TEXTO PARA DISCUSSÃO

No. 544

A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries

Michael McAleer
Marcelo C. Medeiros

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

A MULTIPLE REGIME SMOOTH TRANSITION HETEROGENEOUS AUTOREGRESSIVE MODEL FOR LONG MEMORY AND ASYMMETRIES

Michael McAleer
School of Economics and Commerce
University of Western Australia

Marcelo C. Medeiros
Department of Economics
Pontifical Catholic University of Rio de Janeiro

First version: October 2006
Revised: April 2007

ABSTRACT: In this paper we propose a flexible model to capture nonlinearities and long-range dependence in time series dynamics. The new model is a multiple regime smooth transition extension of the Heterogenous Autoregressive (HAR) model, which is specifically designed to model the behavior of the volatility inherent in financial time series. The model is able to describe simultaneously long memory, as well as sign and size asymmetries. A sequence of tests is developed to determine the number of regimes, and an estimation and testing procedure is presented. Monte Carlo simulations evaluate the finite-sample properties of the proposed tests and estimation procedures. We apply the model to several Dow Jones Industrial Average index stocks using transaction level data from the Trades and Quotes database that covers ten years of data. We find strong support for long memory and both sign and size asymmetries. Furthermore, the new model, when combined with the linear HAR model, is viable and flexible for purposes of forecasting volatility.
KEYWORDS: Realized volatility, smooth transition, heterogeneous autoregression, financial econometrics, leverage, sign and size asymmetries, forecasting, risk management, model combination.

ACKNOWLEDGMENTS: The authors wish to acknowledge the insightful comments and suggestions of two referees, Jiti Gao, Eric Hillebrand and Marcel Scharth. The financial support of the Australian Research Council and CNPq, Brazil is gratefully acknowledged.

1. INTRODUCTION

Given the rapid growth in financial markets and the continual development of new and more complex financial instruments, there is an ever-growing need to understand the theoretical and empirical processes underlying the volatility in financial time series. It is well known that the daily returns of financial assets, especially of stocks, can be extremely difficult to predict, although the volatility of the returns seems to be relatively more straightforward to forecast. Thus, it is hardly surprising that financial econometrics, and particularly the modelling of financial volatility, has played such a central role in modern pricing and risk management theories. Andersen, Bollerslev, Christoffersen, and Diebold (2007) provide a recent overview of the literature.

There is, however, an inherent problem in using models where the volatility measure plays a central role. The conditional variance is latent, and hence is not directly observable. Early classes of volatility models used squared daily returns as a measure of volatility. However, as this measure is very noisy, volatility was specified as a latent variable in different models. Useful and popular examples of such models are the (Generalized) Autoregressive Conditional Heteroskedasticity, or (G)ARCH, model of Engle (1982) and Bollerslev (1986), various stochastic volatility models (see, for example, Taylor (1986)), and the exponentially weighted moving averages (EWMA) approach, as advocated by the Riskmetrics methodology (J. P. Morgan 1996). McAleer (2005) gives a recent exposition of a wide range of univariate and multivariate, conditional and stochastic, models of volatility, and Asai, McAleer, and Yu (2006) provide a review of the rapidly growing literature on multivariate stochastic volatility models. However, as observed by Bollerslev (1987), Malmsten and Teräsvirta (2004), and Carnero, Peña, and Ruiz (2004), among others, most of the latent volatility models have been unable to capture simultaneously several important empirical features that are inherent in financial time series.

An empirical regularity which standard latent volatility models fail to describe adequately is the low, but slowly decreasing, autocorrelations in the squared returns that are associated with the high excess kurtosis of returns. In this sense, the assumption of Gaussian standardized returns has been refuted in many studies, and heavy-tailed distributions have been used instead. Furthermore, there is strong evidence of long-range dependence in the conditional volatility of financial time series. One possible explanation of long memory is aggregation. Volatility is modelled as a sum of different processes, each with low persistence. The aggregation induces long memory, as noted by Granger (1980), LeBaron (2001), Fouque, Papanicolaou, Sircar, and Sølna (2003), Davidson and Sibbertsen (2005), Hyung, Poon, and Granger (2007), and Lieberman and Phillips (2007).

On the other hand, the literature has also documented asymmetric effects in volatility. Starting with Black (1976), it has been observed that there is an asymmetric response of the conditional variance of the series to unexpected news, as represented by shocks. Financial markets become more volatile in response to “bad news” (or negative shocks) than to “good news” (or positive shocks). Goetzmann, Ibbotson, and Peng (2001) found evidence of asymmetric sign effects in volatility as far back as 1857 for the NYSE. They report that unexpected negative shocks in monthly returns of the NYSE from 1857 to 1925 increase volatility almost twice as much as do equivalent positive shocks in returns of a similar magnitude. Similar results were also reported by Schwert (1990). The above mentioned asymmetry has motivated a large number of different asymmetric latent volatility models.

However, most volatility models have been unable to describe simultaneously both nonlinear effects and long memory. The statistical consequences of neglecting or misspecifying nonlinearities have been discussed in the context of structural breaks in the GARCH literature by Diebold (1986), Lamoureux and Lastrapes (1990), Mikosch and Starica (2004), and Hillebrand (2005), and in the literature on long memory models by Lobato and Savin (1998), Diebold and Inoue (2001), Granger and Teräsvirta (2001), Granger and Hyung (2004), and Smith (2005). Neglected changes in levels or persistence induce estimated high persistence, which has often been called "spurious" high persistence (see also Hillebrand and Medeiros (2007) for a recent application).

In the opposite direction, it is also possible to misinterpret data-generating high persistence (in the form of long memory or unit roots) for nonlinearity. Spuriously estimated structural breaks have been reported for unit root processes by Nunes, Kuan, and Newbold (1995) and Bai (1998), and have been extended to long memory processes by Hsu (2001).

The search for an adequate framework for the estimation and prediction of the conditional variance of financial asset returns has led to the analysis of high frequency intraday data. Merton (1980) noted that the variance over a fixed interval can be estimated arbitrarily, although accurately, as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency. More recently, Andersen and Bollerslev (1998) showed that ex post daily foreign exchange volatility is best measured by aggregating 288 squared five-minute returns. The five-minute frequency is a trade-off between accuracy, which is theoretically optimized using the highest possible frequency, and microstructure noise, which can arise through the bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among other factors (see Madhavan (2000) and Biais, Glosten, and Spatt (2005) for recent reviews).

Ignoring the remaining measurement error, which can be problematic, the ex post volatility essentially becomes "observable". Andersen and Bollerslev (1998), Hansen and Lunde (2005), and Patton (2005) used the realized volatility to evaluate the out-of-sample forecasting performance of several latent volatility models. As volatility becomes "observable", it can be modelled directly, rather than being treated as a latent variable. Based on the theoretical results of Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002), several recent studies have documented the properties of realized volatilities that are constructed from high frequency data.

In this paper, we propose a simple model that merges long memory and nonlinearities. The new specification is a multiple regime generalization of the Heterogeneous Autoregression (HAR) that was suggested by Corsi (2004). The HAR model is inspired by the Heterogeneous Market Hypothesis and the asymmetric propagation of volatility between long and short time horizons. The HAR model has been applied with success in modelling and forecasting realized variance (Andersen, Bollerslev, and Diebold 2007). The new model is called the Heterogeneous Autoregression with Multiple-Regime

Smooth Transition (HARST) model, which combines ingredients from the HAR and the Smooth Transition Autoregressive (STAR) models (Chan and Tong 1986, Teräsvirta 1994). The HARST model has the main advantage of modelling simultaneously long-range dependence, as well as incorporating sign and size asymmetries in a simple manner. The choice of the variable that drives the regime switches makes it possible to describe interesting dynamics, such as general asymmetry and leverage. The number of regimes is determined by a simple and easily-implemented sequence of tests that circumvents the identification problem in the nonlinear time series literature, and the model estimation and testing procedure is analysed. A Monte Carlo simulation evaluates the finite-sample properties of the proposed modelling cycle. An empirical application with 16 stocks from the Dow Jones Industrial Average (DJIA) gives strong support in favor of the new model. In particular, evidence is shown of long-range dependence and both sign and size asymmetries in the realized volatility of the series. Finally, the combination of the linear and nonlinear HAR models produces superior one-day-ahead forecasts.

The paper is organized as follows. Section 2 introduces the theoretical foundations and describes the salient features of realized volatility. Section 3 presents the model and discusses estimation issues. A formal test for an additional regime is introduced in Section 4. Section 5 describes the model building procedure, in which the number of regimes is determined by a simple and easily-implemented sequence of tests. Monte Carlo simulations are presented in Section 6. The empirical results are discussed in Section 7. Section 8 gives some concluding comments.

2. Realized Volatility

Suppose that at day t the logarithmic prices of a given asset follow a continuous time diffusion:

$$dp(t+\tau) = \mu(t+\tau) + \sigma(t+\tau)dW(t+\tau),\ 0 \leq \tau \leq 1,\ t = 1, 2, 3, \dots, \tag{1}$$

where $p(t+\tau)$ is the logarithmic price at time $(t+\tau)$, $\mu(t+\tau)$ is the drift component, $\sigma(t+\tau)$ is the instantaneous volatility (or standard deviation), and $dW(t+\tau)$ is standard Brownian motion.

Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002) showed that the daily compound returns, defined as $r_t = p(t) - p(t-1)$, are Gaussian conditionally on $\mathcal{F}_t = \sigma(p(s),\ s \leq t)$, the σ-algebra (information set) generated by the sample paths of p, such that

$$r_t|_{\mathcal{F}_t} \sim \mathsf{N}\left(\int_0^1 \mu(t-1+\tau)d\tau, \int_0^1 \sigma^2(t-1+\tau)d\tau\right). \tag{2}$$

The term $IV_t = \int_0^1 \sigma^2(t-1+\tau)d\tau$ is known as the *integrated variance*, which is a measure of the day-t ex post volatility. In this sense, the integrated variance is the object of interest.

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

ith intra-period return of day t. Realized variance is defined as

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2. \tag{3}$$

Realized volatility is the square-root of RV_t.

The search for asymptotically unbiased, consistent and efficient methods for measuring realized volatility in the presence of microstructure noise has been one of the most active research topics in financial econometrics over the last few years. While early references in the literature, such as Andersen, Bollerslev, Diebold, and Ebens (2001), advocated the simple selection of an arbitrary lower frequency (typically 5-15 minutes) to balance accuracy and the dissipation of microstructure bias, a procedure that is known as sparse sampling, some recent articles have developed estimators that dominate this procedure. These contributions fall in several categories: some examples are the selection of an optimal sampling frequency in sparse sampling, as in Bandi and Russell (2005a, 2005b, 2006), the subsampling method, as in Zhang, Mykland, and Aït-Sahalia (2005), the kernel-based estimators of Zhou (1996), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a, 2006b), and Hansen and Lunde (2006) and MA filtering, as in Hansen, Large, and Lunde (2007). McAleer and Medeiros (2007) review these and other methods, and provide a comparison of the alternative methods.

Three consistent methods of estimation are presently available: the realized kernel estimators of Barndorff- Nielsen, Hansen, Lunde and Shephard (2006a, 2006b), the modified MA filter of Hansen, Large, and Lunde (2007), and the two time scales realized volatility estimator of Zhang, Mykland, and Aït-Sahalia (2005), which is our choice for the empirical part of this paper. Aït-Sahalia, Mykland, and Zhang (2005) show that the estimator works well when the hypothesis of independent microstructure noise is violated, is stable with regard to the choice of grids, and yields estimates that are close to the more efficient but also more computationally-demanding Multi-Scale approach of Zhang (2005).

Several salient features of realized volatility have been identified in the literature:

(1) the unconditional distribution of daily returns exhibits excess kurtosis;
(2) daily returns are not autocorrelated (except for the first order, in some cases);
(3) daily returns that are standardized by the realized variance measure are almost Gaussian;
(4) the unconditional distribution of realized variance and volatility is distinctly non-normal, and is extremely right-skewed;
(5) realized volatility does not seem to have a unit root, but there is strong evidence of fractional integration.

On the other hand, the natural logarithm of the volatility has the following empirical regularities:

(1) the logarithm of realized volatility is close to normal;
(2) the logarithm of realized volatility displays long-range dependence.

The model described in Section 3 aims to model not only long-range dependence found in realized volatility but also describe risk-return asymmetries as documented in the latent volatility literature.

3. Long Memory and Nonlinearity in Realized Volatility

3.1. **A Brief Review of the Literature and Stylized Facts.** As observed in the Introduction, several nonlinear conditional and stochastic volatility models have been proposed in the literature to describe asymmetries in volatility. In most of these models, volatility is a latent variable. Nelson (1991) proposed the Exponential GARCH (EGARCH) model, in which the natural logarithm of the conditional variance is modelled as a nonlinear ARMA model, with a term that introduces asymmetry in the dynamics of the conditional variance, according to the sign of the lagged returns. Glosten, Jagannathan, and Runkle (1993) proposed the GJR model, where the impact of the lagged squared returns on the current conditional variance changes according to the sign of the past return. A similar specification, known as the Threshold GARCH (TGARCH) model, was developed by Rabemanjara and Zakoian (1993) and Zakoian (1994). Ding, Granger, and Engle (1993) discussed the Asymmetric Power ARCH model, which nests several GARCH specifications (see Ling and McAleer (2002) for a derivation of the necessary and sufficient moment conditions).

Engle and Ng (1993) popularized the news impact curve (NIC) as a measure of how new information is incorporated into volatility estimates. The authors also developed formal statistical tests to check the presence of asymmetry in the volatility dynamics. More recently, Fornari and Mele (1997) generalized the GJR model by allowing all the parameters to change according to the sign of the past return. Their proposal is known as the Volatility Switching GARCH (VSGARCH) model. Based on the Smooth Transition AutoRegressive (STAR) model, Hagerud (1997) and Gonzalez-Rivera (1998) proposed the Smooth Transition GARCH (STGARCH) model. While the latter only considered the Logistic STGARCH (LSTGARCH) model, the former discussed both the Logistic and Exponential STGARCH (ESTGARCH) alternatives. In the logistic STGARCH specification, the dynamics of volatility are very similar to those of the GJR model and depends on the sign of the past returns. The difference is that the former allows for a smooth transition between regimes. In the EST-GARCH model, the sign of the past returns does not play any role in the dynamics of the conditional variance, but it is the magnitude of the lagged squared return that is the source of asymmetry.

Anderson, Nam, and Vahid (1999) combined the ideas of Fornari and Mele (1997), Hagerud (1997), and Gonzalez-Rivera (1998) and proposed the Asymmetric Nonlinear Smooth Transition GARCH (ANSTGARCH) model, and found evidence in support of their specification. Medeiros and Veiga (2004) proposed the Flexible Coefficient GARCH (FCGARCH) model, which is a multiple regime generalization of several models in the literature. The authors found strong support of sign and size asymmetries in volatility. Furthermore, an empirical example with ten stock indexes shows evidence of two regimes for six series and three regimes for other four series. Furthermore, for all series with three regimes, the GARCH model associated with the first regime, representing very negative returns ("very bad news"), is explosive. The model in the middle regime, related to tranquil periods, has a slightly lower persistence than the standard estimated GARCH(1,1) models in the literature. Finally, the third regime, representing large positive returns, has an associated GARCH(1,1) specification that is significantly less persistent than the others.

Inspired by the Threshold Autoregressive (TAR) model, Li and Li (1996) proposed the Double Threshold ARCH (DTARCH) model. Liu, Li, and Li (1997) generalized the model and proposed the Double Threshold GARCH (DTGARCH) process to model both the conditional mean and conditional variance as threshold processes. More recently, based on the regression tree literature, Audrino and Bühlmann (2001) proposed the Tree Structured GARCH model to describe multiple limiting regimes in volatility. Caporin and McAleer (2006) developed a dynamic asymmetric univariate GARCH model. When the regime switches are driven by a Markov Chain, the main references are Hamilton and Susmel (1994), Cai (1994), and Gray (1996).

In the class of stochastic volatility (SV) models, several asymmetric models have been developed. One of the first references is Harvey and Shephard (1996). So, Lam, and Li (1998) and Kalimipalli and Susmel (2007) discussed SV models with Markovian regime switches, while So, Li, and Lam (2002) considered a threshold SV specification. Asai and McAleer (2005) proposed a dynamic asymmetric leverage model that accommodates the direct correlation between returns and volatility as well as sign and size threshold effects, and Omori, Chib, Shephard, and Nakajima (2007) developed an SV model with leverage (see also Asai and McAleer (2006, 2007) for different asymmetric SV models). Yu (2005) also considered a SV model with leverage effects. Cappuccio, Lubian, and Davide (2006) provided empirical evidence on asymmetry in financial returns using a simple stochastic volatility model which allows a parsimonious yet flexible treatment of both skewness and heavy tails in the conditional distribution of returns.

With respect to long memory, Baillie, Bollerslev, and Mikkelsen (1996) is one of the main references. The authors proposed the Fractionally Integrated GARCH (FIGARCH) model as a viable alternative to model long range dependence in volatility. Giraitis, Robinson, and Surgailis (2004) considered the Leverage ARCH (LARCH) model and discussed both leverage and long memory effects in volatility. Breidt, Crato, and de Lima (1998), Hurvich and Ray (2003), Jensen (2004), and Deo, Hurvich, and Lu (2006) discussed the specification and estimation of SV models with long memory.

In the realized volatility literature, most of the early contributions considered only linear long memory models. Martens, van Dijk, and de Pooter (2004) were the first to introduce simultaneously long-range dependence, asymmetries and structural breaks into a realized volatility model. The authors also evaluated the relevance of the days of the week and presented a detailed and exhaustive empirical application. Their specification belongs to the class of nonlinear Autoregressive Fractionally Integrated (ARFI) models. However, they did not consider tests of linearity or the presence of more than two limiting regimes. More recently, Scharth and Medeiros (2006) proposed a multiple regime tree structure model to describe the behavior of realized volatility, where past cumulative returns drive the regime switches. Although a formal model building procedure was developed, the proposed specification did not take into account possible long memory that might be caused by aggregation, among other possibilities. The authors considered that the long range dependence is caused by regime switches. Hillebrand and Medeiros (2006) suggested a model that generalizes the approach developed in Martens, van Dijk, and de Pooter (2004) by merging fractionally integrated process with nonlinearity and asymmetry. The authors also considered a volatility-in-mean effect and developed a formal test of linearity following the ideas in van Dijk, Franses, and Paap (2002). However, the estimation of the fractional integration parameter can prove very difficult and noisy.

In this paper we extend the ingredients of Martens, van Dijk, and de Pooter (2004), Scharth and Medeiros (2006), and Hillebrand and Medeiros (2006), and propose a model that accommodates long-range dependence in a very simple manner for straightforward estimation. Asymmetries and nonlinearity are developed in a smooth transition environment. A formal sequence of tests is described in order to determine the number of limiting regimes. Furthermore, external exogenous variables can be incorporated into the model structure in a straightforward way.

3.2. **Model Specification.** The Heterogenous Autoregressive (HAR) model was proposed by Corsi (2004) as an alternative to model and forecast realized volatilities, and is inspired by the Heterogenous Market Hypothesis of Müller, Dacorogna, Dav, Olsen, Pictet, and Ward (1993) and the asymmetric propagation of volatility between long and short horizons. Corsi (2004) defines the partial volatility as the volatility generated by a certain market component, and the model is an additive cascade of different partial volatilities (generated by the actions of different types of market participants). At each level of the cascade (or time scale), the unobserved volatility process is assumed to be a function of the past volatility at the same time scale and the expectation of the next period values of the longer term partial volatilities (due to the asymmetric propagation of volatility). Corsi (2004) showed that by straightforward recursive substitutions of the partial volatilities, this additive volatility cascade leads to a simple restricted linear autoregressive model with the feature of considering volatilities realized over different time horizons. The heterogeneity of the model derives from the fact that at each time scale, the partial volatility is described by a different autoregressive structure.

In this paper, we generalize the HAR model by introducing multiple regime switching. The proposed model is defined as follows.

DEFINITION 1. *Let*

$$y_{t,h} = \frac{y_t + y_{t-1} + y_{t-2} + \cdots + y_{t-h+1}}{h}, \tag{4}$$

$h \in \mathbb{Z}_+$, $\boldsymbol{\iota} = (\iota_1, \ldots, \iota_p)' \in \mathbb{Z}_+^p$ *be a set of indexes where* $\iota_1 < \iota_2 < \cdots < \iota_p$, *and* $\mathbf{x}_t = \left(1, y_{t-1,\iota_1}, \ldots, y_{t-1,\iota_p}\right)' \in \mathbb{R}^{p+1}$. *A time series* $\{y_t\}_{t=1}^T$ *follows a Multiple-Regime Smooth Transition Heterogenous Autoregressive (HARST) model with* $M+1$ *limiting regimes if*

$$y_t = G(\mathbf{x}_t, z_t; \boldsymbol{\psi}) + \varepsilon_t = \boldsymbol{\beta}_0'\mathbf{x}_t + \sum_{m=1}^{M} \boldsymbol{\beta}_m'\mathbf{x}_t f(z_t; \gamma_m, c_m) + \varepsilon_t, \tag{5}$$

where $G(\mathbf{x}_t, z_t; \boldsymbol{\psi})$ *is a nonlinear function of the variables* $\mathbf{x}_t$ *and* z_t, *and is indexed by the vector of parameters* $\boldsymbol{\psi} \in \mathbb{R}^{(M+1)(p+1)+2M}$, $f(z_t; \gamma_m, c_m)$ *is the logistic function given by*

$$f(z_t; \gamma_m, c_m) = \frac{1}{1 + e^{-\gamma_M (z_t - c_m)}}, \tag{6}$$

and ε_t *is a random noise.*

Typical values for the hyper-parameter h in equation (5) are: one (daily volatility), five (weekly volatility), and 22 (monthly volatility). The main advantage of the HARST model is that it can capture both long-range dependence and regime switches (and hence asymmetric effects) in a very simple way. It is clear that $f(z_t;\gamma_m,c_m)$ is a monotonically increasing function, such that $f(z_t;\gamma_m,c_m)\to 1$ as $z_t\to\infty$ and $f(z_t;\gamma_m,c_m)\to 0$ as $z_t\to-\infty$. The parameter γ_m, $m=1,\ldots,M$, is called the *slope parameter* and determines the speed of the transition between two limiting regimes. When $\gamma_m\to\infty$, the logistic function becomes a step function, and the HARST model becomes a threshold-type specification. The variable z_t is known as the *transition variable*. There are several possible choices for z_t. For example, suppose that y_t is the logarithm of the realized volatility and set $z_t=r_{t-1}$, where r_{t-1} is the return of a given asset at time $t-1$. Hence, the differences in the dynamics of the conditional variance are modelled according to the sign and size of the shocks in previous returns, which represent previous "news".

The number of limiting regimes is defined by the hyper-parameter M. For example, suppose that in (5), $M=2$, c_1 is highly negative, and c_2 is very positive, so that the resulting HARST model will have 3 limiting regimes that can be interpreted as follows. The first regime may be related to extremely low negative shocks (or "very bad news") and the dynamics of the volatility are driven by $y_t=\boldsymbol{\beta}_0'\mathbf{x}_t+\varepsilon_t$ as $f(r_{t-1};\gamma_m,c_m)\approx 0$, $m=1,2$. In the the middle regime, which represents low absolute returns (or "tranquil periods"), $y_t=(\boldsymbol{\beta}_0+\boldsymbol{\beta}_1)'\mathbf{x}_t+\varepsilon_t$ as $f(r_{t-1};\gamma_m,c_m)\approx 1$ and $f(r_{t-1};\gamma_2,c_2)\approx 0$. Finally, the third regime is related to high positive shocks (or "very good news") and $y_t=(\boldsymbol{\beta}_0+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)'\mathbf{x}_t+\varepsilon_t$, as $f(r_{t-1};\gamma_i,c_i)\approx 1$, $i=1,2$.

Another interesting choice is $z_t=y_{t-k}$ or $z_t=y_{t-k,t-1}$. In the case where y_t is the logarithm of the realized volatility, this particular choice of transition variable means that regime switches are driven by past volatility. Past cumulated returns are also a suitable candidate for transition variables as discussed in Scharth and Medeiros (2006). As the speed of the transitions between different limiting HAR models is determined by the parameter γ_m, $m=1,2$, the multiple regime interpretation of the HARST specification will become clearer as the transitions ($\gamma_m\gg 0$) become more abrupt[1].

The following examples illustrate interesting situations. The daily return of a given asset is given by r_t, $r_{22,t}$ is the cumulated return over the last 22 days, σ_t is logarithm of the daily volatility, and $\{u_t\}$ is a sequence of independently and normally distributed random variables. Consider the following specifications.

[1]If $z_t=t$, the model accommodates smoothly changing parameters. In the limit $\gamma_m\to\infty$, $m=1,\ldots,M$, we have an HAR model with M structural breaks. However, $z_t=t$ will not be considered in this paper as the asymptotic theory has to be changed.

(1) Example 1:

$$
\begin{aligned}
r_t &= \exp(\sigma_t)u_t, \;\; u_t \sim \mathsf{NID}(0,1) \\
\sigma_t &= 0.01 + 0.95\sigma_{t-1} - \\
&\quad (0.006 + 0.60\sigma_{t-1} - 0.25\sigma_{t-1,5} - 0.15\sigma_{t-1,22})\, f(r_{t-1}; 5, -3.0) + \\
&\quad (0.004 + 0.30\sigma_{t-1} - 0.16\sigma_{t-1,5} - 0.09\sigma_{t-1,22})\, f(r_{t-1}; 5, 2.5) + \varepsilon_t, \\
&\quad \varepsilon_t \sim \mathsf{NID}\left(0, 0.5^2\right).
\end{aligned} \tag{7}
$$

(2) Example 2:

$$
\begin{aligned}
r_t &= \exp(\sigma_t)u_t, \;\; u_t \sim \mathsf{NID}(0,1) \\
\sigma_t &= 0.05 + 0.95\sigma_{t-1} - \\
&\quad (0.035 + 0.58\sigma_{t-1} - 0.27\sigma_{t-1,5} - 0.21\sigma_{t-1,22})\, f(r_{22,t-1}; 4, -10) + \\
&\quad (0.03 + 0.30\sigma_{t-1} - 0.20\sigma_{t-1,5} - 0.18\sigma_{t-1,22})\, f(r_{22,t-1}; 4, 13) + \varepsilon_t, \\
&\quad \varepsilon_t \sim \mathsf{NID}\left(0, 0.25^2\right).
\end{aligned} \tag{8}
$$

In both cases above, current volatility depends on past daily volatility, as well as on weekly and monthly past volatilities. In the first example, when the returns are very negative, the logarithm of the volatility is given by a very persistent first-order autoregressive model and longer lags have no influence in the volatility dynamics, such that $\sigma_t = 0.010 + 0.95\sigma_{t-1} + \varepsilon_t$. During "tranquil periods", the logarithm of the volatility follows an HAR model, where weekly and monthly averages influence current values, namely $\sigma_t = 0.004 + 0.35\sigma_{t-1} + 0.25\sigma_{t-1,5} + 0.15\sigma_{t-1,22} + \varepsilon_t$. When the lagged return is very positive, the effects of the first lag are dominant, such that $\sigma_t = 0.008 + 0.65\sigma_{t-1} + 0.09\sigma_{t-1,5} + 0.06\sigma_{t-1,22} + \varepsilon_t$. In the second example, the monthly returns influence the dynamics of volatility and the regime switches are not as frequent as in Example 1.

Figures 1 and 2 show one realization with 3000 observations of the returns and the logarithm of the volatility when the data are generated as in Examples 1 and 2, respectively. It is clear from the graphs that the generated series have strong volatility clustering and extreme observations. Table 1 shows the descriptive statistics for 1000 replications of equations (7) and (8). The table shows the mean, median, standard deviation, minimum and maximum values of the following statistics: mean, standard deviation, kurtosis, and skewness of the simulated daily returns; sum of the first 500 autocorrelations of the absolute and squared daily returns; the GPH (Geweke and Porter-Hudak 1983) estimator of the fractional difference parameter for the absolute returns, squared returns, and log volatility; and the correlation coefficient between the volatility and the lagged return.

Several interesting facts emerge from Table 1. First, in both examples the returns have excess kurtosis and positive skewness. Note that, even with Gaussian errors, the kurtosis coefficient can be much greater than three. In both cases, the volatility process displays long-range dependence. Note that the average estimate of the d parameter is close to the 0.4 usually documented in the empirical literature.

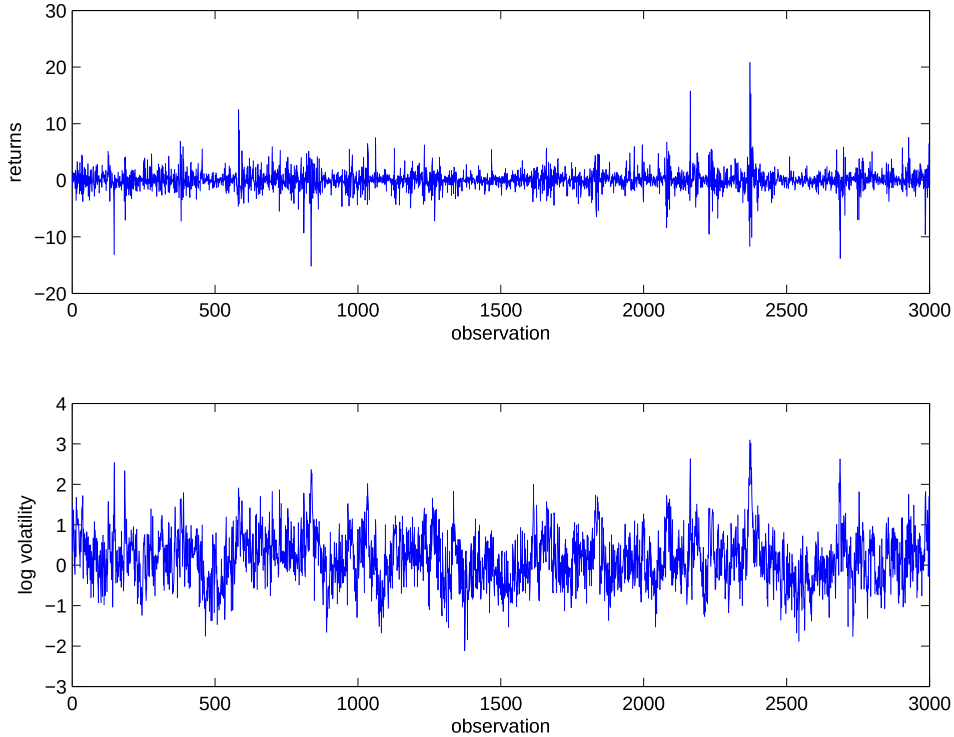

FIGURE 1. Upper panel: one realization of daily returns for Example 1. Lower panel: one realization of the logarithm of the daily volatility for Example 1.

In the first case, there is also a small negative correlation between the lagged return and the volatility process, which indicates the presence of leverage.

3.3. **Probabilistic Properties.** Deriving necessary and sufficient conditions for stationarity and geometric ergodicity of the HARST model is not trivial as it will depend on the particular choice of transition variables and the distribution of the errors. However, it is possible to find a set of sufficient conditions. The core idea is to analyse the HAR model as a restricted (AR) autoregressive model. First, consider the linear HAR specification as follows.

$$y_t = \beta_{00} + \beta_{01} y_{t-1,\iota_1} + \beta_{01} y_{t-1,\iota_2} + \cdots + \beta_{0p} y_{t-1,\iota_p} + \varepsilon_t. \tag{9}$$

It is easy to show that (9) is a restricted AR model given as

$$\begin{aligned} y_t = {} & \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_1 y_{t-\iota_1} + \phi_2 y_{t-(\iota_1+1)} + \cdots + \phi_2 y_{t-\iota_2} \\ & + \phi_3 y_{t-(\iota_2+1)} + \cdots + \phi_3 y_{t-\iota_3} + \cdots + \phi_p y_{t-(\iota_{p-1}+1)} + \cdots + \phi_p y_{t-\iota_p} + \varepsilon_t, \end{aligned} \tag{10}$$

where $\phi_0 = \beta_{00}$ and $\phi_j = \sum_{i=j}^{p} \beta_{0i}$, $j = 1, \ldots, p$.

THEOREM 1. *Suppose that the process $\{y_t\}$ is generated by a HAR model as in (9), where the errors are formed by a sequence $\{\varepsilon_t\}$ of zero mean independent and identically distributed random variables with $\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 < \infty$. $\mathcal{F}_t$ is the σ-algebra formed by the information available to*

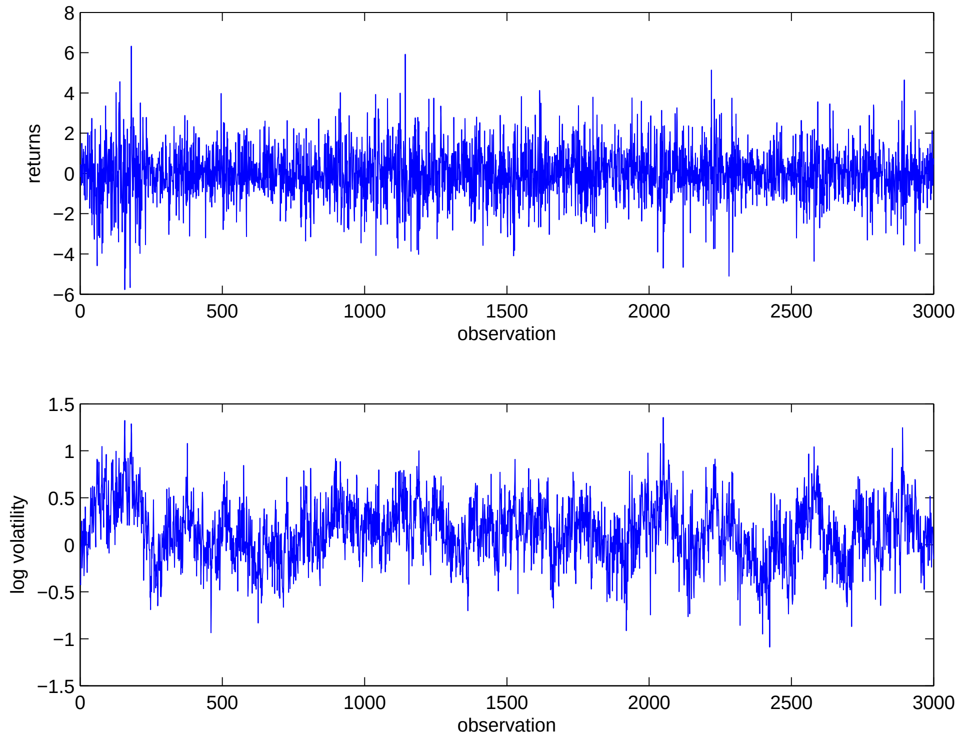

FIGURE 2. Upper panel: one realization of daily returns for Example 2. Lower panel: one realization of the logarithm of the daily volatility for Example 2.

time t. The process $\{y_t\}$ is strictly stationary and geometric ergodic if , and only if, the roots of the polynomial

$$1 - \phi_1 z - \ldots - \phi_1 z^{\iota_1} - \phi_2 z^{\iota_1+1} - \ldots - \phi_2 z^{\iota_2} - \ldots - \phi_p z^{\iota_{p-1}+1} - \ldots - \phi_p z^{\iota_p} = 0$$

are outside the unit circle.

Following the same reasoning as above, the HARST model can be written as a restricted version of the Functional-Coefficient Autoregressive (FAR) model proposed by Chen and Tsay (1993) given by

$$\begin{aligned} y_t = {} & \phi_0(z_t) + \phi_1(z_t)y_{t-1} + \cdots + \phi_1(z_t)y_{t-\iota_1} + \phi_2(z_t)y_{t-(\iota_1+1)} + \cdots + \phi_2(z_t)y_{t-\iota_2} \\ & + \phi_3(z_t)y_{t-(\iota_2+1)} + \cdots + \phi_3(z_t)y_{t-\iota_3} + \cdots + \phi_p(z_t)y_{t-(\iota_{p-1}+1)} + \cdots \\ & + \phi_p(z_t)y_{t-\iota_p} + \varepsilon_t, \end{aligned} \tag{11}$$

where $\phi_0(z_t) = \beta_{00} + \sum_{m=1}^{M} \beta_{m0} f(z_t; \gamma_m, c_m)$ and $\phi_j(z_t) = \sum_{i=j}^{p} \left[\beta_{0i} + \sum_{m=1}^{M} \beta_{mi} f(z_t; \gamma_m, c_m)\right]$, $j = 1, \ldots, p$.

Direct application of Theorem 1.1 in Chen and Tsay (1993) enables us to state the following result.

THEOREM 2. *Suppose that the process $\{y_t\}$ is generated by a HARST model as in (5) where $|\boldsymbol{\beta}_k| < \infty$, $k = 0, \ldots, M$, such that $|\phi_j(z_t)| \leq c_j = \left|\sum_{i=j}^{p}\left(\beta_{0i} + \sum_{m=1}^{M} \beta_{mi}\right)\right| < \infty$, $j = 1, \ldots, p$. Furthermore, assume that the errors are formed by a sequence $\{\varepsilon_t\}$ of zero mean independent and* *identically distributed random variables with* $\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 < \infty$. $\mathcal{F}_t$ *is the* σ*-algebra formed by the information available to time* t. *The process* $\{y_t\}$ *is strictly stationary and geometric ergodic if the roots of the polynomial*

$$1 - c_1 z - \ldots - c_1 z^{\iota_1} - c_2 z^{\iota_1+1} - \ldots - c_2 z^{\iota_2} - \ldots - c_p z^{\iota_{p-1}+1} - \ldots - c_p z^{\iota_p} = 0$$

are outside the unit circle.

It is clear that the condition of Theorem 2 is very strict. However, in order to relax this condition and the assumptions about the error term, it is important to make additional assumptions about the transition variable. Although important, this is beyond the scope of this paper and is left for future research. In practical applications, the estimated model can be checked for stationarity through simulation. In the following sections, we will assume that the process $\{y_t\}$ is stationary and ergodic.

3.4. **Parameter Estimation.** In this section we discuss parameter estimation of the HARST model and the corresponding asymptotic theory. Consider the following assumption about the data generating process (DGP).

ASSUMPTION 1 (Data Generating Process). *The observed sequence of real-valued dependent variable* $\{y_t\}_{t=1}^T$ *is a realization of a stationary and ergodic stochastic process on a complete probability space that can be well approximated by the HARST model, as in (5), such that the sequence* $\{\varepsilon_t\}_{t=1}^T$ *is formed by random variables drawn from an absolutely continuous (with respect to a Lebesgue measure on the real line), positive everywhere distribution with* $\mathbb{E}(\varepsilon_t) = \mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2 < \infty$ *and* $\mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 > 0, \forall\, t$. *Furthermore,* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \sigma_t^2 = \overline{\sigma}^2 < \infty$. $\mathcal{F}_t$ *is the* σ*-algebra formed by the information available to time* t.

Note that only mild restrictions are imposed on the error term, without assuming any particular distribution. However, it is assumed that the conditional mean can be adequately described by a HARST specification.

We make the following assumptions about the vector of parameters.

ASSUMPTION 2 (Parameter Space). *The true parameter vector* $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi} \subseteq \mathbb{R}^{(M+1)(p+1)+2M}$ *is in the interior of* $\boldsymbol{\Psi}$*, a compact and convex parameter space.*

ASSUMPTION 3 (Identifiability). *The parameters* γ_m *and* c_m, $m = 1, \ldots, M$*, satisfy the restrictions:*

(R.1) $\gamma_m > 0$;
(R.2) $-\infty < c_1 < \cdots < c_M < \infty$;
(R.3) *The elements of the vector* $\boldsymbol{\beta}_m$ *do not vanish jointly, for all* $m = 1, \ldots, M$.

Assumption 2 is standard and Assumption 3 guarantees that the HARST model is identified. More specifically, Restriction (R.1) eliminates identification problems caused by the fact that $f(z_t;\gamma_m,c_m) = 1 - f(z_t;-\gamma_m,c_m)$, $m = 1,\ldots,M$, and Restriction (R.2) avoids permutation of the M logistic functions in (5).

The vector of parameters $\boldsymbol{\psi}$ is estimated by nonlinear least squares, which is equivalent to the quasi-maximum likelihood method. The estimator is given by

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\text{argmin}} \mathcal{Q}_T(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\text{argmin}} \frac{1}{T}\sum_{t=1}^{T} q_t(\boldsymbol{\psi}),$$

where $q_t(\boldsymbol{\psi}) = [y_t - G(\mathbf{x}_t, z_t; \boldsymbol{\psi})]^2$.

Define $\mathcal{Q}(\boldsymbol{\psi}) = \mathbb{E}[q_t(\boldsymbol{\psi})]$. In the following theorems, we state the existence, consistency and asymptotic normality of the estimator $\widehat{\boldsymbol{\psi}}$. The existence result is based on Theorem 2.12 in White (1994), which establishes that, under certain conditions of continuity and measurability of the least squares function, $\mathcal{Q}(\boldsymbol{\psi})$ exists.

THEOREM 3 (Existence). *Under Assumptions 1 and 2, $\mathcal{Q}(\boldsymbol{\psi})$ exists, is finite, and is uniquely maximized at $\boldsymbol{\psi}_0$.*

In White (1981) and White and Domowitz (1984), the conditions that guarantee consistency of the nonlinear least squares estimator are established. In the context of stationary time series models, the conditions that ensure the consistency result are established in White (1994) and Wooldridge (1994). In what follows, we state and prove the theorem of consistency of the estimators of the HARST model.

THEOREM 4 (Consistency). *Under Assumptions 1–3, $\widehat{\boldsymbol{\psi}} \xrightarrow{p} \boldsymbol{\psi}_0$.*

The asymptotic normality result is also based on the results in White (1994) and Wooldridge (1994).

THEOREM 5 (Asymptotic Normality). *Under Assumptions 1–3, it follows that*

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

where

$$\mathbf{A}(\boldsymbol{\psi}_0) = \mathbb{E}\left[-\left.\frac{\partial^2 q_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right] \quad \textit{and}$$

$$\mathbf{B}(\boldsymbol{\psi}_0) = \mathbb{E}\left[T\left.\frac{\partial\mathcal{Q}_T(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\left.\frac{\partial\mathcal{Q}_T(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right] \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left.\frac{\partial q_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\left.\frac{\partial q_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right].$$

4. DETERMINING THE NUMBER OF REGIMES

The number of regimes in the HARST model, as represented by the number of transition functions in (5), is not known in advance and should be determined from the data. In this paper we tackle the problem of determining the number of regimes of the HARST model with a "specific-to-general" modelling strategy, but circumvent the problem of identification in a way that controls the significance level of the tests in the sequence and computes an upper bound to the overall significance level.

The following is based on the assumption that the errors ε_t are Gaussian, but the results will be made robust to non-normal errors.

Consider an HARST model as in (5) with M limiting regimes, defined as

$$y_t = \boldsymbol{\beta}_0' \mathbf{x}_t + \sum_{m=1}^{M-1} \boldsymbol{\beta}_m' \mathbf{x}_t f\left(z_t; \gamma_m, c_m\right) + \varepsilon_t. \tag{12}$$

The idea is to test the presence of an additional regime, as represented by an extra term in (12) of the form $\boldsymbol{\beta}_M' \mathbf{x}_t f\left(z_t; \gamma_M, c_M\right)$. A convenient null hypothesis is $\mathcal{H}_0 : \gamma_M = 0$, against the alternative $\mathcal{H}_a : \gamma_M > 0$. Note that model (12) is not identified under the null hypothesis. In order to remedy this problem, we follow Teräsvirta (1994) and expand the logistic function $f\left(z_t; \gamma_M, c_M\right)$ into a third-order Taylor expansion around the null hypothesis $\gamma_M = 0$. After merging terms, the resulting model is [2]

$$y_t = \widetilde{\boldsymbol{\beta}}_0' \mathbf{x}_t + \sum_{m=1}^{M-1} \boldsymbol{\beta}_m' \mathbf{x}_t f\left(z_t; \gamma_m, c_m\right) + \boldsymbol{\alpha}_1' \mathbf{x}_t z_t + \boldsymbol{\alpha}_2' \mathbf{x}_t z_t^2 + \boldsymbol{\alpha}_3' \mathbf{x}_t z_t^3 + \varepsilon_t^*, \tag{13}$$

where $\varepsilon_t^* = \varepsilon_t^* + R(z_t; \gamma_M, c_M)$, $R(z_t; \gamma_M, c_M)$ is the remainder, $\widetilde{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0 + \left(\frac{1}{2} - \frac{\gamma_M c_M}{4} - \frac{\gamma_M^3 c_M^3}{96}\right) \boldsymbol{\beta}_M$, $\boldsymbol{\alpha}_1 = \left(\frac{\gamma_M}{4} + \frac{\gamma_M^3 c_M^2}{32}\right) \boldsymbol{\beta}_M$, $\boldsymbol{\alpha}_2 = -\frac{\gamma_M^3 c_M}{32} \boldsymbol{\beta}_M$, and $\boldsymbol{\alpha}_3 = \frac{\gamma_M^3}{96} \boldsymbol{\beta}_M$.

Consider the following additional assumption.

ASSUMPTION 4 (Moments). *$\mathbb{E}\left(\mathbf{x}_t \mathbf{x}_t' z_t^\delta\right) < \infty$, for $\delta > 6$.*

Under $\mathcal{H}_0$, $R(z_t; \gamma_M, c_M) = 0$ and we can state the following result:

THEOREM 6. *Under Assumptions 1–4, the LM statistic given by*

$$LM = \frac{1}{\widehat{\sigma}^2} \sum_{t=1}^{T} \widehat{\varepsilon}_t \mathbf{v}_t' \left\{ \sum_{t=1}^{T} \mathbf{v}_t \mathbf{v}_t' - \sum_{t=1}^{T} \mathbf{v}_t \widehat{\mathbf{h}}_t' \left[\sum_{t=1}^{T} \widehat{\mathbf{h}}_t \widehat{\mathbf{h}}_t' \right]^{-1} \sum_{t=1}^{T} \widehat{\mathbf{h}}_t \mathbf{v}_t' \right\}^{-1} \sum_{t=1}^{T} \mathbf{v}_t \widehat{\varepsilon}_t, \tag{14}$$

where $\widehat{\sigma}_t^2 = \frac{1}{T} \sum_{t=1}^T \widehat{\varepsilon}_t^2$, $\{\widehat{\varepsilon}_t\}_{t=1}^T$ is the estimated sequence of residuals under the null hypothesis, $\mathbf{v}_t = \left(\mathbf{x}_t' z_t, \mathbf{x}_t' z_t^2, \mathbf{x}_t' z_t^3\right)'$, and

$$\begin{aligned} \widehat{\mathbf{h}}_t = \Bigg(& \mathbf{x}_t', \mathbf{x}_t' f(z_t; \widehat{\gamma}_1, \widehat{c}_1), \ldots, \mathbf{x}_t' f(z_t; \widehat{\gamma}_{M-1}, \widehat{c}_{M-1}), \\ & \widehat{\boldsymbol{\beta}}_1' \mathbf{x}_t \frac{\partial f(z_t; \widehat{\gamma}_1, \widehat{c}_1)}{\partial \gamma_1}, \ldots, \widehat{\boldsymbol{\beta}}_{M-1}' \mathbf{x}_t \frac{\partial f(z_t; \widehat{\gamma}_{M-1}, \widehat{c}_{M-1})}{\partial \gamma_{M-1}}, \\ & \widehat{\boldsymbol{\beta}}_1' \mathbf{x}_t \frac{\partial f(z_t; \widehat{\gamma}_1, \widehat{c}_1)}{\partial c_1}, \ldots, \widehat{\boldsymbol{\beta}}_{M-1}' \mathbf{x}_t \frac{\partial f(z_t; \widehat{\gamma}_{M-1}, \widehat{c}_{M-1})}{\partial c_{M-1}} \Bigg)', \end{aligned}$$

asymptotically has a χ^2 distribution with $3(p+1)$ degrees of freedom under the null hypothesis.

Under the normality assumption, the test can be performed in stages, as follows:

[2] If z_t is an element of $\mathbf{x}_t = \left(1, y_{t-\iota_1,t-1}, \ldots, y_{t-\iota_p,t-1}\right)'$, then the resulting model should be

$$y_t = \widetilde{\boldsymbol{\beta}}_0' \mathbf{x}_t + \sum_{m=1}^{M-1} \boldsymbol{\beta}_m' \mathbf{x}_t f\left(z_t; \gamma_m, c_m\right) + \boldsymbol{\alpha}_1' \widetilde{\mathbf{x}}_t z_t + \boldsymbol{\alpha}_2' \widetilde{\mathbf{x}}_t z_t^2 + \boldsymbol{\alpha}_3' \widetilde{\mathbf{x}}_t z_t^3 + \varepsilon_t^*,$$

where $\widetilde{\mathbf{x}}_t = \left(y_{t-\iota_1,t-1}, \ldots, y_{t-\iota_p,t-1}\right)'$.

(1) Estimate model (5) under $\mathcal{H}_0$ and compute the sequence of residuals $\{\widehat{\varepsilon}_t\}_{t=1}^T$. When the sample size is small, numerical problems in applying the quasi-maximum likelihood algorithm may lead to a solution such that the residual vector is not orthogonal to the gradient matrix of $G(\mathbf{x}_t, z_t; \widehat{\boldsymbol{\psi}})$. This has an adverse effect on the empirical size of the test. In order to circumvent this problem, we regress the residuals $\widehat{\varepsilon}_t$ on $\widehat{\mathbf{h}}_t$ and compute the sum of squared residuals, $SSR_0 = \sum_{t=1}^T \widetilde{\varepsilon}_t^2$. The new residuals, $\{\widetilde{\varepsilon}_t\}_{t=1}^T$, are orthogonal to $\widehat{\mathbf{h}}_t$.
(2) Regress $\widetilde{\varepsilon}_t$ on $\widehat{\mathbf{h}}_t$ and $\mathbf{v}_t$, and compute the sum of squared residuals, SSR_1.
(3) Compute the LM statistic:

$$LM_\chi = T\frac{SSR_0 - SSR_1}{SSR_0}, \tag{15}$$

or the F statistic:

$$LM_F = \frac{(SSR_0 - SSR_1)/3(p+1)}{SSR_1/[T-(3M-5)(p+1)]}. \tag{16}$$

Under $\mathcal{H}_0$, LM_χ is asymptotically distributed as χ^2 with $3(p+1)$ degrees of freedom and LM_F has an asymptotic F distribution with $3(p+1)$ and $T-(3M-5)(p+1)$ degrees of freedom.

Although the test statistic is constructed under the assumption of normality, it is straightforward to follow Lundbergh and Teräsvirta (2002) and consider a robust version of the LM test against non-normal errors. The robust version of the test statistic can be constructed following Procedure 4.1 in Wooldridge (1990). The test statistic can be calculated as follows:

(1) As above.
(2) Regress $\widehat{\mathbf{v}}_t$ on $\widehat{\mathbf{h}}_t$ and compute the residual vectors, $\widehat{\mathbf{r}}_t$, $t = 1, \ldots, T$.
(3) Regress 1 on $\varepsilon_t\widehat{\mathbf{r}}_t$ and compute the residual sum of squares, SSR. The test statistic given by:

$$LM_R = T - SSR \tag{17}$$

has an asymptotic χ^2 distribution with k_x degrees of freedom under the null hypothesis.

5. Model Selection

The modelling cycle of the HARST model involves three steps, namely specification, estimation, and model evaluation. The specification consists of three decisions:

(1) choice of relevant variables;
(2) selection of the transition variable; and
(3) determination of the number of regimes.

In addition to the set of lagged variables as defined in (5), other possible candidate variables are sets of (weakly) exogenous variables. For example, in the context of volatility forecasting, these variables may be dummies for the days of the week and dates of macroeconomic announcements. The set of lags $\boldsymbol{\iota}$ in the HARST model should be determined first. There are several ways of selecting the relevant variables. In the STAR literature, is common to select the set of relevant variables using information criteria, making use of a linear approximation to the true DGP. This is also a possibility for the HARST specification. However, as noted in Pitarakis (2006), this method may have an adverse effect on the final model specification. An alternative approach, which is adopted here, is to consider a k-th order polynomial approximation to the nonlinear component of the DGP, as proposed in Rech, Teräsvirta, and Tschernig (2001), and applied with success in Medeiros, Teräsvirta, and Rech (2006), Medeiros and Veiga (2005), and Suarez-Fariñas, Pedreira, and Medeiros (2004). As the logistic functions in (5) depend only on the scalar variable z_t, the polynomial approximation can be simplified dramatically as follows [3]:

$$y_t = \boldsymbol{\alpha}_0'\mathbf{x}_t + \boldsymbol{\alpha}_1'\mathbf{x}_t z_t + \boldsymbol{\alpha}_2'\mathbf{x}_t z_t^2 + \boldsymbol{\alpha}_3'\mathbf{x}_t z_t^3 + \cdots + \boldsymbol{\alpha}_k'\mathbf{x}_t z_t^k + \varepsilon_t^*, \tag{18}$$

where $\varepsilon_t^* = \varepsilon_t + R(\mathbf{x}_t, z_t; \boldsymbol{\psi})$. In this paper we choose a third-order polynomial approximation.

In equation (18), every product of variables involving at least one redundant variable in $\mathbf{x}_t$ has the coefficient set equal to zero. The idea is to sort out the redundant variables by using this property of (18). In order to do so, we first regress y_t on all the variables on the right-hand side of equation (18), assuming $R(\mathbf{x}_t, z_t; \boldsymbol{\psi}) = 0$, and compute the value of a model selection criterion (MSC), such as AIC or BIC. This leads to the removal of one variable from the original vector, $\mathbf{x}_t$. Then regress y_t on all the remaining terms in the corresponding polynomial, and again compute the value of the MSC. This procedure is repeated sequentially by omitting each variable in turn, and can be continued by simultaneously omitting two regressors in the original model, and proceeding until the vector $\mathbf{x}_t$ is just a constant. The combination of variables is chosen to yield the lowest value of the MSC. Rech, Teräsvirta, and Tschernig (2001) showed that the procedure works well in small samples when compared with well known nonparametric techniques. Furthermore, the procedure can be applied successfully even in large samples when nonparametric model selection is not computationally feasible.

The selection of the transition variable is determined by testing linearity for different possible choices of z_t[4]. We choose the transition variable that minimizes the p-value of the test. Finally, the number of regimes is determined by the sequence of LM tests, as described in Section 4.

We now combine the above procedure into a coherent modelling strategy that involves a sequence of LM tests. The idea is to test a linear HAR model against an alternative HARST model with more than one regime at a λ_1 level of significance. In the event that the null hypothesis is rejected, HARST with two regimes is estimated and than tested against an alternative with more than two regimes. The procedure continues testing J regimes against alternative models with $J^* \geq J+1$ regimes at significance level $\lambda_J = \lambda_1 C^{J-1}$ for some arbitrary constant $0 < C < 1$. The testing sequence is terminated at the first non-rejection outcome, and then the number of additional regimes, M, for the HARST specification is estimated by $\widehat{M} = \bar{J} - 1$, where $\bar{J}$ refers to how many testing runs are necessary to lead to the first non-rejection result. By reducing the significance level at each step of the sequence, it is possible to control the overall level of significance, and hence to avoid excessively large models. The Bonferroni procedure ensures that such a sequence of LM tests is consistent, and that $\sum_{J=1}^{\bar{J}} \lambda_J$ acts as an upper bound on the overall level of significance. As for the determination of the arbitrary constant

[3] Although the motivation is different, this approximation is rather similar to the one used in Section 4.

[4] The transition variable may also be selected by minimizing the MSC in expression (18).

C, it would be sensible practice to perform the sequential testing procedure with different values of C to avoid selecting models that are too parsimonious.

Estimation of the parameters of the model will be determined by nonlinear least squares, which is equivalent to quasi-maximum likelihood estimation, as discussed in Section 3.4.

What follows is evaluation of the final estimated model. Time series models are typically evaluated by their out-of-sample predictive performance. However, a sequence of neglected nonlinearity tests can also be interpreted as model evaluation tests. The construction of tests for serial correlation, in the spirit of Eitrheim and Teräsvirta (1996) and Medeiros and Veiga (2003), is also possible.

6. Monte Carlo Simulation

The goal of this section is to evaluate the finite sample performance of the modelling cycle, as described in the previous section. We simulated two different specifications as follows:

(1) Model 1: HARST (Asymmetric effects)

$$\begin{aligned} r_t &= \exp(\sigma_t)u_t, \;\; u_t \sim \mathsf{NID}(0,1) \\ \sigma_t &= 0.010 + 0.95\sigma_{t-1} - \\ &\quad (0.006 + 0.60\sigma_{t-1} - 0.25\sigma_{t-1,5} - 0.15\sigma_{t-1,22})\, f(r_{t-1}; 5, -3.0) + \\ &\quad (0.004 + 0.30\sigma_{t-1} - 0.16\sigma_{t-1,5} - 0.09\sigma_{t-1,22})\, f(r_{t-1}; 5, 2.5) + \varepsilon_t, \\ &\quad \varepsilon_t \sim \mathsf{NID}\left(0, 0.5^2\right). \end{aligned} \tag{19}$$

(2) Model 2: HARST (Asymmetric effects)

$$\begin{aligned} r_t &= \exp(\sigma_t)u_t, \;\; u_t \sim \mathsf{NID}(0,1) \\ \sigma_t &= 0.05 + 0.95\sigma_{t-1} - \\ &\quad (0.035 + 0.58\sigma_{t-1} - 0.27\sigma_{t-1,5} - 0.21\sigma_{t-1,22})\, f(r_{22,t-1}; 4, -10) + \\ &\quad (0.03 + 0.30\sigma_{t-1} - 0.20\sigma_{t-1,5} - 0.18\sigma_{t-1,22})\, f(r_{22,t-1}; 4, 13) + \varepsilon_t, \\ &\quad \varepsilon_t \sim \mathsf{NID}\left(0, 0.25^2\right). \end{aligned} \tag{20}$$

The simulated models have been analyzed in Examples 1 and 2 in Section 3, and each has three regimes. In the first model the regime switches are more frequent as the transition variable is the past return, while in the second model the switches are less frequent and the model spends a larger fraction of time in each regime. We consider different sample sizes for each model: 300, 500, 1000, 1500, 3000 and 5000. It should be noted that, in financial applications, 300 and 500 observations comprise rather small samples. Most of the datasets, especially those dealing with high frequency data, have more than 2000 observations. We simulate each specification 1000 times, with two different values of the starting significance level of the sequence of tests, namely 0.05 and 0.10, and halve the level of significance at each step. It is important to mention that the tests for the third regime are conducted at the 0.025 and 0.05 levels, respectively.

Table 2 presents the results concerning the determination of the number of regimes. The table shows the frequency of correctly selecting the number of regimes under the correct choice of explanatory variables in the model. The number in parentheses is the frequency of underfitting. Several facts emerge from the table. Both the robust and non-robust sequence of tests seem to be consistent, as the frequency of success increases with the sample size. Furthermore, as expected, the procedure is more accurate when the first model is considered, as the switches are far more frequent. It is also clear that the procedure is conservative as the frequency of underfitting is very high. Finally, the procedure works well for the typical sample sizes that are observed in financial applications.

7. Empirical Application

7.1. **The Data.** The empirical analysis focuses on the realized volatility of sixteen Dow Jones Industrial Average index stocks: Alcoa, American International Group, Boeing, Caterpillar, General Electric, General Motors, Hewlett Packard, IBM, Intel, Johnson and Johnson, Coca-Cola, Microsoft, Merck, Pfizer, Wal-Mart and Exxon. The raw intraday data are constituted of tick-by-tick quotes extracted from the NYSE Trade and Quote (TAQ) database. The period of analysis starts in January 3, 1994, and ends in December 31, 2003. Trading days with abnormally small trading volume and volatility caused by the proximity of holidays (for example, Good Friday) are excluded, leaving a total of 2541 daily observations.

We start by removing non-standard quotes, computing mid-quote prices, filtering possible errors, and obtaining one second returns for the 9:30 am to 4:05 p.m. period. Following the results of Hansen and Lunde (2006), we adopt the *previous tick* method for determining prices at precise time marks. Based on the results of Hasbrouck (1995), who reports a median 92.7% information share at the NYSE for Dow stocks, and Blume and Goldstein (1997), who conclude that NYSE quotes match or determine the best displayed quote most of the time, we use NYSE quotes (or NASDAQ, for Microsoft and Intel) if they are close enough to the time marks in relation to other updates.

In order to estimate our measure of the daily realized volatility, we use the two time scales estimator of Zhang, Mykland, and Aït-Sahalia (2005) with five-minute grids. The final dependent variable is the daily logarithm of the realized volatility. As in Martens, van Dijk, and de Pooter (2004) and Scharth and Medeiros (2006) we also consider dummies for the days of the week and dummies for the following macroeconomic announcements: Federal Open Market Committee meetings (FOM), The Employment Situation Report from the Bureau of Labor Statistics (ESR), CPI and PPI.

Data are used from 1993 to 1999 in order to estimate the models, and from 2000 to 2003 to evaluate the forecasting performance of the different specifications. The estimated models have the following structure.

$$\log(RV_t) = \boldsymbol{\alpha}'\mathbf{w}_t + \boldsymbol{\beta}_0'\mathbf{x}_t + \sum_{m=1}^{M} \boldsymbol{\beta}_m'\mathbf{x}_t f(z_t; \gamma_m, c_m) + \varepsilon_t, \quad (21)$$

where $\log(RV_t)$ is the logarithm of the daily realized volatility computed as described above, $\mathbf{w}_t$ is a vector containing selected dummies for the days-of-the-week and announcement dates, $\mathbf{x}_t = \left(1, \log(RV_{t-1,\iota_1}), \ldots, \log(RV_{t-1,\iota_p})\right)'$, $f(\cdot)$ is the logistic function as in (5), and z_t is the past return (r_{t-1}).

7.2. **Model Specification and Estimation.** We start by selecting the relevant explanatory variables. All the variables are selected according to the procedure described in Section 5 using BIC. In order to keep interpretability of the selected lags and to avoid serious "data mining" problems, we consider the following set of possible lags: $\mathbb{X} = \{1, 2, 5, 10, 15, 22\}$. Table 3 shows the selected variables. Several interesting facts emerge from the table. First, for ten of 16 series, the selected lags are 1, 5, and 22, meaning that daily, weekly, and monthly volatility are highly relevant. Second, announcement effects are selected as explanatory variables in seven cases. The most important announcement seems to be the Federal Open Market Committee meetings. Finally, there is not a clear pattern with respect to the presence of the days-of-the-week dummies in the model.

After selecting the relevant variables, we continue estimating a linear HAR model. Table 4 shows several statistics for the estimated model. The table shows the p-values for the following tests: LM test for residual serial autocorrelation of orders 1, 5, and 10; LM test for ARCH effects of orders 1, 5, and 10; Jarque-Bera test for normality of the residuals; and finally the linearity test against the HARST alternative. As one of our main goals is to model asymmetries and leverage in the volatility dynamics, we fix the transition variable to be the past daily return, r_{t-1}. We report both robust and non-robust versions of the linearity test. We have also tested linearity choosing other transition variables, such as past daily, weekly, and monthly volatilities. However, the best and more significant results are obtained with the past daily return as the transition variable.

According to the results in Table 4 and at a 5% significance level, the linear HAR model fails to account for serial correlation in 8 of the 16 series. In addition, there is evidence of conditional heteroskedasticity in 12 of 16 series (which may be due to nonlinear effects). Furthermore, normality is strongly rejected in all cases. For this reason, we will use the robust sequence of LM tests to specify the HARST model.

Finally, we estimate the HARST model for each series. The dummies for the announcement dates and days-of-the week enter only in the linear part of the model. The results are shown in Table 5, which presents the following diagnostic statistics: p-value of the test of remaining nonlinearity (additional regimes), p-value of the LM residual serial correlation test, p-value of the LM test for ARCH effects, and p-value of the Jarque-Bera test for normality. Only for ALCOA (AA) is there no evidence of more than a single regime. For all the other series there is strong evidence of two regimes, with the exception of Microsoft, where we find evidence of three regimes.

From the results presented in Table 5, there is still some evidence of residual autocorrelation in some cases, although, for most of the series, the HARST model correctly describes the dynamics of the logarithm of the realized volatility. One interesting fact is that now 8 of 16 series do not have conditional heteroskedasticity. However, normality is still strongly rejected.

Figure 3 displays the estimated transition functions. It is interesting to note that in all cases the asymmetry is not around zero returns, as is strongly advocated in the literature. The regime switches are associated with very negative past returns (or "very bad news"). The smoothness of the transition varies according to each series. In some cases, Caterpillar for example, the transition is abrupt. In others, such as General Electric, the transition is very smooth.

7.3. **Forecasting Results.** After estimating the HARST model for each series, the one-day ahead forecasts are computed. The forecasting performance of the HARST model is compared with the following competing specifications: Linear HAR, linear ARFIMA, GARCH, GJR, and EGARCH models. In addition, the forecast combination of a simple model average of the linear HAR and HARST models is examined. As the regime switches are associated with very negative returns, the benefits of using the nonlinear model should become apparent only in periods following very negative returns, such that a combination of forecasts will improve the performance of both models.

The results are reported in Tables 6 and 7. Table 6 presents the mean absolute errors (MAE) and the root mean squared errors (RMSE) for the forecasts from the different models. It can be seen from the table that the forecasting performance of the HARST model is not significantly better than from the linear HAR model in most cases. However, this is likely for the reasons given previously. When the HAR and HARST models are combined, the forecasting performance improves. When compared with the alternative latent volatility models, the performance of both the HAR and HARST models is far superior.

In order to determine if the combination of the linear HAR and HARST models generates more accurate one-step-ahead forecasts than does the linear HAR model, we apply the modified Diebold and Mariano (1995) test of Harvey, Leybourne, and Newbold (1997) to these series of forecasts. In Table 7, the p-values of the test are shown. We compare forecast differences using both the absolute value loss function (MAE) and the quadratic loss function (RMSE). Concerning the absolute errors, the combination of models delivers superior forecasts in six cases. In seven cases, the forecasts are not statistically different, and in only two cases does the linear HAR model perform the best. When squared errors are considered, the combination of models produces better forecasts in six cases, the forecasts are not statistically different in a further six cases, and in three cases the linear HAR has the best performance. In a direct comparison of the linear HAR and HARST models, the forecasts are not statistically different in 12 cases.

8. Conclusion

This paper developed a new flexible nonlinear model that can simultaneously describe long-range dependence and asymmetries in time series dynamics. The model is a generalization of the Heterogenous Autoregression (HAR) model and is called the Multiple Regime Smooth Transition Heterogenous Autoregressive (HARST) model. Following results in the nonlinear time series literature, we developed an estimation and testing procedure, including an easily implemented sequence of Lagrange multiplier

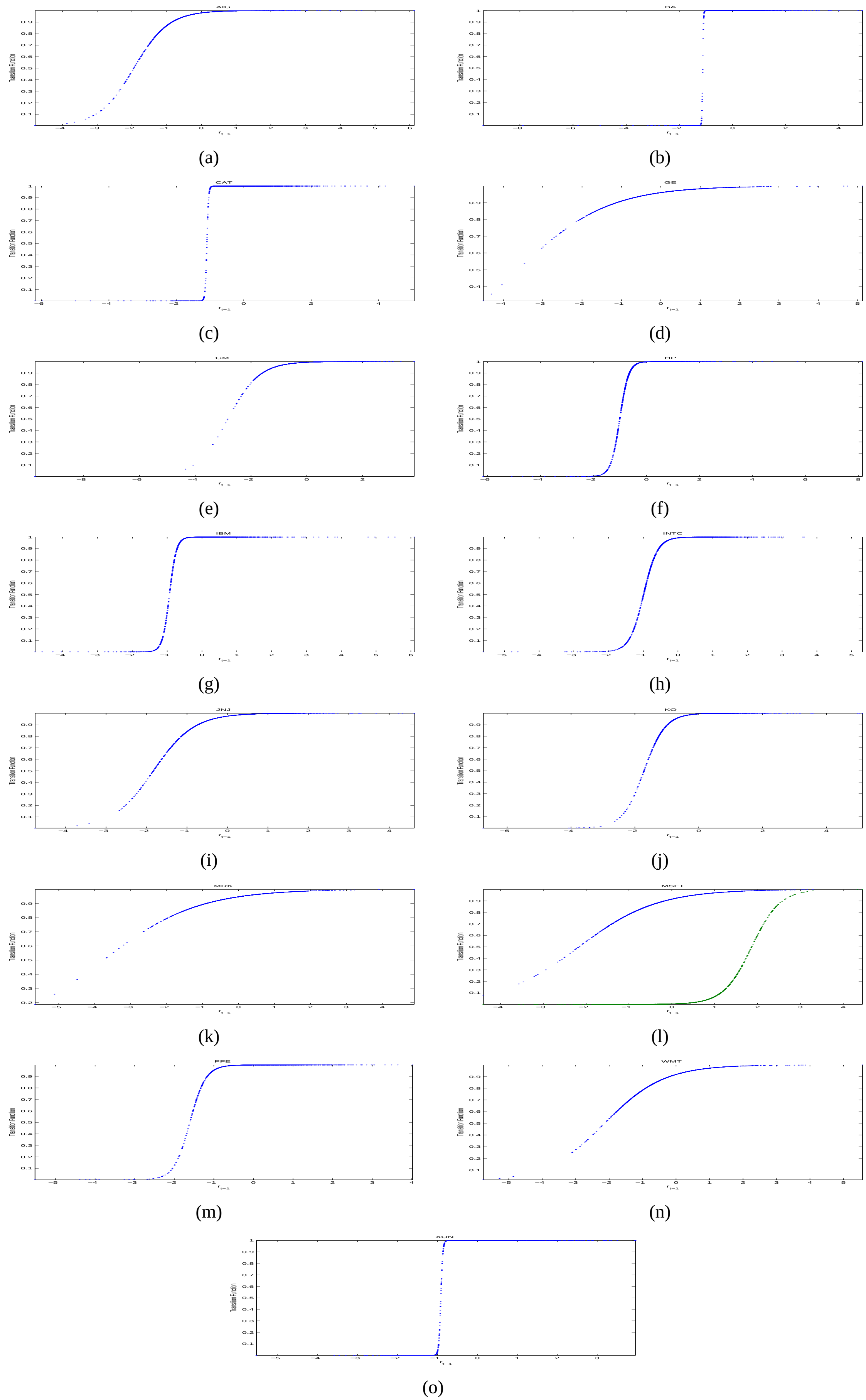

FIGURE 3. Estimated Transition Functions.

tests to determine the number of regimes in the model. A modelling cycle was proposed, and simulations were used to evaluate the finite sample performance of the estimation and testing methods. The new model was used to describe and forecast realized volatility of high frequency financial time series, and the empirical results indicated strong practical support for the model.

REFERENCES

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

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

ANDERSEN, T., T. BOLLERSLEV, AND F. DIEBOLD (2007): "Roughing it up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, forthcoming.

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

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

ANDERSON, H. M., K. NAM, AND F. VAHID (1999): "Asymmetric Nonlinear Smooth Transition GARCH Models," in *Nonlinear Time Series Analysis of Economic and Financial Data*, ed. by P. Rothman, pp. 191–207. Kluwer.

ASAI, M., AND M. MCALEER (2005): "Dynamic Asymmetric Leverage in Stochastic Volatility Models," *Econometric Reviews*, 24, 317–332.

——— (2006): "Asymmetric Multivariate Stochastic Volatility," *Econometric Reviews*, 25, 453–473.

——— (2007): "Non-trading Day Effects in Asymmetric Conditional and Stochastic Volatility Models," *Econometrics Journal*, 10, 113–123.

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

AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2005): "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," NBER Working Papers 11380, National Bureau of Economic Research.

AUDRINO, F., AND P. BÜHLMANN (2001): "Tree-Structured GARCH Models," *Journal of the Royal Statistical Society, Series B*, 63, 727–744.

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

BAILLIE, R., T. BOLLERSLEV, AND H. MIKKELSEN (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

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

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

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

BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2006a): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," Unpublished manuscript, Stanford University.

——— (2006b): "Subsampling realised kernels," Unpublished manuscript, Stanford University.

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

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

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

BLUME, M., AND M. GOLDSTEIN (1997): "Quotes, Order Flow, and Price Discovery," *Journal of Finance*, 52, 221–244.

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

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

BREIDT, F., N. CRATO, AND P. DE LIMA (1998): "The detection and estimation of long memory in stochastic volatility," *Journal of Econometrics*, 83, 325–348.

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

CAPORIN, M., AND M. MCALEER (2006): "Dynamic Asymmetric GARCH," *Journal of Financial Econometrics*, 4, 485–412.

CAPPUCCIO, N., D. LUBIAN, AND R. DAVIDE (2006): "Investigating asymmetry in US stock market indexes: evidence from a stochastic volatility model," *Applied Financial Economics*, 16, 479–490.

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

CHAN, K. S., AND H. TONG (1986): "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis*, 7, 179–190.

CHEN, R., AND R. S. TSAY (1993): "Functional Coefficient AutoRegressive Models," *Journal of the American Statistical Association*, 88, 298–308.

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

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

DEO, R., C. HURVICH, AND Y. LU (2006): "Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment," *Journal of Econometrics*, 131, 29–58.

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

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

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

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

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

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

ENGLE, R. F., AND V. K. NG (1993): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

FORNARI, F., AND A. MELE (1997): "Sign- and Volatility-Switching ARCH Models: Theory and Applications to International Stock Markets," *Journal of Applied Econometrics*, 12, 49–65.

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

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221–238.

GIRAITIS, L., P. ROBINSON, AND D. SURGAILIS (2004): "LARCH, Leverage, and Long Memory," *Journal of Financial Econometrics*, 2, 177–210.

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

GOETZMANN, W. N., R. G. IBBOTSON, AND L. PENG (2001): "A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability," *Journal of Financial Markets*, 4, 1–32.

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

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

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

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

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

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

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

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

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

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

HARVEY, A., AND N. SHEPHARD (1996): "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business and Economic Statistics*, 14, 429–434.

HARVEY, D., S. LEYBOURNE, AND P. NEWBOLD (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291.

HASBROUCK, J. (1995): "One security, many markets: determining the contributions to price discovery," *Journal of Finance*, 50, 1175–1198.

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

HILLEBRAND, E., AND M. MEDEIROS (2006): "Asymmetries, Breaks, and Long-range Dependence in Realized Volatility: A Simultaneous Equations Approach," Discussion paper, Pontifical Catholic University of Rio de Janeiro.

——— (2007): "Estimating and Forecasting GARCH Models in The Presence of Structural Breaks and Regime Switches," in *Forecasting in The Presence of Structural Breaks and Model Uncertainty*, ed. by M. Wohar, and D. Rapach. Elsevier, forthcoming.

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

HURVICH, C., AND B. RAY (2003): "The Local Whittle Estimator of Long-Memory Stochastic Volatility," *Journal of Financial Econometrics*, 1, 445–470.

HWANG, J. T. G., AND A. A. DING (1997): "Prediction Intervals for Artificial Neural Networks," *Journal of the American Statistical Association*, 92, 109–125.

HYUNG, N., S.-H. POON, AND C. GRANGER (2007): "The Source of Long Memory in Financial Market Volatility," in *Forecasting in The Presence of Structural Breaks and Model Uncertainty*, ed. by M. Wohar, and D. Rapach. Elsevier, forthcoming.

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

JENSEN, M. (2004): "Semiparametric Bayesian Inference of Long-Memory Stochastic Volatility Models," *Journal of Time Series Analysis*, 25, 895–922.

KALIMIPALLI, M., AND R. SUSMEL (2007): "Switching Stochastic Volatility and the Short-Term Interest Rates," *Journal of Empirical Finance*, forthcoming.

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

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

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

LIEBERMAN, O., AND P. PHILLIPS (2007): "Refined Inference on Long Memory in Realized Volatility," *Econometric Reviews*, forthcoming.

LING, S., AND M. MCALEER (2002): "Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models," *Econometric Theory*, 18, 722–729.

——— (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

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

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

LUNDBERGH, S., AND T. TERÄSVIRTA (2002): "Evaluating GARCH Models," *Journal of Econometrics*, 110, 417–435.

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

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

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

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

MCALEER, M., AND M. MEDEIROS (2007): "Realized Volatility: A Review," *Econometric Reviews*, forthcoming.

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

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

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

——— (2004): "Modeling Multiple Regimes in Financial Volatility with a Flexible Coefficient GARCH(1,1) Model," Discussion Paper 486, Pontifical Catholic University of Rio de Janeiro.

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

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

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

MÜLLER, U., M. DACOROGNA, R. DAV, R. OLSEN, O. PICTET, AND J. WARD (1993): "Fractals and Intrinsic Time – A Challenge to Econometricians," in *Proceedings of the XXXIXth International AEA Conference on Real Time Econometrics*.

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

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

OMORI, Y., S. CHIB, N. SHEPHARD, AND J. NAKAJIMA (2007): "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, forthcoming.

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

PITARAKIS, J. (2006): "Model Selection Uncertainty and Detection of Threshold Effects," *Studies in Nonlinear Dynamics and Econometrics*, 10, 1–28.

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

RECH, G., T. TERÄSVIRTA, AND R. TSCHERNIG (2001): "A Simple Variable Selection Technique for Nonlinear Models," *Communications in Statistics, Theory and Methods*, 30, 1227–1241.

SCHARTH, M., AND M. MEDEIROS (2006): "Asymmetric Effects and Long Memory in the Volatility of DJIA Stocks," Discussion paper, Pontifical Catholic University of Rio de Janeiro.

SCHWERT, G. W. (1990): "Stock Volatility and the Crash of '87," *Review of Financial Studies*, 3, 77–102.

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

SO, M., K. LAM, AND W. LI (1998): "A stochastic volatility model with Markov switching," *Journal of Business and Economic Statistics*, 16, 244–253.

SO, M., W. LI, AND K. LAM (2002): "A threshold stochastic volatility model," *Journal of Forecasting*, 21, 473–500.

SUAREZ-FARIÑAS, M., C. E. PEDREIRA, AND M. C. MEDEIROS (2004): "Local Global Neural Networks: A New Approach for Nonlinear Time Series Modeling," *Journal of the American Statistical Association*, 99, 1092–1107.

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

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

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

WHITE, H. (1981): "Consequences and Detection of Misspecified Nonlinear Regression Models," *Journal of the American Statistical Association*, 76(374), 419–433.

——— (1994): *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York, NY.

WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143–162.

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

——— (1994): "Estimation and Inference for Dependent Process," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2639–2738. Elsevier Science.

YU, J. (2005): "On Leverage in a Stochastic Volatility Model," *Journal of Econometrics*, 127, 165–178.

ZAKOIAN, J. M. (1994): "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931–955.

ZHANG, L. (2005): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," Unplublished manuscript, Department of Mathematical Sciences, Carnegie Mellon University.

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

ZHOU, B. (1996): "High frequency data and volatility in foreign-exchange rates," *Journal of Business and Economic Statistics*, 14, 45–52.

TABLE 1. DESCRIPTIVE STATISTICS.

The table shows the mean, median, standard deviation, minimum, and maximum of the following statistics: mean, standard deviation, kurtosis, and skewness of the simulated daily returns; sum of the first 500 autocorrelations of the absolute and squared daily returns, $\sum_{j=1}^{500}\rho_j(|r_t|)$ and $\sum_{j=1}^{500}\rho_j(r_t^2)$, respectively; the GPH (Geweke and Porter-Hudak 1983) estimator of the fractional difference parameter for the absolute returns, squared returns, and log volatility, $d(|r_t|)$, $d(r_t^2)$, and $d(\sigma_t)$, respectively; and the correlation coefficient between the volatility and the lagged return, $\rho(\exp(\sigma_t), r_{t-1})$. The number of ordinates in the GPH estimator is set as $l = 3000^{0.5}$.

Example 1

	Mean	Std. Dev.	Kurtosis	Skewness	$\sum_{j=1}^{500}\rho_j(\|r_t\|)$	$\sum_{j=1}^{500}\rho_j(r_t^2)$	$d(\|r_t\|)$	$d(r_t^2)$	$d(\log(\sigma_t))$	$\rho(\exp(\sigma_t), r_{t-1})$
Mean	-0.0012	1.8320	45.954	0.1233	1.6937	0.7408	0.1562	0.0872	0.2261	-0.0929
Median	0.0003	1.7859	21.763	0.0911	1.2936	0.4957	0.1549	0.0751	0.2308	-0.0830
Std. Dev.	0.0325	0.2510	99.464	2.8956	2.0315	1.0996	0.1052	0.1063	0.0963	0.0805
Minimum	-0.1314	1.3649	7.4699	-23.127	-1.8592	-1.2114	-0.2629	-0.2487	-0.0926	-0.6373
Maximum	0.0902	3.7018	1,341.1	30.325	12.3670	7.1483	0.5193	0.5612	0.5125	0.1719

Example 2

	Mean	Std. Dev.	Kurtosis	Skewness	$\sum_{j=1}^{500}\rho_j(\|r_t\|)$	$\sum_{j=1}^{500}\rho_j(r_t^2)$	$d(\|r_t\|)$	$d(r_t^2)$	$d(\log(\sigma_t))$	$\rho(\exp(\sigma_t), r_{t-1})$
Mean	0.0006	1.3429	11.006	0.0620	2.2571	1.7958	0.2690	0.2142	0.3904	-0.0137
Median	0.0011	1.3218	6.0657	0.0543	1.7360	1.4124	0.2676	0.2068	0.3940	-0.0096
Std. Dev.	0.0236	0.1263	20.469	0.7025	2.5185	1.8799	0.1104	0.1184	0.1012	0.0407
Minimum	-0.0924	1.1255	3.8245	-9.2266	-3.1003	-2.0112	-0.0745	-0.1068	0.0301	-0.2990
Maximum	0.0729	2.2681	242.14	8.1517	18.278	13.102	0.8205	0.7679	0.7656	0.1051

TABLE 2. SIMULATION RESULTS: NUMBER OF REGIMES.

Relative frequency of selecting correctly the number of regimes of the model based on 1000 replications with different sample sizes. The number between parenthesis is the frequency of underfitting (selection of fewer regimes).

		Initial significance level: 0.05			
		300 observations		500 observations	
Model	True Value	Non-Robust Test	Robust Test	Non-Robust Test	Robust Test
1	3	0.05 (0.95)	0 (1)	0.07 (0.92)	0.01 (0.99)
2	3	0.02 (0.98)	0.01 (0.99)	0.03 (0.97)	0.02 (0.98)
		1000 observations		1500 observations	
Model	True Value	Non-Robust Test	Robust Test	Non-Robust Test	Robust Test
1	3	0.19 (0.80)	0.06 (0.94)	0.30 (0.69)	0.14 (0.86)
2	3	0.06 90.93)	0.04 (0.96)	0.10 (0.90)	0.04 (0.96)
		3000 observations		5000 observations	
Model	True Value	Non-Robust Test	Robust Test	Non-Robust Test	Robust Test
1	3	0.56 (0.43)	0.41 (0.59)	0.86 (0.12)	0.76 (0.24)
2	3	0.17 (0.82)	0.10 (0.89)	0.28 (0.71)	0.13 (0.87)

		Initial significance level: 0.10			
		300 observations		500 observations	
Model	True Value	Non-Robust Test	Robust Test	Non-Robust Test	Robust Test
1	3	0.07 (0.93)	0.01 (0.99)	0.10 (0.88)	0.02 (0.98)
2	3	0.03 (0.96)	0.01 (0.99)	0.09 (0.90)	0.04 (0.96)
		1000 observations		1500 observations	
Model	True Value	Non-Robust Test	Robust Test	Non-Robust Test	Robust Test
1	3	0.25 (0.73)	0.09 (0.91)	0.34 (0.65)	0.20 (0.80)
2	3	0.12 (0.86)	0.06 (0.93)	0.17 (0.82)	0.09 (0.91)
		3000 observations		5000 observations	
Model	True Value	Non-Robust Test	Robust Test	Non-Robust Test	Robust Test
1	3	0.68 (0.31)	0.52 (0.48)	0.90 (0.09)	0.85 (0.15)
2	3	0.21 (0.78)	0.16 (0.83)	0.34 (0.62)	0.25 (0.73)

TABLE 3. SELECTED VARIABLES.

The table shows for each series the selected lags in ι in (5) and indicates whether or not announcement effects and days-of-the-week dummies are included in the model specification. FOM indicates the dates of the Federal Open Market Committee meetings, ESR is related to the announcement of The Employment Situation Report, and CPI and PPI indicate the dates of the announcement of the CPI and PPI, respectively. We omit the dummy for Friday to avoid perfect collinearity as our model includes a constant in its specification.

Series	Lags	FOM	ESR	CPI	PPI	Monday	Tuesday	Wednesday	Thursday
AA	1, 5, and 22	No	No	No	No	No	Yes	No	No
AIG	1, 5, and 22	No	No	No	No	No	No	No	No
BA	1, 5, and 22	No	No	No	No	No	No	No	No
CAT	1 and 15	Yes	No	No	No	No	No	No	No
GE	1, 5, and 22	Yes	No	No	No	Yes	No	Yes	No
GM	1 and 10	No	No	No	No	No	No	No	No
HP	1, 5, and 22	Yes	No	No	No	No	No	No	Yes
IBM	1, 5, and 22	Yes	No	No	No	No	No	Yes	No
INTC	1, 5, and 22	No	Yes	Yes	Yes	No	No	No	No
JNJ	1, 5, and 15	No	No	No	No	No	No	No	No
KO	1, 5, and 15	No	No	No	No	No	No	No	No
MRK	1, 5, and 22	Yes	No	No	No	No	No	Yes	No
MSF	1, 5, and 22	No	Yes	No	No	No	No	No	No
PFE	1 and 10	No	No	No	No	No	No	No	No
WMT	1, 2, and 10	Yes	No	No	No	No	No	No	No
XON	1, 5, and 22	No	No	No	No	No	No	No	Yes

TABLE 4. LINEAR HAR MODEL: DIAGNOSTIC TESTS.

The table shows for each series the p-values for the following tests: LM test for residual serial autocorrelation of orders 1, 5, and 10; LM test for ARCH effects of order 1, 5, and 10; the Jarque-Bera test for normality; and finally, the linearity test against the HARST alternative using r_{t-1} as transition variable. The table also reports estimates for the residuals kurtosis and skewness.

	Serial Correlation			ARCH			Normality			Linearity Test	
Series	1	5	10	1	5	10	Kurtosis	Skewness	Jarque-Bera	Non-Robust	Robust
AA	0.2005	0.6551	0.6749	0.3076	0.7205	0.8943	4.0196	0.4018	0.0000	0.0021	0.1016
AIG	0.1866	0.2808	0.4959	0.3485	0.1760	0.1198	3.9749	0.0105	0.0000	0.0000	0.0000
BA	0.0294	0.0891	0.2338	0.1540	0.1396	0.2598	5.1427	0.7398	0.0000	0.0000	0.0002
CAT	0.0772	0.0878	0.4366	0.2165	0.4454	0.6576	4.1846	0.3151	0.0000	0.0000	0.0014
GE	0.8184	0.0010	0.0011	0.0000	0.0001	0.0027	4.1731	0.3512	0.0000	0.0000	0.0000
GM	0.0292	0.1359	0.1601	0.0000	0.0000	0.0000	8.4597	0.0494	0.0000	0.0000	0.0004
HP	0.8155	0.8511	0.7802	0.0474	0.0226	0.0292	3.4459	0.2120	0.0000	0.0080	0.0042
IBM	0.3504	0.1704	0.0471	0.0038	0.0588	0.0973	4.1146	0.6374	0.0000	0.0000	0.0000
INTC	0.8656	0.9396	0.9611	0.0000	0.0000	0.0010	4.8292	0.1013	0.0000	0.0000	0.0000
JNJ	0.1568	0.7515	0.7011	0.0006	0.0035	0.0140	4.8042	0.5194	0.0000	0.0000	0.0000
KO	0.0275	0.0493	0.1128	0.0000	0.0000	0.0000	5.4411	0.5049	0.0000	0.0000	0.0006
MRK	0.1103	0.0101	0.0034	0.0000	0.0001	0.0004	4.4371	0.4482	0.0000	0.0000	0.0001
MSF	0.0123	0.0576	0.1785	0.0000	0.0000	0.0004	4.4653	-0.0350	0.0000	0.0000	0.0000
PFE	0.0457	0.1398	0.2206	0.0049	0.0077	0.0325	4.7666	0.4908	0.0000	0.0000	0.0049
WMT	0.0781	0.1209	0.0973	0.0000	0.0000	0.0001	4.0922	0.2602	0.0000	0.0000	0.0000
XON	0.0069	0.1075	0.1432	0.0001	0.0007	0.0016	4.1868	0.4313	0.0000	0.0000	0.0080

TABLE 5. HARST MODEL: DIAGNOSTIC TESTS.

The table shows for each series the p-values for the following tests: LM test for residual serial autocorrelation of orders 1, 5, and 10; LM test for ARCH effects of order 1, 5, and 10; Jarque-Bera test for normality of the residuals; and finally the remaining nonlinearity test (robust version). The table also shows the kurtosis and skewness for the estimated residuals.

	Serial Correlation			ARCH			Normality			Remaining	Number of
Series	1	5	10	1	5	10	Kurtosis	Skewness	Jarque-Bera	Nonlinearity	Regimes
AA	0.2005	0.6551	0.6749	0.3076	0.7205	0.8943	4.0196	0.4018	0.0000	0.1016	1
AIG	0.0643	0.1654	0.3805	0.4938	0.1863	0.1485	4.0691	0.0121	0.0000	0.5553	2
BA	0.0545	0.0590	0.1799	0.1644	0.2947	0.4262	5.1785	0.7496	0.0000	0.1229	2
CAT	0.0529	0.0785	0.3722	0.3529	0.4661	0.6939	4.0072	0.2700	0.0000	0.1364	2
GE	0.0566	0.0066	0.0139	0.8282	0.2193	0.5961	3.8596	0.2246	0.0000	0.0275	2
GM	0.1526	0.7447	0.1778	0.7072	0.9948	0.9990	8.8480	-0.1085	0.0000	0.0370	2
HP	0.1527	0.1693	0.3740	0.0203	0.0155	0.0294	3.3080	0.1683	0.0014	0.2538	2
IBM	0.0611	0.2651	0.0506	0.0909	0.2743	0.3694	4.1327	0.6122	0.0000	0.3097	2
INTC	0.3359	0.5605	0.7286	0.0000	0.0018	0.0206	4.7615	0.0506	0.0000	0.1122	2
JNJ	0.0580	0.1140	0.3196	0.0001	0.0023	0.0145	4.5882	0.4361	0.0000	0.0662	2
KO	0.0573	0.1017	0.3312	0.0000	0.0000	0.0006	5.3710	0.2829	0.0000	0.2093	2
MRK	0.1048	0.1128	0.0810	0.0460	0.0512	0.3181	4.3829	0.3893	0.0000	0.0866	2
MSF	0.0004	0.0067	0.0133	0.0001	0.0008	0.0071	4.6288	-0.0695	0.0000	0.1550	3
PFE	0.0111	0.0564	0.0882	0.0022	0.0129	0.0485	4.7709	0.4197	0.0000	0.0433	2
WMT	0.0556	0.0494	0.1020	0.0169	0.0111	0.0289	3.7307	0.1502	0.0000	0.3956	2
XO	0.0568	0.1612	0.3681	0.0013	0.0086	0.0275	4.1490	0.4080	0.0000	0.3141	2

TABLE 6. FORECASTING RESULTS: MEAN ABSOLUTE ERRORS AND ROOT MEAN SQUARED ERRORS.

The table shows for each series the mean absolute errors (MAE) and the root mean squared errors (RMSE) for the forecasts computed from different models.

			MAE			
Series	HARST	HAR	HARST + HAR	GARCH	EGARCH	GJR
AA	–	**0.4725**	–	0.6170	0.7082	0.5972
AIG	0.3691	0.3671	**0.3653**	0.4648	0.4330	0.4648
BA	0.4164	0.4150	**0.4135**	0.5153	0.5054	0.5297
CAT	0.4069	0.4053	**0.4051**	0.5604	0.5405	0.5879
GE	0.3666	0.3569	**0.3541**	0.4949	0.4363	0.4715
GM	0.4390	0.4282	**0.4267**	0.5001	0.4676	0.4891
HP	0.6456	**0.5999**	0.6189	0.8768	0.8567	0.8716
IBM	0.3424	0.3444	**0.3417**	0.5527	0.5175	0.5499
INTC	0.4890	**0.4776**	0.4812	0.6787	0.6814	0.7411
JNJ	0.3703	0.3679	**0.3641**	0.4718	0.4550	0.4606
KO	0.3414	0.3441	**0.3405**	0.4316	0.4046	0.4145
MRK	0.3726	0.3712	**0.3705**	0.4635	0.4342	0.4628
MSF	0.3695	0.3707	**0.3641**	0.5761	0.5361	0.5780
PFE	0.4207	**0.4186**	0.4190	0.4723	0.5310	0.4758
WMT	0.4168	0.4102	**0.4050**	0.5296	0.5062	0.5194
XON	0.3111	0.3119	**0.3096**	0.4004	0.4052	0.4001

			RMSE			
Series	HARST	HAR	HARST + HAR	GARCH	EGARCH	GJR
AA	–	**0.6808**	–	0.8483	0.9668	0.8041
AIG	0.5516	0.5544	**0.5489**	0.6347	0.6276	0.6264
BA	**0.6132**	0.6208	0.6139	0.7340	0.6973	0.7556
CAT	0.5962	0.5938	**0.5937**	0.7750	0.7460	0.8130
GE	0.5481	0.5423	**0.5329**	0.6869	0.6082	0.6503
GM	0.6731	**0.6538**	0.6547	0.6829	0.6755	0.6733
HP	0.9328	**0.8595**	0.8903	1.1188	1.0995	1.1080
IBM	0.5520	**0.5479**	0.5487	0.7353	0.6671	0.7421
INTC	0.7154	**0.6927**	0.7020	0.9130	0.9151	1.0613
JNJ	0.5847	0.5826	**0.5769**	0.7175	0.7089	0.7103
KO	0.5138	0.5147	**0.5119**	0.6290	0.6013	0.6042
MRK	0.5859	**0.5813**	0.5816	0.6820	0.6538	0.6795
MSF	0.5429	0.5488	**0.5311**	0.7584	0.6870	0.7718
PFE	0.6784	**0.6694**	0.6727	0.7367	0.7995	0.7450
WMT	0.6659	0.6598	**0.6517**	0.8212	0.7906	0.8086
XON	**0.4677**	0.4777	0.4700	0.6148	0.6248	0.6144

TABLE 7. FORECASTING RESULTS: DIEBOLD-MARIANO TEST.

The table shows for each series the p-value of the modified Diebold-Mariano test of equal forecast accuracy. We compare the combination of HAR and HARST models against the HAR model.

	HARST + HAR versus HAR		HARST versus HAR	
Series	MAE	RMSE	MAE	RMSE
AA	–	–	–	–
AIG	0.2002	0.0638	0.7009	0.3452
BA	0.2149	0.0372	0.6434	0.1500
CAT	0.4466	0.5182	0.7714	0.7813
GE	0.1965	0.0403	0.9466	0.7045
GM	0.2913	0.6125	0.9818	0.9965
HP	1.0000	1.0000	1.0000	1.0000
IBM	0.0075	0.7350	0.1732	0.9428
INTC	0.9863	1.0000	0.9998	1.0000
JNJ	0.0603	0.1144	0.6889	0.5862
KO	0.0096	0.1417	0.1668	0.4258
MRK	0.3391	0.5454	0.7103	0.8744
MSF	0.0058	0.0135	0.4281	0.3431
PFE	0.6153	0.9548	0.8058	0.9870
WMT	0.0665	0.0484	0.8653	0.7413
XO	0.0788	0.0092	0.3939	0.0559

APPENDIX A. PROOFS OF THEOREMS

A.1. **Proof of Theorem 1.** This is a standard result and the proof will be omitted.

Q.E.D

A.2. **Proof of Theorem 2.** The result follows directly from the application of Theorem 1.1 in Chen and Tsay (1993).

Q.E.D

A.3. **Proof of Theorem 3.** It is easy to see that $G(\mathbf{x}_t, z_t; \boldsymbol{\psi})$ in (5) is continuous in the parameter vector $\boldsymbol{\psi}$. This follows from the fact that, for each value of $\mathbf{x}_t$ and z_t, $f(z_t; \gamma_m, c_m)$, $m = 1, \ldots, M$, in (5) depend continuously on γ_m and c_m. Similarly, $G(\mathbf{x}_t, z_t; \boldsymbol{\psi})$ is continuous in $\mathbf{x}_t$ and z_t, and therefore measurable, for each fixed value of the parameter vector $\boldsymbol{\psi}$. Again, under stationarity, it is clear that $\mathbb{E}[q_t(\boldsymbol{\psi})] < \infty, \forall\, t$.

Restrictions (R.1)–(R.3) in Assumption 3 guarantee that the HARST model is identifiable, so that $\mathcal{Q}(\boldsymbol{\psi})$ is uniquely maximized at $\boldsymbol{\psi}_0$. This completes the proof.

Q.E.D

A.4. **Proof of Theorem 4.** Following White (1994, page 29), $\boldsymbol{\psi} \xrightarrow{p} \boldsymbol{\psi}_0$ if the following conditions hold:

(1) The parameter space $\boldsymbol{\Psi}$ is compact.
(2) $\mathcal{Q}_T(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Furthermore, $\mathcal{Q}_T(\boldsymbol{\psi})$ is a measurable function of y_t, $t = 1, \ldots, T$, for all $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.
(3) $\mathcal{Q}(\boldsymbol{\psi})$ has a unique maximum at $\boldsymbol{\psi}_0$.
(4) $\mathcal{Q}_T(\boldsymbol{\psi}) \xrightarrow{p} \mathcal{Q}(\boldsymbol{\psi})$.

Condition (1) is satisfied by Assumption 2. Theorem 3 shows that Conditions (2) and (3) are satisfied.

Now set $g(\boldsymbol{\psi}) = q_t(\boldsymbol{\psi}) - \mathbb{E}[q_t(\boldsymbol{\psi})]$. Theorem 3 implies that $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |g(\boldsymbol{\psi})|\right] < \infty$. In addition, because $g(\boldsymbol{\psi})$ is stationary with $\mathbb{E}[g(\boldsymbol{\psi})] = 0$, by Theorem 3.1 in Ling and McAleer (2003) it follows that $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|T^{-1}\sum_{t=1}^{T} g(\boldsymbol{\psi})\right| = o_p(1)$ and Condition (4) is satisfied.

Q.E.D

A.5. **Proof of Theorem 5.** To prove the asymptotically normality of the QMLE, we need the following conditions in addition to those given in the proof of Theorem 4 (see White (1994, page 92)).

(5) The true parameter vector $\boldsymbol{\psi}_0$ is interior to $\boldsymbol{\Psi}$.
(6) The matrix

$$\mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 q_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right)$$

exists and is continuous in $\boldsymbol{\Psi}$.

(7) The matrix $\mathbf{A}_T(\boldsymbol{\psi}) \xrightarrow{p} \mathbf{A}(\boldsymbol{\psi}_0)$, for any sequence $\boldsymbol{\psi}_T$ such that $\boldsymbol{\psi}_T \xrightarrow{p} \boldsymbol{\psi}_0$.

(8) The score vector satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial q_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right) \xrightarrow{D} \mathsf{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\psi}_0)).$$

Condition (5) is satisfied by assumption. Condition (6) follows from the fact that $q_t(\boldsymbol{\psi})$ is differentiable of order two on $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ and the stationarity of the HARST model. Condition (7) is verified by using the same reasoning as in the proof of Theorem 4 and the results of Theorem 3.1 in Ling and McAleer (2003). Furthermore, non-singularity of $\mathbf{A}(\boldsymbol{\psi}_0)$ follows immediately from identification of the HARST model and the non-singularity of $\mathbf{B}(\boldsymbol{\psi}_0)$ (see Hwang and Ding (1997)).

Define

$$\nabla G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}_0\right) \equiv \left.\frac{\partial G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}\right)}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \quad \text{and}$$

$$\nabla^2 G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}_0\right) \equiv \left.\frac{\partial^2 G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}\right)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}.$$

Using Theorem 2.4 from White and Domowitz (1984), the sequence $2\boldsymbol{\xi}'\nabla G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}_0\right)\varepsilon_t$ obeys the Central Limit Theorem (CLT) for some $(r \times 1)$ vector $\boldsymbol{\xi}$, such that $\boldsymbol{\xi}'\boldsymbol{\xi} = 1$. Assumptions A(i) and A(iii) of White and Domowitz (1984) hold because ε_t is a martingale difference sequence. Assumption A(ii) holds with $V = 4\sigma^2\boldsymbol{\xi}'\mathbb{E}\left[\nabla G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}_0\right)\nabla' G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}_0\right)\right]$. Furthermore, since any measurable transformation of mixing processes is itself mixing (see Lemma 2.1 in White and Domowitz (1984)), $2\boldsymbol{\xi}'\nabla G\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}_0\right)\varepsilon_t$ is a strong mixing sequence and obeys the CLT. By using the Cramér-Wold device, $\nabla Q\left(\mathbf{x}_t, z_t; \boldsymbol{\psi}\right)$ also obeys the CLT with covariance matrix $\mathbf{B}(\boldsymbol{\psi}_0)$, which is $O(1)$ and non-singular. This completes the proof.

Q.E.D

A.6. **Proof of Theorem 6.** This is the precise form of the LM test statistic for an additional regime in the HARST model. Under Assumptions 1-4, the asymptotic distribution of the LM statistic is a standard result for nonlinear regression models.

Q.E.D

Departamento de Economia PUC-Rio
Pontifícia Universidade Católica do Rio de Janeiro
Rua Marques de Sâo Vicente 225 - Rio de Janeiro 22453-900, RJ
Tel.(21) 35271078 Fax (21) 35271084
www.econ.puc-rio.br
flavia@econ.puc-rio.br