5), and $K_M > 0$ by the definition of M^*. From Assumption 5(3) $\frac{1}{T}\left[\lambda_T(M^*) - \lambda_T(M)\right] \to 0$ as $T \to \infty$. Therefore, the set

$$
\{\mathcal{Q}(\widehat{\boldsymbol{\psi}}, M^*) - \mathcal{Q}(\widehat{\boldsymbol{\psi}}, M) > \lambda_T(M) - \lambda_T(M^*)\} \to \emptyset.
$$

Since $M^* < \infty$ by definition, it follows from the union bound that $P(B) \to 0$ and the theorem is proved. □

REFERENCES

Amemiya, T.: 1985, *Advanced Econometrics*, Harvard University Press, Cambridge.

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P.: 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P.: 2003, Modeling and forecasting realized volatility, *Econometrica* **71**, 579–625.

Andersen, T., Dobrev, D. and Schaumburg, E.: 2009, Jump-robust volatility estimation using nearest neighbor truncation, *Working paper no. 15533*, National Bureau of Economic Research (NBER).

Andrews, D. W. K. and McDermott, C. J.: 1995, Nonlinear econometric models with deterministically trending variables, *Review of Economic Studies* **62**, 343–360.

Baillie, R. and Kapetanios, G.: 2007, Testing for neglected nonlinearity in long-memory models, *Journal of Business and Economic Statistics* **25**, 447–461.

Baillie, R. and Kapetanios, G.: 2008, Nonlinear models for strongly dependent processes with financial applications, *Journal of Econometrics* **147**, 60–71.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N.: 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 Shephard, N.: 2002, Econometric analysis of realised volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society B* **64**, 253–280.

Billingsley, P.: 1999, *Convergence of Probability Measures*, second edn, Wiley, New York.

Black, F.: 1976, Studies in stock price volatility changes, *Proceedings of The American Statistical Association, Business and Economic Statistics Section* pp. 177–181.

Chen, Y., Härdle, W. and Pigorsch, U.: 2010, Localized realized volatility modeling, *Journal of the American Statistical Association* **105**, 1376–1393.

Corsi, F.: 2009, A simple long memory model of realized volatility, *Journal of Financial Econometrics* **7**, 174–196.

Corsi, F. and Renò, R.: 2012, Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling, *Journal of Business and Economic Statistics* . forthcoming.

Davidson, J.: 1994, *Stochastic Limit Theory*, Oxford University Press, Oxford.

Diebold, F. and Inoue, A.: 2001, Long memory and regime switching, *Journal of Econometrics* **105**, 131–159.

Elck, P. and Makus, L.: 2004, A long range dependent model with nonlinear innovation for simulating daily river flows, *Natural Hazards and Earth System Sciences* **4**, 227–283.

Franses, P. and Ooms, M.: 2001, A seasonal periodic long memory model for monthly riverflows, *Environmental Modelling and Software* **16**, 559–569.

Giacomini, R. and White, H.: 2006, Tests of conditional predictive ability, *Econometrica* **74**, 1545–1578.

Granger, C. and Hyung, N.: 2004, Occasional structural breaks and long memory with an application to the s&p500 absolute stock returns, *Journal of Empirical Finance* **3**, 399–421.

Hamilton, J. D. and Susmel, R.: 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* **64**, 307–333.

Hillebrand, E.: 2005, Neglecting parameter changes in garch models, *Journal of Econometrics* **129**, 121–138.

Hillebrand, E.: 2006, Overlaying time scales in financial volatility data, *in* T. Fomby and D. Terrell (eds), *Advances in Econometrics Vol. 20B: Econometric Analysis of Financial and Economic Time Series*, Elsevier, pp. 153–178.

Lewis, P. and Ray, B.: 1997, Modeling long-range dependence, nonlinearity, and periodic phenomena in sea surface temperatures using TMARS, *Journal of the American Statistical Association* **92**, 881–893.

Ling, S. and McAleer, M.: 2003, Asymptotic theory for a vector ARMA-GARCH model, *Econometric Theory* **19**, 280–310.

Liu, J., Li, W. K. and Li, C. W.: 1997, On a threshold autoregression with conditional heteroscedastic variances, *Journal of Statistical Planning and Inference* **62**(2), 279–300.

Lobato, I. N. and Savin, N. E.: 1998, Real and spurious long-memory properties of stock-market data, *Journal of Business and Economic Statistics* **16**, 261–268.

Martens, M., van Dijk, D. and de Pooter, M.: 2009, Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity, *International Journal of Forecasting* **27**, 282–303.

McAleer, M. and Medeiros, M.: 2008, A multiple regime smooth transition heterogenous autoregressive model for long memory and asymmetries, *Journal of Econometrics* **147**, 104–119.

Medeiros, M. C. and Veiga, A.: 2005, A flexible coefficient smooth transition time series model, *IEEE Transactions on Neural Networks* **16**, 97–113.

Medeiros, M. C. and Veiga, A.: 2009, Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model, *Econometric Theory* **25**, 117–161.

Mendes, E.: 2012a, Consistently estimating the order of a finite mixture-of-experts, *Technical report*, Department of Statistics, Northwestern University.

Mendes, E.: 2012b, Model selection in non-identifiable nested classes, *Technical report*, Department of Statistics, Northwestern University.

Mikosch, T. and Starica, C.: 2004, Nonstationarities in financial time series, the long-range dependence, and the igarch effects, *Review of Economics and Statistics* **86**, 378–390.

Rech, G., Teräsvirta, T. and Tschernig, R.: 2001, A simple variable selection technique for nonlinear models, *Communications in Statistics, Theory and Methods* **30**.

Saikkonen, P. and Choi, I.: 2004, Cointegrating smooth transition regressions, *Econometric Theory* **20**, 301–340.

Scharth, M. and Medeiros, M.: 2009, Asymmetric effects and long memory in the volatility of dow jones stocks, *International Journal of Forecasting* **27**, 304–327.

Strikholm, B. and Teräsvirta, T.: 2006, A sequential procedure for determining the number of regimes in a threshold autoregressive model, *Econometrics Journal* **9**, 472–491.

Teräsvirta, T.: 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

van Dijk, D., Franses, P. and Paap, R.: 2002, A nonlinear long memory model, with an application to US unemployment, *Journal of Econometrics* **110**, 135–165.

Wooldridge, J. M.: 1990, A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17–43.

Wooldridge, J. M.: 1991, On the application of robust, regression-based diagnostics to models of conditional means and conditional variances, *Journal of Econometrics* **47**, 5–46.

Zhang, L., Mykland, P. and Aït-Sahalia, Y.: 2005, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100**, 1394–1411.

Zhang, Q., Lutchen, K. and Suki, B.: 1999, A frequency domain approach to nonlinear and structure identification of long memory systems: Application to lung mechanics, *Annals of Biomedical Engineering* **27**, 1–13.

TABLE 4. FORECASTING RESULTS BHLS (60 SEC): ONE-STEP-AHEAD.

The table displays the ratio of the root mean squared errors (RMSE) for different models. “ARFIMA” is a standard linear long memory model. “STARFIMA I” is a nonlinear long memory model with lagged returns as transition variable. “STARFIMA II” uses time as transition variable. “Ratio I” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the linear HAR-RV model. “Ratio II” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear HAR-RV model. “GW I” is the p-value of the Giacomini and White test of equal predictive ability when the benchmark is the linear HAR-RV model. “GW II” is the p-value of the Giacomini and White test when the benchmark is the nonlinear HAR-RV model.

	ARFIMA				STARFIMA I				STARFIMA II			
	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II
AA	0.995	1.019	0.191	0.084	0.988	1.012	0.039	0.186	1.013	1.038	0.053	0.005
AXP	1.011	1.015	0.073	0.223	1.005	1.009	0.234	0.316	1.015	1.019	0.033	0.157
BA	0.998	1.011	0.306	0.138	0.997	1.010	0.272	0.156	1.009	1.022	0.101	0.020
BAC	1.015	1.009	0.021	0.345	1.014	1.009	0.025	0.354	1.029	1.024	0.002	0.169
CAT	1.001	1.018	0.344	0.055	1.003	1.020	0.211	0.041	1.001	1.018	0.344	0.055
CSCO	1.000	1.021	0.477	0.009	1.005	1.026	0.166	0.004	1.001	1.022	0.439	0.010
CVX	0.998	1.027	0.316	0.085	0.996	1.025	0.206	0.104	1.002	1.031	0.422	0.047
DD	0.999	1.016	0.429	0.088	1.000	1.016	0.464	0.080	1.034	1.051	0.000	0.000
DIS	0.999	1.021	0.440	0.068	0.999	1.021	0.440	0.068	1.068	1.091	0.000	0.000
GE	0.997	1.037	0.345	0.028	0.997	1.037	0.345	0.028	1.095	1.139	0.000	0.000
HD	0.995	1.019	0.150	0.076	0.999	1.023	0.414	0.043	0.998	1.022	0.354	0.038
HPQ	1.002	1.026	0.342	0.001	1.001	1.025	0.388	0.001	1.000	1.024	0.480	0.002
IBM	0.999	1.027	0.421	0.006	0.996	1.024	0.241	0.013	1.028	1.057	0.001	0.000
INTC	1.001	1.013	0.379	0.064	1.001	1.013	0.379	0.064	1.011	1.023	0.030	0.008
JNJ	1.003	1.031	0.287	0.068	1.002	1.030	0.351	0.075	1.016	1.044	0.007	0.018
JPM	1.012	1.045	0.032	0.007	1.010	1.043	0.063	0.009	1.051	1.086	0.000	0.000
KFT	0.996	1.001	0.169	0.446	0.998	1.003	0.357	0.313	1.020	1.025	0.004	0.003
KO	1.000	1.015	0.482	0.089	0.999	1.013	0.392	0.111	1.004	1.019	0.284	0.057
MCD	0.994	1.004	0.109	0.270	0.994	1.004	0.109	0.270	0.994	1.004	0.109	0.270
MMM	0.991	0.997	0.010	0.367	0.990	0.996	0.006	0.336	1.004	1.009	0.310	0.180
MO	0.996	0.998	0.163	0.395	0.995	0.998	0.126	0.355	1.003	1.005	0.298	0.224
MRK	0.993	1.002	0.004	0.364	0.993	1.002	0.004	0.364	0.993	1.002	0.004	0.364
MSFT	1.000	1.015	0.500	0.068	1.001	1.016	0.442	0.057	1.122	1.139	0.000	0.000
PFE	0.991	1.000	0.009	0.472	0.990	0.999	0.005	0.448	0.992	1.001	0.071	0.469
PG	0.993	1.014	0.041	0.152	0.993	1.014	0.041	0.152	1.107	1.130	0.000	0.000
T	0.991	1.004	0.022	0.309	0.991	1.004	0.022	0.309	0.991	1.004	0.022	0.309
UTX	0.999	1.037	0.392	0.007	1.004	1.042	0.249	0.002	1.015	1.053	0.036	0.000
VZ	0.994	1.018	0.101	0.066	0.994	1.018	0.101	0.066	0.995	1.019	0.181	0.054
WMT	0.995	1.018	0.139	0.053	0.994	1.017	0.113	0.052	1.002	1.026	0.362	0.010
XOM	1.000	1.022	0.489	0.138	0.991	1.013	0.099	0.258	1.016	1.038	0.024	0.028

(E. Hillebrand) CREATES, DEPARTMENT OF ECONOMICS AND BUSINESS, AARHUS UNIVERSITY, DENMARK.

E-mail address: ehillebrand@creates.au.dk

(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

TABLE 5. FORECASTING RESULTS BHLS (60 SEC): FIVE- AND TEN-STEPS-AHEAD.

The table displays the ratio of the root mean squared errors (RMSE) for different models. “ARFIMA” is a standard linear long memory model. “STARFIMA I” is a nonlinear long memory model with lagged returns as transition variable. “STARFIMA II” uses time as transition variable. “Ratio I” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the linear HAR-RV model. “Ratio II” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear HAR-RV model. “GW I” is the p-value of the Giacomini and White test of equal predictive ability when the benchmark is the linear HAR-RV model. “GW II” is the p-value of the Giacomini and White test when the benchmark is the nonlinear HAR-RV model.

Five-Steps-Ahead

	ARFIMA				STARFIMA I				STARFIMA II			
	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II
AA	0.973	0.977	0.001	0.014	0.978	0.983	0.004	0.069	0.986	0.990	0.050	0.239
AXP	0.999	0.998	0.467	0.452	1.001	1.001	0.466	0.483	1.006	1.005	0.366	0.384
BA	0.974	0.966	0.000	0.000	0.976	0.967	0.001	0.000	0.978	0.970	0.002	0.001
BAC	0.992	0.994	0.321	0.388	0.999	1.001	0.478	0.487	1.009	1.011	0.289	0.314
CAT	0.971	0.978	0.001	0.029	0.971	0.978	0.001	0.029	0.971	0.978	0.001	0.029
CSCO	0.962	0.949	0.000	0.000	0.962	0.949	0.000	0.000	0.963	0.950	0.000	0.000
CVX	0.949	0.947	0.001	0.002	0.953	0.951	0.000	0.002	0.958	0.956	0.001	0.002
DD	0.969	0.962	0.000	0.001	0.976	0.969	0.000	0.006	0.987	0.980	0.033	0.059
DIS	0.967	0.957	0.000	0.000	0.976	0.966	0.006	0.002	0.999	0.988	0.446	0.191
GE	0.976	0.988	0.040	0.198	0.984	0.997	0.119	0.415	0.998	1.011	0.430	0.226
HD	0.970	0.982	0.001	0.054	0.972	0.983	0.001	0.076	0.974	0.986	0.003	0.124
HPQ	0.973	0.954	0.000	0.000	0.975	0.957	0.001	0.000	0.978	0.960	0.007	0.000
IBM	0.982	0.984	0.032	0.047	0.984	0.986	0.046	0.070	0.988	0.990	0.105	0.150
INTC	0.976	0.969	0.001	0.001	0.977	0.969	0.001	0.001	0.978	0.971	0.002	0.002
JNJ	0.984	0.984	0.084	0.054	0.985	0.985	0.062	0.046	0.992	0.992	0.215	0.212
JPM	0.984	0.970	0.117	0.001	0.997	0.982	0.398	0.043	1.016	1.002	0.143	0.441
KFT	0.987	0.964	0.083	0.051	0.991	0.969	0.216	0.082	1.011	0.988	0.232	0.309
KO	0.981	0.971	0.025	0.011	0.982	0.971	0.031	0.014	0.983	0.973	0.042	0.019
MCD	0.969	0.928	0.001	0.000	0.969	0.928	0.001	0.000	0.969	0.928	0.001	0.000
MMM	0.956	0.961	0.001	0.000	0.965	0.969	0.001	0.000	0.975	0.979	0.002	0.005
MO	0.975	0.971	0.001	0.002	0.976	0.972	0.001	0.002	0.977	0.974	0.002	0.004
MRK	0.977	0.971	0.016	0.010	0.977	0.971	0.016	0.010	0.977	0.971	0.016	0.010
MSFT	0.967	0.967	0.000	0.002	0.978	0.978	0.011	0.025	0.996	0.995	0.334	0.355
PFE	0.961	0.909	0.000	0.000	0.961	0.909	0.000	0.000	0.962	0.910	0.000	0.000
PG	0.958	0.929	0.000	0.000	0.965	0.936	0.000	0.000	0.978	0.948	0.004	0.000
T	0.969	0.959	0.000	0.001	0.969	0.959	0.000	0.001	0.969	0.959	0.000	0.001
UTX	0.975	0.977	0.000	0.011	0.978	0.980	0.001	0.029	0.982	0.985	0.009	0.096
VZ	0.970	0.970	0.001	0.002	0.971	0.970	0.001	0.002	0.971	0.971	0.001	0.002
WMT	0.967	0.967	0.000	0.000	0.968	0.968	0.000	0.000	0.971	0.970	0.001	0.001
XOM	0.965	0.956	0.000	0.000	0.971	0.962	0.000	0.000	0.979	0.969	0.003	0.000

Ten-Steps-Ahead

	ARFIMA				STARFIMA I				STARFIMA II			
	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II
AA	0.983	0.971	0.121	0.005	0.987	0.974	0.157	0.013	0.992	0.979	0.252	0.048
AXP	1.008	0.984	0.379	0.204	1.012	0.988	0.333	0.256	1.017	0.992	0.273	0.343
BA	0.979	0.970	0.034	0.028	0.980	0.971	0.040	0.032	0.982	0.973	0.054	0.039
BAC	1.004	1.009	0.439	0.379	1.010	1.016	0.352	0.305	1.020	1.025	0.236	0.207
CAT	0.978	0.978	0.048	0.068	0.978	0.978	0.048	0.068	0.978	0.978	0.048	0.068
CSCO	0.964	0.957	0.012	0.039	0.964	0.958	0.012	0.039	0.965	0.958	0.012	0.040
CVX	0.934	0.908	0.024	0.017	0.938	0.911	0.023	0.016	0.942	0.915	0.024	0.017
DD	0.967	0.954	0.003	0.003	0.972	0.960	0.004	0.006	0.980	0.967	0.016	0.018
DIS	0.976	0.961	0.050	0.008	0.990	0.975	0.257	0.056	1.026	1.010	0.073	0.288
GE	0.978	0.974	0.164	0.046	0.984	0.981	0.244	0.091	0.994	0.990	0.391	0.239
HD	0.975	0.972	0.038	0.035	0.976	0.974	0.048	0.043	0.979	0.976	0.068	0.061
HPQ	0.974	0.966	0.010	0.001	0.977	0.969	0.017	0.004	0.980	0.972	0.037	0.012
IBM	0.992	0.986	0.293	0.156	0.993	0.987	0.319	0.176	0.996	0.990	0.385	0.228
INTC	0.985	0.989	0.096	0.181	0.985	0.989	0.095	0.178	0.986	0.990	0.105	0.191
JNJ	1.002	0.997	0.455	0.427	0.997	0.993	0.438	0.293	1.003	0.998	0.424	0.452
JPM	0.989	0.969	0.292	0.023	1.005	0.985	0.409	0.163	1.029	1.009	0.137	0.341
KFT	1.004	0.978	0.411	0.196	1.007	0.982	0.358	0.245	1.034	1.008	0.089	0.402
KO	0.992	0.989	0.290	0.203	0.992	0.989	0.291	0.205	0.993	0.989	0.298	0.212
MCD	0.978	0.968	0.025	0.012	0.978	0.968	0.025	0.012	0.978	0.968	0.025	0.012
MMM	0.959	0.955	0.049	0.034	0.966	0.962	0.051	0.035	0.974	0.970	0.063	0.045
MO	0.972	0.966	0.006	0.006	0.973	0.967	0.007	0.006	0.974	0.969	0.010	0.008
MRK	0.978	0.984	0.049	0.084	0.978	0.984	0.049	0.084	0.978	0.984	0.049	0.084
MSFT	0.975	0.980	0.030	0.062	0.981	0.986	0.059	0.117	0.991	0.996	0.231	0.362
PFE	0.961	0.937	0.002	0.002	0.962	0.938	0.002	0.002	0.964	0.940	0.005	0.003
PG	0.953	0.922	0.028	0.001	0.958	0.926	0.020	0.001	0.965	0.934	0.019	0.001
T	0.965	0.963	0.004	0.009	0.965	0.963	0.004	0.009	0.965	0.963	0.004	0.009
UTX	0.982	0.968	0.065	0.034	0.986	0.972	0.085	0.042	0.991	0.977	0.178	0.071
VZ	0.972	0.968	0.036	0.013	0.972	0.968	0.036	0.013	0.972	0.968	0.036	0.013
WMT	0.967	0.969	0.004	0.004	0.969	0.971	0.005	0.006	0.971	0.973	0.010	0.011
XOM	0.961	0.943	0.006	0.013	0.967	0.950	0.008	0.014	0.975	0.957	0.017	0.019

SUPPLEMENT TO NONLINEARITY, BREAKS, AND LONG-RANGE DEPENDENCE IN TIME-SERIES MODELS

ERIC HILLEBRAND AND MARCELO C. MEDEIROS

1. Introduction

In this supplement we present additional results both for the simulations and the empirical application. With respect to the simulations, we present results concerning the parameter estimates when the sample size consists of 500 observations. We also report, for different values of the parameter C, the performance of the sequence of LM tests in determining the number of nonlinear terms in the model. Concerning the empirical example, we report descriptive statistics as well as forecasting results when RVmed is considered.

2. Additional Simulation Results

TABLE 1. PARAMETER ESTIMATES UNDER CORRECT SPECIFICATION: 500 OBSERVATIONS.

The table reports average bias and mean-squared error (MSE) of parameter estimates from 1000 simulations of the models listed in the paper with $T = 500$. The regime and lag structure M and p are assumed to be known in the estimation.

	Short-Memory Models																	
	Linear						Nonlinear I						Nonlinear II					
	Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ϕ_{10}	0.04	0.01	0.02	0.01	0.05	0.03	0.00	0.01	0.10	0.04	0.09	0.04	0.02	0.01	0.08	0.03	-0.07	0.03
ϕ_{11}	–	–	–	–	–	–	0.00	0.01	-0.10	0.05	-0.02	0.02	-0.06	0.01	0.01	0.01	0.01	0.02
ϕ_{12}	–	–	–	–	–	–	0.04	0.01	0.40	0.10	0.07	0.05	0.01	0.02	0.04	0.02	0.21	0.12
ϕ_{20}	-0.05	0.01	0.02	0.01	0.01	0.02	-0.06	0.02	-0.06	0.04	0.06	0.05	0.06	0.03	-0.02	0.02	-0.05	0.08
ϕ_{21}	–	–	–	–	–	–	0.04	0.02	0.10	0.05	0.15	0.08	0.02	0.01	-0.10	0.06	-0.05	0.04
ϕ_{22}	–	–	–	–	–	–	0.10	0.08	-0.03	0.02	0.14	0.03	0.01	0.01	-0.14	0.10	-0.02	0.02
γ_1	–	–	–	–	–	–	-5.13	904.12	-5.00	150.12	-4.74	65.42	-53.82	56.35	-50.31	78.53	-57.32	145.84
γ_2	–	–	–	–	–	–	-15.10	803.13	2.38	79.24	-13.63	87.34	-65.73	73.34	-65.42	98.64	-137.84	746.31
c_1	–	–	–	–	–	–	0.02	0.02	0.06	0.02	-0.04	0.02	0.02	0.01	0.03	0.02	0.10	0.09
c_2	–	–	–	–	–	–	-0.03	0.01	0.10	0.09	0.23	0.13	0.01	0.01	0.10	0.08	0.09	0.12
d	0.01	0.01	0.00	0.01	0.00	0.01	0.09	0.05	0.03	0.02	0.10	0.15	0.10	0.05	0.10	0.12	0.01	0.03
σ_u	0.00	0.00	0.00	0.01	0.20	0.24	0.00	0.01	0.10	0.34	0.20	0.13	0.00	0.00	0.01	0.01	0.20	0.34

	Long-Memory Models																	
	Linear						Nonlinear I						Nonlinear II					
	Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH		Gaussian		Fat-Tailed		GARCH	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ϕ_{10}	0.02	0.02	0.02	0.02	0.06	0.02	0.01	0.01	0.11	0.04	0.07	0.03	0.01	0.02	0.08	0.03	0.04	0.04
ϕ_{11}	–	–	–	–	–	–	0.01	0.01	0.14	0.05	0.01	0.01	-0.03	0.02	0.01	0.01	0.02	0.05
ϕ_{12}	–	–	–	–	–	–	0.04	0.01	0.33	0.10	0.04	0.04	0.01	0.02	0.04	0.02	0.18	0.15
ϕ_{20}	0.07	0.02	0.02	0.02	0.02	0.01	0.05	0.01	0.06	0.04	0.04	0.05	0.01	0.04	0.02	0.02	0.06	0.05
ϕ_{21}	–	–	–	–	–	–	-0.03	0.01	0.18	0.05	0.18	0.03	0.07	0.02	0.10	0.06	0.04	0.04
ϕ_{22}	–	–	–	–	–	–	0.08	0.09	0.03	0.02	0.14	0.09	0.04	0.02	0.14	0.10	0.05	0.03
γ_1	–	–	–	–	–	–	-8.45	944.93	6.94	463.36	53.13	98.53	89.13	95.36	50.31	78.53	84.64	184.84
γ_2	–	–	–	–	–	–	-19.76	933.34	5.46	162.84	15.25	92.35	84.94	97.84	65.42	98.64	97.42	137.36
c_1	–	–	–	–	–	–	0.01	0.01	0.06	0.01	0.04	0.02	0.03	0.02	0.03	0.02	0.11	0.10
c_2	–	–	–	–	–	–	0.02	0.02	0.10	0.07	0.23	0.13	0.02	0.02	0.10	0.08	0.19	0.18
d	0.02	0.01	0.01	0.02	0.02	0.02	0.10	0.05	0.03	0.01	0.10	0.15	0.13	0.08	0.10	0.12	0.02	0.04
σ_u	0.00	0.00	0.00	0.01	0.12	0.18	0.00	0.01	0.10	0.24	0.20	0.13	0.01	0.01	0.01	0.01	0.21	0.47

TABLE 2. FREQUENCY OF CORRECT SPECIFICATION: LM TESTS AND DIFFERENT VALUES OF C.

The table reports the proportion of correctly determined numbers of regimes in 1000 simulations of the models listed in the paper for different values of the significance-level adjusting parameter C. In the simulations the lag structure is assumed to be known and fixed at $p = 2$.

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
Short-Memory Models: 500 observations									
$C = 1/2$	1.00	1.00	1.00	0.16	0.20	0.06	0.42	0.50	0.52
$C = 2/3$	1.00	1.00	1.00	0.20	0.24	0.10	0.50	0.62	0.65
$C = 1/3$	1.00	1.00	1.00	0.16	0.20	0.06	0.42	0.50	0.52
Short-Memory Models: 1000 observations									
$C = 1/2$	1.00	1.00	1.00	0.37	0.42	0.21	0.60	0.73	0.69
$C = 2/3$	1.00	1.00	1.00	0.48	0.51	0.32	0.71	0.84	0.78
$C = 1/3$	1.00	1.00	1.00	0.37	0.42	0.21	0.60	0.73	0.69
Long-Memory Models: 500 observations									
$C = 1/2$	1.00	0.96	0.98	0.08	0.10	0.10	0.56	0.50	0.44
$C = 2/3$	1.00	0.90	0.91	0.16	0.21	0.24	0.64	0.61	0.58
$C = 1/3$	1.00	0.96	0.98	0.08	0.10	0.10	0.56	0.50	0.44
Long-Memory Models: 1000 observations									
$C = 1/2$	1.00	1.00	1.00	0.19	0.32	0.30	0.97	0.90	0.88
$C = 2/3$	1.00	1.00	1.00	0.30	0.48	0.51	1.00	1.00	0.99
$C = 1/3$	1.00	1.00	1.00	0.19	0.32	0.30	0.97	0.90	0.88

3. Additional Empirical Application Results

Table 3. Full Sample Results: Descriptive Statistics.

The table displays the sample moments of the realized volatility time series considered in this paper.

	Date Sample	Date Forecasts	Sample	BHLS (60 sec)				RVmed			
	Starts	Starts	Size	average	std. dev.	skewness	kurtosis	average	std. dev.	skewness	kurtosis
AA	01/03/2000	01/03/2006	2409	1.25	0.91	0.77	3.89	1.43	0.88	0.85	3.71
AXP	01/03/2000	01/03/2006	2409	0.79	1.33	0.20	2.47	0.96	1.25	0.28	2.48
BA	01/03/2000	01/03/2006	2409	0.78	0.88	0.37	2.98	0.97	0.87	0.35	2.72
BAC	01/03/2000	01/03/2006	2409	0.56	1.42	0.81	3.41	0.71	1.38	0.87	3.47
CAT	01/03/2000	01/03/2006	2409	0.86	0.86	0.71	3.64	0.99	0.87	0.71	3.34
CSCO	01/03/2000	01/03/2006	2298	1.23	0.98	0.40	2.75	1.40	0.97	0.54	3.00
CVX	10/10/2001	01/03/2006	1966	0.41	0.86	0.85	4.78	0.51	0.85	0.90	4.87
DD	01/03/2000	01/03/2006	2409	0.68	0.93	0.42	2.96	0.89	0.87	0.50	2.92
DIS	01/03/2000	01/03/2006	2409	0.78	0.97	0.31	2.88	1.15	0.93	0.29	2.42
GE	01/03/2000	01/03/2006	2409	0.54	1.17	0.49	3.06	0.81	1.10	0.53	2.92
HD	01/03/2000	01/03/2006	2409	0.92	0.94	0.34	2.95	1.12	0.89	0.40	2.80
HPQ	01/03/2000	01/03/2006	2409	1.12	1.00	0.22	2.79	1.32	0.91	0.38	2.91
IBM	01/03/2000	01/03/2006	2409	0.42	0.99	0.48	3.02	0.59	0.91	0.59	3.10
INTC	01/03/2000	01/03/2006	2409	1.27	0.95	0.28	2.50	1.51	0.85	0.44	2.64
JNJ	01/03/2000	01/03/2006	2409	-0.04	0.97	0.33	3.16	0.20	0.87	0.42	3.08
JPM	01/03/2000	01/03/2006	2409	0.93	1.27	0.34	2.74	1.14	1.18	0.41	2.71
KFT	06/13/2001	01/03/2006	2045	0.14	0.80	0.46	3.79	0.22	0.75	0.78	4.05
KO	01/03/2000	01/03/2006	2409	0.17	0.92	0.41	3.25	0.42	0.85	0.53	2.99
MCD	01/03/2000	01/03/2006	2409	0.62	0.83	0.41	3.34	0.90	0.85	0.47	2.83
MMM	01/03/2000	01/03/2006	2409	0.35	0.88	0.59	3.52	0.47	0.80	0.73	3.76
MO	01/03/2000	01/03/2006	2409	0.37	0.94	0.46	3.39	0.66	1.00	0.73	3.20
MRK	01/03/2000	01/03/2006	2409	0.61	0.86	0.62	3.99	0.79	0.78	0.74	4.07
MSFT	01/03/2000	01/03/2006	2409	0.70	1.01	0.16	2.62	0.99	0.88	0.37	2.63
PFE	01/03/2000	01/03/2006	2409	0.53	0.88	0.40	3.32	0.90	0.81	0.56	2.87
PG	01/03/2000	01/03/2006	2409	0.09	0.92	0.64	3.74	0.32	0.85	0.67	3.34
T	01/03/2000	01/03/2006	2402	0.80	0.98	0.24	2.99	1.16	0.95	0.42	2.70
UTX	01/03/2000	01/03/2006	2409	0.57	0.93	0.51	3.22	0.70	0.83	0.68	3.51
VZ	01/03/2000	01/03/2006	2409	0.60	0.95	0.34	3.12	0.84	0.85	0.45	3.09
WMT	01/03/2000	01/03/2006	2409	0.50	0.93	0.47	2.92	0.69	0.88	0.54	2.72
XOM	01/03/2000	01/03/2006	2409	0.43	0.84	0.72	4.46	0.64	0.79	0.80	4.40

(E. Hillebrand) CREATES, Department of Economics and Business, Aarhus University, Denmark.

E-mail address: ehillebrand@creates.au.dk

(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

TABLE 4. FULL SAMPLE RESULTS: BIC.

The table displays the BIC for the nonlinear ARFIMA models when the transition variable is either the past return or the time period. Four different criteria are considered. LM is the sequence of robust LM tests. BIC and HQIC consists of the cases when either BIC or HQIC is used as a criterium to determine the number of regimes.

	Transition Variable: Past Return						Transition Variable: Time					
	BHLS (60 sec)			RVmed			BHLS (60 sec)			RVmed		
	LM	BIC	HQIC	LM	BIC	HQIC	LM	BIC	HQIC	LM	BIC	HQIC
AA	-1.524	-1.524	-1.521	-1.783	-1.783	-1.783	-1.524	-1.524	-1.524	-1.783	-1.783	-1.783
AXP	-1.350	-1.351	-1.351	-1.710	-1.710	-1.712	-1.350	-1.350	-1.350	-1.708	-1.710	-1.712
BA	-1.449	-1.452	-1.449	-1.788	-1.788	-1.785	-1.452	-1.452	-1.452	-1.788	-1.788	-1.788
BAC	-1.410	-1.410	-1.403	-1.616	-1.620	-1.615	-1.412	-1.412	-1.397	-1.619	-1.619	-1.607
CAT	-1.553	-1.553	-1.553	-1.785	-1.785	-1.785	-1.553	-1.553	-1.553	-1.785	-1.785	-1.782
CSCO	-1.585	-1.585	-1.585	-1.703	-1.703	-1.703	-1.581	-1.581	-1.581	-1.703	-1.703	-1.703
CVX	-1.595	-1.597	-1.597	-1.773	-1.776	-1.745	-1.594	-1.595	-1.594	-1.775	-1.775	-1.757
DD	-1.534	-1.534	-1.530	-1.871	-1.871	-1.869	-1.531	-1.534	-1.531	-1.871	-1.871	-1.871
DIS	-1.455	-1.455	-1.455	-1.774	-1.774	-1.774	-1.456	-1.456	-1.449	-1.770	-1.774	-1.770
GE	-1.370	-1.372	-1.372	-1.677	-1.681	-1.681	-1.372	-1.372	-1.368	-1.682	-1.682	-1.682
HD	-1.500	-1.501	-1.500	-1.801	-1.801	-1.801	-1.501	-1.501	-1.501	-1.797	-1.797	-1.797
HPQ	-1.299	-1.299	-1.299	-1.555	-1.555	-1.551	-1.293	-1.299	-1.299	-1.555	-1.555	-1.550
IBM	-1.520	-1.521	-1.520	-1.893	-1.894	-1.897	-1.514	-1.521	-1.521	-1.894	-1.894	-1.897
INTC	-1.678	-1.678	-1.678	-1.916	-1.916	-1.912	-1.678	-1.678	-1.678	-1.908	-1.916	-1.916
JNJ	-1.332	-1.332	-1.332	-1.826	-1.828	-1.828	-1.328	-1.331	-1.328	-1.824	-1.826	-1.824
JPM	-1.381	-1.387	-1.387	-1.721	-1.721	-1.711	-1.384	-1.384	-1.384	-1.717	-1.717	-1.704
KFT	-1.079	-1.091	-1.090	-1.267	-1.267	-1.266	-1.077	-1.091	-1.090	-1.260	-1.267	-1.259
KO	-1.526	-1.526	-1.526	-1.896	-1.896	-1.896	-1.526	-1.526	-1.526	-1.896	-1.896	-1.896
MCD	-1.311	-1.311	-1.310	-1.536	-1.536	-1.535	-1.311	-1.311	-1.310	-1.532	-1.536	-1.531
MMM	-1.445	-1.445	-1.445	-1.775	-1.780	-1.780	-1.438	-1.442	-1.438	-1.775	-1.775	-1.771
MO	-1.010	-1.010	-1.010	-1.273	-1.273	-1.273	-1.007	-1.010	-1.007	-1.272	-1.273	-1.272
MRK	-1.192	-1.192	-1.186	-1.507	-1.507	-1.504	-1.192	-1.192	-1.192	-1.507	-1.507	-1.507
MSFT	-1.499	-1.499	-1.494	-1.808	-1.808	-1.805	-1.494	-1.499	-1.494	-1.804	-1.808	-1.806
PFE	-1.345	-1.345	-1.345	-1.753	-1.753	-1.753	-1.345	-1.345	-1.345	-1.753	-1.753	-1.750
PG	-1.400	-1.400	-1.396	-1.822	-1.822	-1.822	-1.400	-1.400	-1.400	-1.822	-1.822	-1.822
T	-1.270	-1.270	-1.270	-1.528	-1.528	-1.525	-1.270	-1.270	-1.270	-1.528	-1.528	-1.525
UTX	-1.464	-1.464	-1.462	-1.852	-1.852	-1.849	-1.461	-1.464	-1.461	-1.852	-1.852	-1.848
VZ	-1.384	-1.385	-1.384	-1.770	-1.770	-1.770	-1.385	-1.385	-1.385	-1.766	-1.766	-1.766
WMT	-1.556	-1.558	-1.556	-1.876	-1.876	-1.875	-1.558	-1.558	-1.558	-1.869	-1.876	-1.876
XOM	-1.559	-1.565	-1.565	-1.834	-1.834	-1.829	-1.563	-1.563	-1.563	-1.829	-1.830	-1.818

TABLE 5. FORECASTING RESULTS RVMED: ONE-STEP-AHEAD.

The table displays the ratio of the root mean squared errors (RMSE) for different models. "ARFIMA" is a standard linear long memory model. "STARFIMA I" is a nonlinear long memory model with lagged returns as transition variable. "STARFIMA II" uses time as transition variable. "Ratio I" is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the linear HAR-RV model. "Ratio II" is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear HAR-RV model. "GW I" is the p-value of the Giacomini and White test of equal predictive ability when the benchmark is the linear HAR-RV model. "GW II" is the p-value of the Giacomini and White test when the benchmark is the nonlinear HAR-RV model.

	ARFIMA				STARFIMA I				STARFIMA II			
	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II
AA	0.994	1.018	0.159	0.132	0.993	1.017	0.171	0.156	1.019	1.043	0.020	0.004
AXP	1.006	1.009	0.215	0.329	1.005	1.008	0.254	0.348	1.026	1.029	0.007	0.074
BA	0.999	1.012	0.445	0.163	1.000	1.013	0.472	0.165	1.029	1.043	0.002	0.001
BAC	1.008	1.012	0.139	0.295	1.008	1.012	0.136	0.293	1.026	1.030	0.003	0.094
CAT	1.001	1.014	0.385	0.092	1.001	1.013	0.427	0.101	1.001	1.014	0.404	0.090
CSCO	0.994	1.014	0.130	0.077	0.996	1.016	0.241	0.057	1.010	1.030	0.125	0.003
CVX	0.995	1.017	0.182	0.201	0.995	1.017	0.180	0.201	0.997	1.019	0.313	0.174
DD	0.997	1.009	0.257	0.232	0.992	1.004	0.082	0.363	1.030	1.043	0.001	0.001
DIS	0.998	1.019	0.365	0.097	0.998	1.019	0.365	0.097	1.058	1.080	0.001	0.000
GE	1.000	1.039	0.483	0.043	1.000	1.038	0.497	0.046	1.067	1.108	0.000	0.000
HD	0.996	1.020	0.266	0.068	0.995	1.018	0.215	0.092	0.996	1.019	0.263	0.057
HPQ	1.001	1.025	0.436	0.002	1.000	1.024	0.482	0.002	1.008	1.032	0.162	0.001
IBM	1.001	1.027	0.412	0.011	0.999	1.024	0.433	0.019	1.026	1.052	0.023	0.001
INTC	1.000	1.019	0.496	0.036	1.001	1.019	0.456	0.030	1.069	1.089	0.000	0.000
JNJ	1.001	1.031	0.425	0.066	1.001	1.031	0.425	0.066	1.024	1.054	0.002	0.005
JPM	1.011	1.047	0.037	0.002	1.052	1.090	0.000	0.000	1.056	1.094	0.000	0.000
KFT	1.006	1.011	0.075	0.092	1.005	1.010	0.092	0.100	1.026	1.031	0.000	0.001
KO	0.995	1.006	0.203	0.274	0.995	1.006	0.203	0.274	0.995	1.006	0.203	0.274
MCD	0.999	1.009	0.432	0.096	0.999	1.009	0.427	0.098	1.052	1.063	0.010	0.003
MMM	0.988	0.994	0.003	0.285	0.988	0.995	0.004	0.299	0.993	0.999	0.184	0.476
MO	0.996	0.999	0.215	0.442	0.994	0.996	0.098	0.305	0.998	1.001	0.348	0.465
MRK	0.992	1.000	0.001	0.496	0.992	1.000	0.001	0.496	1.003	1.011	0.289	0.062
MSFT	1.002	1.015	0.391	0.077	1.004	1.017	0.284	0.058	1.142	1.158	0.000	0.000
PFE	0.992	0.999	0.046	0.456	0.992	0.999	0.046	0.456	1.018	1.026	0.013	0.006
PG	0.990	1.006	0.014	0.332	0.990	1.006	0.014	0.332	0.990	1.006	0.014	0.332
T	0.991	1.009	0.046	0.184	0.991	1.009	0.046	0.184	1.091	1.111	0.000	0.000
UTX	0.997	1.031	0.303	0.022	0.997	1.031	0.303	0.022	1.002	1.036	0.395	0.009
VZ	0.992	1.017	0.095	0.102	0.993	1.018	0.128	0.088	0.991	1.016	0.087	0.116
WMT	0.994	1.015	0.124	0.109	0.991	1.013	0.069	0.132	1.077	1.100	0.000	0.000
XOM	0.998	1.018	0.366	0.190	0.991	1.011	0.153	0.313	1.032	1.053	0.001	0.006

TABLE 6. FORECASTING RESULTS RVMED: FIVE- AND TEN-STEPS-AHEAD.

The table displays the ratio of the root mean squared errors (RMSE) for different models. “ARFIMA” is a standard linear long memory model. “STARFIMA I” is a nonlinear long memory model with lagged returns as transition variable. “STARFIMA II” uses time as transition variable. “Ratio I” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the linear HAR-RV model. “Ratio II” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear HAR-RV model. “GW I” is the p-value of the Giacomini and White test of equal predictive ability when the benchmark is the linear HAR-RV model. “GW II” is the p-value of the Giacomini and White test when the benchmark is the nonlinear HAR-RV model.

Five-Steps-Ahead

	ARFIMA				STARFIMA I				STARFIMA II			
	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II
AA	0.967	0.974	0.001	0.007	0.974	0.981	0.004	0.063	0.984	0.991	0.057	0.272
AXP	0.987	0.984	0.213	0.193	0.991	0.989	0.283	0.257	1.004	1.002	0.384	0.455
BA	0.971	0.969	0.003	0.009	0.976	0.974	0.015	0.033	0.983	0.981	0.085	0.115
BAC	0.984	0.993	0.160	0.378	0.989	0.999	0.247	0.483	1.000	1.010	0.487	0.312
CAT	0.967	0.975	0.001	0.029	0.967	0.975	0.000	0.027	0.967	0.975	0.000	0.026
CSCO	0.955	0.939	0.000	0.000	0.960	0.944	0.000	0.000	0.966	0.950	0.000	0.000
CVX	0.952	0.947	0.000	0.000	0.952	0.947	0.000	0.000	0.953	0.948	0.000	0.000
DD	0.967	0.968	0.000	0.002	0.976	0.977	0.001	0.033	0.988	0.989	0.102	0.225
DIS	0.972	0.946	0.006	0.000	0.979	0.953	0.030	0.002	0.992	0.966	0.253	0.023
GE	0.979	0.990	0.085	0.237	0.985	0.996	0.163	0.391	0.995	1.006	0.384	0.347
HD	0.969	0.968	0.000	0.003	0.971	0.971	0.001	0.006	0.975	0.974	0.003	0.018
HPQ	0.971	0.958	0.001	0.000	0.975	0.962	0.004	0.001	0.980	0.966	0.023	0.007
IBM	0.971	0.963	0.003	0.001	0.974	0.966	0.005	0.002	0.980	0.972	0.024	0.012
INTC	0.965	0.943	0.000	0.000	0.973	0.951	0.001	0.000	0.984	0.962	0.048	0.012
JNJ	0.976	0.965	0.030	0.001	0.978	0.966	0.037	0.002	0.981	0.969	0.062	0.005
JPM	0.982	0.966	0.070	0.004	0.996	0.979	0.358	0.048	1.015	0.998	0.110	0.444
KFT	0.983	0.981	0.025	0.081	0.991	0.989	0.126	0.203	1.015	1.013	0.132	0.225
KO	0.972	0.939	0.002	0.000	0.972	0.939	0.002	0.000	0.972	0.939	0.002	0.000
MCD	0.979	0.926	0.017	0.000	0.982	0.929	0.039	0.000	0.991	0.937	0.191	0.001
MMM	0.950	0.957	0.001	0.000	0.956	0.963	0.001	0.000	0.963	0.970	0.001	0.001
MO	0.975	0.963	0.000	0.000	0.976	0.964	0.000	0.000	0.978	0.966	0.001	0.001
MRK	0.967	0.955	0.001	0.001	0.970	0.958	0.004	0.002	0.978	0.966	0.034	0.012
MSFT	0.973	0.965	0.004	0.005	0.981	0.973	0.027	0.022	0.997	0.989	0.376	0.207
PFE	0.959	0.913	0.000	0.002	0.966	0.919	0.000	0.005	0.975	0.928	0.003	0.015
PG	0.955	0.905	0.000	0.000	0.955	0.905	0.000	0.000	0.955	0.905	0.000	0.000
T	0.974	0.947	0.002	0.000	0.981	0.954	0.012	0.001	0.994	0.966	0.253	0.014
UTX	0.971	0.964	0.000	0.000	0.973	0.965	0.000	0.001	0.976	0.969	0.001	0.005
VZ	0.972	0.965	0.002	0.001	0.973	0.966	0.003	0.002	0.975	0.968	0.005	0.002
WMT	0.964	0.963	0.000	0.000	0.969	0.968	0.000	0.000	0.979	0.978	0.003	0.009
XOM	0.962	0.964	0.000	0.000	0.976	0.978	0.005	0.009	0.997	0.999	0.405	0.469

Ten-Steps-Ahead

	ARFIMA				STARFIMA I				STARFIMA II			
	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II	Ratio I	Ratio II	GW I	GW II
AA	0.983	0.977	0.143	0.025	0.987	0.982	0.199	0.054	0.994	0.988	0.327	0.157
AXP	1.004	0.984	0.440	0.192	1.010	0.990	0.338	0.285	1.028	1.008	0.125	0.335
BA	0.983	0.983	0.144	0.146	0.987	0.986	0.211	0.216	0.992	0.992	0.319	0.323
BAC	1.002	1.013	0.465	0.331	1.008	1.018	0.380	0.259	1.020	1.031	0.210	0.130
CAT	0.968	0.979	0.021	0.083	0.968	0.979	0.019	0.081	0.968	0.979	0.019	0.081
CSCO	0.952	0.939	0.007	0.012	0.957	0.943	0.006	0.012	0.962	0.948	0.006	0.013
CVX	0.941	0.922	0.040	0.036	0.941	0.923	0.042	0.037	0.942	0.923	0.044	0.038
DD	0.967	0.957	0.003	0.005	0.975	0.965	0.011	0.019	0.986	0.975	0.114	0.090
DIS	0.989	0.969	0.285	0.058	0.993	0.973	0.353	0.078	1.004	0.984	0.427	0.207
GE	0.981	0.983	0.215	0.141	0.985	0.987	0.262	0.197	0.990	0.993	0.349	0.326
HD	0.975	0.967	0.030	0.008	0.976	0.968	0.031	0.010	0.978	0.970	0.040	0.015
HPQ	0.984	0.981	0.098	0.042	0.986	0.983	0.138	0.079	0.990	0.987	0.212	0.149
IBM	0.981	0.979	0.086	0.046	0.982	0.980	0.094	0.055	0.987	0.985	0.169	0.121
INTC	0.973	0.976	0.015	0.058	0.980	0.983	0.038	0.152	0.989	0.992	0.162	0.332
JNJ	0.989	0.985	0.305	0.192	0.990	0.986	0.308	0.194	0.991	0.987	0.328	0.214
JPM	0.990	0.972	0.281	0.026	1.005	0.987	0.393	0.164	1.025	1.007	0.098	0.312
KFT	0.997	1.014	0.405	0.236	1.001	1.017	0.463	0.178	1.032	1.049	0.071	0.031
KO	0.983	0.972	0.086	0.047	0.983	0.972	0.086	0.047	0.983	0.972	0.086	0.047
MCD	0.988	0.956	0.186	0.026	0.989	0.957	0.216	0.031	0.995	0.963	0.369	0.054
MMM	0.947	0.946	0.045	0.033	0.952	0.951	0.046	0.035	0.959	0.958	0.049	0.040
MO	0.976	0.968	0.019	0.002	0.977	0.968	0.022	0.002	0.978	0.970	0.030	0.003
MRK	0.971	0.979	0.038	0.048	0.969	0.977	0.034	0.040	0.976	0.984	0.087	0.126
MSFT	0.985	0.988	0.156	0.197	0.988	0.991	0.175	0.236	0.995	0.998	0.342	0.440
PFE	0.962	0.976	0.003	0.280	0.969	0.983	0.007	0.347	0.978	0.992	0.045	0.430
PG	0.951	0.930	0.025	0.002	0.951	0.930	0.025	0.002	0.951	0.930	0.025	0.002
T	0.977	0.963	0.041	0.013	0.983	0.968	0.073	0.019	0.991	0.977	0.239	0.060
UTX	0.978	0.966	0.032	0.036	0.981	0.969	0.037	0.042	0.985	0.973	0.064	0.058
VZ	0.974	0.972	0.039	0.019	0.974	0.972	0.040	0.019	0.975	0.972	0.042	0.021
WMT	0.964	0.968	0.001	0.001	0.966	0.969	0.001	0.001	0.970	0.973	0.004	0.002
XOM	0.961	0.957	0.007	0.003	0.975	0.972	0.042	0.015	0.998	0.994	0.461	0.367

Research Papers 2012

2012-14: Niels Haldrup, Robinson Kruse, Timo Teräsvirta and Rasmus T. Varneskov: Unit roots, nonlinearities and structural breaks

2012-15: Matt P. Dziubinski and Stefano Grassi: Heterogeneous Computing in Economics: A Simplified Approach

2012-16: Anders Bredahl Kock and Laurent A.F. Callot: Oracle Inequalities for High Dimensional Vector Autoregressions

2012-17: Eric Hillebrand, Huiyu Huang, Tae-Hwy Lee and Canlin Li: Using the Yield Curve in Forecasting Output Growth and Inflation

2012-18: Eric Hillebrand and Tae-Hwy Lee: Stein-Rule Estimation and Generalized Shrinkage Methods for Forecasting Using Many Predictors

2012-19: Bent Jesper Christensen, Morten Ørregaard Nielsen and Jie Zhu: The impact of financial crises on the risk-return tradeoff and the leverage effect

2012-20: Hendrik Kaufmann, Robinson Kruse and Philipp Sibbertsen: On tests for linearity against STAR models with deterministic trends

2012-21: Andrey Launov, Olaf Posch and Klaus Wälde: On the estimation of the volatility-growth link

2012-22: Peter O. Christensen and Zhenjiang Qin: Information and Heterogeneous Beliefs: Cost of Capital, Trading Volume, and Investor Welfare

2012-23: Zhenjiang Qin: Heterogeneous Beliefs, Public Information, and Option Markets

2012-24: Zhenjiang Qin: Continuous Trading Dynamically Effectively Complete Market with Heterogeneous Beliefs

2012-25: Heejoon Han and Dennis Kristensen: Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Non-Stationary Covariates

2012-26: Lei Pan, Olaf Posch and Michel van der Wel: Measuring Convergence using Dynamic Equilibrium Models: Evidence from Chinese Provinces

2012-27: Lasse Bork and Stig V. Møller: Housing price forecastability: A factor analysis

2012-28: Johannes Tang Kristensen: Factor-Based Forecasting in the Presence of Outliers: Are Factors Better Selected and Estimated by the Median than by The Mean?

2012-29: Anders Rahbek and Heino Bohn Nielsen: Unit Root Vector Auto-regression with volatility Induced Stationarity

2012-30: Eric Hillebrand and Marcelo C. Medeiros: Nonlinearity, Breaks, and Long-Range Dependence in Time-Series Models