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 time series models and apply our modeling framework to daily realized measures of integrated variance. 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 in financial volatility. An out-of-sample analysis shows that modeling these effects can improve forecast performance.

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

JEL CODES: C22.

1. INTRODUCTION

This paper studies a time series model that displays long memory and nonlinear behavior in the short-memory component. Long memory on the one hand and nonlinear effects on the other, such as parameter changes in time or parameter dependence on a state variable, have long been recognized as confounding effects. Granger and Teräsvirta (2001), Granger and Hyung (2004), and Diebold and Inoue (2001) show that short-memory models with occasional breaks induce long-memory characteristics. Hsu (2001) shows that the confounding also works in the other direction: Long memory can lead to the spurious detection of change-points. In financial volatility data, other common nonlinear influences are the leverage and feedback effects (Black 1976), whereby volatility depends on past returns.

In this study, nonlinearity is modeled as a smooth transition function that depends on a state variable. We focus particularly on time as a state variable, since smooth parameter changes in time pose challenges to asymptotic theory. We show consistency and asymptotic normality of the nonlinear least-squares estimator in a triangular-array approach following Andrews and McDermott (1995) and Saikkonen and Choi (2004). The number of nonlinear terms and the lag order of the short memory components are determined by a sequence of Lagrange multiplier tests following Teräsvirta (1994) and an information criteria search, which we show to be consistent. A simulation study illustrates our results. We apply the model to logarithmic realized measures of integrated variance for the 30 stocks of the Dow Jones Industrial Average (DJIA) from 2000 to 2009. Across all stocks, we find evidence of nonlinear dependence of parameters both on time and on past returns. For the majority of stocks, at least one parameter transition occurred in the vicinity of the 2007/2008 crisis. Accounting for the nonlinear terms in a forecasting exercise yields performance gains in multi-period ahead predictions compared to the HAR-RV model of Corsi (2009) and its nonlinear extension in Corsi and Renò (2012).

Other models that combine long memory with nonlinear features have been proposed, for example, in van Dijk et al. (2002) for unemployment data, and in Martens et al. (2009) for daily S&P500 volatility data. The articles most closely related to our study are Baillie and Kapetanios (2007, 2008) and McAleer and Medeiros (2008). In distinction to Baillie and Kapetanios (2007, 2008), we explicitly consider time as a state variable in the nonlinear transition function and develop triangular-array asymptotic estimation theory for this case. Further, in Baillie and Kapetanios (2008) a single nonlinear transition function is considered for daily exchange rate data, whereas we consider the determination of multiple nonlinear terms in realized measures of integrated variance obtained from high-frequency stock data, and we conduct a pseudo-out-of-sample forecasting exercise. McAleer and Medeiros (2008) consider a heterogeneous autoregressive (HAR) approximation to long memory applied to 16 stocks of the DJIA from 1994 to

2003, whereas this study considers fractional differentiation in the sense of Granger and Joyeux (1980), Granger (1980), and Hosking (1981) on all 30 stocks of the DJIA between 2000 and 2009. McAleer and Medeiros (2008) do not consider parameter changes in time, but only dependence on past returns.

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 simulations and empirical 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 $y_t := \log(RM_t) - \mu$, where RM_t is any consistent estimator of daily integrated variance and $\mu = \mathbb{E}[\log(RM_t)] < \infty$.[1] RM stands for realized measure. 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 $\Phi(\boldsymbol{s}_t;\boldsymbol{\xi}) = 1 - \phi_1(\boldsymbol{s}_t;\boldsymbol{\xi}_1)L - \ldots - \phi_p(\boldsymbol{s}_t;\boldsymbol{\xi}_p)L^p$, and $t \in \mathbb{Z}$. 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 $\boldsymbol{s}_t \in \mathbb{R}^{k_s}$. For any vector x, we denote the number of elements by k_x. The 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 and L being the lag operator. $\Theta(L) =$

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

$(1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q)$ is a moving average (MA) lag polynomial and the error process u_t has zero mean. Further assumptions on u_t will be made in Section 2.3.2.[2]

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). *1 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 a short-memory linear ARMA process.*

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

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

The parameter γ controls the smoothness of the transition. In the limit $\gamma\to\infty$, the model becomes an ARFIMA model with a structural break at $t=c$.

EXAMPLE 3 (General Nonlinear ARFIMA). *This specification leaves the state variable $\boldsymbol{s}_t$ general and allows for several transition functions. Let*

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

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 Equation (1). The model put forward in Baillie and Kapetanios (2008) is also nested in Equation (1).

[2] In this paper we only consider nonlinearity in the AR term as this is the most common approach in the literature.

2.3. **Parameter Estimation.** We denote the vector of parameters of the entire model as $\boldsymbol{\zeta} = (d, \boldsymbol{\xi}', \boldsymbol{\theta}', \sigma_u^2)' \in \mathbb{R}^{k_\zeta}$. Here, $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_p'\right)' \in \mathbb{R}^{k_\xi}$ 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.* We employ in-fill asymptotics, as in Andrews and McDermott (1995), Saikkonen and Choi (2004), and Amado and Teräsvirta (2013), for example. Let T_0 be a constant positive number. For any sequence $\{x_t\}$, $t = 1, \ldots, T$, define $x_{tT} := (T_0/T)x_t$.

Asymptotic theory considers $T \to \infty$, that is, it considers T not as a constant, but as a variable. Implicitly, this defines a triangular array, where $\{T\}_{T=1}^{\infty} = \{1, 2, \ldots, T_0, \ldots\}$ is a sequence, $T_0 \in \mathbb{N}$ a typical element, and $(t) = (1, \ldots, T)$ is a T-tuple for any given T in the sequence.

In-fill asymptotics address the problem that a regime of finite length in untransformed time will become negligible as $T \to \infty$, that is, its length as a fraction of T will converge to zero. Thereby, any parameters governing that regime will become econometrically unrecoverable. The time transformation $x_{tT} = (T_0/T)x_t$ ensures that the length of any regime remains a constant fraction of the sample size T as $T \to \infty$ and that the transition function retains its qualitative shape. 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).

As an example, consider the logistic function $f[\gamma(t-c)] = [1 - \exp(\gamma(t-c))]^{-1}$, where c is the locus of the transition and γ determines its slope. Then, under the transformation, $x_t = t$ becomes $x_{tT} = tT_0/T$, and

$$f\left[\gamma\left(\frac{T_0}{T}t - c\right)\right] = f\left[\frac{T_0}{T}\gamma\left(t - \frac{T}{T_0}c\right)\right]. \tag{3}$$

The constant T_0 can be chosen to be any positive number. It is particularly instructive to set it equal to the constant sample size in a given data set. Suppose a given time series has 100 observations and the locus of a logistic tansition is at a fraction $\lambda \in [0,1]$ of the sample, that is, at $c = 100\lambda$. Then, setting $T_0 = 100$, the right-hand side of Equation (3) shows that the time transformation achieves two things. (1) The locus of the transition in the asymptotic analysis is at $cT/T_0 = \lambda T$, the same fraction of the now variable sample size T as in the given time series of given length T_0. As $T \to \infty$, the relative size of the first regime will remain equal to λ and not converge to zero. (2) The slope of the transition in the asymptotic analysis is $\gamma T_0/T$, that is, the slope diminishes with T, such that as T grows, the logistic function retains its shape and does not converge to the heavyside function.

Alternatively, setting $T_0 = 1$ normalizes the scale of the time axis to one, since now tT_0/T ranges from $1/T$ to T/T. Considering the right-hand side of Equation (3), the time transition occurs at fraction $c = \lambda T$ and with slope γ/T. With growing T and $t = 1, \ldots, T$, the unit interval is populated by a finer and finer mesh of points and the transition locus remains at fraction $\lambda \in [0,1]$.

With this time transformation, model (1) is embedded in a sequence of models:

$$\{\Phi_{tT}(L)v_{tT}(d) = \Theta(L)u_t\}_{T=1}^{\infty},$$

where $\Phi_{tT}(L) = 1 - \phi_1(\boldsymbol{s}_{tT};\boldsymbol{\xi}_1)L - \ldots - \phi_p(\boldsymbol{s}_{tT};\boldsymbol{\xi}_p)L^p$, $v_{tT}(d) = (1-L)^d y_{tT}$, $\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$, and u_t is a zero-mean error with assumptions to be specified in the next subsection. Note that u_t, $t = 1, \ldots, T$ for any given T, is not time-transformed.

2.3.2. *Assumptions.* We denote the data-generating 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 $u_t(\boldsymbol{\psi}) = \Theta^{-1}(L)\left[\Phi_{tT}(L)v_{tT}\right]$. 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) $\Theta_*(L)$ *has all roots outside the unit circle.*

ASSUMPTION 4 (Autoregressive Nonlinear Function). *Let* A *be a constant such that* $|A| < \infty$ *and consider the following assumptions.*

(1) *The transition functions are identified (see Example 4).*

(2) *Denote* $\Phi_{tT,*}(z) = 1 - \phi_1(\boldsymbol{s}_{tT}; \boldsymbol{\xi}_{*,1})z - \ldots - \phi_p(\boldsymbol{s}_{tT}; \boldsymbol{\xi}_{*,p})z^p$. *For all* $\boldsymbol{s}_{tT} \in \mathbb{R}^{k_s}$, *the roots of* $\Phi_{tT,*}(z)$ *are outside the unit circle.*

(3) *For* $i = 1, \ldots, p$, *and for all* $\boldsymbol{\xi}_i$, $\frac{\partial}{\partial \boldsymbol{\xi}_i}\phi_i(\cdot; \boldsymbol{\xi}_i) \leq A < \infty$.

(4) *For* $i = 1, \ldots, p$, *and for all* $\boldsymbol{\xi}_i$, $\frac{\partial^2}{\partial \boldsymbol{\xi}_i \partial \boldsymbol{\xi}_i'}\phi_i(\cdot; \boldsymbol{\xi}_i) \leq A < \infty$.

EXAMPLE 4 (for Assumption 4 (1): Logistic Transition). *If there are* $M + 1$ *different regimes of variance 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 is standard. Assumption 2(1) allows for conditional heteroskedasticity of the error term. Although conditional heteroskedasticity is not directly modeled here, our estimation theory and test sequence in the specification search allows for time-varying conditional second moments (Wooldridge 1990, 1991). This is important for applications to realized measures of integrated variance, since there is evidence of volatility of volatility (Corsi et al. 2008). Finite fourth-order unconditional moments of the error process are required in Assumption 2(2) and on the short-memory component and the state variables in Assumption 3(1). Early papers in the realized volatility literature, such as Andersen, Bollerslev, Diebold and Ebens (2001) or Andersen, Bollerslev, Diebold, and Labys (2001a,b), present evidence that logarithmic realized variance series are nearly log-normal. If this was the case, the moment conditions on v_t would be clearly satisfied. In Section 5, in particular in Tables 4 through 7, we report skewness and kurtosis of the estimated processes $\hat{v}_t$ and $\hat{u}_t$ from the model. It can be seen that for our data set of logarithmic realized measures, there is non-zero skewness and kurtosis roughly between three and six, but assuming finiteness does not seem too restrictive.

Primitive and necessary conditions for Assumption 3(1) to hold are difficult to find in the general case, however, sufficient conditions for specific cases are available. For instance, Chen and Tsay (1993) show that the conditions in Assumptions 2 and 4(2) are sufficient for weak stationarity of the process v_t. Assumptions 4(3) and 4(4) require that the functions $\phi_i(\boldsymbol{s}_{tT};\boldsymbol{\xi}_i)$, $i = 1, \ldots, p$, be bounded and two times differentiable with respect to $\boldsymbol{\xi}_i$. This is the case for the examples and models considered in Sections 2.2 and 5, as well as for a large class of sigmoid transition functions. 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.

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 $\widehat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} u_t^2(\widehat{\boldsymbol{\psi}})$. The proofs of the following three statements are given in the appendix.

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 p and the number of nonlinear terms M. We study two ways to determine M, either through an information criterion (IC) search or by a sequence of tests for remaining nonlinearity.

3.1. **Information Criteria.** In this section we show the consistency of an IC for joint determination of the number of autoregressive lags p and of nonlinear terms M in our model. Similar results can be found in Sin and White (1996), Baillie and Kapetanios (2008) and Mendes (2012a,b). Define $p \in \{0, 1, 2, \ldots, \overline{p}\}$, $M \in \{0, 1, 2, \ldots, \overline{M}\}$, and denote the parameter vector

$\boldsymbol{\psi}(p, M) \in \mathbb{R}^{k_{\boldsymbol{\psi}(p,M)}}$ as defined as in Section 2.1; its length $k_{\boldsymbol{\psi}(p,M)}$ depends on (p, M). Denote (p^*, M^*) as the data-generating pair. Our goal is to estimate (p, M) by minimizing the following information criterion:

$$\mathsf{IC}(p, M) = Q(\boldsymbol{\psi}(p, M)) + \lambda_T(p, M), \tag{4}$$

where $Q(\boldsymbol{\psi}(p, M)) = \sum_{t=1}^{T} u_t^2(\boldsymbol{\psi}(p, M))$ and $\lambda_T(p, M)$ is a penalty term that depends on p, M, and T. In the empirical application in this paper, we use the BIC, where $\lambda_T(p, M) = k_{\boldsymbol{\psi}(p,M)} \log T$. Define the sets

$$\boldsymbol{\Psi}^*_{(p,M)} = \left\{\boldsymbol{\psi} \in \boldsymbol{\psi} : \boldsymbol{\psi} = \arg\min \mathbb{E}[Q(\boldsymbol{\psi}(p, M))]\right\}, \text{ and}$$

$$\boldsymbol{\Psi}_{(p,M)} = \left\{\boldsymbol{\psi} \in \boldsymbol{\psi} : \boldsymbol{\psi} = \arg\min Q(\boldsymbol{\psi}(p, M))\right\}.$$

ASSUMPTION 5 (Information Criterion and Penalty Function).

(1) $\frac{1}{T}\lambda_T(p, M) \to 0$ *and* $\lambda_T(p, M) \to \infty$*, as* $T \to \infty$ *for every* (p, M)*;*
(2) $\lambda_T(p, M) - \lambda_T(p^*, M^*) \to \infty$ *as* $T \to \infty$ *if* $p > p^*$ *and* $m > M^*$*.*

Note that by Theorem 1, $\mathbb{E}[Q(\boldsymbol{\psi}(p, M))]$ attains a minimum on $\boldsymbol{\psi}_*(p^*, M^*)$. We assume that the data-generating specification (p^*, M^*) is in the basket of considered models; the IC attains its minimum on (p^*, M^*) by the information inequality (White 1994, Sin and White 1996, Hong and Preston 2012). Assumption 5(1) and (2) are standard for IC criteria, see for example Baillie and Kapetanios (2008). They do not cover the AIC, but the Hannan-Quinn IC, and the BIC employed in the empirical application in this paper, since $\lambda_T(p, M)/T = k_{\boldsymbol{\psi}(p,M)} \log T/T \to 0$ and $\lambda_T(p, M) = k_{\boldsymbol{\psi}(p,M)} \log T \to \infty$. Since $k_{\boldsymbol{\psi}(p,M)} > k_{\boldsymbol{\psi}(p*,M*)}$ for $p > p*$ and $M > M*$, we also have that $\lambda_T(p, M) - \lambda_T(p^*, M^*) = (k_{\boldsymbol{\psi}(p,M)} - k_{\boldsymbol{\psi}(p*,M*)}) \log T \to \infty$.

The proof of the following statement is presented in the appendix.

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

3.2. **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:

$$v_t = \boldsymbol{\phi}_0' \boldsymbol{V}_{t-1} + \sum_{m=1}^{M^*} \boldsymbol{\phi}_m' \boldsymbol{V}_{t-1} f\left[\gamma_m\left(s_t - c_m\right)\right] + \sum_{m=M^*+1}^{M} \boldsymbol{\phi}_m' \boldsymbol{V}_{t-1} f\left[\gamma_m\left(s_t - c_m\right)\right] + u_t. \quad (5)$$

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

$$\mathbb{H}_0 : \gamma_{\tilde{M}+1} = \gamma_{\tilde{M}+2} = \cdots = \gamma_M = 0. \quad (6)$$

Model (5) is identified only under the alternative, since under the null the nonlinear transition functions at locations $c_{\tilde{M}+i}$ do not exist. Therefore, 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:

$$v_t = \widetilde{\boldsymbol{\phi}}_0' \boldsymbol{V}_{t-1} + \sum_{m=1}^{\tilde{M}} \widetilde{\boldsymbol{\phi}}_m' \boldsymbol{V}_{t-1} f\left[\gamma_m\left(s_t - c_m\right)\right] + \boldsymbol{\rho}_1' \boldsymbol{V}_{t-1} s_t + \boldsymbol{\rho}_2' \boldsymbol{V}_{t-1} s_t^2 + \boldsymbol{\rho}_3' \boldsymbol{V}_{t-1} s_t^3 + u_t^*, \quad (7)$$

where $u_t^* = u_t + R_3$ and R_3 is the remainder in the Taylor expansion. The null hypothesis (6) 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 (7) 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 $\widehat{\boldsymbol{H}} = \left(\widehat{\boldsymbol{h}}_1', \ldots, \widehat{\boldsymbol{h}}_T'\right)'$.

Wooldridge (1990, 1991) proposed a simple modification to the standard regression approach in order to robustify the test to non-normality and unknown forms of heteroskedasticity. 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 $\widetilde{\boldsymbol{H}}$, which is $\widehat{\boldsymbol{H}}$ computed with the new sequence of residuals.

(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 (Wooldridge 1990, Theorem 2.1 and Procedure 2.1). See also the exposition in Teräsvirta et al. (2010, p. 70), who refer to this test procedure as the "TR^2-form." Robustness to heteroskedasticity is important in financial applications, where the errors are rarely normal and homoskedastic. Although standard in the linearity testing literature, the sixth-order moment assumption above may appear strong in the context of financial data. In Section 4 and Tables 2 and 3, we evaluate violations of the assumptions in simulations. A similar study can be found in Medeiros and Veiga (2009).

We combine the procedure above into a sequence of tests. We start by testing a linear model against a model with one nonlinear term at an α_1-level of significance. In case $\mathbb{H}_0$ (linearity) is rejected, the nonlinear term is added, the new model is re-estimated and then tested against an alternative with two nonlinear terms. 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 at (Lütkepohl 2005, p. 144)

$$\alpha^* = 1 - \prod_{J=1}^{\bar{J}}(1 - \alpha_J) = 1 - \prod_{J=1}^{\bar{J}}(1 - \alpha_1\, C^{J-1}). \tag{8}$$

Table 1 presents some upper bounds for different values of C and initial significance level α_1. In the applications considered in this paper, we set $C = 1/2$ and $\alpha_1 = 0.05$, yielding an upper bound of 0.0967. As can be seen from the table, small values of α_1 and C are conservative, large values liberal. In Equation (8), the parametrization $\alpha_J = \alpha_1\, C^{J-1}$ is chosen simply for computational convenience, others are possible. The autoregressive order is specified by BIC search (Rech et al. 2001).

For the asymptotic theory to hold we just need the functions $\phi(\cdot)$ to be second-order differentiable. On the other hand, for the remaining nonlinearity test to hold we need as much derivatives as used in the test.

4. Monte-Carlo Evidence

We consider three distinct data-generating processes (DGP) for u_t. In the simplest case, u_t is independent and normally distributed with zero mean and variance 0.25, $u_t \sim \mathsf{NID}(0, 0.25)$, where NID stands for identically and independently normally distributed. 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 short-memory (SM) data-generating processes are given by:

$$y_t = 0.04 + 0.55y_{t-1} + 0.34y_{t-2} + \sigma_t\varepsilon_t. \tag{9}$$

$$\begin{aligned} y_t = {} & 0.55y_{t-1} + 0.34y_{t-2} - (0.4y_{t-1} + 0.2y_{t-2})\, f\left[12\left(r_{t-1} + 0.5\right)\right] \\ & + (0.4y_{t-1} + 0.2y_{t-2})\, f\left[4\left(r_{t-1} - 1\right)\right] + \sigma_t\varepsilon_t. \end{aligned} \tag{10}$$

$$\begin{aligned} y_t = {} & 0.55y_{t-1} + 0.34y_{t-2} - (0.4y_{t-1} + 0.2y_{t-2})\, f\left[200\left(t/T + 0.25\right)\right] \\ & + (0.4y_{t-1} + 0.2y_{t-2})\, f\left[100\left(t/T - 0.75\right)\right] + \sigma_t\varepsilon_t. \end{aligned} \tag{11}$$

The long memory models are defined as $v_t = (1-L)^{0.4}y_t$ and the short memory component v_t is given by

$$v_t = 0.55v_{t-1} + 0.34v_{t-2} + \sigma_t\varepsilon_t. \tag{12}$$

$$\begin{aligned} v_t = {} & 0.55v_{t-1} + 0.34v_{t-2} - (0.4v_{t-1} + 0.2v_{t-2})\, f\left[12\left(r_{t-1} + 0.5\right)\right] \\ & + (0.4v_{t-1} + 0.2v_{t-2})\, f\left[4\left(r_{t-1} - 1\right)\right] + \sigma_t\varepsilon_t. \end{aligned} \tag{13}$$

$$\begin{aligned} v_t = {} & 0.55v_{t-1} + 0.34v_{t-2} - (0.4v_{t-1} + 0.2v_{t-2})\, f\left[200\left(t/T + 0.25\right)\right] \\ & + (0.4v_{t-1} + 0.2v_{t-2})\, f\left[100\left(t/T - 0.75\right)\right] + \sigma_t\varepsilon_t. \end{aligned} \tag{14}$$

The first DGP is a simple short-memory AR specification. The next two DGPs are nonlinear short-memory processes, while the remaining specifications are all long-memory models. We generate 1000 replications of the models above with $T = 500$ and $T = 1000$ observations. The specification and estimation results are reported in Tables 2 and 3.

Tables 2 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 (Teräsvirta et al. 2010, p. 381). In order to estimate γ precisely, it is necessary (1) to have a large number of observations close to the location parameter c, which is rarely the case, and (2) the transitions should be rather smooth, i.e. γ should be small. However, noisy estimates of γ do not seem to impede the estimation of the remaining parameters.

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 3, where C is set to 0.50. 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 3. First, in the linear case both methodologies work well. Second, the selection of p is accurate in almost all of the cases. The number of regimes is underestimated by both LM tests and BIC, but the performance improves as the sample size increases. As expected, the sequence of robust LM test 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.

5. Empirical Application

5.1. **Long Memory and Nonlinearity in Realized Measures.** Andersen et al. (2001, 2003) and Barndorff-Nielsen and Shephard (2002) have pioneered the use of intraday data to construct realized measures. Several studies have proposed extensions of the basic realized variance (RV) estimator (the sum of squared intra-day returns) that are robust to microstructure noise (Zhang et al. 2005, Barndorff-Nielsen et al. 2008) and the presence of jumps (Andersen et al. 2012). The daily dynamics of realized variance 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, variance 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 and dependence on lagged returns (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). Dependence on lagged returns and the corresponding leverage and feedback effects are discussed in, for example, McAleer and Medeiros (2008), Scharth and Medeiros (2009), and Chen et al. (2010). Parameter changes in time, in particular if they affect the unconditional variance, can mask long memory, and vice versa, as studied, for example, in Lamoureux and Lastrapes (1990), Diebold and Inoue (2001), Granger and Hyung (2004), Hillebrand and Medeiros (2008), Craioveanu and Hillebrand (2012).

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. *Integrated Variance Estimation.* In estimating daily integrated variance, 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 use the MedRV statistic proposed in Andersen et al. (2012), and the results, reported in the supplemental material, are very similar. The BHLS estimator is a consistent estimator of quadratic variation and the MedRV is robust to jumps in estimating integrated variance. We discard transactions outside trading hours, considering transactions between 9.30 a.m. through 4.00 p.m. 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.** In this subsection, we report in-sample results for the time series of realized measures of the 30 stocks for the full sample 3-Jan-2000 through 31-Dec-2009. In total, two different nonlinear specifications (past returns and time transitions) are estimated on two different realized measures (BHLS and MedRV) for each of 30 different stocks, resulting in 120 estimated models. We focus mainly on the results for BHLS. The results for MedRV are in the supplemental material and qualitatively very similar.

Tables 4 and 5 show the results for models specified by the sequence of LM tests, Tables 6 and 7 show the results for the BIC specification search. The sequence of LM tests starts with $\alpha_1 = 0.05$ as the initial significance level and $C = 1/2$, resulting in an upper bound of 0.097 for the significance level of the entire sequence. The tables report the selected order pair (p, M), the p-value of the test for remaining non-linearity (RN), the p-value of a Ljung-Box test for the first two autocorrelation coefficients of the residuals $\hat{u}_t$ (AC), and the p-value of a first-order ARCH LM-test. $K(v_t)$ and $S(v_t)$ are, respectively, kurtosis and skewness of the fractional filtered series, $v_t = (1-L)^{\widehat{d}} y_t$. $K(u_t)$ and $S(u_t)$ are, respectively, kurtosis and skewness of the residuals.

Several noteworthy results emerge from the tables. For most of the stocks, the p-values of the remaining nonlinearity test are quite high, indicating that the results are reasonably robust to the choice of the initial significance level as well as the constant C. Both methods, BIC search and LM-test sequence, select a small number of regimes, either one or two regimes for all stocks. The BIC tends to select a smaller number of regimes than the LM tests. However, there are a number of cases where both criteria agree. The autoregressive order is also very small, no larger than three and equal to one for most of the series. Models with time as transition variable have, on average, fewer regimes than the models with past returns as transition variable. With regard to the diagnostic tests, there is no evidence of residual autocorrelation. On the other hand, the ARCH LM-test indicates clear evidence of conditional heteroskedasticity. As pointed out earlier, the sequence of LM tests is robust to deviations from normality and to the presence of conditional heteroskedasticity (Wooldridge 1990, Teräsvirta et al. 2010). Despite the statistical evidence of conditional heteroskedasticity, both the filtered series $\hat{v}_t$ and the residuals $\hat{u}_t$ have excess kurtosis within a reasonable range of three to six. This supports the assumption of finite fourth moments in Section 2.3.2.

In the remainder of this subsection, we will discuss two out of the 120 estimated models in detail for illustration. We focus on one stock (KFT) that is fairly representative for the entire set and show the estimated model with time as transition variable on the MedRV series, and the estimated model with past returns as transition variable on the BHLS series. The sample period for KFT is Jun 13, 2001, through Dec 31, 2009. Figure 1, Panel (a), shows the MedRV estimator (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.

The estimated specification for the MedRV estimator and time transitions is (standard errors in parentheses):

$$\begin{aligned}
y_t &= \log(RM_t) - 0.2218,\ v_t = (1-L)^{\underset{(0.0036)}{0.4416}} y_t, \\
v_t &= \left[\underset{(0.0055)}{-0.1230} + \underset{(0.0166)}{0.2295} f(\gamma, c)\right] v_{t-1} - \left[\underset{(0.0032)}{0.0851} - \underset{(0.0151)}{0.1375} f(\gamma, c)\right] v_{t-2} + u_t, \qquad (15) \\
f(\gamma, c) &= \{1 + \exp[\gamma(t/T - c)]\}^{-1} = \left\{1 + \exp\left[\underset{(42973)}{69.30}\left(t/T - \underset{(1.26e-4)}{0.7661}\right)\right]\right\}^{-1}.
\end{aligned}$$

The time transition captures the change in variance 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 $1/(1 - [-0.1230 - 0.0851]) \approx 0.8$ days to persistence at a scale of $1/(1 - [-0.1230 - 0.0851] - [0.2295 + 0.1375]) \approx 1.2$ days. This corresponds to the short decorrelation scale in stock variance reported, for example, in Fouque et al. (2003) and Hillebrand (2006).

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

$$\begin{aligned}
y_t &= \log(RM_t) - 0.1356,\ v_t = (1-L)^{\underset{(0.0033)}{0.4900}} y_t, \\
v_t &= \left[\underset{(0.0215)}{-0.0678} - \underset{(0.0193)}{0.0873} f(\gamma, c)\right] v_{t-1} + \left[\underset{(0.0468)}{0.2804} - \underset{(0.0473)}{0.3834} f(\gamma, c)\right] v_{t-2} \\
&\quad - \left[\underset{(0.0330)}{0.1543} - \underset{(0.0345)}{0.0837} f(\gamma, c)\right] v_{t-3} + u_t, \qquad (16) \\
f(\gamma, c) &= \left(1 + \exp\left\{\gamma\left[100\left(\log\frac{S_{t-1}}{S_{t-2}}\right) - c\right]\right\}\right)^{-1} \\
&= \left(1 + \exp\left\{\underset{(436.69)}{7.5875}\left[100\left(\log\frac{S_{t-1}}{S_{t-2}}\right) + \underset{(9.9e-5)}{2.2422}\right]\right\}\right)^{-1}.
\end{aligned}$$

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 variance on shorter scales (Medeiros and Veiga 2009).

These examples demonstrate how the model in Equation (1) can disentangle long memory, short memory, and non-linear terms, for example in time or in past returns. In the next section, we show that the model can also be used for forecasting purposes, having at the same time the advantage of producing economically interpretable estimators.

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 $y_t = \log(RM_t) - \mu$, for both realized measures RM=MedRV and RM=BHLS. 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 lagged daily returns as transition variable (“STARFIMA I”); (3) nonlinear ARFIMA with time as transition variable (“STARFIMA II”); (4) a short memory STAR model with time as transition variable (“STAR”). The latter model is motivated by the confounding effects of breaks and long memory and is used to check if simply accounting for changes in the unconditional mean can improve forecasts sufficiently. The benchmark models are: (1) the linear HAR-RV model of Corsi (2009) (“Ratio I”) and (2) the nonlinear HAR-RV of Corsi and

Renò (2012) (“Ratio II”), which allows for leverage. Table 8 reports the results for one-step-ahead prediction, while Table 9 shows the results for five- and ten-steps ahead prediction.

The reporting format are ratios of root-mean square forecast errors (RMSE) of the proposed model to the RMSE of the benchmark model. The RMSE for a given model, a given realized measure, and a given forecast horizon h is

$$RMSE = \left[\frac{1}{P}\sum_{t=1}^{P} e_{t,h}^2\right]^{\frac{1}{2}},\ e_{t,h} = y_t - \hat{y}_{t|t-h},$$

where $\hat{y}_{t|t-h}$ is the h-period ahead forecast of y_t, and P is the size of the forecast sample. 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 with small Giacomini-White p-values. 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 predictions. For one-step-ahead prediction the benchmarks are superior. More specifically, for five-days-ahead, the ARFIMA is statistically superior to 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 nonlinear ARFIMA with asymmetry is similar. For five-days-ahead, it is superior to 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. The short-memory plus-breaks STAR specification substantially underperforms in comparison to the nonlinear ARFIMA models. This provides evidence that even after accounting for structural breaks, there is enough long-range dependence left in the data that an explicit long-memory model has superior forecast performance.

In summary, with regard to forecasting financial variance, we recommend the model proposed in this paper for forecast horizons longer than a single day. The model 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

Nonlinearities such as structural breaks are usually 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, an information criteria search and a sequence of LM tests.

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 time series of realized measures for integrated variance. A forecast competition indicates that a specification with long memory and asymmetry can outperform the standard and non-linear HAR-RV models 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

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

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

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

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

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

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(2),

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

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 showing that $\mathbb{E}\left(\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left|\frac{\partial^2 q_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right|\right) < \infty$. 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. Here, we will consider the second derivative of q_t with respect to d, since this is the most involved one.

First, note that

$$\frac{\partial^2}{\partial d^2}(1-L)^d = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left[\left(\sum_{i=0}^{j-1} \frac{1}{d-i} \right)^2 - \sum_{i=0}^{j-1} \left(\frac{1}{d-i} \right)^2 \right] \prod_{i=0}^{j-1} (d-i) L^j, \tag{17}$$

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

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 (Sin and White 1996, Baillie and Kapetanios 2008, Mendes 2012a,b). Consider the event

$$
\begin{aligned}
\{(\hat{p}, \hat{M}) \neq (p^*, M^*)\} &= \{(\hat{p}, \hat{M}) \neq (p^*, M^*) \cap (p > p^* \cap M > M^*)\} \cup \\
&\quad \{(\hat{p}, \hat{M}) \neq (p^*, M^*) \cap (p < p^* \cup M < M^*)\}, \\
&\Leftrightarrow \{(\hat{p} > p^* \cap \hat{M} > M^*) \cap (p > p^* \cap M > M^*)\} \cup \\
&\quad \{(\hat{p} < p^* \cup \hat{M} < M^*) \cap (p < p^* \cup M < M^*)\} \\
&\Leftrightarrow \{IC(\hat{p}, \hat{M}) < IC(p^*, M^*) \cap (p > p^* \cap M > M^*)\} \cup \\
&\quad \{IC(\hat{p}, \hat{M}) < IC(p^*, M^*) \cap (p < p^* \cup M < M^*)\} \\
&\Leftrightarrow: A \cup B. \qquad (18)
\end{aligned}
$$

We show that $\mathbb{P}(A \cup B) \to 0$ as $T \to \infty$. By the definition of the IC in equation (4), we have

$$IC(p, M) < IC(p^*, M^*) \Leftrightarrow \left(Q(\hat{\boldsymbol{\psi}}(p^*, M^*)) - Q(\hat{\boldsymbol{\psi}}(p, M)) > \lambda_T(p, M) - \lambda_T(p^*, M^*) \right). \tag{19}$$

Consider first the case $(p > p^*)$ and $(M > M^*)$. Apply Markov's inequality to the right-hand side of equation (19):

$$\begin{aligned} &\mathbb{P}\left[Q(\hat{\boldsymbol{\psi}}(p^*, M^*)) - Q(\hat{\boldsymbol{\psi}}(p, M)) > \lambda_T(p, M) - \lambda_T(p^*, M^*) \right] \\ &\leq \frac{\mathbb{E}\left| Q(\hat{\boldsymbol{\psi}}(p^*, M^*)) - Q(\hat{\boldsymbol{\psi}}(p, M)) \right|}{\lambda_T(p, M) - \lambda_T(p^*, M^*)}. \end{aligned} \tag{20}$$

Apply the triangular inequality to the numerator of the right-hand side:

$$\begin{aligned} &\mathbb{E}|Q(\hat{\boldsymbol{\psi}}(p^*, M^*)) - Q(\hat{\boldsymbol{\psi}}(p, M))| \\ &\leq \max_{\hat{\boldsymbol{\psi}} \in \boldsymbol{\Psi}_{(p^*, M^*)}} \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}^*_{(p,M)}} \mathbb{E}|Q(\hat{\boldsymbol{\psi}}(p^*, M^*)) - Q(\boldsymbol{\psi}(p^*, M^*))|, \\ &+ \max_{\hat{\boldsymbol{\psi}} \in \boldsymbol{\Psi}_{(p,M)}} \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}^*_{(p,M)}} \mathbb{E}|Q(\hat{\boldsymbol{\psi}}(p, M)) - Q(\boldsymbol{\psi}(p, M))|, \\ &+ \min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}^*_{(p,M)}} \mathbb{E}|Q(\boldsymbol{\psi}(p, M)) - Q(\boldsymbol{\psi}(p^*, M^*))|, \\ &=: A_1 + A_2 + A_3. \end{aligned} \tag{21}$$

The first two terms A_1 and A_2 are expectations of differences

$$Q(\hat{\boldsymbol{\psi}}) - Q(\boldsymbol{\psi}) = \Delta Q(\boldsymbol{\psi})(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + \frac{1}{2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})' \Delta^2 Q(\bar{\boldsymbol{\psi}})(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}),$$

for a $\bar{\boldsymbol{\psi}}$ between $\boldsymbol{\psi}$ and $\hat{\boldsymbol{\psi}}$ and a given pair (p, M). We have that $\Delta Q(\boldsymbol{\psi}) = 0$ because $\boldsymbol{\psi} \in \boldsymbol{\Psi}^*_{(p,M)}$. By Theorem 2 and Proposition 1, $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} = O_p(T^{-\frac{1}{2}})$ and $\Delta^2 Q(\bar{\boldsymbol{\psi}}) = O_p(T)$, so that $Q(\hat{\boldsymbol{\psi}}) - Q(\boldsymbol{\psi}) = O_p(1)$. By Assumption 2(2) on u_t, $Q(\boldsymbol{\psi})$ is uniformly integrable, and thus

$\mathbb{E}[Q(\hat{\boldsymbol{\psi}})-Q(\boldsymbol{\psi})] = O(1)$. Therefore, $A_1+A_2 = O(1)$. The last term $A_3 = 0$, since by definition

$$\min_{\boldsymbol{\psi}\in\boldsymbol{\Psi}^*_{(p,M)}} \mathbb{E}Q(\boldsymbol{\psi}(p,M)) = \min_{\boldsymbol{\psi}\in\boldsymbol{\Psi}^*_{(p,M)}} \mathbb{E}Q(\boldsymbol{\psi}(p^*,M^*)).$$

This is the case $(p > p^*)$ and $(M > M^*)$, and the denominator of equation (20) grows large for large T according to Assumption 5(2). Applying the union bound on event A from equation (18), we obtain

$$\mathbb{P}(A) \leq \sum_{p=p^*+1}^{\overline{p}} \sum_{M=M^*+1}^{\overline{M}} \left[\frac{C}{\lambda_T(p,M)-\lambda_T(p^*,M^*)}\right] \underset{T\to\infty}{\to} 0,$$

for some finite C from the fact that $A_1 + A_2 = O(1)$ and Assumption 5(2).

Consider now the case $(p < p^*)$ and/or $(M < M^*)$. For $\boldsymbol{\psi} \in \boldsymbol{\Psi}^*_{(p,M)}$, we have

$$\begin{aligned}
&\frac{1}{T}\left[Q(\hat{\boldsymbol{\psi}}(p^*,M^*)) - Q(\hat{\boldsymbol{\psi}}(p,M))\right] \\
&= \frac{1}{T}\left[Q(\hat{\boldsymbol{\psi}}(p^*,M^*)) - \mathbb{E}Q(\hat{\boldsymbol{\psi}}(p^*,M^*))\right] - \frac{1}{T}\left[Q(\hat{\boldsymbol{\psi}}(p,M)) - \mathbb{E}Q(\hat{\boldsymbol{\psi}}(p,M))\right] \\
&\quad + \frac{1}{T}\mathbb{E}\left[Q(\hat{\boldsymbol{\psi}}(p^*,M^*)) - Q(\boldsymbol{\psi}(p^*,M^*))\right] - \frac{1}{T}\mathbb{E}\left[Q(\hat{\boldsymbol{\psi}}(p,M)) - Q(\boldsymbol{\psi}(p,M))\right] \\
&\quad - \frac{1}{T}\mathbb{E}\left[Q(\boldsymbol{\psi}(p,M)) - Q(\boldsymbol{\psi}(p^*,M^*))\right] \\
&= o_p(1) - K_{(p,M)}, \qquad (22)
\end{aligned}$$

where $K_{(p,M)} := \frac{1}{T}\mathbb{E}\left[Q(\boldsymbol{\psi}(p,M)) - Q(\boldsymbol{\psi}(p^*,M^*))\right]$. The first term on the right-hand side of equation (22) is $o_p(1)$ according to Assumption 2(2) on u_t. The second term is $o_p(1)$ according to the ergodic theorem (Beran et al. 2013, p. 546, for example). The third and fourth terms are $o_p(1)$ by the arguments outlined for A_1 and A_2 for the case $(p > p^*)$ and $(M > M^*)$ above. The fifth and last term $K_{(p,M)} > 0$ by the definition of (p^*, M^*), and because this is case $(p < p^*)$

and/or $(M < M^*)$. From Assumption 5(1),

$$\frac{1}{T}\left[\lambda_T(p, M) - \lambda_T(p^*, M^*)\right] \underset{T\to\infty}{\to} 0.$$

Therefore, the set

$$\left\{Q(\hat{\boldsymbol{\psi}}(p^*, M^*)) - Q(\hat{\boldsymbol{\psi}}(p, M)) > \lambda_T(p, M) - \lambda_T(p^*, M^*)\right\} \underset{T\to\infty}{\to} \emptyset.$$

Since (p^*, M^*) finite, the union bound on event B from equation (18) yields

$$\mathbb{P}(B) \underset{T\to\infty}{\to} 0.$$

□

REFERENCES

Amado, C. and Teräsvirta, T.: 2013, Modelling volatility by variance decomposition, *Journal of Econometrics* **175**, 142–153.

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

Andersen, T., Bollerslev, T., Diebold, F. X. and Ebens, H.: 2001, The distribution of realized stock return volatility, *Journal of Financial Economics* **61**, 43–76.

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P.: 2001a, 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.: 2001b, Exchange rate returns standardized by realized volatility are (nearly) Gaussian, *Multinational Finance Journal* . forthcoming.

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.: 2012, Jump-robust volatility estimation using nearest neighbor truncation, *Journal of Econometrics* **169**, 75–93.

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.

Beran, J., Feng, Y., Ghosh, S. and Kulik, R.: 2013, *Long-Memory Processes*, Springer, Berlin.

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, R. and Tsay, R. S.: 1993, Functional coefficient autoregressive models, *Journal of the American Statistical Association* **88**, 298–308.

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., Mittnik, S., Pigorsch, C. and Pigorsch, U.: 2008, Volatility of realized volatility, *Econometric Reviews* **27**, 46–78.

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 Economics Statistics* **30**, 368–330.

Craioveanu, M. and Hillebrand, E.: 2012, Level shifts in volatility models, *Annals of Finance* **8**, 277–308.

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.

Fouque, J. P., Papanicolaou, G., Sircar, R. and Sølna, K.: 2003, Short time-scale in S&P 500 volatility, *Journal of Computational Finance* **6**, 1–23.

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.

Granger, C. and Joyeux, R.: 1980, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* **1(1)**, 15–29.

Granger, C. W. J.: 1980, Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* **14**, 227–238.

Granger, C. W. J. and Teräsvirta, T.: 2001, A simple nonlinear time series model with misleading linear properties, *Economics Letters* **62**, 161–165.

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

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.

Hillebrand, E. and Medeiros, M.: 2008, Estimating and forecasting GARCH models in the presence of structural breaks and regime switches, *in* M. Wohar and D. Rapach (eds), *Forecasting in The Presence of Structural Breaks and Model Uncertainty*, Elsevier, pp. 303–328.

Hong, H. and Preston, B.: 2012, Bayesian averaging, prediction, and nonnested model selection, *Journal of Econometrics* **167**, 358–369.

Hosking, J.: 1981, Fractional differencing, *Biometrika* **65**, 165–176.

Hsu, C.-C.: 2001, Change-point estimation in regressions with i(d) variables, *Economics Letters* **70**, 147–155.

Lamoureux, C. G. and Lastrapes, W. D.: 1990, Persistence in variance, structural change, and the garch model, *Journal of Business and Economic Statistics* **8**, 225–234.

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.

Lütkepohl, H.: 2005, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

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.

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.

Sin, C.-Y. and White, H.: 1996, Information criteria for selecting possibly misspecified parametric models, *Journal of Econometrics* **71**, 207–225.

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.

Teräsvirta, T., Tjøstheim, D. and Granger, C.: 2010, *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford, UK.

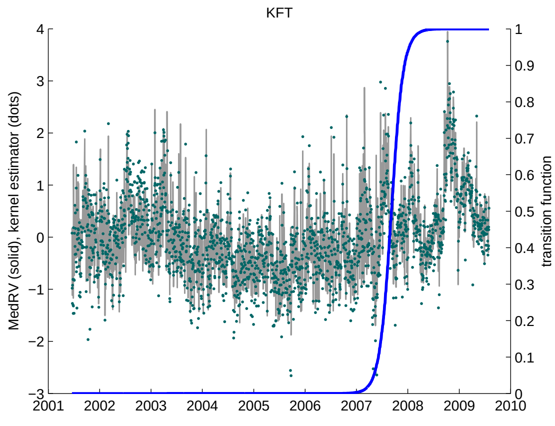
van Dijk, D., Franses, P. and Paap, R.: 2002, A nonlinear long memory model, with an application to US unemployment, *Journal of Econometrics* **110**, 135–165.

White, H.: 1994, *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York, NY.

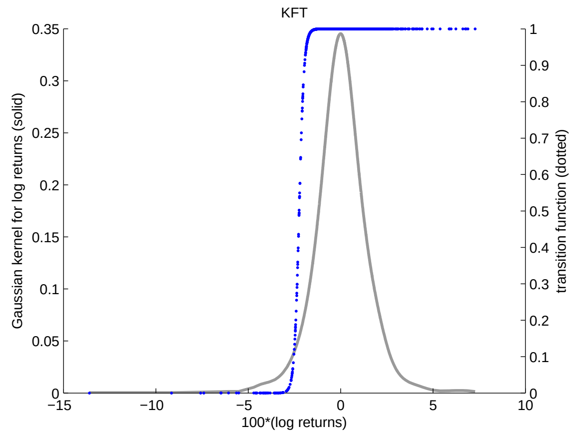
Wooldridge, J. M.: 1990, A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17–43.

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.



(a)



(b)

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

TABLE 1. UPPER SIGNIFICANCE BOUNDS.

The table displays upper bounds for the sequence of LM tests for different values of the initial significance level α_1 and for the constant C.

$C \backslash \alpha_1$	**0.1**	**0.05**	**0.01**
$\mathbf{1/4}$	0.1298	0.0658	0.0133
$\mathbf{1/3}$	0.1444	0.0736	0.0149
$\mathbf{1/2}$	0.1870	0.0967	0.0199
$\mathbf{2/3}$	0.2662	0.1413	0.0296

TABLE 2. 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 3. 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. Nonlinear I features a transition function with past returns as state variable; Nonlinear II has time as state variable. 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 $T = 1000$, $C = 0.50$.

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)

TABLE 4. FULL SAMPLE RESULTS: SEQUENCE OF LM TESTS AND PAST RETURNS.

The table displays statistics for models with past returns as transition variables and number of regimes determined by the LM tests with 5% as the initial significance level and $C = 1/2$; p is the autoregressive order. M is the number of nonlinear terms. RN, AC, and $ARCH$ are, respectively, the p-values of the remaining nonlinearity test, the Ljung-Box autocorrelation test up to the second lag and the first-order ARCH LM-test. $K(v_t)$ and $S(v_t)$ are, respectively, the kurtosis and skewness of the fractional filtered series. $K(u_t)$ and $S(u_t)$ are, respectively, the kurtosis and skewness of the residuals.

	Transition Variable: Past Return								
	p	M	RN	AC	ARCH	$K(v_t)$	$S(v_t)$	$K(u_t)$	$S(u_t)$
AA	1	1	0.102	0.777	0.349	3.808	0.133	3.828	0.145
AXP	1	0	0.212	0.132	0.292	4.723	0.069	4.814	0.080
BA	1	1	0.192	0.246	0.284	3.619	0.118	3.621	0.091
BAC	2	0	0.230	0.976	0.048	4.387	0.516	4.318	0.484
CAT	2	0	0.229	0.994	0.059	3.888	0.130	3.904	0.155
CSCO	1	1	0.226	0.814	0.001	3.918	0.187	3.921	0.145
CVX	1	0	0.281	0.427	0.007	3.911	0.196	3.946	0.203
DD	1	0	0.345	0.998	0.000	3.968	0.129	3.979	0.124
DIS	1	0	0.378	0.166	0.001	4.383	0.288	4.418	0.295
GE	1	0	0.138	0.554	0.000	3.700	0.188	3.760	0.197
HD	1	1	0.156	0.769	0.132	3.904	0.050	3.943	0.040
HPQ	1	0	0.087	0.403	0.002	4.382	0.181	4.381	0.169
IBM	1	1	0.241	0.893	0.594	4.407	0.028	4.505	0.029
INTC	1	0	0.119	0.958	0.000	3.727	0.049	3.729	0.052
JNJ	2	1	0.722	0.897	0.000	4.651	0.209	4.589	0.166
JPM	1	0	0.095	0.490	0.000	4.144	0.192	4.142	0.191
KFT	3	1	0.157	0.964	0.495	4.602	0.295	4.737	0.319
KO	1	0	0.199	0.918	0.035	4.562	0.193	4.678	0.192
MCD	1	0	0.669	0.547	0.231	4.474	0.307	4.508	0.316
MMM	1	1	0.269	0.947	0.008	4.760	0.117	4.888	0.099
MO	1	0	0.223	0.621	0.006	4.841	0.553	4.897	0.570
MRK	2	0	0.227	0.988	0.399	6.556	0.680	6.715	0.710
MSFT	1	0	0.338	0.989	0.467	4.562	0.037	4.560	0.030
PFE	3	0	0.474	0.999	0.004	4.810	0.381	4.883	0.394
PG	1	0	0.199	0.291	0.000	5.176	0.193	5.241	0.209
T	1	0	0.357	0.990	0.016	4.104	0.276	4.144	0.270
UTX	1	0	0.054	0.954	0.003	4.496	0.150	4.530	0.162
VZ	1	1	0.082	0.229	0.008	4.034	0.259	4.033	0.226
WMT	1	1	0.436	0.847	0.615	4.035	0.138	4.042	0.156
XOM	1	0	0.087	0.818	0.006	3.928	0.214	4.003	0.226

TABLE 5. FULL SAMPLE RESULTS: SEQUENCE OF LM TESTS AND TIME.

The table displays statistics for models with time as transition variables and number of regimes determined by the LM tests with 5% as the initial significance level and $C = 1/2$; p is the autoregressive order. M is the number of nonlinear terms. RN, AC, and $ARCH$ are, respectively, the p-values of the remaining nonlinearity test, the Ljung-Box autocorrelation test up to the second lag and the first-order ARCH LM-test. $K(v_t)$ and $S(v_t)$ are, respectively, the kurtosis and skewness of the fractional filtered series. $K(u_t)$ and $S(u_t)$ are, respectively, the kurtosis and skewness of the residuals.

	Transition Variable: Time								
	p	M	RN	AC	ARCH	$K(v_t)$	$S(v_t)$	$K(u_t)$	$S(u_t)$
AA	1	1	0.903	0.756	0.751	3.808	0.133	3.855	0.137
AXP	1	0	0.052	0.132	0.292	4.722	0.069	4.814	0.080
BA	1	0	0.405	0.477	0.112	3.617	0.117	3.646	0.119
BAC	2	1	0.364	0.951	0.202	4.387	0.515	4.284	0.458
CAT	2	0	0.051	0.994	0.059	3.887	0.130	3.904	0.155
CSCO	1	0	0.256	0.938	0.000	3.917	0.186	3.916	0.185
CVX	1	1	0.443	0.717	0.008	3.910	0.196	3.898	0.200
DD	1	1	0.329	0.853	0.001	3.968	0.128	3.947	0.114
DIS	1	1	0.046	0.055	0.005	4.383	0.287	4.458	0.287
GE	1	1	0.195	0.923	0.007	3.699	0.187	3.752	0.174
HD	1	0	0.244	0.897	0.057	3.903	0.050	3.942	0.055
HPQ	1	1	0.320	0.303	0.006	4.382	0.180	4.388	0.166
IBM	1	1	0.090	0.971	0.461	4.407	0.027	4.459	0.044
INTC	1	0	0.073	0.958	0.000	3.727	0.048	3.729	0.052
JNJ	2	1	0.814	0.994	0.001	4.650	0.208	4.567	0.180
JPM	1	1	0.578	0.859	0.004	4.143	0.192	4.131	0.173
KFT	3	1	0.088	0.996	0.510	4.602	0.295	4.751	0.327
KO	1	0	0.476	0.918	0.035	4.561	0.192	4.678	0.192
MCD	1	0	0.077	0.547	0.231	4.473	0.307	4.508	0.316
MMM	1	1	0.835	0.905	0.000	4.760	0.117	4.811	0.120
MO	1	1	0.086	0.574	0.031	4.851	0.556	4.917	0.566
MRK	2	0	0.199	0.988	0.399	6.556	0.680	6.715	0.710
MSFT	1	1	0.113	0.922	0.334	4.562	0.037	4.574	0.017
PFE	3	0	0.493	0.999	0.004	4.809	0.381	4.883	0.394
PG	1	0	0.220	0.291	0.000	5.175	0.193	5.241	0.209
T	1	0	0.147	0.990	0.016	4.104	0.276	4.144	0.270
UTX	1	1	0.650	0.798	0.002	4.495	0.149	4.496	0.147
VZ	1	0	0.204	0.341	0.021	4.033	0.259	4.094	0.263
WMT	1	0	0.217	0.983	0.184	4.034	0.138	4.133	0.170
XOM	1	1	0.214	0.973	0.002	3.927	0.214	3.886	0.208

TABLE 6. FULL SAMPLE RESULTS: BIC SEARCH AND PAST RETURNS.

The table displays statistics for models with past returns as transition variables and number of regimes determined by the LM tests with 5% as the initial significance level and $C = 1/2$; p is the autoregressive order. M is the number of nonlinear terms. AC and $ARCH$ are, respectively, the p-values of the Ljung-Box autocorrelation test up to the second lag and the first-order ARCH LM-test. $K(v_t)$ and $S(v_t)$ are, respectively, the kurtosis and skewness of the fractional filtered series. $K(u_t)$ and $S(u_t)$ are, respectively, the kurtosis and skewness of the residuals.

	Transition Variable: Past Return							
	p	M	AC	ARCH	$K(v_t)$	$S(v_t)$	$K(u_t)$	$S(u_t)$
AA	1	0	0.996	0.211	3.808	0.133	3.874	0.157
AXP	1	1	0.328	0.399	4.722	0.069	4.807	0.046
BA	1	0	0.477	0.112	3.617	0.117	3.646	0.119
BAC	2	0	0.976	0.048	4.387	0.515	4.318	0.484
CAT	2	0	0.994	0.059	3.887	0.130	3.904	0.155
CSCO	1	1	0.814	0.001	3.917	0.186	3.921	0.145
CVX	1	1	0.662	0.002	3.910	0.196	4.061	0.183
DD	1	0	0.998	0.000	3.968	0.128	3.979	0.124
DIS	1	0	0.166	0.001	4.383	0.287	4.418	0.295
GE	1	1	0.822	0.018	3.699	0.187	3.671	0.150
HD	1	0	0.897	0.057	3.903	0.050	3.942	0.055
HPQ	1	0	0.403	0.002	4.382	0.180	4.381	0.169
IBM	1	0	0.933	0.399	4.407	0.027	4.474	0.038
INTC	1	0	0.958	0.000	3.727	0.048	3.729	0.052
JNJ	2	1	0.897	0.000	4.650	0.208	4.589	0.166
JPM	1	1	0.833	0.018	4.143	0.192	4.105	0.141
KFT	3	0	0.975	0.548	4.602	0.295	4.753	0.332
KO	1	0	0.918	0.035	4.561	0.192	4.678	0.192
MCD	1	0	0.547	0.231	4.473	0.307	4.508	0.316
MMM	1	1	0.947	0.008	4.760	0.117	4.888	0.099
MO	1	0	0.621	0.006	4.840	0.553	4.897	0.570
MRK	2	0	0.988	0.399	6.556	0.680	6.715	0.710
MSFT	1	0	0.989	0.467	4.562	0.037	4.560	0.030
PFE	3	0	0.999	0.004	4.809	0.381	4.883	0.394
PG	1	0	0.291	0.000	5.175	0.193	5.241	0.209
T	1	0	0.990	0.016	4.104	0.276	4.144	0.270
UTX	1	0	0.954	0.003	4.495	0.149	4.530	0.162
VZ	1	0	0.341	0.021	4.033	0.259	4.094	0.263
WMT	1	0	0.983	0.184	4.034	0.138	4.133	0.170
XOM	1	1	0.935	0.001	3.927	0.214	4.007	0.204

TABLE 7. FULL SAMPLE RESULTS: BIC SEARCH AND TIME.

The table displays statistics for models with time as transition variables and number of regimes determined by the LM tests with 5% as the initial significance level and $C = 1/2$; p is the autoregressive order. M is the number of nonlinear terms. AC and $ARCH$ are, respectively, the p-values of the Ljung-Box autocorrelation test up to the second lag and the first-order ARCH LM. $K(v_t)$ and $S(v_t)$ are, respectively, the kurtosis and skewness of the fractional filtered series. $K(u_t)$ and $S(u_t)$ are, respectively, the kurtosis and skewness of the residuals.

			Transition Variable: Time					
	p	M	AC	ARCH	$K(v_t)$	$S(v_t)$	$K(u_t)$	$S(u_t)$
AA	1	0	0.996	0.211	3.808	0.133	3.874	0.157
AXP	1	0	0.132	0.292	4.722	0.069	4.814	0.080
BA	1	0	0.477	0.112	3.617	0.117	3.646	0.119
BAC	2	0	0.976	0.048	4.387	0.515	4.318	0.484
CAT	2	0	0.994	0.059	3.887	0.130	3.904	0.155
CSCO	1	0	0.938	0.000	3.917	0.186	3.916	0.185
CVX	1	0	0.427	0.007	3.910	0.196	3.946	0.203
DD	1	0	0.998	0.000	3.968	0.128	3.979	0.124
DIS	1	1	0.055	0.005	4.383	0.287	4.458	0.287
GE	1	1	0.923	0.007	3.699	0.187	3.752	0.174
HD	1	0	0.897	0.057	3.903	0.050	3.942	0.055
HPQ	1	0	0.403	0.002	4.382	0.180	4.381	0.169
IBM	1	0	0.933	0.399	4.407	0.027	4.474	0.038
INTC	1	0	0.958	0.000	3.727	0.048	3.729	0.052
JNJ	2	0	0.998	0.002	4.650	0.208	4.737	0.224
JPM	1	1	0.859	0.004	4.143	0.192	4.131	0.173
KFT	3	0	0.975	0.548	4.602	0.295	4.753	0.332
KO	1	0	0.918	0.035	4.561	0.192	4.678	0.192
MCD	1	0	0.547	0.231	4.473	0.307	4.508	0.316
MMM	1	0	0.994	0.001	4.760	0.117	4.825	0.118
MO	1	0	0.621	0.006	4.840	0.553	4.897	0.570
MRK	2	0	0.988	0.399	6.556	0.680	6.715	0.710
MSFT	1	0	0.989	0.467	4.562	0.037	4.560	0.030
PFE	3	0	0.999	0.004	4.809	0.381	4.883	0.394
PG	1	0	0.291	0.000	5.175	0.193	5.241	0.209
T	1	0	0.990	0.016	4.104	0.276	4.144	0.270
UTX	1	0	0.954	0.003	4.495	0.149	4.530	0.162
VZ	1	0	0.341	0.021	4.033	0.259	4.094	0.263
WMT	1	0	0.983	0.184	4.034	0.138	4.133	0.170
XOM	1	1	0.973	0.002	3.927	0.214	3.886	0.208

TABLE 8. FORECASTING RESULTS BHLS: 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 daily returns as transition variable. "STARFIMA II" uses time as transition variable. "STAR" is a short-memory model with past returns 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. "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. "Ratio II" is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear 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				STAR			
	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II
AA	0.995	0.191	1.019	0.084	0.988	0.03	1.0129	0.186	1.013	0.053	1.038	0.005	1.634	0.00	1.6740	0.000
AXP	1.011	0.073	1.015	0.223	1.005	0.23	1.0094	0.316	1.015	0.033	1.019	0.157	1.834	0.00	1.8410	0.000
BA	0.998	0.306	1.011	0.138	0.997	0.27	1.0102	0.156	1.009	0.101	1.022	0.020	1.605	0.00	1.6260	0.000
BAC	1.015	0.021	1.009	0.345	1.014	0.02	1.0095	0.354	1.029	0.002	1.024	0.169	1.970	0.00	1.9590	0.000
CAT	1.001	0.344	1.018	0.055	1.003	0.21	1.0201	0.041	1.001	0.344	1.018	0.055	1.702	0.00	1.7310	0.000
CSCO	1.000	0.477	1.021	0.009	1.005	0.16	1.0266	0.004	1.001	0.439	1.022	0.010	1.685	0.00	1.7210	0.000
CVX	0.998	0.316	1.027	0.085	0.996	0.20	1.0256	0.104	1.002	0.422	1.031	0.047	1.779	0.00	1.8310	0.000
DD	0.999	0.429	1.016	0.088	1.000	0.46	1.0164	0.080	1.034	0.000	1.051	0.000	1.714	0.00	1.7430	0.000
DIS	0.999	0.440	1.021	0.068	0.999	0.44	1.0210	0.068	1.068	0.000	1.091	0.000	1.689	0.00	1.7260	0.000
GE	0.997	0.345	1.037	0.028	0.997	0.34	1.0375	0.028	1.095	0.000	1.139	0.000	1.784	0.00	1.8560	0.000
HD	0.995	0.150	1.019	0.076	0.999	0.41	1.0234	0.043	0.998	0.354	1.022	0.038	1.668	0.00	1.7080	0.000
HPQ	1.002	0.342	1.026	0.001	1.001	0.38	1.0258	0.001	1.000	0.480	1.024	0.002	1.663	0.00	1.7040	0.000
IBM	0.999	0.421	1.027	0.006	0.996	0.24	1.0241	0.013	1.028	0.001	1.057	0.000	1.673	0.00	1.7210	0.000
INTC	1.001	0.379	1.013	0.064	1.001	0.37	1.0139	0.064	1.011	0.030	1.023	0.008	1.616	0.00	1.6360	0.000
JNJ	1.003	0.287	1.031	0.068	1.002	0.35	1.0301	0.075	1.016	0.007	1.044	0.018	1.644	0.00	1.6900	0.000
JPM	1.012	0.032	1.045	0.007	1.010	0.06	1.0433	0.009	1.051	0.000	1.086	0.000	1.948	0.00	2.0120	0.000
KFT	0.996	0.169	1.001	0.446	0.998	0.35	1.0037	0.313	1.020	0.004	1.025	0.003	1.431	0.00	1.4370	0.000
KO	1.000	0.482	1.015	0.089	0.999	0.39	1.0132	0.111	1.004	0.284	1.019	0.057	1.592	0.00	1.6160	0.000
MCD	0.994	0.109	1.004	0.270	0.994	0.10	1.0049	0.270	0.994	0.109	1.004	0.270	1.552	0.00	1.5670	0.000
MMM	0.991	0.010	0.997	0.367	0.990	0.00	0.9966	0.336	1.004	0.310	1.009	0.180	1.605	0.00	1.6140	0.000
MO	0.996	0.163	0.998	0.395	0.995	0.12	0.9986	0.355	1.003	0.298	1.005	0.224	1.494	0.00	1.4980	0.000
MRK	0.993	0.004	1.002	0.364	0.993	0.00	1.0024	0.364	0.993	0.004	1.002	0.364	1.520	0.00	1.5350	0.000
MSFT	1.000	0.500	1.015	0.068	1.001	0.44	1.0162	0.057	1.122	0.000	1.139	0.000	1.600	0.00	1.6240	0.000
PFE	0.991	0.009	1.000	0.472	0.990	0.00	0.9995	0.448	0.992	0.071	1.001	0.469	1.619	0.00	1.6330	0.000
PG	0.993	0.041	1.014	0.152	0.993	0.04	1.0141	0.152	1.107	0.000	1.130	0.000	1.576	0.00	1.6090	0.000
T	0.991	0.022	1.004	0.309	0.991	0.02	1.0042	0.309	0.991	0.022	1.004	0.309	1.676	0.00	1.6970	0.000
UTX	0.999	0.392	1.037	0.007	1.004	0.24	1.0429	0.002	1.015	0.036	1.053	0.000	1.636	0.00	1.6990	0.000
VZ	0.994	0.101	1.018	0.066	0.994	0.10	1.0181	0.066	0.995	0.181	1.019	0.054	1.655	0.00	1.6960	0.000
WMT	0.995	0.139	1.018	0.053	0.994	0.11	1.0173	0.052	1.002	0.362	1.026	0.010	1.654	0.00	1.6940	0.000
XOM	1.000	0.489	1.022	0.138	0.991	0.09	1.0139	0.258	1.016	0.024	1.038	0.028	1.812	0.00	1.8510	0.000

TABLE 9. FORECASTING RESULTS BHLS: FIVE-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 daily returns as transition variable. “STARFIMA II” uses time as transition variable. “STAR” is a short-memory model with past returns 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. “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. “Ratio II” is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear 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				STAR			
	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II
AA	0.973	0.001	0.977	0.014	0.978	0.004	0.983	0.069	0.986	0.050	0.990	0.239	1.6266	0.000	1.6340	0.000
AXP	0.999	0.467	0.998	0.452	1.001	0.466	1.001	0.483	1.006	0.366	1.005	0.384	1.7158	0.000	1.7146	0.000
BA	0.974	0.000	0.966	0.000	0.976	0.001	0.967	0.000	0.978	0.002	0.970	0.001	1.5517	0.000	1.5387	0.000
BAC	0.992	0.321	0.994	0.388	0.999	0.478	1.001	0.487	1.009	0.289	1.011	0.314	1.7259	0.000	1.7287	0.000
CAT	0.971	0.001	0.978	0.029	0.971	0.001	0.978	0.029	0.971	0.001	0.978	0.029	1.6702	0.000	1.6828	0.000
CSCO	0.962	0.000	0.949	0.000	0.962	0.000	0.949	0.000	0.963	0.000	0.950	0.000	1.4522	0.000	1.4321	0.000
CVX	0.949	0.001	0.947	0.002	0.953	0.000	0.951	0.002	0.958	0.001	0.956	0.002	1.4099	0.000	1.4071	0.000
DD	0.969	0.000	0.962	0.001	0.976	0.000	0.969	0.006	0.987	0.033	0.980	0.059	1.6202	0.000	1.6089	0.000
DIS	0.967	0.000	0.957	0.000	0.976	0.006	0.966	0.002	0.999	0.446	0.988	0.191	1.5744	0.000	1.5582	0.000
GE	0.976	0.040	0.988	0.198	0.984	0.119	0.997	0.415	0.998	0.430	1.011	0.226	1.6611	0.000	1.6829	0.000
HD	0.970	0.001	0.982	0.054	0.972	0.001	0.983	0.076	0.974	0.003	0.986	0.124	1.6383	0.000	1.6579	0.000
HPQ	0.973	0.000	0.954	0.000	0.975	0.001	0.957	0.000	0.978	0.007	0.960	0.000	1.5138	0.000	1.4847	0.000
IBM	0.982	0.032	0.984	0.047	0.984	0.046	0.986	0.070	0.988	0.105	0.990	0.150	1.5690	0.000	1.5714	0.000
INTC	0.976	0.001	0.969	0.001	0.977	0.001	0.969	0.001	0.978	0.002	0.971	0.002	1.4528	0.000	1.4419	0.000
JNJ	0.984	0.084	0.984	0.054	0.985	0.062	0.985	0.046	0.992	0.215	0.992	0.212	1.5610	0.000	1.5609	0.000
JPM	0.984	0.117	0.970	0.001	0.997	0.398	0.982	0.043	1.016	0.143	1.002	0.441	1.6374	0.000	1.6141	0.000
KFT	0.987	0.083	0.964	0.051	0.991	0.216	0.969	0.082	1.011	0.232	0.988	0.309	1.3341	0.000	1.3041	0.000
KO	0.981	0.025	0.971	0.011	0.982	0.031	0.971	0.014	0.983	0.042	0.973	0.019	1.5750	0.000	1.5584	0.000
MCD	0.969	0.001	0.928	0.000	0.969	0.001	0.928	0.000	0.969	0.001	0.928	0.000	1.5038	0.000	1.4411	0.000
MMM	0.956	0.001	0.961	0.000	0.965	0.001	0.969	0.000	0.975	0.002	0.979	0.005	1.5387	0.000	1.5457	0.000
MO	0.975	0.001	0.971	0.002	0.976	0.001	0.972	0.002	0.977	0.002	0.974	0.004	1.4066	0.000	1.4012	0.000
MRK	0.977	0.016	0.971	0.010	0.977	0.016	0.971	0.010	0.977	0.016	0.971	0.010	1.4567	0.000	1.4472	0.000
MSFT	0.967	0.000	0.967	0.002	0.978	0.011	0.978	0.025	0.996	0.334	0.995	0.355	1.5386	0.000	1.5384	0.000
PFE	0.961	0.000	0.909	0.000	0.961	0.000	0.909	0.000	0.962	0.000	0.910	0.000	1.5808	0.000	1.4962	0.000
PG	0.958	0.000	0.929	0.000	0.965	0.000	0.936	0.000	0.978	0.004	0.948	0.000	1.5133	0.000	1.4678	0.000
T	0.969	0.000	0.959	0.001	0.969	0.000	0.959	0.001	0.969	0.000	0.959	0.001	1.4783	0.000	1.4634	0.000
UTX	0.975	0.000	0.977	0.011	0.978	0.001	0.980	0.029	0.982	0.009	0.985	0.096	1.5786	0.000	1.5822	0.000
VZ	0.970	0.001	0.970	0.002	0.971	0.001	0.970	0.002	0.971	0.001	0.971	0.002	1.4885	0.000	1.4878	0.000
WMT	0.967	0.000	0.967	0.000	0.968	0.000	0.968	0.000	0.971	0.001	0.970	0.001	1.4811	0.000	1.4804	0.000
XOM	0.965	0.000	0.956	0.000	0.971	0.000	0.962	0.000	0.979	0.003	0.969	0.000	1.4366	0.000	1.4232	0.000

TABLE 10. FORECASTING RESULTS BHLS: 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 daily returns as transition variable. "STARFIMA II" uses time as transition variable. "STAR" is a short-memory model with past returns 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. "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. "Ratio II" is the RMSE of the linear and nonlinear ARFIMA models divided by the RMSE of the nonlinear 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				STAR			
	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II	Ratio I	GW I	Ratio II	GW II
AA	0.983	0.121	0.971	0.005	0.987	0.157	0.974	0.013	0.992	0.252	0.979	0.048	1.7039	0.000	1.6825	0.000
AXP	1.008	0.379	0.984	0.204	1.012	0.333	0.988	0.256	1.017	0.273	0.992	0.343	1.8831	0.000	1.8384	0.000
BA	0.979	0.034	0.970	0.028	0.980	0.040	0.971	0.032	0.982	0.054	0.973	0.039	1.5777	0.000	1.5629	0.000
BAC	1.004	0.439	1.009	0.379	1.010	0.352	1.016	0.305	1.020	0.236	1.025	0.207	1.8216	0.000	1.8309	0.000
CAT	0.978	0.048	0.978	0.068	0.978	0.048	0.978	0.068	0.978	0.048	0.978	0.068	1.6467	0.000	1.6481	0.000
CSCO	0.964	0.012	0.957	0.039	0.964	0.012	0.958	0.039	0.965	0.012	0.958	0.040	1.4298	0.000	1.4202	0.000
CVX	0.934	0.024	0.908	0.017	0.938	0.023	0.911	0.016	0.942	0.024	0.915	0.017	1.3181	0.000	1.2806	0.000
DD	0.967	0.003	0.954	0.003	0.972	0.004	0.960	0.006	0.980	0.016	0.967	0.018	1.6065	0.000	1.5858	0.000
DIS	0.976	0.050	0.961	0.008	0.990	0.257	0.975	0.056	1.026	0.073	1.010	0.288	1.6796	0.000	1.6535	0.000
GE	0.978	0.164	0.974	0.046	0.984	0.244	0.981	0.091	0.994	0.391	0.990	0.239	1.7806	0.000	1.7742	0.000
HD	0.975	0.038	0.972	0.035	0.976	0.048	0.974	0.043	0.979	0.068	0.976	0.061	1.7059	0.000	1.7010	0.000
HPQ	0.974	0.010	0.966	0.001	0.977	0.017	0.969	0.004	0.980	0.037	0.972	0.012	1.5014	0.000	1.4899	0.000
IBM	0.992	0.293	0.986	0.156	0.993	0.319	0.987	0.176	0.996	0.385	0.990	0.228	1.6106	0.000	1.6006	0.000
INTC	0.985	0.096	0.989	0.181	0.985	0.095	0.989	0.178	0.986	0.105	0.990	0.191	1.5253	0.000	1.5317	0.000
JNJ	1.002	0.455	0.997	0.427	0.997	0.438	0.993	0.293	1.003	0.424	0.998	0.452	1.5982	0.000	1.5905	0.000
JPM	0.989	0.292	0.969	0.023	1.005	0.409	0.985	0.163	1.029	0.137	1.009	0.341	1.7141	0.000	1.6798	0.000
KFT	1.004	0.411	0.978	0.196	1.007	0.358	0.982	0.245	1.034	0.089	1.008	0.402	1.2642	0.000	1.2325	0.000
KO	0.992	0.290	0.989	0.203	0.992	0.291	0.989	0.205	0.993	0.298	0.989	0.212	1.6323	0.000	1.6263	0.000
MCD	0.978	0.025	0.968	0.012	0.978	0.025	0.968	0.012	0.978	0.025	0.968	0.012	1.4409	0.000	1.4266	0.000
MMM	0.959	0.049	0.955	0.034	0.966	0.051	0.962	0.035	0.974	0.063	0.970	0.045	1.5167	0.000	1.5098	0.000
MO	0.972	0.006	0.966	0.006	0.973	0.007	0.967	0.006	0.974	0.010	0.969	0.008	1.3825	0.000	1.3748	0.000
MRK	0.978	0.049	0.984	0.084	0.978	0.049	0.984	0.084	0.978	0.049	0.984	0.084	1.4162	0.000	1.4248	0.000
MSFT	0.975	0.030	0.980	0.062	0.981	0.059	0.986	0.117	0.991	0.231	0.996	0.362	1.5897	0.000	1.5974	0.000
PFE	0.961	0.002	0.937	0.002	0.962	0.002	0.938	0.002	0.964	0.005	0.940	0.003	1.5446	0.000	1.5064	0.000
PG	0.953	0.028	0.922	0.001	0.958	0.020	0.926	0.001	0.965	0.019	0.934	0.001	1.4962	0.000	1.4474	0.000
T	0.965	0.004	0.963	0.009	0.965	0.004	0.963	0.009	0.965	0.004	0.963	0.009	1.4923	0.000	1.4892	0.000
UTX	0.982	0.065	0.968	0.034	0.986	0.085	0.972	0.042	0.991	0.178	0.977	0.071	1.6066	0.000	1.5837	0.000
VZ	0.972	0.036	0.968	0.013	0.972	0.036	0.968	0.013	0.972	0.036	0.968	0.013	1.5263	0.000	1.5204	0.000
WMT	0.967	0.004	0.969	0.004	0.969	0.005	0.971	0.006	0.971	0.010	0.973	0.011	1.4405	0.000	1.4434	0.000
XOM	0.961	0.006	0.943	0.013	0.967	0.008	0.950	0.014	0.975	0.017	0.957	0.019	1.3631	0.000	1.3382	0.000

(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

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 $p = 2$.

Short-Memory Models: 500 observations

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
$C = 1/2$	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

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
$C = 1/2$	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

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
$C = 1/2$	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

	Linear			Nonlinear I			Nonlinear II		
	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH	Gaussian	Fat-Tailed	GARCH
$C = 1/2$	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 Starts	Date Forecasts Starts	Sample Size	BHLS (60 sec) average	BHLS (60 sec) std. dev.	BHLS (60 sec) skewness	BHLS (60 sec) kurtosis	RVmed average	RVmed std. dev.	RVmed skewness	RVmed 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