Essex Finance Centre
Working Paper Series

Working Paper No 77: 06-2022

“Choosing between persistent and stationary volatility”

“Ilias Chronopoulos, Liudas Giraitis, George Kapetanios”

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

CHOOSING BETWEEN PERSISTENT AND STATIONARY VOLATILITY

BY ILIAS CHRONOPOULOS[1], LIUDAS GIRAITIS[2] AND GEORGE KAPETANIOS[3]

[1]*Essex Business School, University of Essex, ilias.chronopoulos@essex.ac.uk*

[2]*School of Economics and Finance, Queen Mary University of London, l.giraitis@qmul.ac.uk*

[3]*King's Business School, King's College London, george.kapetanios@kcl.ac.uk*

This paper suggests a multiplicative volatility model where volatility is decomposed into a stationary and a non-stationary persistent part. We provide a testing procedure to determine which type of volatility is prevalent in the data. The persistent part of volatility is associated with a nonstationary persistent process satisfying some smoothness and moment conditions. The stationary part is related to stationary conditional heteroskedasticity. We outline theory and conditions that allow the extraction of the persistent part from the data and enable standard conditional heteroskedasticity tests to detect stationary volatility after persistent volatility is taken into account. Monte Carlo results support the testing strategy in small samples. The empirical application of the theory supports the persistent volatility paradigm, suggesting that stationary conditional heteroskedasticity is considerably less pronounced than previously thought.

1. Introduction. Two important issues widely discussed in the statistical and finance literature, over the last 25 years, are structural change and volatility modelling. Starting with the seminal work of [12], volatility modelling has developed into a large topic of study. Most work has produced volatility models that are stationary and allow for time variation in the conditional variance. There are two important groups of parametric models used to model volatility. The first group represents the conditional variance as a function of observables and includes autoregressive conditional heteroskedasticity (ARCH) and generalised ARCH (GARCH) models. The second group, where the conditional variance is treated as a latent variable and may depend on more than one innovation processes, includes stochastic volatility models.

Empirical work though, has repeatedly concluded that volatility can exhibit extreme persistence. Such persistence is not easily accommodated by stationary volatility models. The challenge is revealed through the *integrated GARCH effect*, see e.g. [23], when parameter estimates are observed to lie near the boundary of stationarity. This effect can be caused by smooth or abrupt structural change in the unconditional variance over time. So it is possible that once allowed for, volatility can be best characterised by persistent, and possibly nonstationary processes. There is a growing literature that tries to characterise volatility using parameter processes that allow for gradual change in the unconditional variance. First, we succinctly summarise the main ways this problem is addressed in the literature, and then present our main contributions.

The first line of recent research on structural change has focused on paradigms coming from the statistical literature, such as the work of [25] and [6], where parameter processes are smooth deterministic functions of time. [8] proposed the locally stationary time-varying ARCH model, that is globally nonstationary. Along the same lines [28] proposed another

Keywords and phrases: ARCH effect, persistence, volatility, time-varying coefficient models, non-parametric estimation.

model with deterministic smoothly varying parameters, where volatility is multiplicatively decomposed into a stationary and nonstationary part. The assumption that a nonstationary part could drive volatility, has recently permeated in standard ARCH (stationary) models as well. Specifically, [13] and [4] propose using splines, [22] use the Fourier Flexible form of [14] in their periodic volatility model, [20] consider the generalized GARCH model and, in a series of papers [1] and [2], suggest to model volatility as a linear combination of logistic functions.

While the above characterisations provide an avenue to describe and estimate stationary and nonstationary volatility processes, there is no clear way to separate the two kinds of volatility. Further, the above characterisations are either tied to parametric forms or assumed to be smooth deterministic functions of time.

This paper makes a number of contributions. We suggest a multiplicative specification of a volatility process that allows a stationary and a nonstationary part. The latter can potentially account for the extreme persistence of volatility observed in data. We use ideas from the recent literature on structural change to show how persistence, perhaps surprisingly, allows kernel estimation of the unobserved stochastic persistent part of volatility without strong parametric assumptions and the requirement to be a smooth, deterministic function of time, see e.g. [25], [6], [21] and [28], among others. While smooth deterministic functions for the persistent part of volatility are still allowed, including stochastic elements in persistent volatility modelling can provide a richer representation of volatility.

Recent work by [16] shows that as long as a parameter process satisfies some smoothness and moment or boundedness conditions, it can be stochastic but still estimable using a kernel estimator. Such processes may adequately fit the observed behaviour of volatility, as they are clearly more persistent than stationary processes. In fact, persistence is their most distinctive characteristic. [17] essentially ask the following question: Assuming a decomposition of the form $y_t = h_t u_t$, for some observed process y_t and unobserved stationary process u_t, what properties should h_t have, so that h_t^2 can be consistently estimated by, essentially, a rolling window form, mean estimate of y_t^2? They show that h_t has to change slowly, in the sense that $|h_t - h_s|$ has to be small when t and s are close, and thus, stationary processes do not qualify. A normalised random walk provides a canonical example for the sort of processes we have in mind.

We demonstrate in this paper how the uniform consistent estimation of h_t leads to a strategy of separation between the stationary and persistent parts of volatility. Basically, if the persistent part can be uniformly estimated, then the rescaled series of residuals $\widehat{|u_t|^\gamma} = (\widehat{h_t^\gamma})^{-1}|y_t|^\gamma$, $\gamma > 0$, can be used to test for ARCH effects (conditional heteroskedasticity) or the presence of a stationary volatility in u_t. If only persistent volatility is present, standard ARCH tests will not detect ARCH effects in residuals. If the persistent part h_t is absent, the normalisation by $\widehat{h_t^\gamma}$ will not distort the residuals and stationary volatility in u_t will be detected. Our specification allows for both the persistent and stationary parts of volatility to co-exist. Moreover, they can be extracted from the data. After testing for ARCH effects in u_t, is performed, in a second step, a stationary volatility model can be fitted to u_t. This extension is beyond the scope of the current paper.

In this paper, we discuss, in detail, conditions needed for consistent estimation of the persistent part, h_t, of volatility and further, conditions that enable the use of standard ARCH tests to separate persistent volatility from stationary volatility. We provide illustrative Monte Carlo results that support our approach on testing, in small samples. We proceed and present extensive empirical evidence clearly supporting the persistent volatility paradigm, suggesting that stationary time varying conditional volatility is less pronounced than previously thought and, further, conditional second moments of asset returns are very persistent and change slowly.

The remainder of this paper is organised as follows. Section 2 presents our statistical procedure and theoretical results. Section 3 contains the simulation study. Section 4 reports the empirical results from implementing our testing strategy in financial data. Section 5 concludes. Proofs are relegated to the Supplement. Below $\to_D$, $\to_P$ stand for convergence in distribution and probability.

2. Theoretical considerations. We consider the following white noise model for a series of uncorrelated random variables

$$y_t = h_t u_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\{u_t\}$ is a stationary sequence of uncorrelated random variables with $Eu_t = 0$, $Eu_t^2 = 1$, and $h_t > 0$ is the persistent part of volatility (stochastic or deterministic). Formally, $y_t = y_{tn}$, $t = 1, ..., T$ and $h_t = h_{tn}$, $t = 1, ..., T$ are triangular arrays but it is unnecessary to add the additional index in what follows. We assume that sequences $\{u_t\}$ and $\{h_t\}$ are mutually independent. Then

$$\operatorname{cov}(y_t, y_s) = E[h_t h_s]E[u_t u_s] = 0 \quad \text{for } t \neq s.$$

First, we establish notions of persistent and stationary volatility. We will assume that h_t is measurable with respect to the information set $\mathcal{F}_{t-1}$ at time $t-1$, and $E(u_t|\mathcal{F}_{t-1}) = 0$. Then, the conditional variance of y_t is defined by

$$\operatorname{var}(y_t|\mathcal{F}_{t-1}) = h_t^2 E(u_t^2|\mathcal{F}_{t-1}) = h_t^2\sigma_t^2. \tag{2}$$

To specify the properties of persistence for h_t, we introduce below Assumption M. Overall, the notion of persistence of h_t simplifies to

$$\lim_{T\to\infty} (h_{t,T}^2 - h_{t-1,T}^2) =_P 0, \tag{3}$$

for any $1 \le t = t_T \le T$. For example, if h_t is a deterministic function, the property (3) will imply that the unconditional variance $\operatorname{var}(y_t) = h_t^2 E u_t^2$ changes smoothly when T increases. Other processes, such as locally stationary and stochastic unit root processes, h_t, satisfy this property as well, as we discuss below. We refer to h_t^2 as the persistent part of volatility.

There is a vast body of literature on modeling stationary volatility. We define the stationary part of volatility as a conditional variance

$$\sigma_t^2 = \operatorname{var}(u_t|\mathcal{F}_{t-1}),$$

with respect to the information set $\mathcal{F}_{t-1}$ at time $t-1$. Here, both $\{\sigma_t^2\}$ and $\{\sigma_t^2 - \sigma_{t-1}^2\}$ are stationary processes and, thus, the persistence property $\sigma_t^2 - \sigma_{t-1}^2 = o_P(1)$, does not hold. The main two classes of stationary volatility models for σ_t^2 are the autoregressive conditional heteroskedastic ARCH and GARCH models and stochastic volatility models.

Our objective is to test whether the conditional variance $\operatorname{var}(y_t|\mathcal{F}_{t-1})$ contains a stationary component σ_t^2. A simple general specification of such a hypothesis is given by

$$H_0: \operatorname{var}(y_t|\mathcal{F}_{t-1}) = h_t^2 \quad vs \quad H_1: \operatorname{var}(y_t|\mathcal{F}_{t-1}) = h_t^2\sigma_t^2, \tag{4}$$

where $\{\sigma_t^2\}$ is a stationary sequence of dependent random variables. To construct a feasible testing procedure, we further assume that, under H_0, $\{u_t\}$ is a sequence of independent identically distributed (i.i.d.) random variables and y_t is generated according to the following processes:

$$\begin{aligned} &y_t = h_t\varepsilon_t && \text{under } H_0, \\ &y_t = h_t u_t, \quad u_t = \sigma_t\varepsilon_t && \text{under } H_1, \end{aligned} \tag{5}$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence with $E\varepsilon_t = 0$, $\mathrm{var}(\varepsilon_t) = 1$ and such that $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$. This model specification implies (4).

According to the specification (5), $\{u_t\}$ is a white noise process. Essentially, a standard test for ARCH effect is a white noise test, for the squared time series $\{u_t^2\}$ and requires the existence of six finite moments for u_t. To relax this assumption, in this paper we consider tests for ARCH effects, based on $\{|u_t|^\gamma\}$, $\gamma > 0$, which is basically a white-noise test for the time series $\{|u_t|^\gamma\}$. The analysis covers the case $\gamma = 2$.

Since $\{u_t\}$ is not observed, we base our test for ARCH effects on the power transform of residuals

$$\widehat{|u_t|^\gamma} \quad \text{where} \quad \widehat{|u_t|^\gamma} = (\widehat{h_t^\gamma})^{-1}|y_t|^\gamma \tag{6}$$

and $\widehat{h_t^\gamma} > 0$ is an estimate of h_t^γ. Such testing requires uniformly consistent estimation of h_t^γ by $\widehat{h_t^\gamma}$ and thus, stronger conditions on $\{h_t, u_t\}$ are needed, than for consistent point estimation of h_t at time t. This is reflected in the Assumptions M and H, we make for u_t and h_t. In particular, we impose the assumption of mutual independence between $\{h_t^2\}$ and $\{u_t^2\}$, which clearly holds for a deterministic volatility factor h_t.

Assumption M (α-mixing)

1. $\{u_t\}$ is a stationary white-noise ergodic sequence with $Eu_t = 0$, $Eu_t^2 = 1$, $Eu_tu_s = 0$ for $t \neq s$.
2. $\{u_t\}$ is α-mixing with mixing coefficients $\alpha_k \leq c\phi^k$, $k \geq 1$, for some $0 < \phi < 1$ and $c > 0$.

Assumption H (Smoothness)

1. For some $\nu \in (1/2, 1]$,

$$\begin{aligned} &|h_t - h_j| \leq C(|t-j|/T)^\nu, \ \ t, j = 1, ..., T, \quad \text{or} \\ &|h_t - h_j| \leq (|t-j|/T)^\nu \xi_{tj}, \end{aligned} \tag{7}$$

   where $C > 0$ does not depend on t, T, and for some $0 < \alpha < \infty$, $c > 0$,

$$\max_{t,j=1,\ldots,T} E[\exp(c|\xi_{tj}|^\alpha)] \leq C < \infty, \qquad \max_{t=1,\ldots,T} E[\exp(c|h_t|^\alpha)] \leq C < \infty. \tag{8}$$

2. There exists $a > 0$ such that $h_t \geq a > 0$ a.s. for all $t \geq 1$.
3. $\{h_t\}$ and $\{u_t\}$ are mutually independent.

The model specification (1) abstracts from the general case of a time series with a specified conditional mean. It is possible to generalize our test for ARCH effects to a time series with a non-zero conditional mean

$$y_t = \mu_t + h_t u_t, \quad \mu_t = E\left(y_t|\mathcal{F}_{t-1}\right), \quad t = 1, \ldots, T.$$

The smoothness condition (7) in Assumption H implies that the persistent component of volatility, h_t, drifts slowly in time, which essentially rules out abrupt or explosive behaviour for h_t. This assumption is widely used in the statistical and econometric literature. It allows the use of both deterministic and stochastic time-varying processes h_t and implies the persistence property (3).

The deterministic specification $h_t = g(t/T)$, $t = 1, ..., T$, where $g(\cdot)$ is a Lipschitz smooth function with parameter $1/2 < \nu \leq 1$, i.e. $|g(x) - g(y)| \leq C|x - y|^\nu$, is a standard assumption in the work of Dahlhaus on locally stationary processes (see, e.g. [6] or [7]). It implies $|h_t - h_s| \leq C(|t-s|/T)^\nu$.

The stochastic specification $h_t = T^{-\nu}|\sum_{j=1}^{t} v_j|$, $t = 1, ..., T$ of h_t, where $\{v_j\}$ is a stationary sequence with zero mean, was proposed by [16, 15, 17], to allow for a persistent process h_t that can be presented as non-stationary random walks, see Example 1 below.

A combination of the two, satisfying (7) with parameter ν can be summarised as

$$h_t = |T^{-\nu}(v_1 + ... + v_t) + g(t/T)| + a, \quad t = 1, ..., T, \qquad (a > 0). \tag{9}$$

Our testing procedures will still work for the case of y_t with a non-zero conditional mean; in this case a first step estimator for the mean will be required, see e.g. [5].

EXAMPLE 1. Let $\{v_j\}$ be a stationary Gaussian ARFIMA(p, d, q) sequence with parameter $d \in (0, 1/2)$ and zero mean, see e.g. Chapter 7 in [18]. Then $h_t = T^{-\nu}\big|\sum_{j=1}^{t} v_j\big|$, $t = 1, ..., T$ satisfies (7) of Assumption H with $\nu = 1/2 + d$ and $\alpha = 2$. Indeed, for $t > s$,

$$\begin{aligned} |h_t - h_s| &= \Big|T^{-\nu}\big|\textstyle\sum_{j=1}^{t} v_j\big| - T^{-\nu}\big|\sum_{j=1}^{s} v_j\big|\Big| \\ &\le T^{-\nu}|\textstyle\sum_{j=s+1}^{t} v_j| \le (|t-s|/T)^{\nu}|\xi_{ts}|, \quad \xi_{ts} = (t-s)^{-\nu}\sum_{j=s+1}^{t} v_j. \end{aligned}$$

Here, ξ_{ts} is a Gaussian variable, and by Proposition 3.3.1 in [18] and stationarity of $\{v_j\}$, the variance $\text{var}(\xi_{ts}) = \text{var}(\xi_{t-s,0}) \to v^2 < \infty$ as $t - j \to \infty$. Hence, ξ_{ts} and h_t satisfy (8) with $\alpha = 2$.

2.1. *Volatility estimation.* To extract residuals $\widehat{|u_t|}^{\gamma} = (\widehat{h_t^{\gamma}})^{-1}|y_t|^{\gamma}$, required for the testing of ARCH effects, we need an estimate for h_t^{γ} in $|y_t|^{\gamma} = h_t^{\gamma}|u_t|^{\gamma}$. Without loss of generality we assume that h_t is rescaled so that $E|u_t|^{\gamma} = 1$. We will show that in model (1), under Assumptions H and M, h_t^{γ} can be consistently estimated by a kernel type estimate

$$\widehat{h_t^{\gamma}} = K_t^{-1}\sum_{j=1}^{T} b_{H,|t-j|}|y_j|^{\gamma}, \quad K_t = \sum_{j=1}^{T} b_{H,|t-j|}, \quad t = 1, ..., T, \tag{10}$$

where $b_{H,|t-j|} = K(|t-j|/H)$ are kernel weights. $K(\cdot)$ is assumed to be a non-negative and bounded function, with piecewise bounded derivative, and H is a bandwidth parameter that satisfies $H = o(T)$, as $T \to \infty$. Commonly used examples of $K(x)$ include:

$$K(x) = (1/2)I(|x| \le 1), \quad \text{flat kernel},$$

$$K(x) = (3/4)(1 - x^2)I(|x| \le 1), \quad \text{Epanechnikov kernel},$$

$$K(x) = (1/\sqrt{2\pi})e^{-x^2/2}, \quad \text{Gaussian kernel}.$$

The first two kernel functions have finite support, whereas the Gaussian kernel has infinite support. We further assume that on its support,

$$K(x) \le C(1 + x^g)^{-1}, \ |(d/dx)K(x)| \le C(1 + x^g)^{-1}, \ x \ge 0 \ \text{ with } g > 4, \ C > 0. \tag{11}$$

Under this setup, in Lemma 7.2 in the Supplement we show the pointwise consistency of this estimate:

$$|\widehat{h_t^{\gamma}} - h_t^{\gamma}| = O_p\left((H/T)^{\nu} + H^{-1/2}\right). \tag{12}$$

Similar results for vector autoregressive models were derived in [17]. In Lemma 7.2, using the results of [11], we establish the uniform convergence

$$\max_{t=1,...,T}|\widehat{h_t^{\gamma}} - h_t^{\gamma}| = o_P(1). \tag{13}$$

This uniform convergence result will prove useful in our testing procedure for the distinction between the persistent and stationary parts of volatility, that follows.

2.2. *Testing.* In this subsection we outline how our strategy to discriminate between the persistent and stationary components of volatility works under (1), where u_t is not observed. Basically, it will be seen that the test of hypothesis (4) reduces to a white noise test for $\{|u_t|^\gamma\}$. First, we briefly summarise the basic tests for ARCH effects when a white noise time series $\{u_t\}$ is observed.

The Lagrange Multiplier (LM) test by [12] is the most commonly used standard ARCH test to detect autoregressive conditional heteroskedasticity (or stationary volatility) in $\{u_t\}$. We simply fit to u_t^2 an AR(p), $p \geq 1$, model

$$u_t^2 = \beta_0 + \beta_1 u_{t-1}^2 + ... + \beta_p u_{t-p}^2 + \eta_t, \tag{14}$$

where $\beta_0 > 0$ and test the following null hypothesis:

$$H_0: \ \beta_1 = \beta_2 = \ldots = \beta_p = 0 \quad vs \quad H_1: \ \beta_j \neq 0 \quad \text{for some} \quad j = 1, ..., p. \tag{15}$$

The null hypothesis H_0 implies absence of correlation in the first p lags of the series $\{u_t^2\}$ and *vice versa*. Basically, this ARCH LM test is equivalent to testing for absence of autocorrelation in $\{u_t^2\}$.

The test statistic of the ARCH LM test by [12] is defined as TR^2, where T is the sample size and R^2 is the coefficient of determination of the AR regression (14). Under H_0, when $\{u_t\}$ is an i.i.d. sequence with finite fourth moment, the LM statistic follows asymptotically a χ_p^2 distribution. Further tests, such as the Wald and Likelihood ratio, have been shown to be asymptotically equivalent to the LM test.

Through testing, the literature mainly addresses two distinct problems: the misspecification of the conditional mean, see e.g. the discussion in [3], and the correct specification of the volatility process. Our work naturally falls in the second category by addressing the question of whether allowing for a persistent component h_t in (1), provides a better specification for the volatility process. By ARCH effects in $\{u_t\}$ we mean the presence of correlation in a sequence $\{|u_t|^\gamma\}$, where $\gamma > 0$ is selected in advance. We will test for ARCH effects in the unobserved component u_t of $y_t = h_t u_t$ in (1) by fitting to $|u_t|^\gamma$ an $AR(p)$, $p \geq 1$ model

$$|u_t|^\gamma = \beta_0 + \beta_1 |u_{t-1}|^\gamma + ... + \beta_p |u_{t-p}|^\gamma + \eta_t \tag{16}$$

and then testing the hypothesis (15) on $\beta_1, .., \beta_p$. We replace the unobserved variables $|u_t|^\gamma$ by residuals

$$\widehat{|u_t|^\gamma} = (\widehat{h_t^\gamma})^{-1} |y_t|^\gamma, \tag{17}$$

where $\widehat{h_t^\gamma}$ is the kernel estimate (10) of the γ-power h_t^γ of the persistent factor h_t of volatility. Our aim is to show that asymptotically, it is equivalent to test for ARCH effects using the residuals $\widehat{\boldsymbol{u}^\gamma} = \left[\widehat{|u_1|^\gamma}, \widehat{|u_2|^\gamma}, \ldots, \widehat{|u_T|^\gamma}\right]'$, instead of $\boldsymbol{u}^\gamma = \left[|u_1|^\gamma, |u_2|^\gamma, \ldots, |u_T|^\gamma\right]'$.

In addition, such an equivalence implies that the residuals, $\widehat{u}_t = \widehat{h}_t^{-1} y_t$ obtained using $\widehat{h}_t$ with $\gamma = 1$, should behave as a white noise. In Theorems 2.2 and 2.3 we show that both the ARCH LM test based on regression on u_t and the correlogram of $\widehat{u}_1, ..., \widehat{u}_T$ can be used to test for absence of correlation in $\{u_t\}$.

The ARCH test using regression (14) for powers, $|u_t|^\gamma$ obtained in Theorem 2.1, shows that, that if a stationary process σ_t^2 (co-)drives the volatility via u_t, then the normalisation by $\widehat{h_t^\gamma}$ in (17) will not corrupt the properties of testing.

In our setup and for $p \geq 1$, we consider $TS\big(\boldsymbol{u}^\gamma\big) = TR^2$, the test statistic where R^2 is the coefficient of determination of the AR regression (16) based on $\boldsymbol{u}^\gamma$, and $TS\left(\widehat{\boldsymbol{u}^\gamma}\right) = T\widehat{R^2}$ where $\widehat{R^2}$ is based on the residuals, $\widehat{\boldsymbol{u}^\gamma}$, as described above. The formulas of $S\big(\boldsymbol{u}^\gamma\big)$ and $S\big(\widehat{\boldsymbol{u}^\gamma}\big)$ are given in (6.6) and (6.7) of the Supplement.

The following Theorem proven in the Supplement, gives a sufficient condition for LM test for ARCH effects to be asymptotically valid when applied to $\widehat{\boldsymbol{u}^{\gamma}}$, instead of $\boldsymbol{u}^{\gamma}$.

Denote by $\gamma_k = \mathrm{cov}(|u_k|^{\gamma}, |u_0|^{\gamma})$, $k \geq 0$ the covariance function of a stationary sequence $\{|u_t|^{\gamma}\}$, $\gamma > 0$, and define the $p \times p$ matrix $\boldsymbol{\Gamma}_p$ and $p \times 1$ vector $\boldsymbol{\gamma}_p$ by setting

$$\boldsymbol{\Gamma}_p = \big(\gamma_{|j-k|}\big)_{j,k=1,\dots,p}, \qquad \boldsymbol{\gamma}_p = \big(\gamma_1, ..., \gamma_p\big)'.$$

Denote by $\boldsymbol{\beta}_p$ the $p \times 1$ vector of parameters $(\beta_1, ..., \beta_p)$ which will appear in testing under the alternative hypothesis H_1 and set

$$\boldsymbol{\beta}_p = (\beta_1, ..., \beta_p)' = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p, \qquad \sigma_p^2 = \mathrm{var}(|u_{p+1}|^{\gamma} - \beta_1|u_p|^{\gamma} - ... - \beta_p|u_1|^{\gamma}).$$

Notice that the existence of $\boldsymbol{\Gamma}_p^{-1}$ follows from Lemma 3.1(i) in [10] because the stationary sequence $\{u_t^2\}$ has bounded a spectral density. The latter follows from the absolute summability of the covariance function γ_k, see (7.8) in the Supplement.

Recall notation ν of the smoothness parameter of h_t appearing in (7).

THEOREM 2.1. *Let $\{y_t,\ t = 1, ..., T\}$, follow (1), Assumptions H and M hold, and H satisfies*

$$T^{1/2+a} \leq H = o(T^{1-(1/4\nu)}) \qquad \textit{(for some } a > 0\textit{)}. \tag{18}$$

Assume that $\gamma > 0$ and $E|u_t|^{3\gamma} < \infty$. Then the ARCH LM *test statistic based on regression (16) on $|u_t|^{\gamma}$, has the following properties. As $T \to \infty$, for any $p \geq 1$,*

$$S(\widehat{\boldsymbol{u}^{\gamma}}) = S(\boldsymbol{u}^{\gamma}) + o_P(1) = \sigma_p^{-2}\boldsymbol{\beta}_p'\boldsymbol{\Gamma}_p\boldsymbol{\beta}_p + o_P(1). \tag{19}$$

In addition, if $\{u_t\}$ is an i.i.d. sequence, then $\boldsymbol{\beta}_p = 0$, and

$$TS(\widehat{\boldsymbol{u}^{\gamma}}) = TS(\boldsymbol{u}^{\gamma}) + o_P(1) \to_D \chi_p^2. \tag{20}$$

Result (19) implies that testing for ARCH effects based on regression (16) is equivalent to testing for the white noise for the series $\{|u_t|^{\gamma}\}$. Indeed, the matrix $\boldsymbol{\Gamma}_p$ is positive definite, and therefore its smallest and largest eigenvalues obey $0 < \lambda_{min} \leq \lambda_{max} < \infty$. Notice that $\boldsymbol{\beta}_p'\boldsymbol{\Gamma}_p\boldsymbol{\beta}_p = \boldsymbol{\gamma}_p'\boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p$, and

$$\boldsymbol{\beta}_p'\boldsymbol{\Gamma}_p\boldsymbol{\beta}_p \geq ||\boldsymbol{\beta}_p||^2\lambda_{min}, \quad \boldsymbol{\gamma}_p'\boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p \geq ||\boldsymbol{\gamma}_p||^2\lambda_{max}^{-1},$$

where $||\boldsymbol{\beta}_p||$ denotes the Euclidean norm of $\boldsymbol{\beta}_p$. Hence, $||\boldsymbol{\beta}_p|| = 0$ implies $||\boldsymbol{\gamma}_p|| = 0$ and *vice versa* which proves the above claim.

Theorem 2.1 implies that a test for ARCH effects in $\{u_t\}$ based on statistic $TS(\widehat{\boldsymbol{u}^{\gamma}})$ has the same asymptotic size and power properties as a test based on $TS(\boldsymbol{u}^{\gamma})$ applied on unobserved $\boldsymbol{u}^{\gamma}$. If the hypothesis H_0 is not rejected, then this implies the absence of correlation in $\{|u_t|^{\gamma}\}$ up to lag p. Conversely, the alternative H_1 is detected with a rate T.

Notice that the value $\boldsymbol{\beta}_p = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p$ appearing in (19) is the same as the "true" value of the parameter $\boldsymbol{\beta}_p$ estimated by the OLS method in regression (14) when η_t is a white noise sequence.

REMARK 1. If Assumption H is satisfied with $\nu = 1$, for example, which is the case for deterministic weights $h_t = g(t/T)$, where g is a continuous piecewise differentiable function with a bounded derivative, then assumption (18) on the bandwidth H becomes

$$T^{1/2+a} \leq H = o(T^{3/4}). \tag{21}$$

Note that lower values of ν imply narrower permitted intervals for bandwidth H. Given that ν is unknown, a theoretical recommendation for the choice of H is that it should be larger than $T^{1/2}$ and less than $T^{3/4}$. In practice, we recommend trying different values for H, within the interval $T^{1/2}$ to $T^{3/4}$, to establish test robustness. This matter is further discussed in the Monte Carlo study.

REMARK 2. The moment condition $E|u_t|^{3\gamma} < \infty$ of Theorem 2.1 implies that testing for ARCH effects based on regression (14) using u_t^2, $|u_t|$ or $|u_t|^{1/2}$ which corresponds to $\gamma = 2, 1, 1/2$, requires the existence of $E|u_t|^6$, $E|u_t|^3$ and $E|u_t|^{3/2}$, respectively. If $E|u_t|^{3\gamma+\epsilon} < \infty$ for some $\epsilon > 0$, then (18) can be replaced by

$$T^{1/2} = o(H), \quad H = o(T^{1-(1/4\nu)}).$$

We also suggest two methods to examine whether the unobserved time series $\{u_t\}$ in the model (1) is a white noise sequence. The first approach is similar to testing for ARCH effects. It amounts to fitting to u_t an $AR(p)$, $p \geq 1$ model

$$u_t = \beta_0 + \beta_1 u_{t-1} + \ldots + \beta_p u_{t-p} + \eta_t \tag{22}$$

and testing the hypothesis H_0 and H_1 on $(\beta_1, \ldots, \beta_p)$ in (14). The test statistics $TS(\boldsymbol{u})$ and $TS(\widehat{\boldsymbol{u}})$ in Theorem 2.2 correspond to the AR regression (22) on u_t and $\widehat{u}_t = (\widehat{h}_t)^{-1} y_t$. They satisfy the asymptotic results of Theorem 2.1 with and $\boldsymbol{\beta}_p$, $\boldsymbol{\Gamma}_p$ defined as in (19) using

$$\gamma_k = \mathrm{cov}(u_k, u_0), \quad \sigma_p^2 = \mathrm{var}(u_{p+1} - \beta_1 u_p - \ldots - \beta_p u_1). \tag{23}$$

THEOREM 2.2. *Let $\{y_t, t = 1, \ldots, T\}$ be as in (1) and $E|u_t|^3 < \infty$. Suppose that Assumptions M, H and (18) hold. Then statistics $TS(\boldsymbol{u})$ and $TS(\widehat{\boldsymbol{u}})$ based on the AR regression (22), have the following properties.*

For any $p \geq 1$, as $T \to \infty$,

$$S(\widehat{\boldsymbol{u}}) = S(\boldsymbol{u}) + o_P(1) = \sigma_p^{-2} \boldsymbol{\beta}_p' \boldsymbol{\Gamma}_p \boldsymbol{\beta}_p + o_P(1). \tag{24}$$

In addition, if $\{u_t\}$ is an i.i.d. sequence, then $\boldsymbol{\beta}_p = 0$, and

$$TS(\widehat{\boldsymbol{u}}) = TS(\boldsymbol{u}) + o_P(1) \to_D \chi_p^2. \tag{25}$$

The same argument as used below Theorem 2.1, implies that the hypothesis $H_0: \beta_1 = \ldots = \beta_p = 0$ is equivalent to the absence of correlation up to lag p in $\{u_t\}$.

Alternatively, absence of correlation in $\{u_t\}$ can be tested using the correlogram of residuals $\widehat{u}_t = (\widehat{h}_t)^{-1} y_t$. This important step of data analysis allows one to verify the model specification (1) for y_t since standard tests for white noise based on y_t might not be applicable.

For $k = 0, 1, \ldots$, denote

$$\begin{aligned} \widehat{r}_k &= T^{-1} \textstyle\sum_{t=k+1}^T (\widehat{u}_t - \overline{\widehat{u}})(\widehat{u}_{t-k} - \overline{\widehat{u}}), \\ \widetilde{r}_k &= T^{-1} \textstyle\sum_{t=k+1}^T (u_t - Eu_t)(u_{t-k} - Eu_{t-k}). \end{aligned} \tag{26}$$

THEOREM 2.3. *Suppose that assumptions of Theorem 2.2 are satisfied. Then, as $T \to \infty$,*

$$\widehat{r}_k = \widetilde{r}_k + o_P(1) = \mathrm{cov}(u_k, u_0) + o_P(1), \quad k \geq 0. \tag{27}$$

In addition, if $\{u_t\}$ is an i.i.d. sequence, then

$$T^{1/2} \widehat{r}_k = T^{1/2} \widetilde{r}_k + o_P(1) \to \mathcal{N}(0, (Eu_1^2)^2) \quad k \geq 1. \tag{28}$$

Denote $\widehat{\rho}_k = \widehat{r}_k/\widehat{r}_0$, $k = 0, 1, 2, ...$ If $\{u_t\}$ is an i.i.d. sequence then (28) of Theorem 2.3 implies that

$$T^{1/2}(\widehat{\rho}_1, ..., \widehat{\rho}_m) \to_D \mathcal{N}(0, I_m), \quad m \geq 1. \tag{29}$$

This shows that using residuals $\widehat{u}_t$, we can perform standard tests for the absence of correlation in $\{u_t\}$ at individual lag k and Ljung-Box tests, for the cumulative lag m as if variables $\{u_t\}$ were observed.

Notice that in our setting no ARCH effects in $\{u_t\}$ imply no correlation in $\{|u_t|^\gamma\}$. This is a slightly weaker property than the i.i.d. assumption on $\{u_t\}$ under H_0. The latter property leads to standard approximations (20) and (29) for the test statistics, which are not guaranteed for a non i.i.d. white noise $\{u_t\}$.

REMARK 3. In financial and economic applications, the common choice for volatility modelling is a stationary GARCH type model. Given this, it is also relevant to test the hypothesis

$$H_0: \ y_t = u_t \quad vs \quad H_1: \ y_t = h_t u_t,$$

for absence of persistent component h_t ($h_t = 1$) in the model (1), where $\{u_t\}$ is a stationary sequence of uncorrelated random variables. In general, this is equivalent to testing mean stability of series $y_t^2 = h_t^2 E u_t^2 + h_t^2(u_t - E u_t^2)$. Tests for detection of alternatives with deterministic h_t were developed in [9].

3. Simulation Study. In this section, we use simulations to verify the theoretical properties of the test statistics $TS(\widehat{\boldsymbol{u}^\gamma})$ for ARCH effects in $\{u_t\}$, and explore its finite-sample size and power performance. In particular, we examine the impact of the three types of persistent volatility h_t (constant, deterministic, stochastic) and the choice of the bandwidth parameter H on the size and power of the test, and how crucial the moment condition $E|u_t|^{3\gamma} < \infty$ is.

We generate an array of samples

$$y_t = h_t u_t, \quad u_t = \sigma_t \varepsilon_t, \quad t = 1, ..., T, \tag{30}$$

where $\{\varepsilon_t\}$ is an i.i.d. $N(0, 1)$ noise. For σ_t^2 we use stationary ARCH(1) and GARCH$(1, 1)$ models:

$$\sigma_t^2 = 1 + \beta u_{t-1}^2, \quad \beta = 0, 0.2, 0.4, \quad \text{ARCH(1) model;} \tag{31}$$
$$\sigma_t^2 = 1 + 0.2u_{t-1}^2 + 0.7\sigma_{t-1}^2, \quad \text{GARCH}(1,1) \text{ model.}$$

The case $\beta = 0$ leads to $\sigma_t^2 = 1$, or H_0, and

$$y_t = h_t \varepsilon_t. \tag{32}$$

We use (32) to study the empirical size of our test for ARCH effects in unobserved u_t at lag p, in particular, to determine whether the size of this test is robust to the choice of H. For a persistent volatility component h_t, with parameter $\nu = 1$, by Theorem 2.1, under (32),

$$TS(\widehat{\boldsymbol{u}^\gamma}) \sim TS(\boldsymbol{u}^\gamma) \sim \chi_p^2 \qquad \text{if} \ \ T^{0.5} = o(H), \ H = o(T^{0.75}).$$

Ideally, we expect both the test statistic $TS(\widehat{\boldsymbol{u}^\gamma})$ based on residuals, and $TS(\boldsymbol{u}^\gamma)$ based on u_t, to exhibit similar empirical size. Further, for $\beta > 0$ in ARCH(1) model or GARCH$(1, 1)$ model in (30)-(31), a stationary volatility component is present, and then for the above choice of H, we expect them to achieve similar empirical power, as shown in Theorem 2.1.

We set the significance level to $\alpha = 5\%$ and conduct 5000 replications. We consider $T = 200, 400, 800, 1600$, $\gamma = 2, 1, 1/2$ and we conduct testing for ARCH effects on $|u_t|^\gamma$ at lags 1, 5 and 10 for $\gamma = 2, 1, 1/2$.

In the tables below, we examine the impact of bandwidth H and parameter γ on testing. The shaded grey range of $H = [T^{0.60}, \ldots, T^{0.70}]$ denotes theoretically permissible values of H for smoothness parameter $\nu = 1$. As a reference point, we include a third shaded line reporting the size and power for statistics $TS(\boldsymbol{u}^\gamma)$. Theoretically, as the sample size increases, size and power in the shaded area should approach the benchmark.

We use three models for $\{y_t\}$.

MODEL 1. We set $h_t = 1$, and σ_t^2 follows (31).

This model does not involve a persistent time-varying volatility component. It includes a stationary volatility component except for the case $\beta = 0$, where $\sigma_t = 1$ and $y_t = \varepsilon_t$.

The left panel of Table 1 confirms that both tests $TS(\widehat{\boldsymbol{u}^2})$ and $TS(\boldsymbol{u}^2)$ for ARCH effects at the lag $p = 5$ based on squares ($\gamma = 2$) achieve similar empirical size and power when H is chosen from the permitted range $[T^{0.6}, T^{0.7}]$.

MODEL 2. We set $h_t = \sin(2\pi t/T) + 2,\ t = 1, ..., T$, and σ_t^2 follows (31).

This model contains a persistent deterministic volatility component h_t, which satisfies Assumption H with parameter $\nu = 1$. For $\beta = 0$, the model $y_t = h_t \varepsilon_t$ includes no stationary volatility (ARCH effects) in u_t. From the right panel of Table 1 we conclude, that overall the empirical size and power of the test $TS(\widehat{\boldsymbol{u}}^2)$ are comparable to those of $TS(\boldsymbol{u}^2)$ as long as $H \in [T^{0.6}, T^{0.7}]$, but the use of a permissible bandwidth H plays an essential role here. In real data, u_t^2 is not observed and the test $TS(\boldsymbol{u}^2)$ cannot be performed. Not surprisingly, in the model $y_t = h_t \varepsilon_t$ ($\beta = 0$), the empirical size of the test $TS(\boldsymbol{y}^2)$ applied on squares y_t^2 of the data is close to 100%, i.e. it falsely suggests the presence of a stationary volatility in y_t.

MODEL 3. We set $h_t = T^{-(d-1/2)}|I_{d,t}| + 1,\ t = 1, ..., T$, where $I_{d,t}$ is a non-stationary ARFIMA$(0, d, 0)$ process. We consider the values $d = 1.2, 1.4, 1.5$, and σ_t^2 follows (31).

Here, we assume that $\{h_t\}$ and $\{\varepsilon_t\}$ are mutually independent. Such a stochastic persistent process h_t satisfies Assumption H with parameter $\nu = d - 1/2 > 1/2$, see Example 1. Table 3 shows, that overall the performance of the test for ARCH effects on squares of residuals at the lag $p = 5$ exhibits similar patterns as for Model 2, although the lower degree of persistence of h_t in Model 3 results in a somewhat smaller rate of detection of spurious presence of stationary volatility in y_t by $TS(\boldsymbol{y}^2)$.

In sum, to test for ARCH effects in $\{u_t\}$, we have used the statistics $S(\widehat{\boldsymbol{u}^\gamma})$ based on residuals $\widehat{|u_t|}^\gamma$. We report testing results for the values $\gamma = 2$, while the results for $\gamma = 1, 1/2$ can be found in the Supplement. Testing results for ARCH effects at lags $p = 1, 10$ produce similar patterns as for $p = 5$, and are available upon request. We report additional testing results for errors $\varepsilon_t \sim t(4)$, in the Supplement. They show that the lack of finite $E|u_t|^{3\gamma}$ moment has an impact on testing results.

Finally, in Figure 1 we explore the impact of bandwidth H on the size of the test by plotting the MC average of the $TS(\widehat{\boldsymbol{u}^\gamma})$ test statistic for various values of H under the null hypothesis. We consider data $y_t = h_t \varepsilon_t$ produced by Model 2 for $T = 1600$. From theory, the Monte Carlo average of a well behaved test statistic should approach the number of degrees of freedom $p = E\chi_p^2$ marked by the black dashed line. We see that this is indeed the case for bandwidths $H \in [T^{0.6}, T^{0.7}]$, suggesting that such bandwidth values perform well in small samples and meet the requirements of our theoretical analysis.

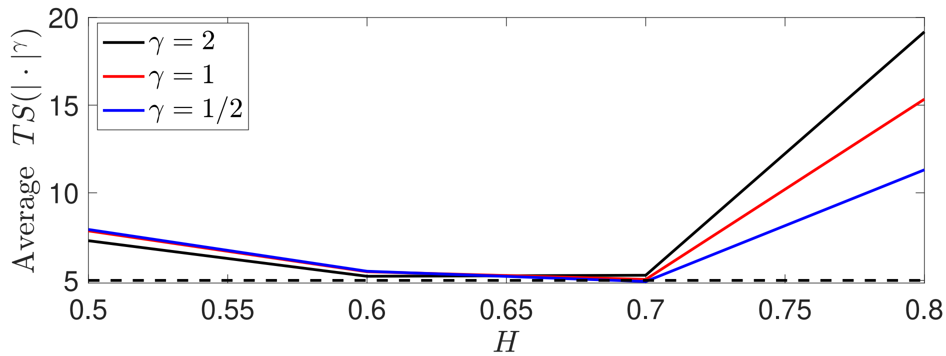

FIG 1. *Average of $TS(\widehat{\boldsymbol{u}^{\gamma}})$ test statistic for various values of H under H_0, $y_t = h_t\varepsilon_t$.*

TABLE 1
Testing for ARCH effects on squares at the lag $p = 5$ in Model 1 and Model 2. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power).

			Model 1 (with $h_t = 1$)				Model 2 (with deterministic h_t)			
T	H	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$
200	$T^{0.5}$	$\widehat{u_t^2}$	12.98	19.04	48.82	7.22	10.50	17.96	49.14	8.02
	$T^{0.6}$		4.90	23.30	60.82	28.66	3.20	24.42	62.84	34.96
	$T^{0.7}$		3.72	30.14	69.60	52.42	7.64	40.68	74.92	66.44
	$T^{0.8}$		3.60	34.64	74.32	65.52	25.10	60.12	84.08	83.34
		u_t^2	4.34	40.30	78.94	77.12	4.34	40.30	78.96	77.12
		y_t^2	4.34	40.30	78.94	77.12	73.66	86.78	93.62	95.22
400	$T^{0.5}$	$\widehat{u_t^2}$	13.18	42.80	89.26	33.86	12.10	42.62	89.30	35.12
	$T^{0.6}$		4.80	52.00	94.12	76.14	3.58	53.36	94.34	79.00
	$T^{0.7}$		3.62	60.04	96.08	90.56	7.90	68.78	96.86	95.10
	$T^{0.8}$		3.80	64.70	97.14	94.94	40.86	86.58	98.64	98.60
		u_t^2	4.68	69.06	97.60	97.22	4.68	69.04	97.60	97.22
		y_t^2	4.68	69.06	97.60	97.22	95.40	99.14	99.80	99.92
800	$T^{0.5}$	$\widehat{u_t^2}$	15.32	80.58	99.92	89.00	14.42	80.62	99.92	89.58
	$T^{0.6}$		5.60	87.42	100.00	99.38	4.74	88.16	100.00	99.52
	$T^{0.7}$		4.38	91.38	100.00	99.88	7.78	93.88	100.00	99.96
	$T^{0.8}$		4.18	92.96	100.00	99.92	60.80	98.98	100.00	100.00
		u_t^2	4.74	94.14	100.00	99.98	4.74	94.14	100.00	99.98
		y_t^2	4.74	94.14	100.00	99.98	99.98	100.00	100.00	100.00
1600	$T^{0.5}$	$\widehat{u_t^2}$	16.44	99.04	100.00	99.96	15.96	99.08	100.00	99.96
	$T^{0.6}$		6.20	99.72	100.00	100.00	5.68	99.72	100.00	100.00
	$T^{0.7}$		5.08	99.82	100.00	100.00	7.06	99.84	100.00	100.00
	$T^{0.8}$		5.16	99.84	100.00	100.00	80.06	100.00	100.00	100.00
		u_t^2	5.44	99.86	100.00	100.00	5.44	99.86	100.00	100.00
		y_t^2	5.44	99.86	100.00	100.00	100.00	100.00	100.00	100.00

TABLE 2

Testing for ARCH effects on squares at the lag $p = 5$ in Model 3. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power).

Model 3 (with stochastic h_t)											
T	H	data	$\beta = 0$			$\beta = 0.2$			$\beta = 0.4$		
			$d = 1.2$	$d = 1.4$	$d = 1.5$	$d = 1.2$	$d = 1.4$	$d = 1.5$	$d = 1.2$	$d = 1.4$	$d = 1.5$
200	$T^{0.50}$	$\widehat{u_t^2}$	11.20	11.80	12.14	19.12	19.48	19.38	49.26	49.00	49.14
	$T^{0.60}$		3.90	4.06	4.30	23.40	23.30	23.36	62.06	61.40	61.52
	$T^{0.70}$		3.92	3.82	3.68	31.40	30.00	29.70	71.22	70.86	70.70
	$T^{0.80}$		6.32	5.04	4.78	38.68	37.14	36.54	77.26	76.54	76.38
		u_t^2	4.52	4.52	4.52	39.76	39.76	39.76	79.92	79.92	79.92
		y_t^2	22.58	20.24	19.36	56.08	54.18	53.76	84.12	83.46	83.34
400	$T^{0.50}$	$\widehat{u_t^2}$	12.24	13.02	13.10	43.94	43.90	43.86	89.96	90.02	89.88
	$T^{0.60}$		4.08	4.58	4.68	54.08	53.34	53.50	94.64	94.54	94.42
	$T^{0.70}$		4.52	4.20	4.04	63.70	62.56	62.22	96.80	96.60	96.56
	$T^{0.80}$		8.34	6.62	6.00	71.02	69.42	68.80	97.92	97.70	97.68
		u_t^2	5.12	5.12	5.12	70.10	70.10	70.10	97.84	97.84	97.84
		y_t^2	36.40	32.48	30.64	83.80	81.96	81.70	98.78	98.68	98.52
	$T^{0.50}$	$\widehat{u_t^2}$	14.22	15.06	15.30	81.62	81.34	81.34	99.92	99.92	99.92
	$T^{0.60}$		4.86	5.22	5.38	88.68	88.30	88.22	99.98	99.98	99.98
	$T^{0.70}$		4.66	4.32	4.52	92.58	92.10	91.94	99.98	99.98	99.98
	$T^{0.80}$		9.46	6.28	5.74	95.22	94.10	93.82	99.98	99.98	99.98
		u_t^2	5.34	5.34	5.34	94.46	94.46	94.46	99.98	99.98	99.98
		y_t^2	51.20	45.04	43.28	98.02	97.68	97.46	99.98	99.98	99.98
	$T^{0.50}$	$\widehat{u_t^2}$	15.44	15.94	16.10	99.28	99.26	99.24	100.00	100.00	100.00
	$T^{0.60}$		5.32	5.64	5.92	99.72	99.68	99.66	100.00	100.00	100.00
	$T^{0.70}$		5.76	5.14	5.06	99.86	99.84	99.80	100.00	100.00	100.00
	$T^{0.80}$		11.98	7.16	6.22	99.92	99.90	99.88	100.00	100.00	100.00
		u_t^2	5.28	5.28	5.28	99.88	99.88	99.88	100.00	100.00	100.00
		y_t^2	65.84	57.46	55.22	99.98	99.96	99.96	100.00	100.00	100.00

Since the absence of a stationary volatility part can be detected, it is intriguing to conduct a comparison of volatility forecasts for a specific case of the model (1):

$$y_t = h_t \varepsilon_t, \quad t = 1, \ldots, T, \tag{33}$$

with a persistent non-parametric volatility $\mathrm{var}\,(y_t|\mathcal{F}_{t-1}) = h_t^2$ and an i.i.d.$(0,1)$ noise $\{\varepsilon_t\}$. This model does not includes a stationary volatility component σ_t^2, see (2). Our primary interest is to verify whether the kernel forecasting method $\widehat{h}^2_{T+1|T} := \widehat{h}^2_T$ of the volatility h^2_{T+1} outperforms the 1-step ahead forecast $\widehat{h}^2_{j;T+1|T}$ formed from parametric GARCH, GJR, SV, GARCH-t and APARCH volatility models denoted by "j" and defined in the Supplement. In Monte Carlo simulations h_t^2 is known. We use two volatility proxies, $p_t = h_t^2$ and y_t^2.

For the given volatility proxy p_t, the best forecasting method j minimizes the average quadratic loss

$$\mathrm{MSFE}_j = (T - T_0)^{-1} \sum_{t=T_0+1}^{T} \left(p_t - \widehat{h}^2_{j;t|t-1}\right)^2, \tag{34}$$

over $t \in (T_0, T]$. We set $T = 1000$ and $T_0 = 200$. Forecasting of h_t^2 with the kernel predictor $\widehat{h}^2_{t|t-1}$ is performed with fixed bandwidths $H = t^{0.60}, t^{0.65}$, $t^{0.70}$, and with a cross-validated bandwidth $H_{CV,t}$ which minimises

$$\sum_{s=t-t_0}^{t} \left(p_s - \widehat{h}^2_{s|s-1}\right)^2 \tag{35}$$

over $H = t^{0.55}, t^{0.60}, ..., t^{0.75}$ where $t_0 = 50$.

To make comparisons across different forecasting methods, we use the benchmark $\mathrm{MSFE}_{\mathrm{GARCH}}$ of the parametric GARCH$(1,1)$ volatility model and calculate the relative root quadratic loss, $\mathrm{RMSFE}_j = \frac{(\mathrm{MSFE}_j)^{1/2}}{(\mathrm{MSFE}_{\mathrm{GARCH}})^{1/2}}$.

Tables 3 and 4 report the average value of the relative RMSFE_j over 1000 replications for data generating Models 2 and 3 of y_t, and two proxies, $p_t = h_t^2$ and y_t^2. The smaller is the entry (< 1), the better the forecast.

Table 3 reports RMSFE results where the forecasting performance is evaluated using the "optimal" proxy $p_t = h_t^2$. The kernel forecasting methods produce the smallest values of RMSFE and clearly outperform the parametric forecasting methods of volatility in both Model 2 and 3, and the cross-validated bandwidth $H_{\mathrm{CV},t}$ outperforms forecasts with a fixed bandwidth. For comparison, we report RMSE for the kernel forecast with a feasible cross-validated bandwidth $H^*_{\mathrm{CV},t}$ that minimises (35) with the commonly used volatility proxy $p_t^* = y_t^2$, see [26]. It works well for Model 2 but slightly less well for Model 3.

Table 4 reports results of the same experiments as Table 3, using the imperfect, but observed, proxy $p_t = y_t^2$, see [24]. It is still noticeable that the kernel forecasts outperform the parametric forecasting methods, but the difference becomes marginal. It is clear that the choice of proxy for the cross-validation in empirical analysis is crucial. We use realised variance as a feasible volatility proxy in our empirical forecasting example below.

TABLE 3
Comparison of forecasting methods. Table reports RMSFE with the (true) volatility proxy $p_t = h_t^2$.

Model 2 (with $h_t = deterministic$)						Model 3 (with $h_t = persistent$, $d = 1.4$)					
H	Kernel	GJR	SV	GARCH-t	APARCH	H	Kernel	GJR	SV	GARCH-t	APARCH
$t^{0.55}$	0.776	1.009	1.309	0.992	1.084	$t^{0.55}$	1.021	1.041	1.626	1.008	1.095
$t^{0.60}$	0.700					$t^{0.60}$	0.887				
$t^{0.65}$	0.660					$t^{0.65}$	0.790				
$t^{0.70}$	0.668					$t^{0.70}$	0.738				
$t^{0.75}$	0.738					$t^{0.75}$	0.742				
H_{CV}	0.544					H_{CV}	0.651				
H^*_{CV}	0.664					H^*_{CV}	0.894				

TABLE 4
Comparison of forecasting methods. Table reports RMSFE with the volatility proxy $p_t = y_t^2$.

Model 2 (with $h_t = deterministic$)						Model 3 (with $h_t = persistent$, $d = 1.4$)					
H	Kernel	GJR	SV	GARCH-t	APARCH	H	Kernel	GJR	SV	GARCH-t	APARCH
$t^{0.55}$	0.988	1.001	1.026	1.000	1.005	$t^{0.55}$	1.000	1.001	1.023	1.000	1.003
$t^{0.60}$	0.985					$t^{0.60}$	0.997				
$t^{0.65}$	0.983					$t^{0.65}$	0.995				
$t^{0.70}$	0.984					$t^{0.70}$	0.995				
$t^{0.75}$	0.986					$t^{0.75}$	0.995				
H_{CV}	0.984					H_{CV}	0.997				

4. An Empirical Example. In this section, we illustrate the practical applicability of our testing methodology for the detection of stationary volatility.

We use weekly stock returns for a group of 254 companies in the S&P 500 over the period Jan 1994 to Dec 2019, obtained from *Bloomberg*. After data cleaning, it contains $T = 1340$ observations. In particular, we split the historical weekly returns into three subperiods: the pre-crisis period (Jan 1994 - Dec 2007), the period covering the global financial crisis (Jan 2005 - Dec 2012), and the post-crisis period (Jan 2011 - Dec 2019).

In line with the finance literature, we assume that weekly returns of an individual company stock follow the model:

$$r_t = \mu_t + y_t, \qquad \text{where} \quad y_t = r_t - \mu_t = h_t u_t,$$

where $\mu_t = E[r_t|\mathcal{F}_{t-1}]$ is the conditional mean and y_t is a white noise process. We test for the presence or absence of stationary volatility $\sigma_t^2 = \text{var}[r_t|\mathcal{F}_{t-1}]$ in u_t. If ARCH effects in u_t are not detected, then $r_t = \mu_t + h_t\varepsilon_t$, where $\{\varepsilon_t\}$ is an i.i.d. noise.

In Table 5, we report the proportion of stock returns (in %) exhibiting no ARCH effects, according to our test, in u_t (among 254 stocks). Testing for ARCH effects is based on residuals $\widehat{|u_t|}^\gamma = (\widehat{h_t^\gamma})^{-1}|y_t|^\gamma$ where $y_t = r_t - \widehat{\mu}_t$, and conducted at the 5% significance level. To obtain an estimate for μ_t, we use the single index model $\mu_t = r_{f,t} + \beta_1(R_{m,t} - r_{f,t})$, where $R_{m,t}$ is the market factor and $r_{f,t}$ is the risk free rate. Specifically, we would like to understand the impact of the choice for the bandwidth H, lag p, subperiods, and values of γ on testing. The shaded grey range of $H = [T^{0.60}, \ldots, T^{0.70}]$ denotes theoretically permissible values of H for smoothness parameter $\nu = 1$.

The empirical testing results can be summarised as follows: Across lags $1, 5, 10$, $\gamma = 2, 1$ and three subperiods, in the recommended (shaded grey) range of $H = [T^{0.60}, \ldots, T^{0.70}]$, the vast majority ($\sim 80\%$) of the stock returns have no ARCH effects. Further, in the more volatile subperiod $2005 - 2012$, the proportion of stock returns with no ARCH effects falls slightly. However, when persistent volatility is not taken into account and testing is performed directly on the powers of y_t, the number of stock returns with no ARCH effects drops sharply.

Table 5, shows clear robustness of the testing results across values $\gamma = 1, 2$, which is a reassuring finding. Recall that estimation with $\gamma = 2$ requires at least sixth moments of u_t, whereas estimation with $\gamma = 1$ requires at least three. We include $\gamma = 1$, since by following the methodology of [27], we find that the majority of the stock returns have finite fifth moments, but not sixth. More details on this, and empirical testing results for $\gamma = 1/2$ can be found in the Supplement.

Overall, these empirical results are in line with the Monte Carlo experiments and provide clear evidence that stationary volatility might be considerably less pronounced in the data, than previously thought.

TABLE 5

Proportion of stock returns with no ARCH effects. Testing based on squares ($\gamma = 2$) and absolutes ($\gamma = 1$), for different bandwidths H, lags p, and subperiods as defined in the main text. Testing at the 5% significance level. μ_t estimated using the single index model.

Testing on squares ($\gamma = 2$)										
H	data	$p = 1$			$p = 5$			$p = 10$		
		1994-2007	2005-2012	2011-2019	1994-2007	2005-2012	2011-2019	1994-2007	2005-2012	2011-2019
$T^{0.50}$	$\widehat{u_t^2}$	93.31	97.24	97.64	97.64	96.06	96.46	94.88	91.73	92.91
$T^{0.55}$		89.76	94.88	96.85	94.49	96.85	97.24	96.85	96.06	98.03
$T^{0.60}$		86.61	90.55	95.67	90.55	92.13	96.06	93.31	93.70	98.43
$T^{0.65}$		81.10	82.28	93.31	86.61	85.43	95.28	87.80	85.43	97.24
$T^{0.70}$		76.77	74.02	92.52	77.95	72.44	92.13	77.17	71.26	93.31
$T^{0.75}$		68.50	65.75	92.52	64.17	57.09	90.16	62.21	54.72	91.73
	y_t^2	39.37	42.91	78.35	22.05	33.47	80.32	20.87	29.53	77.17

Testing on absolutes ($\gamma = 1$)										
H	data	$p = 1$			$p = 5$			$p = 10$		
		1994-2007	2005-2012	2011-2019	1994-2007	2005-2012	2011-2019	1994-2007	2005-2012	2011-2019
$T^{0.50}$	$\widehat{\lvert u_t \rvert}$	89.76	94.49	94.49	88.58	84.25	86.22	80.32	72.44	69.69
$T^{0.55}$		86.61	93.70	93.31	90.95	91.34	90.16	89.76	87.80	83.07
$T^{0.60}$		81.50	87.80	94.09	87.40	89.37	93.70	86.61	87.80	89.37
$T^{0.65}$		73.62	80.71	91.73	80.71	83.47	94.09	84.25	82.68	92.91
$T^{0.70}$		62.60	70.87	90.16	68.11	70.87	89.76	70.47	70.87	92.52
$T^{0.75}$		53.15	59.06	89.37	51.18	56.30	88.58	52.76	53.54	90.55
	$\lvert y_t \rvert$	12.60	30.71	72.05	2.36	24.41	71.65	1.97	18.11	72.05

Next, we consider the problem of forecasting of persistent volatility of the weekly log returns of some major stock indices and exchange rate series using data from the database 'Oxford-Man Institute's realised library' version 0.1, produced by [19], see also [26], over the period from January 3, 1999 to December 23, 2007 (469 observations).

We consider the same forecasting methods of volatility as in the simulation study section. In order for these forecasting methods to be applied on returns with persistent volatility as in the model (33), we employ our test for ARCH effects from Section 2 on demeaned returns r_t of assets and select four stock indices and three exchange rates which returns do not exhibit ARCH effects in squares. Instead of using a noisy proxy of squared returns for the volatility, we use the proxy p_t of "realised variance", see [24].

Table 6 presents the values of relative RMSFE$_j$ introduced in Section 3. Kernel prediction uses cross-validated bandwidth $H_{CV,t}$. The quadratic loss and the cross-validated bandwidth $H_{CV,t}$ are derived using the proxy of realised variance in (34) and (35). We set $T_0 = 200$, $t_0 = 50$. The results suggest that for the majority of asset returns, the kernel forecasting method with cross-validated bandwidth $H_{CV,t}$, significantly outperforms stationary alternatives.

TABLE 6
Comparison of the forecasting methods. Table reports RMSFE with the realised variance proxy p_t.

Asset	Kernel	GJR	SV	GARCH-t	APARCH
Dow Jones	0.654	0.904	0.665	1.001	0.966
Nasdaq 100	0.696	1.238	1.086	1.137	1.311
Nikkei 250	0.831	1.008	0.971	0.996	1.117
S&P 500	0.705	0.901	0.934	1.085	0.851
USD British pound	0.989	1.038	1.116	1.002	1.160
USD Euro	0.915	1.013	1.043	1.000	1.042
USD Swiss franc	0.925	0.986	0.886	1.006	0.998

5. Discussion. This paper contributes to the literature in three ways. First, we introduce a setup for persistent processes, that can provide a general approximation to the volatility process of a time series. Second, we develop a consistent uniform estimation theory for the unobserved volatility processes, without strong parametric assumptions, and third, we suggest a testing strategy that enables the separation of stationary volatility from its persistent counterpart. To prove our main results, the uniform bounds for kernel type estimates obtained in [17] and based on Bernstein inequalities for dependent random variables, were used.

Testing results on U.S. stock returns provides extensive support for the persistent volatility paradigm, suggesting that the role of stationary conditional heteroskedasticity is not as outstanding in the data, as was previously thought. In addition, forecasting results on persistent volatility of log returns of stock indices and exchange rates provide evidence in favour of using kernel forecasting methods.

There are a number of interesting avenues for future work, in particular, the extension of our testing procedure to multivariate time series.

This is a distinct problem rather than a simple generalization of the univariate case. In general, a multivariate model for a $p \times 1$ process

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{u}_t, \quad t = 1, ..., T$$

can include a $p \times p$ matrix, $\boldsymbol{H}_t$, of persistent volatility and a stationary $p \times 1$ white noise process $\boldsymbol{u}_t$. The objective would be to test whether components of this white noise exhibit stationary conditional heteroskedasticity (ARCH effects). The estimation of $\boldsymbol{H}_t$ that would enable such testing could be undertaken using the work by [11].

Acknowledgements. The first author was supported by ESRC grant ES/P000703/1.

SUPPLEMENTARY MATERIAL

Supplement to: "Choosing between persistent and Stationary volatility"
Provides proofs of all results given in the main paper and supplemental material for simulation study and empirical exercise.

REFERENCES

[1] AMADO, C. and TERÄSVIRTA, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics* **175** 142–153.

[2] AMADO, C. and TERÄSVIRTA, T. (2017). Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Reviews* **36** 421–446.

[3] BERA, A. K. and HIGGINS, M. L. (1993). ARCH models: properties, estimation and testing. *Journal of Economic Surveys* **7** 305–366.

[4] BROWNLEES, C. T. and GALLO, G. M. (2009). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* **8** 29–56.

[5] CHRONOPOULOS, I., KAPETANIOS, G. and PETROVA, K. (2021). Kernel-based volatility generalised least squares. *Econometrics and Statistics* **20** 2–11.

[6] DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics* **28** 1762-1794.

[7] DAHLHAUS, R. and POLONIK, W. (2006). Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics* **34** 2790-2824.

[8] DAHLHAUS, R., RAO, S. S. et al. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics* **34** 1075–1114.

[9] DALLA, V., GIRAITIS, L. and PHILLIPS, P. C. B. (2015). Testing mean stability of heteroskedastic time series. *Cowles Foundation discussion paper no 2006*.

[10] DALLA, V., GIRAITIS, L. and PHILLIPS, P. C. B. (2020). Robust tests for white noise and cross-correlation. *Econometric Theory, published online*.

[11] DENDRAMIS, Y., GIRAITIS, L. and KAPETANIOS, G. (2021). Estimation of Time-Varying Covariance Matrices for Large Datasets. *Econometric Theory, published online*.

[12] ENGLE, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50** 987–1008.

[13] ENGLE, R. F. and RANGEL, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies* **21** 1187–1222.

[14] GALLANT, A. R. (1984). The Fourier flexible form. *American Journal of Agricultural Economics* **66** 204–208.

[15] GIRAITIS, L., KAPETANIOS, G., WETHERILT, A. and ŽIKEŠ, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. *Journal of Applied Econometrics* **31** 58–84.

[16] GIRAITIS, L., KAPETANIOS, G. and YATES, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics* **179** 46–65.

[17] GIRAITIS, L., KAPETANIOS, G. and YATES, T. (2018). Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis* **39** 129–149.

[18] GIRAITIS, L., KOUL, H. and SURGAILIS, D. (2012). *Large Sample Inference for Long memory Processes*. Imperial College Press, London.

[19] HEBER, G., LUNDE, A., SHEPHARD, N. and SHEPPARD, K. (2009). Oxford-Man Institute's realized library, version 0.1, https://realized.oxford-man.ox.ac.uk/data/download.

[20] JIANG, F., LI, D. and ZHU, K. (2021). Adaptive inference for a semiparametric generalized autoregressive conditional heteroskedasticity model. *Journal of Econometrics* **224** 306–329.

[21] KAPETANIOS, G. and YATES, T. (2008). An analysis of inflation persistence and the Great Moderation using a model of deterministic structural change. *Mimeo, Bank of England*.

[22] MAZUR, B. and PIPIEŃ, M. (2012). On the empirical importance of periodicity in the volatility of financial returns-time varying GARCH as a second order APC (2) process. *Central European Journal of Economic Modelling and Econometrics* **4** 95–116.

[23] MIKOSCH, T. and STĂRICĂ, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics* **86** 378–390.

[24] PATTON, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160** 246–256.

[25] PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Series B* **27** 204–237.

[26] SHEPHARD, N. and SHEPPARD, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **25** 197–231.

[27] TRAPANI, L. (2016). Testing for (in) finite moments. *Journal of Econometrics* **191** 57–67.

[28] VAN BELLEGEM, S. and VON SACHS, R. (2004). Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting* **20** 611–627.

SUPPLEMENT TO "CHOOSING BETWEEN PERSISTENT AND STATIONARY VOLATILITY"

BY ILIAS CHRONOPOULOS[1], LIUDAS GIRAITIS[2] AND GEORGE KAPETANIOS[3]

[1]*Essex Business School, University of Essex,*

[2]*School of Economics and Finance, Queen Mary University of London,*

[3]*King's Business School, King's College London,*

This Supplement provides proofs of the results given in the text of the main paper. It is organised as follows: Section 6 provides proofs of the main theorems. Section 7 contains auxiliary technical lemmas. Section 8 contains supplementary simulation material and Section 9 supplementary empirical material.

Formula numbering in this supplement includes the section number, e.g. (6.1), and references to lemmas are signified as "Lemma 6.#", "Lemma 7.#", e.g. Lemma 6.1. Theorem references to the main paper include section number and are signified, e.g. as Theorem 2.1, while equation references do not include section number, e.g. (1), (2).

In the proofs, C stands for a generic positive constant which may assume different values in different contexts.

6. Proof of Theorems 2.1, 2.2 and 2.3. This section contains the proofs of the results of Section 2 of the main paper on the asymptotic properties of the test statistics $TS(\boldsymbol{u}^\gamma)$ and $TS(\widehat{\boldsymbol{u}^\gamma})$.

In the proof of Theorem 2.1 without loss of generality we assume that $E|u_t|^\gamma = 1$. In addition, we use the claim that $\{h_t^\gamma, t = 1, ..., T\}$ satisfies Assumption H which is verified the next proposition.

PROPOSITION 6.1. Let $\{h_t, t = 1, ..., T\}$ satisfy Assumption H with parameters ν and α. Then $\{h_t^\gamma, t = 1, ..., T\}$ satisfies Assumption H with parameters ν and $\alpha^* = \alpha / \max(\gamma, 1)$:

$$|h_t^\gamma - h_j^\gamma| \le (|t-j|/T)^\nu \xi_{tj}^*, \ \ t, j = 1, ..., T \tag{6.1}$$

and there exist $c > 0$ such that

$$\max_{t,j=1,...,T} E[\exp(c|\xi_{tj}^*|^{\alpha^*})] \le C, \quad \max_{t=1,...,T} E[\exp(c|h_t^\gamma|^{\alpha^*})] \le C, \tag{6.2}$$

where $C < \infty$ does not depend on T.

Proof of Proposition 6.1. By assumption (7) and (8), h_t has the following properties:

$$|h_t^\gamma - h_j^\gamma| \le (|t-j|/T)^\nu \xi_{tj}, \ \ t, j = 1, ..., T \tag{6.3}$$

$$\max_{t,j=1,...,T} E[\exp(c|\xi_{tj}|^\alpha)] \le C, \quad \max_{t=1,...,T} E[\exp(c|h_t|^\alpha)] \le C, \tag{6.4}$$

where $C < \infty$ does not depend on T. Therefore, h_j^γ satisfies the second claim in (6.2).

Keywords and phrases: ARCH effects, persistence, volatility, time-varying coefficient models, non-parametric estimation.

Let $\gamma \geq 1$ and $0 < a \leq x < y$. Then, by the mean value theorem,

$$\begin{aligned}(y^\gamma - x^\gamma) &\leq \gamma y^{\gamma-1}(y-x), \quad \text{if } \gamma \geq 1,\\ &\leq \gamma a^{1-\gamma}(y-x), \quad \text{if } 0<\gamma<1.\end{aligned}$$

This together with (6.3) implies

$$|h_t^\gamma - h_j^\gamma| \leq \gamma(h_t^{\gamma-1} + h_j^{\gamma-1})|h_t - h_j| \leq (|t-j|/T)^\nu \xi^*_{tj}, \tag{6.5}$$

where $\xi^*_{tj} = \gamma(h_t^{\gamma-1} + h_j^{\gamma-1})\xi_{tj} \leq \gamma(h_t^\gamma + h_j^\gamma + 2\xi_{tj}^\gamma)$. Then (6.4) implies (6.2) for ξ^*_{tj}. □

Proof of Theorem 2.1. In Theorem 2.1 we analyse the Wald version of the test for the null hypothesis of absence of ARCH effects in u_t. First recall the definitions of $S(\boldsymbol{u}^\gamma)$ and $S(\widehat{\boldsymbol{u}^\gamma})$. Denote

$$z_t = |u_t|^\gamma, \qquad \overline{z} = T^{-1}\sum_{t=1}^T z_t.$$

Given data $\boldsymbol{u}^\gamma = [z_1, z_2, \ldots, z_T]$, we define the test statistic for testing H_0 as follows:

$$S(\boldsymbol{u}^\gamma) = \widetilde{\sigma}_p^{-2}\widetilde{\boldsymbol{\beta}}_p'(X'X)\widetilde{\boldsymbol{\beta}}_p, \quad \widetilde{\boldsymbol{\beta}}_p = (X'X)^{-1}X'Y, \quad \widetilde{\sigma}_p^2 = (Y - X\widetilde{\boldsymbol{\beta}}_p)'(Y - X\widetilde{\boldsymbol{\beta}_p}), \tag{6.6}$$

where Y is a $(T-p)\times 1$ vector and X is a $(T-p)\times p$ design matrix:

$$Y = (z_{p+1} - \overline{z}, ..., z_T - \overline{z})',$$

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,p}\\ x_{2,1} & x_{2,2} & \ldots & x_{2,p}\\ \ldots & \ldots & \ldots & \ldots\\ x_{T-p,1} & x_{T-p,2} & \ldots & x_{T-p,p}\end{bmatrix} = \begin{bmatrix} z_p - \overline{z} & z_{p-1} - \overline{z} & \ldots & z_1 - \overline{z}\\ z_{p+1} - \overline{z} & z_p - \overline{z} & \ldots & z_2 - \overline{z}\\ \ldots & \ldots & \ldots & \ldots\\ z_{T-1} - \overline{z} & z_{T-2} - \overline{z} & \ldots & z_{T-p} - \overline{z}\end{bmatrix}.$$

Here $\widetilde{\boldsymbol{\beta}}_p = (\widetilde{\beta}_1, ..., \widetilde{\beta}_p)'$ denotes the OLS estimate of regression coefficients in regression

$$z_t = \beta_0 + \beta_1 z_{t-1} + ... + \beta_p z_{t-p} + \eta_t.$$

Similarly, let

$$\widehat{z}_t = \widehat{|u_t|}^\gamma, \qquad \overline{\widehat{z}} = T^{-1}\sum_{t=1}^T \widehat{z}_t, \qquad \text{where} \quad \widehat{|u_t|}^\gamma = (\widehat{h_t^\gamma})^{-1}|y_t|^\gamma.$$

Denote by $\widehat{\boldsymbol{\beta}}_p = (\widehat{\beta}_1, ..., \widehat{\beta}_p)'$ the OLS estimate of regression coefficients in

$$\widehat{z}_t = \beta_0 + \beta_1 \widehat{z}_{t-1} + ... + \beta_p \widehat{z}_{t-p} + \eta_t.$$

Then

$$S(\widehat{\boldsymbol{u}^\gamma}) = \widehat{\sigma}_p^{-2}\widehat{\boldsymbol{\beta}}_p'(\widehat{X}'\widehat{X})\widehat{\boldsymbol{\beta}}_p, \quad \widehat{\boldsymbol{\beta}}_p = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'\widehat{Y}, \quad \widehat{\sigma}_p^2 = (\widehat{Y} - \widehat{X}\widehat{\boldsymbol{\beta}}_p)'(\widehat{Y} - \widehat{X}\widehat{\boldsymbol{\beta}}_p) \tag{6.7}$$

where

$$\widehat{Y} = (\widehat{z}_{p+1} - \overline{\widehat{z}}, ..., \widehat{z}_T - \overline{\widehat{z}})',$$

$$\widehat{X} = \begin{bmatrix} \widehat{x}_{1,1} & \widehat{x}_{1,2} & \ldots & \widehat{x}_{1,p}\\ \widehat{x}_{2,1} & \widehat{x}_{2,2} & \ldots & \widehat{x}_{2,p}\\ \ldots & \ldots & \ldots & \ldots\\ \widehat{x}_{T-p,1} & \widehat{x}_{T-p,2} & \ldots & \widehat{x}_{T-p,p}\end{bmatrix} = \begin{bmatrix} \widehat{z}_p - \overline{\widehat{z}} & \widehat{z}_{p-1} - \overline{\widehat{z}} & \ldots & \widehat{z}_1 - \overline{\widehat{z}}\\ \widehat{z}_{p+1} - \overline{\widehat{z}} & \widehat{z}_p - \overline{\widehat{z}} & \ldots & \widehat{z}_2 - \overline{\widehat{z}}\\ \ldots & \ldots & \ldots & \ldots\\ \widehat{z}_{T-1} - \overline{\widehat{z}} & \widehat{z}_{T-2} - \overline{\widehat{z}} & \ldots & z_{T-p} - \overline{\widehat{z}}\end{bmatrix}.$$

Observe that we can write

$$T^{-1}(X'X)=(g_{ij})_{i,j=1,\dots,p}, \quad T^{-1}(X'Y)=(g_{0j})_{j=1,\dots,p}, \quad \text{where} \tag{6.8}$$
$$g_{ij}=T^{-1}\textstyle\sum_{t=p+1}^{T}(z_{t-i}-\overline{z})(z_{t-j}-\overline{z}).$$

Similarly,

$$T^{-1}(\widehat{X}'\widehat{X})=(\widehat{g}_{ij})_{i,j=1,\dots,p}, \quad T^{-1}(\widehat{X}'\widehat{Y})=(\widehat{g}_{0j})_{j=1,\dots,p}, \quad \text{where} \tag{6.9}$$
$$\widehat{g}_{ij}=T^{-1}\textstyle\sum_{t=p+1}^{T}(\widehat{z}_{t-i}-\overline{\widehat{z}})(\widehat{z}_{t-j}-\overline{\widehat{z}}).$$

The proof of Theorem 2.1 is based on Lemmas 6.2 and 6.3 below. Auxiliary results used to prove these lemmas are placed in Section 7. Denote by

$$\gamma_k=\mathrm{cov}(z_k,z_0), \quad k\ge 0, \tag{6.10}$$
$$\widetilde{\gamma}_k=T^{-1}\sum_{t=k+1}^{T}\big(z_t-Ez_t\big)\big(z_{t-k}-Ez_{t-k}\big),$$

the autocovariance and sample autocovariance functions of $\{z_t\}$. Define

$$\mathbf{\Gamma}_p=(\gamma_{|i-j|})_{i,j=1,\dots,p}, \quad \boldsymbol{\gamma}_p=(\gamma_1,\dots,\gamma_p)', \quad \boldsymbol{\beta}_p=\mathbf{\Gamma}_p^{-1}\boldsymbol{\gamma}_p;$$
$$\widetilde{\mathbf{\Gamma}}_p=(\widetilde{\gamma}_{|i-j|})_{i,j=1,\dots,p}, \quad \widetilde{\boldsymbol{\gamma}}_p=(\widetilde{\gamma}_1,\dots,\widetilde{\gamma}_p)', \quad \widetilde{\boldsymbol{\beta}}_p=\widetilde{\mathbf{\Gamma}}_p^{-1}\widetilde{\boldsymbol{\gamma}}_p$$

and set $\sigma_p^2=\mathrm{var}(|u_{p+1}|^\gamma-\beta_1|u_p|^\gamma-\dots-\beta_p|u_1|^\gamma)$.

To prove the theorem we will derive the following results: As $T\to\infty$,

$$T^{-1}(\widehat{X}'\widehat{X})=\widetilde{\mathbf{\Gamma}}_p+o_P(1), \quad T^{-1}(\widehat{X}'\widehat{Y})=\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \tag{6.11}$$

$$T^{-1}(X'X)=\widetilde{\mathbf{\Gamma}}_p+o_P(1), \quad T^{-1}(X'Y)=\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \tag{6.12}$$

$$\widetilde{\mathbf{\Gamma}}_p\to_p\mathbf{\Gamma}_p, \quad \widetilde{\boldsymbol{\gamma}}_p\to_P\boldsymbol{\gamma}_p, \quad \widetilde{\boldsymbol{\beta}}_p\to\boldsymbol{\beta}_p, \tag{6.13}$$

$$T^{-1}\widehat{\sigma}_p^2\to_P\sigma_p^2, \quad T^{-1}\widetilde{\sigma}_p^2\to_P\sigma_p^2. \tag{6.14}$$

If, in addition, $\{u_t\}$ is an i.i.d. sequence, then it holds

$$T^{-1/2}(\widehat{X}'\widehat{Y})=T^{-1/2}\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \tag{6.15}$$

$$T^{-1/2}(X'Y)=T^{-1/2}\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \tag{6.16}$$

$$T^{-1/2}\widetilde{\boldsymbol{\gamma}}_p\to_D\mathcal{N}(0,\boldsymbol{I}_p\gamma_0), \quad \gamma_0=\mathrm{var}(z_1). \tag{6.17}$$

In turn, properties (6.11), (6.12) and (6.13) follow by applying in (6.8) and (6.9) the asymptotic relations (6.18), (6.19) and (6.20), shown in Lemma 6.2.

The convergence (6.14) follows from the definitions of $\widehat{\sigma}_p^2$ and $\widetilde{\sigma}_p^2$, using (6.11)-(6.14), noting that $T^{-1}\widehat{Y}'\widehat{Y}=\widehat{g}_{00}\to_P\gamma_0$, $T^{-1}Y'Y=g_{00}\to_P\gamma_0$, and using the equality $\sigma_p^2=\gamma_0-2\boldsymbol{\gamma}_p'\boldsymbol{\beta}_p+\boldsymbol{\beta}_p'\mathbf{\Gamma}_p\boldsymbol{\beta}_p$.

Subsequently, (6.15) and (6.16) follow from (6.21) and (6.22) of Lemma 6.2. Finally, observe that the i.i.d. property of $\{u_t\}$ implies that $\{z_t\}$ is also an i.i.d. sequence. It remains to note, that the validity of convergence (6.17) for an i.i.d. r.v. z_t-Ez_t is a well known fact.

From the definition of $S(\widehat{\boldsymbol{u}^\gamma})$ and $S(\boldsymbol{u}^\gamma)$, given in (6.7) and (6.6), using relations (6.11)-(6.14), we obtain

$$\widehat{\boldsymbol{\beta}}_p\to_P\boldsymbol{\beta}_p=\mathbf{\Gamma}_p^{-1}\boldsymbol{\gamma}_p, \quad \widetilde{\boldsymbol{\beta}}_p\to_P\boldsymbol{\beta}_p,$$
$$S(\widehat{\boldsymbol{u}^\gamma})=S(\boldsymbol{u}^\gamma)+o_P(1)=\sigma_p^{-2}\boldsymbol{\beta}_p'\mathbf{\Gamma}_p\boldsymbol{\beta}_p+o_P(1),$$

which proves the claim (19) of the theorem.

In addition, if $\{u_t\}$ is an i.i.d. sequence, then it holds that $\boldsymbol{\Gamma}_p = \gamma_0 \boldsymbol{I}_p$ and $\sigma_p^2 = \gamma_0$. Then, using (6.15)–(6.17), we obtain

$$\widehat{\sigma}_p^{-2}(\widehat{X}'\widehat{X}) = \gamma_0^{-2}\boldsymbol{I}_p(1+o_P(1)), \quad \widetilde{\sigma}_p^{-2}(\widetilde{X}'\widetilde{X}) = \gamma_0^{-2}\boldsymbol{I}_p(1+o_P(1)),$$
$$T^{-1/2}\widehat{\boldsymbol{\beta}}_p \to_D \gamma_0^{1/2}\mathcal{N}(0,\boldsymbol{I}_p), \quad T^{-1/2}\widetilde{\boldsymbol{\beta}}_p \to_D \gamma_0^{1/2}\mathcal{N}(0,\boldsymbol{I}_p).$$

This together with the definitions of $S(\widehat{\boldsymbol{u}^\gamma})$ and $S(\boldsymbol{u}^\gamma)$ implies

$$TS(\widehat{\boldsymbol{u}^\gamma}) = TS(\boldsymbol{u}^\gamma) + o_P(1) \to_D \chi_p^2,$$

which proves (20). This completes the proof of the theorem. □

Proof of Theorems 2.2. The proof follows the same line of arguments as in Theorem 2.1. □

Proof of Theorem 2.3. The claims (27) and (28) follow using a similar reasoning as in the proof of (6.26) and (6.27) of Lemma 6.3, noting that for i.i.d. r.v.'s the convergence $T^{1/2}\widetilde{r}_k \to \mathcal{N}(0,(Eu_1^2)^2)$ is well-known. □

LEMMA 6.2. *(a) Suppose that $\{h_t, u_t\}$ satisfy Assumptions M and H, and the bandwidth H satisfies (18). Then for $i,j = 0,1,...,p$, as $T \to \infty$,*

$$\widehat{g}_{ij} = \widetilde{\gamma}_{|i-j|} + o_P(1), \tag{6.18}$$

$$g_{ij} = \widetilde{\gamma}_{|i-j|} + o_P(1), \tag{6.19}$$

$$\widetilde{\gamma}_k \to_P \gamma_k, \quad k \geq 0. \tag{6.20}$$

(b) In addition, if $\{u_t\}$ is an i.i.d. sequence, then for $j = 1,...,p$,

$$T^{1/2}\widehat{g}_{0j} = T^{1/2}\widetilde{\gamma}_j + o_P(1), \tag{6.21}$$

$$T^{1/2}g_{0j} = T^{1/2}\widetilde{\gamma}_j + o_P(1). \tag{6.22}$$

Proof of Lemma 6.2.
Proof of (6.18). It suffices to verify (6.18) for $i \leq j$. Denote

$$\widehat{\gamma}_k = T^{-1}\sum_{t=k+1}^{T}\big(\widehat{z}_t - \overline{\widehat{z}}\big)\big(\widehat{z}_{t-k} - \overline{\widehat{z}}\big), \quad k = 0,1,2,.... \tag{6.23}$$

Then setting $k = j - i$, we can write

$$\widehat{g}_{ij} = T^{-1}\sum_{t=p+1-i}^{T-i}(\widehat{z}_t - \overline{z})(\widehat{z}_{t-k} - \overline{\widehat{z}}) = T^{-1}\sum_{t=k+1+(p-j)}^{T-i}(\widehat{z}_t - \overline{\widehat{z}})(\widehat{z}_{t-k} - \overline{\widehat{z}})$$
$$= \widehat{\gamma}_{|j-i|} - \delta_{ij}, \quad \delta_{ij} = T^{-1}[\sum_{t=k+1}^{k+p-j} + \sum_{t=T-i+1}^{T}](\widehat{z}_t - \overline{\widehat{z}})(\widehat{z}_{t-k} - \overline{\widehat{z}}).$$

So,

$$\widehat{g}_{ij} = \widetilde{\gamma}_{|i-j|} + (\widehat{\gamma}_{|i-j|} - \widetilde{\gamma}_{|i-j|}) - \delta_{ij}. \tag{6.24}$$

By (6.26) of Lemma 6.3, $\widehat{\gamma}_{|i-j|} - \widetilde{\gamma}_{|i-j|} = o_P(1)$. On the other hand, straightforward use of (7.1) and (7.3) of Lemma 7.1 implies that

$$\delta_{ij} = O_P(T^{-1}), \tag{6.25}$$

which together with (6.24) proves (6.18): $\widehat{g}_{ij} = \widetilde{\gamma}_{|i-j|} + o_P(1)$.
Proof of (6.19). Property (6.19) follows from the proof of (6.18) as a special case correponding to $\widehat{h_t^\gamma} = h_t^\gamma$, which implies $\widehat{z}_t = z_t$.
Proof of (6.20). By Assumption M, $\{u_t\}$ is a stationary ergodic sequence, and we assume that $Ez_t^3 = E|u_t|^{3\gamma} < \infty$. Then the sequence $\omega_t = \big(z_t - Ez_t\big)\big(z_{t-k} - Ez_{t-k}\big)$ is stationary and ergodic with $E|\omega_t| < \infty$, which implies that $\widetilde{\gamma}_k \to_P E\omega_k = \mathrm{cov}(z_k, z_0) = \gamma_k$.
Proof of (6.21). By (6.24), (6.25) and (6.27), for $j = 1, ..., p$,

$$T^{1/2}\widehat{g}_{0j} = T^{1/2}\widetilde{\gamma}_j + o_P(1).$$

Proof of (6.22). This claim follows from (6.21) by setting $\widehat{h_t^\gamma} = h_t^\gamma$. □

LEMMA 6.3. *(a) Suppose that $\{h_t, u_t\}$ satisfy Assumptions M and H, and the bandwidth H satisfies (18). Then, as $T \to \infty$,*

$$\widehat{\gamma}_k - \widetilde{\gamma}_k = o_P(1), \quad k \geq 0. \tag{6.26}$$

(b) In addition, if $\{u_t\}$ is an i.i.d. sequence, then

$$\widehat{\gamma}_k - \widetilde{\gamma}_k = o_P(T^{-1/2}), \quad k \geq 1. \tag{6.27}$$

Proof of Lemma 6.3. Denote

$$\widehat{\gamma}_k^* = T^{-1} \sum_{t=k+1}^{T} \big(\widehat{z}_t - Ez_t\big)\big(\widehat{z}_{t-k} - Ez_{t-k}\big), \quad k = 0, 1, 2, ... \tag{6.28}$$

Then

$$\widehat{\gamma}_k - \widetilde{\gamma}_k = (\widehat{\gamma}_k^* - \widetilde{\gamma}_k) + (\widehat{\gamma}_k - \widehat{\gamma}_k^*).$$

Thus, to prove (6.26), it suffices to show

$$\widehat{\gamma}_k^* - \widetilde{\gamma}_k = o_P(1), \quad k \geq 1, \tag{6.29}$$

$$\widehat{\gamma}_k - \widehat{\gamma}_k^* = o_P(1), \quad k \geq 0, \tag{6.30}$$

$$\widehat{\gamma}_0^* - \widetilde{\gamma}_0 = o_P(1). \tag{6.31}$$

In turn, to prove (6.27), we show in addition that for an i.i.d. sequence $\{u_t\}$ it holds

$$\widehat{\gamma}_k^* - \widetilde{\gamma}_k = o_P(T^{-1/2}), \quad k \geq 1, \tag{6.32}$$

$$\widehat{\gamma}_k - \widehat{\gamma}_k^* = o_P(T^{-1/2}), \quad k \geq 1. \tag{6.33}$$

Proof of (6.29). Recall that $Ez_t = 1$. We have,

$$\begin{aligned}
&(\widehat{z}_t - Ez_t)(\widehat{z}_{t-k} - Ez_{t-k}) - (z_t - Ez_t)(z_{t-k} - Ez_{t-k}) \\
&\quad = \{(\widehat{z}_t - z_t) + (z_t - 1)\}\{(\widehat{z}_{t-k} - z_{t-k}) + (z_{t-k} - 1)\} - (z_t - 1)(z_{t-k} - 1) \\
&\quad = (\widehat{z}_t - z_t)(\widehat{z}_{t-k} - z_{t-k}) + (\widehat{z}_t - z_t)(z_{t-k} - 1) + (z_t - 1)(\widehat{z}_{t-k} - z_{t-k}).
\end{aligned}$$

Hence,

$$\widehat{\gamma}_k^* - \widetilde{\gamma}_k = S_{T,1} + S_{T,2} + S_{T,3}, \tag{6.34}$$

where

$$\begin{aligned}
S_{T,1} &= T^{-1} \textstyle\sum_{t=k+1}^{T} (\widehat{z}_t - z_t)(\widehat{z}_{t-k} - z_{t-k}), \\
S_{T,2} &= T^{-1} \textstyle\sum_{t=k+1}^{T} (\widehat{z}_t - z_t)(z_{t-k} - 1), \\
S_{T,3} &= T^{-1} \textstyle\sum_{t=k+1}^{T} (z_t - 1)(\widehat{z}_{t-k} - z_{t-k}).
\end{aligned} \tag{6.35}$$

To prove (6.29), it remains to show that $S_{T,\ell}$ in (6.35) satisfy

$$S_{T,\ell} = o_P(1), \quad \ell = 1,2,3. \tag{6.36}$$

Notice that

$$\widehat{z}_t - z_t = (h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1)z_t. \tag{6.37}$$

Therefore,

$$|S_{T,1}| \le T^{-1}\sum_{t=k+1}^{T} \big|(h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1)(h_{t-k}^{\gamma}/\widehat{h_{t-k}^{\gamma}} - 1)\big| z_t z_{t-k}$$

$$\le m_T^2 T^{-1}\sum_{t=1}^{T} z_t z_{t-k}, \qquad m_T := \max_{t=1,\dots,T} |h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1|. \tag{6.38}$$

By (7.11) and (7.10) of Lemma 7.2, for some $\delta > 0$,

$$m_T \le \{\max_{t=1,\dots,T} (\widehat{h_t^{\gamma}})^{-1}\}\{\max_{t=1,\dots,T} |h_t^{\gamma} - \widehat{h_t^{\gamma}}|\} = O_P(1)O_P(T^{-\delta}) = o_P(1). \tag{6.39}$$

By Assumption M, $\{u_t\}$ is a stationary sequence, and $Ez_t z_{t-k} \le Ez_t^2 = E|u_t|^{2\gamma} = E|u_1|^{2\gamma} < \infty$. Therefore, $E[T^{-1}\sum_{t=k+1}^{T} z_t z_{t-k}] = Ez_1^2 < \infty$. Thus, by (6.38),

$$|S_{T,1}| \le o_p(1)T^{-1}\sum_{t=k+1}^{T} z_t z_{t-k} = o_P(1).$$

The proof of (6.36) for $S_{T,2},\ S_{T,3}$ is similar to that for $S_{T,1}$. This completes the proof of (6.29).

Proof of (6.30). We have,

$$\begin{aligned}
&(\widehat{z}_t - \overline{\widehat{z}})(\widehat{z}_{t-k} - \overline{\widehat{z}}) - (\widehat{z}_t - 1)(\widehat{z}_{t-k} - 1)\\
&\quad = \{(\widehat{z}_t - 1) + (1 - \overline{\widehat{z}})\}\{(\widehat{z}_{t-k} - 1) + (1 - \overline{\widehat{z}})\} - (\widehat{z}_t - 1)(\widehat{z}_{t-k} - 1)\\
&\quad = (\overline{\widehat{z}} - 1)^2 + (1 - \overline{\widehat{z}})(\widehat{z}_{t-k} - 1) + (\widehat{z}_t - 1)(1 - \overline{\widehat{z}}).
\end{aligned}$$

Notice that

$$\begin{aligned}
&T^{-1}\textstyle\sum_{t=k+1}^{T}(\widehat{z}_t - 1) = (\overline{\widehat{z}} - 1) - T^{-1}\sum_{t=1}^{k}(\widehat{z}_t - 1),\\
&T^{-1}\textstyle\sum_{t=k+1}^{T}(\widehat{z}_{t-k} - 1) = (\overline{\widehat{z}} - 1) - T^{-1}\sum_{t=T-k+1}^{T}(\widehat{z}_t - 1).
\end{aligned}$$

Hence,

$$\begin{aligned}
\widehat{\gamma}_k - \widehat{\gamma}_k^* &= T^{-1}\textstyle\sum_{t=k+1}^{T}\{(\overline{\widehat{z}} - 1)^2 - (\overline{\widehat{z}} - 1)(\widehat{z}_{t-k} - 1) - (\widehat{z}_t - 1)(\overline{\widehat{z}} - 1)\}\\
&= (\overline{\widehat{z}} - 1)^2\{T^{-1}(T-k) - 2)\\
&\quad + (\overline{\widehat{z}} - 1)\{T^{-1}\textstyle\sum_{t=1}^{k}(\widehat{z}_t - 1) + T^{-1}\sum_{t=T-k+1}^{T}(\widehat{z}_t - 1)\}.
\end{aligned} \tag{6.40}$$

By Assumption M, $\{u_t\}$ is a stationary α-mixing sequence. Applying in (6.40) the bounds (7.1) and (7.2) of Lemma 7.1, we obtain (6.30):

$$\widehat{\gamma}_k - \widehat{\gamma}_k^* = o_P(1), \quad k \ge 0.$$

Proof of (6.31). By Assumption M, $\{u_t\}$ is a stationary α-mixing sequence. Using the equality $a^2 - b^2 = (a-b)^2 + (a-b)2b$ with $a = \widehat{z}_t - 1$, $b = z_t - 1$ we obtain

$$\begin{aligned}
\widehat{\gamma}_0^* - \widetilde{\gamma}_0 &= T^{-1}\textstyle\sum_{t=1}^{T}\{(\widehat{z}_t - 1)^2 - (z_t - 1)^2\}\\
&= T^{-1}\textstyle\sum_{t=1}^{T}\{(\widehat{z}_t - z_t)^2 + (\widehat{z}_t - z_t)2(z_t - 1)\}.
\end{aligned}$$

Using (6.37) and (6.39), we obtain

$$|\widehat{\gamma}_0^* - \widetilde{\gamma}_0| \le m_T^2\big(T^{-1}\sum_{t=1}^{T} z_t^2\big) + 2m_T\big(T^{-1}\sum_{t=1}^{T} z_t|z_t-1|\big)$$
$$= o_P(1)\,T^{-1}\sum_{t=1}^{T}(z_t^2 + z_t|z_t-1|) = o_P(1),$$

since $E[z_t^2 + z_t|z_t-1|] = E[z_1^2 + z_1|z_1-1|] < \infty$ implies that $E[T^{-1}\sum_{t=1}^{T}(z_t^2 + z_t|z_t - 1|)] = E[z_1^2 + z_1|z_1-1|] = O(1)$. This proves (6.31).

Proof of (6.32). By (6.34), to prove (6.32), it remains to show that

$$S_{T,\ell} = o_P(T^{-1/2}), \quad \ell = 1, 2, 3. \tag{6.41}$$

First we evaluate $S_{T,1}$. Using in the definition (6.35) of $S_{T,1}$ the equality (6.37), we can bound

$$|S_{T,1}| \le T^{-1}\sum_{t=k+1}^{T}\big|(h_t^{\gamma} - \widehat{h_t^{\gamma}})(h_{t-k}^{\gamma} - \widehat{h_{t-k}^{\gamma}})\big| z_t z_{t-k}.$$

Applying Hölders inequality with $p_1 = p_2 = p_3 = 1/3$, we obtain

$$|S_{T,1}| \le T^{-1}\{\sum_{t=k+1}^{T}|h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1|^3\}^{1/3}\{\sum_{t=k+1}^{T}|h_{t-k}^{\gamma}/\widehat{h_{t-k}^{\gamma}} - 1|^3\}^{1/3}\{\sum_{t=k+1}^{T}|z_t z_{t-k}|^3\}^{1/3}$$
$$\le \{T^{-1}\sum_{t=k+1}^{T}|h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1|^3\}^{2/3}\{T^{-1}\sum_{t=k+1}^{T}|z_t z_{t-k}|^3\}^{1/3}. \tag{6.42}$$

Denote

$$s_{T,0} = \max_{t=1,\dots,T}(\widehat{h_t^{\gamma}})^{-1}, \quad s_{T,1} = T^{-1}\sum_{t=1}^{T}|h_t^{\gamma} - \widehat{h_t^{\gamma}}|^3, \quad s_{T,2} = T^{-1}\sum_{t=k+1}^{T}|z_t z_{t-k}|^3.$$

Then,

$$|S_{T,1}| \le \{s_{T,0}^3 s_{T,1}\}^{2/3}\, s_{T,2}^{1/3}. \tag{6.43}$$

In (7.11) of Lemma 7.2 it is shown that $s_{T,0} = O_P(1)$, while by (7.13),

$$Es_{T,1} = T^{-1}\sum_{t=1}^{T}E|h_t^{\gamma} - \widehat{h_t^{\gamma}}|^3 = O_P\big((H/T)^{3\nu} + H^{-3/2}\big).$$

By the assumption of the theorem, $\{z_t\}$ are non-negative i.i.d. random variables, and $Ez_t^3 < \infty$. Therefore, $Es_{T,2} = E[z_t^3 z_{t-k}^3] = (E[z_1^3])^2 < \infty$, which implies that $s_{T,2} = O_p(1)$. This together with (6.43) yields

$$|S_{T,1}| = O_P\big((H/T)^{2\nu} + H^{-1}\big) = O_P(T^{-1/2}), \tag{6.44}$$

where the last equality holds because of assumption (18) on H. So, $S_{T,1} = o_P(T^{-1/2})$. This proves (6.41) for $S_{T,1}$.

Next we evaluate $S_{T,2}$. Write,

$$S_{T,2} = T^{-1}\sum_{t=k+1}^{T}(h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1)\zeta_t, \quad \zeta_t = z_t^2(z_{t-k} - 1). \tag{6.45}$$

Write

$$\widehat{h_t^\gamma} = h_t^\gamma + (\widehat{h_t^\gamma} - h_t^\gamma) = h_t^\gamma(1 + x_t), \ x_t = \frac{\widehat{h_t^\gamma} - h_t^\gamma}{h_t^\gamma}.$$

Then

$$\begin{aligned}\frac{h_t^\gamma}{\widehat{h_t^\gamma}} &= \frac{1}{1+x_t} = 1 - x_t + \frac{x_t^2}{1+x_t} \\ &= 1 - x_t + \frac{h_t^\gamma}{\widehat{h_t^\gamma}} x_t^2 = 1 - x_t + \frac{(\widehat{h_t^\gamma} - h_t^\gamma)^2}{\widehat{h_t^\gamma} h_t^\gamma}.\end{aligned}$$

Then, by (6.45),

$$\begin{aligned}S_{T,2} &= -T^{-1} \sum_{t=k+1}^{T} x_t \zeta_t + T^{-1} \sum_{t=k+1}^{T} \frac{(\widehat{h}_t^2 - h_t^2)^2}{\widehat{h}_t^2 h_t^2} \zeta_t \\ &=: q_{T,1} + q_{T,2}.\end{aligned}$$

To prove (6.41) for $S_{T,2}$, we verify that

$$q_{T,\ell} = o_P(T^{-1/2}), \quad \ell = 1, 2. \tag{6.46}$$

In (7.14) of Lemma 7.2 it is shown $E|q_{T,1}| = o(T^{-1/2})$ which proves (6.46) for $q_{T,1}$.

Next, we bound $q_{T,2}$. By Assumption H, $h_t \geq a > 0$ a.s. Therefore, using Hölders inequality with $p_1 = 3/2$, $p_2 = 1/3$, we obtain the same type bound as in (6.42):

$$|q_{T,2}| \leq a^{-2\gamma} \{T^{-1} \sum_{t=k+1}^{T} |h_t^\gamma / \widehat{h_t^\gamma} - 1|^3\}^{2/3} \{T^{-1} \sum_{t=k+1}^{T} |\zeta_t|^3\}^{1/2}.$$

Notice that for an i.i.d. sequence $\{u_t\}$ it holds $E|\zeta_t|^3 = Ez_t^3 E|z_{t-k} - 1|^3 = Ez_1^3 E|z_{1-k} - 1|^3 < \infty$. Therefore the same argument as we used to obtain the bound (6.44) for the r.h.s. of (6.42) implies that

$$q_{T,2} = O_P\big((H/T)^{2\nu} + H^{-1}\big) = o_P(T^{-1/2}),$$

under assumption (18) on H, which proves (6.46). This verifies (6.41) for $S_{T,2}$. The proof of (6.41) for $S_{T,3}$ is similar to the proof for $S_{T,2}$. This completes the proof of (6.32).

Proof of (6.33) Using in (6.40) the bounds (7.1) and (7.2) of Lemma 7.1 we obtain

$$\widehat{\gamma}_k - \widehat{\gamma}_k^* = O_p\big(((H/T)^\nu + H^{-1/2})^2\big) + O_p\big(((H/T)^\nu + H^{-1/2})T^{-1}\big) = o_P(T^{-1/2}).$$

This proves (6.33) and completes the proof of the Lemma 6.3. □.

7. Auxiliary results. This section contains auxiliary lemmas used in the proofs of Section 6.

LEMMA 7.1. *(a) Under assumptions of Lemma 6.2(a), for any fixed $k \geq 1$,*

$$\bar{\widehat{z}} - 1 = O_P\big((H/T)^\nu + H^{-1/2}\big), \tag{7.1}$$

$$\textstyle\sum_{t=1}^{k} |\widehat{z}_t - 1| = O_p(1), \quad \sum_{t=T-k+1}^{T} |\widehat{z}_t - 1| = O_p(1). \tag{7.2}$$

$$\textstyle\sum_{t=1}^{k} (\widehat{z}_t + \widehat{z}_t^2) = O_p(1), \quad \sum_{t=T-k+1}^{T} (\widehat{z}_t + \widehat{z}_t^2) = O_p(1). \tag{7.3}$$

Proof of Lemma 7.1. *Proof of (7.1)*. We have

$$\bar{\widehat{z}} - 1 = T^{-1}\sum_{t=1}^{T}(\widehat{z}_t - 1) = T^{-1}\sum_{t=1}^{T}(\widehat{z}_t - z_t) + T^{-1}\sum_{t=1}^{T}(z_t - 1)$$

$$=: Q_{1,T} + Q_{2,T}. \tag{7.4}$$

We will show

$$Q_{1,T} = O_P\big((H/T)^{\nu} + H^{-1/2}\big), \qquad Q_{2,T} = O_P(T^{-1/2}) \tag{7.5}$$

which proves (7.1). We have,

$$|Q_{1,T}| = T^{-1}\Big|\sum_{t=1}^{T}(h_t^{\gamma}/\widehat{h_t^{\gamma}} - 1)z_t\Big| \le \{\max_{t=1,\dots,T}(\widehat{h_t^{\gamma}})^{-1}\}T^{-1}\sum_{t=1}^{T}|\widehat{h_t^{\gamma}} - h_t^{\gamma}|z_t \tag{7.6}$$

$$\le \{\max_{t=1,\dots,T}(\widehat{h_t^{\gamma}})^{-1}\}\{T^{-1}\sum_{t=1}^{T}(\widehat{h_t^{\gamma}} - h_t^{\gamma})^2\}^{1/2}\{T^{-1}\sum_{t=1}^{T}z_t^2\}^{1/2}$$

$$= O_P(1)\{T^{-1}\sum_{t=1}^{T}(\widehat{h_t^{\gamma}} - h_t^{\gamma})^2\}^{1/2},$$

since $\max_{t=1,\dots,T}(\widehat{h_t^{\gamma}})^{-1} = O_P(1)$ by (7.11) of Lemma 7.2, noting that $Ez_t^2 = Ez_1^2 < \infty$ implies $E[T^{-1}\sum_{t=1}^{T}z_t^2] = O(1)$. By (7.12) of Lemma 7.2, we have $E[(\widehat{h_t^{\gamma}} - h_t^{\gamma})^2] \le C\big((H/T)^{2\nu} + H^{-1}\big)$ which yields

$$E\big[T^{-1}\sum_{t=1}^{T}(\widehat{h_t^{\gamma}} - h_t^{\gamma})^2\big] \le C\big((H/T)^{2\nu} + H^{-1}\big).$$

Hence,

$$|Q_{1,T}| = O_P\big(\{(H/T)^{2\nu} + H^{-1}\}^{1/2}\big) = O_P\big((H/T)^{\nu} + H^{-1/2}\big), \tag{7.7}$$

which proves the first claim in (7.5).

Under Assumption M, $\{u_j\}$ is a stationary α-mixing sequence. Therefore, the sequence $\{z_j = |u_j|^{\gamma}\}$ is also a stationary α-mixing sequence which satisfies the α-mixing Assumption M, see Theorem 14.1 in [4]. Then, by Conclusion 2.2 in [3] (for more details see (A.11) in [5]), the stationary α-mixing sequence $z_t - Ez_t$ has the following property

$$\sum_{k=0}^{\infty}|\mathrm{cov}(z_k, z_0)| < \infty. \tag{7.8}$$

Therefore,

$$EQ_{2,T}^2 = E\big(T^{-1}\sum_{t=1}^{T}(z_t - Ez_t)\big)^2 = T^{-2}\sum_{k,j=1}^{T}\mathrm{cov}(z_k, z_j)$$

$$\le T^{-1}\sum_{k=-\infty}^{\infty}|\mathrm{cov}(z_k, z_0)| \le CT^{-1},$$

which proves the second claim in (7.5). This completes the proof of (7.1).

Proof of (7.2). We have

$$\sum_{t=1}^{k} |\widehat{z}_t - 1| \le \sum_{t=1}^{k} \widehat{z}_t + k \le k^{1/2} (\sum_{t=1}^{k} \widehat{z}_t^{\,2})^{1/2}.$$

We can bound

$$\sum_{t=1}^{k} \widehat{z}_t^{\,2} = \sum_{t=1}^{k} (h_t^{\gamma}/\widehat{h_t^{\gamma}})^2 z_t^2 \le \{\max_{t=1,\dots,T} (\widehat{h_t^{\gamma}})^{-2}\} \sum_{t=1}^{k} h_t^{2\gamma} z_t^2 = O_P(1) \tag{7.9}$$

by (7.11) of Lemma 7.2, and noting that $E[h_t^{2\gamma} z_t^2] \le E[h_t^{2\gamma}] E[z_t^2] < \infty$. This proves the first claim in (7.2). The proof of the second claim is similar.
Proof of (7.3). We can bound

$$\sum_{t=1}^{k} (\widehat{z}_t + \widehat{z}_t^{\,2}) \le k^{1/2} \{\sum_{t=1}^{k} \widehat{z}_t^{\,2}\}^{1/2} + \sum_{t=1}^{k} \widehat{z}_t^{\,2} = O_P(1)$$

by (7.9) which implies the first claim in (7.3). The proof of the second claim is similar. This completes the proof of the lemma. □

LEMMA 7.2. *(a) Under assumptions of Lemma 6.2(a),*

$$\max_{t=1,\dots,T} |h_t^{\gamma} - \widehat{h_t^{\gamma}}| = o_P(T^{-\delta}), \quad (\exists \delta > 0), \tag{7.10}$$

$$\max_{t=1,\dots,T} \widehat{h}_t^{-1} = O_P(1), \tag{7.11}$$

$$E(\widehat{h_t^{\gamma}} - h_t^{\gamma})^2 \le C\big((H/T)^{2\nu} + H^{-1}\big), \tag{7.12}$$

where C does not depend on t, H, T.

(b) In addition, if $\{u_t\}$ is an i.i.d. sequence, then

$$E|\widehat{h_t^{\gamma}} - h_t^{\gamma}|^3 \le C\big((H/T)^{3\nu} + H^{-3/2}\big), \tag{7.13}$$

$$\begin{aligned} E\Big|T^{-1} \textstyle\sum_{t=k+1}^{T} h_t^{-\gamma} (\widehat{h_t^{\gamma}} - h_t^{\gamma}) z_t (z_{t-k} - 1)\Big| & \\ = o\big(T^{-1/2}\big), \quad \text{for } k \ge 1,& \end{aligned} \tag{7.14}$$

where C does not depend on t, H, T.

Proof of Lemma 7.2. *Proof of (7.10).* By definition,

$$\widehat{h_t^{\gamma}} = K_t^{-1} \sum_{j=1}^{T} b_{H,|t-j|} |y_j|^{\gamma} = K_t^{-1} \sum_{j=1}^{T} b_{H,|t-j|} h_j^{\gamma} z_j.$$

Therefore,

$$\begin{aligned} h_t^{\gamma} - \widehat{h_t^{\gamma}} &= K_t^{-1} \textstyle\sum_{j=1}^{T} b_{H,|t-j|} (h_t^{\gamma} - h_j^{\gamma} z_j) \\ &= K_t^{-1} \textstyle\sum_{j=1}^{T} b_{H,|t-j|} (h_t^{\gamma} - h_j^{\gamma}) z_j - K_t^{-1} h_t^{\gamma} \textstyle\sum_{j=1}^{T} b_{H,|t-j|} (z_j - 1) \\ &=: p_t - r_t. \end{aligned} \tag{7.15}$$

We will show that for some $\delta > 0$,

$$\max_{t=1,\dots,T} |r_t| = O_P(T^{-\delta}), \quad \max_{t=1,\dots,T} |p_t| = O_P(T^{-\delta}), \tag{7.16}$$

which implies (7.10)

First, notice that by Assumption H and Proposition 6.1,

$$|h_t^\gamma - h_j^\gamma| \le (|t-j|/T)^\nu \xi_{tj} = (H/T)^\gamma (|t-j|/H)^\nu \xi_{tj}, \tag{7.17}$$

where $\{\xi_{tj}\}$ and $\{h_t^\gamma\}$ satisfy the condition (6.2) of existence of a finite exponential moment with parameter $\alpha^* > 0$. Moreover, properties (11) of the kernel function K imply

$$\max_{t=1,\dots,T} K_t^{-1} \le CH^{-1}, \quad \max_t b_{H,t} \le C < \infty, \tag{7.18}$$

where C does not depend on T.

Denote

$$\begin{aligned} &R_{T,t} = H^{-1} \textstyle\sum_{j=1}^T b_{H,|t-j|}(z_j - 1), \\ &R'_{T,t} = H^{-1} \textstyle\sum_{j=1}^T b^*_{H,|t-j|} z_j, \quad \text{where} \quad b^*_{H,|t-j|} = b_{H,|t-j|}(|t-j|/H)^\nu. \end{aligned} \tag{7.19}$$

Thus, we can bound

$$\begin{aligned} |p_t| &\le K_t^{-1} \sum_{j=1}^T b_{H,|t-j|} |h_t^\gamma - h_j^\gamma| z_j \le CH^{-1} \sum_{j=1}^T b_{H,|t-j|} (|t-j|/T)^\nu \xi_{t,j} z_j \\ &\le C(H/T)^\nu \{\max_{t,j=1,..,T} \xi_{t,j}\} R'_{T,t}, \\ |r_t| &\le C\{\max_{t=1,..,T} h_t\} R_{T,t}. \end{aligned} \tag{7.20}$$

By (v) of Lemma C1 in the online supplement of [5], we can bound

$$\max_{1\le t\le T} h_t^\gamma = O_P\big((\log T)^{2/\alpha^*})\big), \quad \max_{1\le t,j\le T} |\xi_{tj}| = O_P\big((\log T)^{2/\alpha^*}\big). \tag{7.21}$$

Under assumption (18) on H, we have $(H/T)^\nu = O(T^{-\delta})$ for some $\delta > 0$. In view of (7.20) and (7.21), to prove (7.16), it suffices to show that for some $\delta > 0$,

$$\max_{t=1,\dots,T} |R_t| = O_P(T^{-\delta}), \quad \max_{t=1,\dots,T} |R'_t| = O_P(1). \tag{7.22}$$

We start with the first claim in (7.22). As we concluded above, $\{z_j\}$ is a stationary α-mixing sequence which satisfies the α-mixing Assumption M, and we assume that $E|z_1|^3 < \infty$. Under these assumptions, Corollary 6(b) of [5] implies that for any $\varepsilon > 0$,

$$\max_{t=1,\dots,T} |R_t| = O_P\big(H^{-1/2}\sqrt{\log T} + (HT)^{1/3} H^{\varepsilon - 1}\big). \tag{7.23}$$

By assumption (18), we have that $H \ge T^{1/2+a}$ for some $a > 0$. For such H, the r.h.s. of (7.23) is of order $O_P(T^{-\delta})$ for some $\delta > 0$ when ε is selected sufficiently small which proves (7.22) for R_t.

To prove the second claim, write

$$R'_{T,t} = H^{-1} \textstyle\sum_{j=1}^T b^*_{H,|t-j|}(z_j - 1) + H^{-1} \sum_{j=1}^T b^*_{H,|t-j|} =: r_{1,t} + r_{2,t}. \tag{7.24}$$

First we evaluate $r_{1,t}$. By definition, the kernel weights $b^*_{H,|t|} = K^*(|j|/H)$, where $K^*(x) = K(x)|x|^\nu$ satisfy properties (11) imposed on the kernel function $K(x)$ with g replaced by $g - 1 \ge 3$. Moreover,

$$H^{-1} \sum_{j=1}^T b^*_{H,|t-j|} \le C, \tag{7.25}$$

where $C<\infty$ does not depend on t,T. Therefore, likewise as $K_{H,|j|}$, the kernel weights $b^*_{H,|t|}$ satisfy the conditions of Corollary 6(b) of [5] which, as in (7.23), implies

$$\max_{t=1,\dots,T}|r_{1,t}|=O_P\big(H^{-1/2}\sqrt{\log T}+(HT)^{1/3}H^{\varepsilon-1}\big)=O_P(T^{-\delta}). \tag{7.26}$$

On the other hand, it is easy to verify that, as $T\to\infty$,

$$\max_{t=1,\dots,T}|r_{2,t}|=O(1). \tag{7.27}$$

This implies

$$\max_{t=1,\dots,T}|R'_{T,t}|=\max_{t=1,\dots,T}|r_{1,t}|+\max_{t=1,\dots,T}|r_{2,t}|=O_P(T^{-\delta})+O(1)=O_P(1)$$

which completes the proof of (7.22) and (7.10).

Proof of (7.11). Finally, by Assumption H, $h_t\ge a>0$ a.s. Thus,

$$\min_{t=1,\dots,T}\widehat{h_t^\gamma}=\min_{t=1,\dots,T}\big(h_t^\gamma-(h_t^\gamma-\widehat{h_t^\gamma})\big)\ge\min_{t=1,\dots,T}h_t^2-\max_{t=1,\dots,T}|h_t^\gamma-\widehat{h_t^\gamma}|\ge a-O_P(T^{-\delta})$$

by (7.10), which proves (7.11).

Proof of (7.12). By (7.15),

$$(h_t^\gamma-\widehat{h}_t^\gamma)^2=(p_t-r_t)^2\le 2(p_t^2+r_t^2). \tag{7.28}$$

We will show that

$$Ep_t^2\le C(H/T)^{2\nu},\quad Er_t^2\le CH^{-1}, \tag{7.29}$$

which together with (7.28) proves (7.12).

We have

$$\begin{aligned}Ep_t^2&=E\big(K_t^{-1}\textstyle\sum_{j=1}^T b_{H,|t-j|}(h_t^\gamma-h_j^\gamma)z_j\big)^2\\ &\le K_t^{-2}\textstyle\sum_{j,k=1}^T b_{H,|t-j|}b_{H,|t-k|}E[(h_t^\gamma-h_j^\gamma)(h_t^\gamma-h_k^\gamma)z_jz_k].\end{aligned} \tag{7.30}$$

Recall that $\{h_t\}$ and $\{z_t\}$ are mutually independent, and $E[z_jz_k]\le(Ez_k^2Ez_j^2)^{1/2}=Ez_1^2<\infty$. By Proposition 6.1,

$$|h_t^\gamma-h_j^\gamma|\le C(|t-j|/T)^\nu\xi_{\gamma,tj},\quad \max_{tj}E\xi^k_{\gamma,tj}<\infty\quad(\text{for any }\ k\ge1).$$

Therefore,

$$\begin{aligned}\big|E[(h_t^\gamma-h_j^\gamma)(h_t^\gamma-h_k^\gamma)z_jz_k]\big|&=\big|E[(h_t^\gamma-h_j^\gamma)(h_t^\gamma-h_k^\gamma)]E[z_jz_k]\big|\\ &\le C(|t-j|/T)^\nu(|t-k|/T)^{2\nu}E[\xi_{\gamma,tj}\xi_{\gamma,tk}]\le C(|t-j|/T)^\nu(|t-k|/T)^\nu,\end{aligned}$$

where C does not depend on t,j,k,T. Hence, by (7.30),

$$\begin{aligned}Ep_t^2&\le C\textstyle\sum_{j,k=1}^T b_{H,|t-j|}b_{H,|t-k|}(|t-j|/T)^\nu(|t-k|/T)^\nu\\ &\le C(H/T)^{2\nu}\big(K_t^{-1}\sum_{j=1}^T b^*_{H,|t-j|}\big)^2,\end{aligned} \tag{7.31}$$

where C does not depend on t,j,k,T, and $b^*_{H,|j|}=b_{H,|j|}(|j|/H)^\nu$.

Together with (7.18) and (7.25), this implies the first claim in (7.29):

$$Ep_t^2\le C(H/T)^{2\nu}\big(K_t^{-1}\sum_{j=1}^T b^*_{H,|t-j|}\big)^2\le C(H/T)^{2\nu}. \tag{7.32}$$

To bound $E|r_t|^2$, notice that

$$E|r_t|^2 \le K_t^{-2} E[h_t^{2\gamma}] E\big|\sum_{j=1}^T b_{H,|t-j|}(z_j-1)\big|^2$$
$$\le C K_t^{-2} E[h_t^{2\gamma}] \sum_{j,s=1}^T b_{H,|t-j|} b_{H,|t-s|} \mathrm{cov}(z_j, z_s).$$

By assumption, the sequences $\{h_t\}$ and $\{u_t\}$ are mutually independent, and $\max_{1\le t\le T} E[h_t^{2\gamma}] = O(1)$. By (7.8), the covariance function $\mathrm{cov}(z_j, z_s) = \mathrm{cov}(z_{j-s}, z_0)$ of $\{z_t\}$ has the property: $\sum_{s=1}^T |\mathrm{cov}(z_{j-s}, z_0)| \le \sum_{k=-\infty}^{\infty} |\mathrm{cov}(z_k, z_0)| < \infty$. Therefore,

$$Er_t^2 \le C K_t^{-2} \sum_{j,s=1}^T b_{H,|t-j|} |\mathrm{cov}(z_{j-s}, z_0)|$$
$$\le C K_t^{-1} \big(K_t^{-1} \sum_{j=1}^T b_{H,|t-j|} \{ \sum_{s=1}^T |\mathrm{cov}(z_{j-s}, z_0)| \} \big) \le C K_t^{-1} \le C H^{-1},$$

by (7.18). This proves (7.29) for r_t^2 and completes the proof of (7.12).

Proof of (7.13). By (7.15),

$$|h_t^\gamma - \widehat{h}_t^\gamma|^3 = |p_t - r_t|^3 \le 3(|p_t|^3 + |r_t|^3). \tag{7.33}$$

We will show that

$$E|p_t|^3 \le C(H/T)^{3\nu}, \quad E|r_t|^3 \le C H^{-3/2}. \tag{7.34}$$

These bounds together with (7.33) imply (7.13).

We have

$$E|p_t|^3 \le E\big(K_t^{-1} \sum_{j=1}^T b_{H,|t-j|} |h_t^\gamma - h_j^\gamma| z_j \big)^3$$
$$\le K_t^{-3} \sum_{j,k,\ell=1}^T b_{H,|t-j|} b_{H,|t-k|} b_{H,|t-\ell|} E[|h_t^\gamma - h_j^\gamma|\,|h_t^\gamma - h_k^\gamma|\,|h_t^\gamma - h_\ell^\gamma|] E[z_j z_k z_\ell].$$

Notice that $E[z_j z_k z_\ell] \le E z_1^3 < \infty$. Therefore the same argument as that we used to bound the r.h.s. of (7.30) implies

$$E|p_t|^3 \le C(H/T)^{3\nu} \big(K_t^{-1} \sum_{j=1}^T b^*_{H,|t-j|} \big)^3 \le C(H/T)^{3\nu}. \tag{7.35}$$

This proves the first claim in (7.34).

Next we bound $E|r_t|^3$. We have

$$E|r_t|^3 \le K_t^{-3} E[h_t^{3\gamma}] E\big|\sum_{j=1}^T b_{H,|t-j|}(z_j - 1)\big|^3. \tag{7.36}$$

Observe that by Proposition 6.1, $\max_t E[h_t^{3\gamma}] < \infty$. To evaluate the r.h.s. of (7.36), we will use the following bound. If an i.i.d. sequence $\{\xi_t\}$ has zero mean and $E|\xi_t|^p < \infty$ for some $p \ge 2$, then

$$E\big|\sum_{j=1}^T d_j \xi_j\big|^p \le C\big(\sum_{j=1}^T d_j^2\big)^{p/2} \tag{7.37}$$

for any non-random d_j's, where $C < \infty$ does not depend on d_j's and T, see Corollary 2.5.1. in [7]. Since the i.i.d. variables $\xi_j = z_j - 1$ have zero mean and by assumption, $Ez_j^3 < \infty$, then, by using (7.37) in (7.36), we obtain

$$E|r_t|^3 \le C K_t^{-3} \big(\sum_{j=1}^T b^2_{H,|t-j|} \big)^{3/2}$$
$$\le C K_t^{-3} \big(\sum_{j=1}^T b_{H,|t-j|} \big)^{3/2} = C K_t^{-3} K_t^{3/2} \le C H^{-3/2}, \tag{7.38}$$

by (7.18). This verifies the second claim in (7.34) and completes the proof of (7.13).
Proof of (7.14). Denote $\zeta_t = z_t(z_{t-k}-1)$ where $k \geq 1$. Write

$$|y_j|^\gamma - h_t^\gamma = h_j^\gamma z_j - h_t^\gamma = \{h_j^\gamma - h_t^\gamma\} + \{h_j^\gamma(z_j-1)\}.$$

Then

$$\begin{aligned}\widehat{h_t^\gamma} - h_t^\gamma &= K_t^{-1}\sum_{j=1}^T b_{H,|t-j|}(|y_j|^\gamma - h_t^\gamma)\\ &= K_t^{-1}\sum_{j=1}^T b_{H,|t-j|}(h_j^\gamma - h_t^\gamma) + K_t^{-1}\sum_{j=1}^T b_{H,|t-j|}h_j^\gamma(z_j-1)\\ &=: d_{1,t} + d_{2,t}.\end{aligned}$$

Hence, we obtain

$$\begin{aligned}&T^{-1}\textstyle\sum_{t=k+1}^T h_t^{-\gamma}(\widehat{h}_t^\gamma - h_t^\gamma)\zeta_t\\ &\quad= T^{-1}\textstyle\sum_{t=k+1}^T h_t^{-\gamma}d_{1,t}\zeta_t + T^{-1}\sum_{t=k+1}^T h_t^{-\gamma}d_{2,t}\zeta_t\\ &\quad= Q_{1,T} + Q_{2,T}.\end{aligned}$$

We will show that

$$E|Q_{1,T}| = o(T^{-1/2}), \quad E|Q_{2,T}| = O\big(H^{-1}\big). \tag{7.39}$$

Under assumption (18), it holds $H^{-1} = o(T^{-1/2})$ which proves (7.14).

Notice that the sequence $\{h_t^{-\gamma}d_{1,t}\}$ depends only on $\{h_t\}$, and $\{h_t\}$ is independent of an i.i.d sequence $\{u_t\}$. Moreover, $E[\zeta_t\zeta_s] = 0$ for $t \neq s$, and $E\zeta_j^2 = E\zeta_1^2 = Ez_1^2E(z_{1-k}-1)^2 < \infty$. Recall that by Assumption H, $\min_t h_t^2 \geq a > 0$ $a.s.$.

Therefore,

$$\begin{aligned}EQ_{1,T}^2 &= T^{-2}\sum_{t,s=k+1}^T E[h_t^{-\gamma}h_s^{-\gamma}d_{1,t}d_{1,ts}]E[\zeta_t\zeta_s]\\ &= T^{-2}\sum_{t=k+1}^T E[h_t^{-2\gamma}d_{1,t}^2]E\zeta_t^2 \leq a^{-2\gamma}E[\zeta_1^2]T^{-2}\sum_{t=k+1}^T E[d_{1,t}^2].\end{aligned}$$

The same argument is in the proof of the bound $Ep_t^2 \leq C(H/T)^{2\nu}$ in (7.29) implies that

$$Ed_{1,t}^2 \leq C(H/T)^{2\nu},$$

where C does not t, T. Hence, $EQ_{1,T}^2 \leq CT^{-1}(H/T)^{2\nu}$, and therefore,

$$E|Q_{1,T}| \leq (EQ_{1,T}^2)^{1/2} \leq CT^{-1/2}(H/T)^\nu = o(T^{-1/2})$$

which proves the required bound in (7.39).

On the other hand, to evaluate $E|Q_{2,T}|$, write

$$\begin{aligned}Q_{2,T} &= T^{-1}K_t^{-1}\sum_{t=k+1}^T\sum_{j=1}^T b_{H,|t-j|}h_t^{-\gamma}\zeta_t h_j^\gamma(z_j-1)\\ &= T^{-1}K_t^{-1}\sum_{t=k+1}^T\sum_{j=1:\,j>t+3k}^T[...] + T^{-1}K_t^{-1}\sum_{t=k+1}^T\sum_{j=1:\,j<t-3k}^T[...]\end{aligned}$$

$$+T^{-1}K_t^{-1}\sum_{t=k+1}^{T}\sum_{j=1:\,|j-t|\le 3k}^{T}[...]$$
$$= R_{1,T}+R_{2,T}+R_{3,T}.$$

We will show that

$$E|R_{\ell,T}| = O(H^{-1}), \quad \ell = 1,2,3, \tag{7.40}$$

which implies $E|Q_{2,T}| \le CH^{-1}$ and proves (7.39) for $Q_{2,T}$.

First we will evaluate $ER_{1,T}^2$. Since $\{h_t\}$ and $\{u_t\}$ are mutually independent and u_t's are i.i.d. random variables, then for any $j > t+3k$, $j' > t'+3k$,

$$E[\{h_t^{-\gamma}\zeta_t h_j^{\gamma}(z_j-1)\}\{h_{t'}^{-\gamma}\zeta_{t'}h_{j'}^{\gamma}(z_{j'}-1)\}]$$
$$= E[h_t^{-\gamma}h_j^{\gamma}h_{t'}^{-\gamma}h_{j'}^{\gamma}]E[\zeta_{t'}\zeta_t]E[(z_j-1)(z_{j'}-1)] = 0 \quad \text{if } t \ne t' \text{ or } j \ne j';$$
$$= E[h_t^{-2\gamma}h_j^{2\gamma}]E[\zeta_t^2]E[(z_j-1)^2] \text{ if } \quad t=t',\ j=j'.$$

Hence,

$$ER_{1,T}^2 = T^{-2}K_t^{-2}\sum_{t=k+1}^{T}\sum_{j=1}^{T} b_{|t-j|}^2 E[h_t^{-2\gamma}h_j^{2\gamma}]E[\zeta_t^2]E[(z_j-1)^2]$$
$$\le E[\zeta_1^2]E[(z_1-1)^2](\max_{t,j} E[h_t^{-2\gamma}h_j^{2\gamma}])T^{-2}K_t^{-2}\sum_{t=k+1}^{T}\sum_{j=1}^{T} b_{|t-j|}^2.$$

Under assumption $Ez_1^3 < \infty$, $E[\zeta_1^2] = E[z_1^2(z_{1-k}-1)^2] = E[z_1^2]E[(z_{1-k}-1)^2] < \infty$, Assumption H implies $h_t^{-2\gamma} \le a^{-2\gamma} < \infty$ and $\max_j E[h_j^{2\gamma}] < \infty$. Moreover, $\max_j |b_{H,|j|}| < \infty$. Thus,

$$ER_{1,T}^2 \le CT^{-2}K_t^{-1}\sum_{t=k+1}^{T}(K_t^{-1}\sum_{j=1}^{T} b_{H,|t-j|}) \le CT^{-1}K_t^{-1} \le CT^{-1}H^{-1} \le CH^{-2},$$

in view of (7.18). Then $E|R_{1,T}| \le (ER_{1,T}^2)^{-1/2} \le CH^{-1}$ which proves (7.40). The proof of (7.40) for $R_{2,T}$ is similar to the proof for $R_{1,T}$.
Finally, using similar arguments as above, we obtain

$$E|R_{3,T}| \le T^{-1}K_t^{-1}\sum_{t=k+1}^{T}\sum_{j=1:|j-t|\le 3k}^{T} b_{H,|t-j|}E[h_t^{-2}h_j^2]E|\zeta_t(u_t^2-1)|$$
$$\le (\max_{t,k} E[h_t^{-2}h_j^2])E|\zeta_1(u_1^2-1)|(\max_j b_{H,|j|})K_t^{-1}(6k+1) \le CH^{-1},$$

which proves (7.40) for $R_{3,T}$. This completes the proof of (7.14) and the lemma. □

8. Simulation Study. Supplementary material .

8.1. *Supplement to Testing for ARCH effects.* We consider Models 1-3 defined in the main paper with Gaussian noise ε_t. In the main paper we study the performance of our test for ARCH effects on squares ($\gamma = 2$). Tables 3 – 4 below report size and power of the test applied on absolutes, $\gamma = 1$. The shaded grey range of $H = [T^{0.60}, \ldots, T^{0.70}]$ denotes theoretically permissible values of H for smoothness parameter $\nu = 1$. Respectively, Tables 5 – 6 contain testing results for power transform, $\gamma = 1/2$, which show that the size and power of our test in the shaded area are close to the reference values of size and power corresponding to testing on $|u_t|$ and $|u_t|^{1/2}$, as the sample size increases. Overall, for $\gamma = 1,\ 1/2$, we observe similar size and power patterns as for $\gamma = 2$. This suggests that testing is robust across the considered values of γ.

Further, in Tables 7-8 we investigate the size and power of the test for ARCH effects in Model 2, when $\varepsilon_t \sim t(4)$. Such ε_t does not has finite sixth moment, but its fourth moment is finite. Following our theoretical results, testing imposes condition that $E|\varepsilon_t|^{3\gamma} < \infty$. We expected, testing that is based on $\gamma = 2$ to be affected. From the left panel of Table 7 it is easy to see that the test becomes undersized. As expected testing based on $\gamma = 1, 1/2$ is not affected, see Tables 7 (the right panel) and 8. Results for Models 1 and 2 when $\varepsilon_t \sim t(4)$ suggest similar patters and are available upon request.

Finally, in Figure 2 we examine the impact of the existence of the moment $E|\varepsilon_t|^{3\gamma}$ of the noise on the distribution of the test statistic under the null hypothesis. We consider data $y_t = h_t \varepsilon_t$ produced by Model 2 for $T = 1600$, when $\varepsilon_t \sim t(4)$ is an i.i.d. noise. We plot the Monte Carlo average of the $TS(\widehat{\boldsymbol{u}^{\gamma}})$ test statistic for various values of H. From theory, if $E|\varepsilon_t|^{3\gamma} < \infty$, then $TS(\widehat{\boldsymbol{u}^{\gamma}}) \sim \chi^2_p$ for permissible values of H, and the Monte Carlo average of a well behaved test statistic should approach $E\chi^2_p = p$ (illustrated using the colour magenta). We observe this for $\gamma = 1, 1/2$ and for bandwidths $H \in [T^{0.60}, T^{0.70}]$. Testing with these values of H meets the requirements of our theoretical analysis and performs well in small samples. On the contrary, for $\gamma = 2$ the size of our test in the left panel of Table 7 is affected by the absence of the sixth moment, and the Monte Carlo average of the test statistic in Figure 2 drops below $p = 5$, suggesting the failure of the approximation $TS(\widehat{\boldsymbol{u}^{\gamma}}) \sim \chi^2_p$.

8.2. *Supplement to Forecasting of Volatility.* In comparing performance of different volatility forecasting methods in the main paper, we consider the following stationary volatility models for σ_t^2 in $u_t = \sigma_t \varepsilon_t$. We set $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,1)$ in all models except Model 4.

1. The stationary GARCH $(1,1)$ model of [1],

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2, \quad t = 1, \ldots, T. \tag{8.1}$$

   This model is the benchmark in our volatility forecasting.
2. The GJR-GARCH $(1,1)$ model of [8]:

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \gamma u_{t-1}^2 I\left(u_{t-1} < 0\right) + \beta \sigma_{t-1}^2, \quad t = 1, \ldots, T.$$

   This model enables simulation of the leverage effect, which can be important in forecasting volatility.
3. The APARCH $(1,1,1)$ model of [6]:

$$\sigma_t^\delta = \omega + \alpha\left(|u_{t-1}| - \gamma u_{t-1}\right)^\delta + \beta \sigma_{t-1}^\delta, \quad t = 1, \ldots, T.$$

   It adds flexibility capturing volatility dynamics and asymmetries via parameter δ.
4. The GARCH-t $(1,1)$ of [2]. It uses (8.1) and assumes that ε_t follows Student $t(\nu)$ distribution with $v > 2$ (unknown) degrees of freedom. It allows us to assess whether the choice of an $i.i.d$ ε_t noise has impact on forecasting.

5. The stochastic volatility (SV) model of [9]. The presence of two separate generating noises provides to this model extra flexibility.

Parameters of these models are estimated using the quasi maximum likelihood method. Using data $y_1, ..., y_{t-1}$ we estimate σ^2_{t-1} for all five models $j = 1, ..., 5$ and define the corresponding one-step ahead forecast of σ^2_t, $\widehat{\sigma}^2_{j,t|t-1}$, as $\widehat{\sigma}^2_{j;t-1}$ with estimated parameters.

9. An Empirical Example. Supplementary material. Implementation of our test for ARCH effects based on γ-powers hinges on the assumption $E|u_t|^{3\gamma} < \infty$, being satisfied. If this condition is not satisfied, this can lead to size distortions, see the left panel of Table 7.

To investigate the number of finite moments of stock returns in our empirical example, we consider the subperiod from $Jan1994$ to $Dec2007$. First, among 254 stock returns, we select a subset $\mathcal{S}$ of returns that contain no stationary volatility component. For that, we fit to a stock the model $r_t = \mu_t + y_t = \mu_t + h_t u_t$ and test for the absence of ARCH effects in u_t, using the absolute values ($\gamma = 1$) of residuals $\widehat{u}_t = (\widehat{h}_t)^{-1}(r_t - \widehat{\mu}_t)$, at lag $p = 5$. Such testing and thus, the set $\mathcal{S} = \mathcal{S}_H$, depends on H. Next, from each stock from $\mathcal{S}_H$, we test for the existence of a finite moment $E|u_t|^\kappa$, $\kappa = 1, 2, ..., 12$ using specification (17) of Trapani's test in [10]. Table 1 reports the proportion (in %) of stock returns in $\mathcal{S}$ with finite k-th moment, for each bandwidth value H.

From the table we can conclude that there exist four finite moments in the majority of the returns from $\mathcal{S}_H$, but not for six. This suggests that caution is needed when running a test for ARCH effects on squares u^2_t ($\gamma = 2$) since the finite sixth moment of returns may not exist, and illustrates the benefit of a test based on $\gamma = 1$ or $1/2$, which involves a more relaxed moment condition, $E|u_t|^3 < \infty$.

In Figure 1 we report the average values of the test statistic $TS(\widehat{\boldsymbol{u}^\gamma})$ for $p = 5$ over stock returns from $\mathcal{S}_H$ for different bandwidths H. According to our setting, the test statistic for stock returns from $\mathcal{S}_H$, is expected to have the property $TS(\widehat{\boldsymbol{u}^\gamma}) \sim \chi^2_5$ for some range of H if the moment $E|\varepsilon_t|^{3\gamma}$ is finite and ARCH effects in u_t are absent. The average for these values of H should approach to $E\chi^2_5 = 5$. Overall, for $\gamma = 1, 1/2$ the average is close to 5 for bandwidths $H \in [T^{.55}, T^{0.65}]$. This confirms indirectly that the moment $E|\varepsilon_t|^{3\gamma}$ exists and the approximation by χ^2_5 is valid for bandwidths $H \in [T^{0.55}, T^{0.65}]$. This range of bandwidths also meets the theoretical requirements (18) of Theorem 2.1. Hence, these bandwidths should be used in testing for ARCH effects in our empirical example.

Figure 1 also shows that for $\gamma = 2$, the average of the test statistic falls below 5 and exhibits similar patterns as the Monte Carlo average in the case where $u_t \sim t(4)$ and the sixth moment does not exist, see Figure 2. This confirms our testing results above that the stock returns may not have six finite moments. Hence, in our empirical exercise the powers $\gamma = 1, 1/2$ should be used.

Finally, we check the robustness across the proportions of stock returns with no ARCH effect, for different values of $\gamma = 2, 1, 1/2$. Table 1 of the main paper reports the proportions of stock returns with no ARCH effect using testing on powers ($\gamma = 2$, 1) which requires six and three finite moments of u_t, respectively. We need to check these proportions for consistency across γ. Table 2 below reports the proportions on $\gamma = 1/2$ which requires finite $3/2$-th moments of u_t. We see reasonable robustness of proportions with respect to $\gamma = 1/2$ and $\gamma = 1, 2$, which is a positive feature of the test.

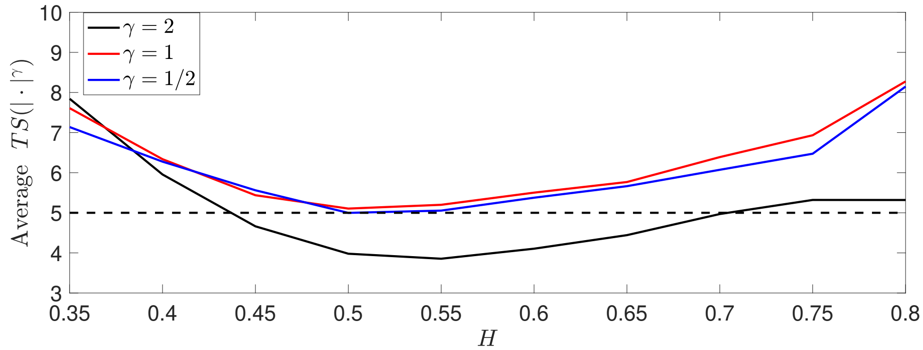

FIG 1. *Average of $TS(\widehat{\boldsymbol{u}^{\gamma}})$ test statistic for various values of H under H_0, from the empirical application, using the single index model for first subperiod 1994-2007.*

TABLE 1

Proportion of stock returns with κ-th finite moment $E|\widehat{u}_t|^{\kappa}$ and $E|y_t|^{\kappa}$ in the subperiod $[1994-2007]$. Trapani's test for finite moments is used. Testing at the 5% significance level. Estimation of h_t based on absolutes $|y_t|$ and bandwidth H.

H	data	$\kappa =$	1	2	3	4	5	6	7	8	9	10	11	12
$T^{0.50}$	$\widehat{u}_t$		100.00	100.00	100.00	99.56	70.67	9.78	8.44	5.33	3.56	8.00	4.89	3.56
$T^{0.55}$			100.00	100.00	100.00	98.70	63.20	9.52	6.93	6.49	6.93	2.60	5.19	4.33
$T^{0.60}$			100.00	100.00	100.00	98.65	63.06	12.61	7.66	6.76	4.05	4.95	4.05	3.60
$T^{0.65}$			100.00	100.00	100.00	98.54	58.54	6.83	6.83	7.32	4.39	7.32	3.41	6.83
$T^{0.70}$			100.00	100.00	100.00	97.11	57.80	8.09	6.94	5.20	4.05	8.09	6.36	5.78
$T^{0.75}$			100.00	100.00	100.00	96.15	53.08	6.15	4.62	11.54	5.38	6.15	9.23	8.46
	y_t		100.00	100.00	100.00	100.00	83.33	16.67	0.00	16.67	0.00	0.00	0.00	0.00

TABLE 2

Proportion of stock returns with no ARCH effects. Testing based on $\gamma = 1/2$ powers, for different bandwidths H, lags p, and subperiods as defined in the main text. Testing at the 5% significance level. μ_t estimated using the single index model.

H	data	$p=1$			$p=5$			$p=10$		
		1994-2007	2005-2012	2011-2019	1994-2007	2005-2012	2011-2019	1994-2007	2005-2012	2011-2019
$T^{0.50}$	$\widehat{\lvert u_t\rvert}^{1/2}$	90.55	94.49	92.91	87.01	81.50	81.89	75.20	68.11	66.93
$T^{0.55}$		88.19	94.49	94.49	87.80	87.80	92.13	87.01	81.89	79.13
$T^{0.60}$		82.68	90.55	94.49	86.22	90.55	95.28	87.40	87.40	88.58
$T^{0.65}$		78.74	85.83	94.88	82.68	85.83	95.28	85.04	84.65	91.73
$T^{0.70}$		72.44	79.13	92.13	72.84	77.95	92.91	78.35	75.59	90.95
$T^{0.75}$		61.42	68.50	90.95	57.09	64.57	90.55	57.87	62.21	90.16
	$\lvert y_t\rvert^{1/2}$	16.14	39.76	74.80	3.54	31.89	71.26	3.94	28.35	70.87

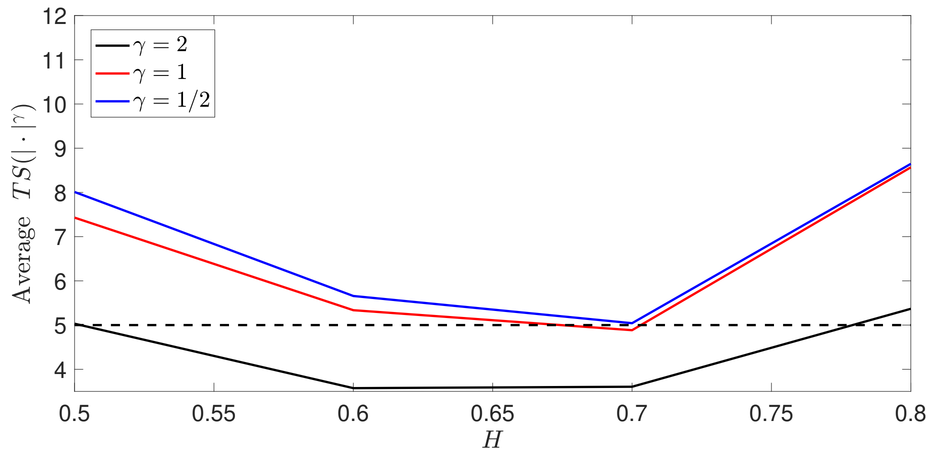

FIG 2. *Average of* $TS(\widehat{\boldsymbol{u}^{\gamma}})$ *test statistic for various values of H under* H_0, $y_t = h_t\varepsilon_t,\ \varepsilon_t \sim t(4)$

TABLE 3
Testing for ARCH effects on squares at the lag $p = 5$ *in Model 1 and Model 2. Rejection frequencies (in* $\%$*) at the* 5% *significance level* ($\beta = 0$ *size,* $\beta > 0$ *power).* $\varepsilon_t \sim \mathcal{N}(0,1)$

		Model 1 (with $h_t = 1$)					Model 2 (with deterministic h_t)				
T	H	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$
200	$T^{0.5}$	$\widehat{\|u_t\|}$	22.88	27.34	59.94	8.52	$\widehat{\|u_t\|}$	20.58	26.36	59.76	8.56
	$T^{0.6}$		9.48	24.72	67.16	24.38		6.26	24.38	68.06	28.62
	$T^{0.7}$		5.98	27.92	72.94	46.94		5.88	35.74	78.76	60.98
	$T^{0.8}$		5.20	31.04	76.70	60.04		21.16	59.12	88.36	82.88
		$\|u_t\|$	4.64	34.70	80.34	73.62	$\|u_t\|$	4.64	34.72	80.36	73.62
		$\|y_t\|$	4.64	34.70	80.34	73.62	$\|y_t\|$	88.94	95.56	98.72	97.46
400	$T^{0.5}$	$\widehat{\|u_t\|}$	21.40	46.36	92.76	30.16	$\widehat{\|u_t\|}$	20.50	46.00	92.64	30.78
	$T^{0.6}$		8.52	49.06	95.90	71.64		6.72	49.40	96.10	74.24
	$T^{0.7}$		5.88	54.44	97.14	88.44		5.60	63.12	98.02	93.00
	$T^{0.8}$		5.12	58.56	97.84	93.38		34.12	86.82	99.62	98.58
		$\|u_t\|$	5.06	61.82	98.32	96.26	$\|u_t\|$	5.06	61.82	98.32	96.26
		$\|y_t\|$	5.06	61.82	98.32	96.26	$\|y_t\|$	99.58	99.88	100.00	100.00
800	$T^{0.5}$	$\widehat{\|u_t\|}$	21.60	78.56	99.96	85.16	$\widehat{\|u_t\|}$	21.14	78.64	99.96	85.48
	$T^{0.6}$		8.08	84.42	99.98	98.98		7.16	84.68	99.98	99.06
	$T^{0.7}$		5.24	88.22	99.98	99.74		5.18	91.98	100.00	99.92
	$T^{0.8}$		4.66	89.80	99.98	99.94		51.20	98.90	100.00	100.00
		$\|u_t\|$	4.46	91.40	99.98	99.94	$\|u_t\|$	4.46	91.40	99.98	99.94
		$\|y_t\|$	4.46	91.40	99.98	99.94	$\|y_t\|$	100.00	100.00	100.00	100.00
1600	$T^{0.5}$	$\widehat{\|u_t\|}$	21.08	98.82	100.00	99.94	$\widehat{\|u_t\|}$	20.94	98.82	100.00	99.94
	$T^{0.6}$		8.06	99.42	100.00	100.00		7.32	99.38	100.00	100.00
	$T^{0.7}$		5.48	99.60	100.00	100.00		5.20	99.80	100.00	100.00
	$T^{0.8}$		4.82	99.72	100.00	100.00		68.52	100.00	100.00	100.00
		$\|u_t\|$	4.78	99.74	100.00	100.00	$\|u_t\|$	4.78	99.74	100.00	100.00
		$\|y_t\|$	4.78	99.74	100.00	100.00	$\|y_t\|$	100.00	100.00	100.00	100.00

TABLE 4

Testing for ARCH effects on absolutes at the lag $p = 5$ in Model 3. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). $\varepsilon_t \sim \mathcal{N}(0,1)$.

Model 3 (with h_t stochastic)											
T	H	data	$\beta = 0$			$\beta = 0.2$			$\beta = 0.4$		
			$d = 1.2$	$d = 1.4$	$d = 1.5$	$d = 1.2$	$d = 1.4$	$d = 1.5$	$d = 1.2$	$d = 1.4$	$d = 1.5$
200	$T^{0.50}$	$\widehat{\lvert u_t\rvert}$	21.32	21.82	22.12	27.10	27.24	27.50	61.02	60.90	61.12
	$T^{0.60}$		8.86	9.46	9.66	25.44	25.12	25.06	68.08	67.90	68.04
	$T^{0.70}$		6.02	6.18	6.30	29.10	28.10	27.88	75.42	74.84	74.98
	$T^{0.80}$		5.86	5.82	5.74	34.82	33.42	32.88	80.28	79.36	79.38
		$\lvert u_t\rvert$	5.70	5.70	5.70	34.94	34.94	34.94	81.84	81.84	81.84
		$\lvert y_t\rvert$	21.74	19.44	18.52	53.38	51.80	50.86	87.86	87.22	87.18
400	$T^{0.50}$	$\widehat{\lvert u_t\rvert}$	20.76	21.76	22.02	46.88	47.04	46.94	93.26	93.30	93.26
	$T^{0.60}$		7.26	8.32	8.40	51.04	50.46	50.38	96.14	96.18	96.10
	$T^{0.70}$		5.06	5.32	5.62	58.66	57.00	56.46	97.76	97.68	97.64
	$T^{0.80}$		6.42	5.32	5.24	66.26	63.46	62.70	98.54	98.46	98.38
		$\lvert u_t\rvert$	5.00	5.00	5.00	63.78	63.78	63.78	98.46	98.46	98.46
		$\lvert y_t\rvert$	35.94	31.98	30.48	82.36	79.76	78.78	99.38	99.30	99.26
	$T^{0.50}$	$\widehat{\lvert u_t\rvert}$	20.38	21.46	21.52	79.58	79.74	79.64	99.98	99.98	99.98
	$T^{0.60}$		7.66	8.44	8.56	85.90	85.60	85.64	99.98	99.98	99.98
	$T^{0.70}$		5.24	5.62	5.60	90.52	89.54	89.32	100.00	100.00	100.00
	$T^{0.80}$		7.42	5.34	5.20	93.78	92.26	91.90	100.00	100.00	100.00
		$\lvert u_t\rvert$	5.12	5.12	5.12	91.94	91.94	91.94	100.00	100.00	100.00
		$\lvert y_t\rvert$	49.54	43.22	40.98	98.06	97.12	96.82	100.00	100.00	100.00
	$T^{0.50}$	$\widehat{\lvert u_t\rvert}$	20.10	21.02	21.22	98.88	98.84	98.86	100.00	100.00	100.00
	$T^{0.60}$		7.24	8.14	8.14	99.44	99.34	99.34	100.00	100.00	100.00
	$T^{0.70}$		5.82	5.74	5.90	99.66	99.62	99.60	100.00	100.00	100.00
	$T^{0.80}$		9.58	6.36	5.98	99.80	99.76	99.76	100.00	100.00	100.00
		$\lvert u_t\rvert$	5.60	5.60	5.60	99.68	99.68	99.68	100.00	100.00	100.00
		$\lvert y_t\rvert$	63.04	55.22	52.76	99.96	99.94	99.94	100.00	100.00	100.00

TABLE 5

Testing for ARCH effects on squares at the lag $p = 5$ in Model 1 and Model 2. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). $\varepsilon_t \sim \mathcal{N}(0,1)$

			Model 1 (with $h_t = 1$)				Model 2 (with deterministic h_t)				
T	H	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$
200	$T^{0.5}$	$\widehat{\lvert u_t\rvert^{1/2}}$	25.26	25.64	50.18	8.36	$\widehat{\lvert u_t\rvert^{1/2}}$	23.70	24.46	49.60	8.30
	$T^{0.6}$		10.34	18.48	53.04	15.78		7.98	17.52	53.88	17.52
	$T^{0.7}$		6.42	18.40	59.02	33.48		4.90	22.86	65.08	46.54
	$T^{0.8}$		5.56	19.58	63.02	46.84		14.22	43.02	79.88	71.90
		$\lvert u_t\rvert^{1/2}$	5.18	22.32	67.32	61.16	$\lvert u_t\rvert^{1/2}$	5.18	22.34	67.34	61.16
		$\lvert y_t\rvert^{1/2}$	5.18	22.32	67.32	61.16	$\lvert y_t\rvert^{1/2}$	83.28	92.32	97.50	95.00
400	$T^{0.5}$	$\widehat{\lvert u_t\rvert^{1/2}}$	23.56	36.10	83.44	17.96	$\widehat{\lvert u_t\rvert^{1/2}}$	22.94	35.96	83.34	18.20
	$T^{0.6}$		9.96	33.64	88.56	53.98		8.34	33.56	88.86	56.42
	$T^{0.7}$		6.40	36.52	91.68	76.84		4.78	43.42	93.80	84.40
	$T^{0.8}$		5.24	39.40	92.80	85.36		22.28	72.72	98.22	96.44
		$\lvert u_t\rvert^{1/2}$	4.72	42.80	94.18	90.42	$\lvert u_t\rvert^{1/2}$	4.72	42.80	94.18	90.42
		$\lvert y_t\rvert^{1/2}$	4.72	42.80	94.18	90.42	$\lvert y_t\rvert^{1/2}$	99.12	99.76	100.00	99.98
800	$T^{0.5}$	$\widehat{\lvert u_t\rvert^{1/2}}$	23.10	61.12	99.48	66.30	$\widehat{\lvert u_t\rvert^{1/2}}$	22.86	60.92	99.48	66.48
	$T^{0.6}$		9.36	65.04	99.74	94.96		8.36	65.64	99.80	95.28
	$T^{0.7}$		6.22	70.08	99.90	98.78		4.94	76.32	99.94	99.42
	$T^{0.8}$		5.36	72.92	99.92	99.44		33.60	94.94	100.00	99.96
		$\lvert u_t\rvert^{1/2}$	4.98	75.66	99.92	99.74	$\lvert u_t\rvert^{1/2}$	4.98	75.64	99.92	99.74
		$\lvert y_t\rvert^{1/2}$	4.98	75.66	99.92	99.74	$\lvert y_t\rvert^{1/2}$	100.00	100.00	100.00	100.00
1600	$T^{0.5}$	$\widehat{\lvert u_t\rvert^{1/2}}$	21.42	92.50	100.00	99.42	$\widehat{\lvert u_t\rvert^{1/2}}$	21.24	92.46	100.00	99.44
	$T^{0.6}$		7.82	95.32	100.00	100.00		7.48	95.38	100.00	100.00
	$T^{0.7}$		5.50	96.64	100.00	100.00		4.90	97.80	100.00	100.00
	$T^{0.8}$		5.06	97.38	100.00	100.00		45.48	99.84	100.00	100.00
		$\lvert u_t\rvert^{1/2}$	4.88	97.70	100.00	100.00	$\lvert u_t\rvert^{1/2}$	4.88	97.70	100.00	100.00
		$\lvert y_t\rvert^{1/2}$	4.88	97.70	100.00	100.00	$\lvert y_t\rvert^{1/2}$	100.00	100.00	100.00	100.00

TABLE 6

Testing for ARCH effects on $\gamma = 1/2$-powers at the lag $p = 5$ in Model 3. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). $\varepsilon_t \sim \mathcal{N}(0,1)$.

Model 3 (with h_t stochastic)											
T	H	data	$\beta = 0$			$\beta = 0.2$			$\beta = 0.4$		
			$d = 1.2$	$d = 1.4$	$d = 1.5$	$d = 1.2$	$d = 1.4$	$d = 1.5$	$d = 1.2$	$d = 1.4$	$d = 1.5$
200	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	25.04	25.56	25.72	24.82	25.32	25.30	50.82	50.98	51.00
	$T^{0.60}$		10.56	11.40	11.70	18.74	18.78	18.80	54.88	54.48	54.30
	$T^{0.70}$		6.82	7.18	7.30	19.66	19.26	19.28	60.68	60.36	60.22
	$T^{0.80}$		5.90	5.94	5.92	22.74	21.62	21.22	66.60	65.02	64.86
		$\lvert u_t\rvert^{1/2}$	5.86	5.86	5.86	22.42	22.42	22.42	68.06	68.06	68.06
		$\lvert y_t\rvert^{1/2}$	17.68	15.82	15.06	39.74	38.30	37.36	77.68	76.44	75.98
400	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	23.38	23.88	24.10	35.72	36.04	36.10	83.26	83.28	83.24
	$T^{0.60}$		8.28	8.96	9.22	34.04	34.20	34.22	88.60	88.62	88.60
	$T^{0.70}$		5.38	5.56	5.86	39.96	38.56	38.18	92.52	92.06	92.02
	$T^{0.80}$		5.92	5.22	5.08	47.02	44.08	43.42	94.74	94.18	94.04
		$\lvert u_t\rvert^{1/2}$	5.24	5.24	5.24	44.18	44.18	44.18	94.76	94.76	94.76
		$\lvert y_t\rvert^{1/2}$	29.44	26.06	24.70	68.84	65.92	64.66	97.52	97.12	97.08
	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	21.86	22.86	22.92	61.58	61.78	61.88	99.34	99.32	99.32
	$T^{0.60}$		8.50	9.02	9.24	65.98	65.70	65.58	99.76	99.76	99.72
	$T^{0.70}$		5.40	5.86	5.90	73.02	71.20	71.00	99.92	99.90	99.90
	$T^{0.80}$		5.92	5.10	5.02	80.24	77.12	76.52	99.98	99.96	99.96
		$\lvert u_t\rvert^{1/2}$	5.28	5.28	5.28	76.52	76.52	76.52	99.96	99.96	99.96
		$\lvert y_t\rvert^{1/2}$	41.72	36.96	34.86	92.34	90.46	90.00	100.00	100.00	100.00
	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	21.48	22.32	22.36	91.92	91.88	91.84	100.00	100.00	100.00
	$T^{0.60}$		7.44	7.94	8.06	95.24	95.12	95.18	100.00	100.00	100.00
	$T^{0.70}$		5.54	5.56	5.62	97.18	96.58	96.42	100.00	100.00	100.00
	$T^{0.80}$		7.30	5.74	5.56	98.44	97.86	97.68	100.00	100.00	100.00
		$\lvert u_t\rvert^{1/2}$	5.36	5.36	5.36	97.38	97.38	97.38	100.00	100.00	100.00
		$\lvert y_t\rvert^{1/2}$	55.66	48.64	46.08	99.70	99.54	99.40	100.00	100.00	100.00

TABLE 7

Testing for ARCH effects on squares at the lag $p = 5$ in Model 2 (with deterministic h_t). Rejection frequencies (in $\%$) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). $\varepsilon_t \sim t(4)$.

		Testing on squares ($\gamma = 2$)					Testing on absolutes ($\gamma = 1$)				
T	H	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$GARCH$
200	$T^{0.5}$	$\widehat{u_t^2}$	2.48	18.84	50.14	15.92	$\widehat{\lvert u_t\rvert}$	14.74	45.28	84.08	19.78
	$T^{0.6}$		1.00	31.34	63.38	56.30		4.18	52.32	89.92	69.58
	$T^{0.7}$		3.18	42.16	70.20	79.16		3.84	61.94	93.56	91.52
	$T^{0.8}$		8.92	50.74	74.02	87.94		9.62	72.02	95.72	97.16
		u_t^2	4.98	49.90	75.88	91.56	$\lvert u_t\rvert$	3.92	66.94	94.96	98.18
		y_t^2	24.46	61.62	78.82	93.00	$\lvert y_t\rvert$	53.86	90.64	98.56	99.28
400	$T^{0.5}$	$\widehat{u_t^2}$	2.16	50.42	88.28	66.72	$\widehat{\lvert u_t\rvert}$	14.80	82.26	99.38	80.94
	$T^{0.6}$		1.44	62.20	90.90	94.76		4.50	87.60	99.84	99.20
	$T^{0.7}$		4.08	69.66	91.62	98.54		4.24	92.26	99.88	99.96
	$T^{0.8}$		10.52	74.92	91.84	98.92		15.02	96.20	99.96	99.98
		u_t^2	6.08	73.32	91.12	98.98	$\lvert u_t\rvert$	4.44	93.78	99.94	100.00
		y_t^2	31.66	81.60	92.16	99.12	$\lvert y_t\rvert$	82.78	99.70	99.96	100.00
800	$T^{0.5}$	$\widehat{u_t^2}$	1.92	87.54	99.58	99.36	$\widehat{\lvert u_t\rvert}$	16.58	99.08	100.00	99.98
	$T^{0.6}$		1.54	89.30	99.16	99.96		5.24	99.60	100.00	100.00
	$T^{0.7}$		4.10	90.22	98.14	99.96		4.30	99.80	100.00	100.00
	$T^{0.8}$		10.92	91.30	97.54	99.94		18.88	99.90	100.00	100.00
		u_t^2	5.56	89.46	97.10	99.86	$\lvert u_t\rvert$	4.46	99.92	100.00	100.00
		y_t^2	40.24	93.16	97.12	99.90	$\lvert y_t\rvert$	98.06	100.00	100.00	100.00
1600	$T^{0.5}$	$\widehat{u_t^2}$	2.34	99.58	100.00	100.00	$\widehat{\lvert u_t\rvert}$	17.58	100.00	100.00	100.00
	$T^{0.6}$		2.12	98.88	100.00	100.00		6.08	100.00	100.00	100.00
	$T^{0.7}$		4.38	97.96	99.66	100.00		4.76	100.00	100.00	100.00
	$T^{0.8}$		11.84	97.36	99.36	100.00		25.98	100.00	100.00	100.00
		u_t^2	6.32	96.62	98.84	100.00	$\lvert u_t\rvert$	5.08	100.00	100.00	100.00
		y_t^2	53.24	97.38	98.84	99.98	$\lvert y_t\rvert$	99.92	100.00	100.00	100.00

TABLE 8

Testing for ARCH effects on $\gamma=1/2$-powers at the lag $p=5$ in Model 2. Rejection frequencies (in %) at the 5% significance level ($\beta=0$ size, $\beta>0$ power). $\varepsilon_t \sim t(4)$

Model 2 (with h_t deterministic)						
T	H	data	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$GARCH$
200	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	23.56	44.54	83.64	14.04
	$T^{0.60}$		7.80	46.70	88.92	60.40
	$T^{0.70}$		4.56	54.78	93.02	88.00
	$T^{0.80}$		9.06	68.28	96.08	96.04
		$\lvert u_t\rvert^{1/2}$	4.50	59.54	94.50	97.60
		$\lvert y_t\rvert^{1/2}$	64.96	93.06	99.42	99.08
400	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	22.46	76.50	99.30	68.86
	$T^{0.60}$		7.16	82.36	99.74	98.42
	$T^{0.70}$		4.54	88.08	99.92	99.88
	$T^{0.80}$		13.82	95.16	99.98	99.96
		$\lvert u_t\rvert^{1/2}$	4.80	89.56	99.98	100.00
		$\lvert y_t\rvert^{1/2}$	93.36	99.94	100.00	100.00
800	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	22.52	98.12	100.00	99.90
	$T^{0.60}$		7.70	99.10	100.00	100.00
	$T^{0.70}$		4.74	99.60	100.00	100.00
	$T^{0.80}$		19.20	99.90	100.00	100.00
		$\lvert u_t\rvert^{1/2}$	4.86	99.76	100.00	100.00
		$\lvert y_t\rvert^{1/2}$	99.84	100.00	100.00	100.00
1600	$T^{0.50}$	$\widehat{\lvert u_t\rvert^{1/2}}$	22.58	100.00	100.00	100.00
	$T^{0.60}$		7.92	100.00	100.00	100.00
	$T^{0.70}$		5.08	100.00	100.00	100.00
	$T^{0.80}$		26.80	100.00	100.00	100.00
		$\lvert u_t\rvert^{1/2}$	5.26	100.00	100.00	100.00
		$\lvert y_t\rvert^{1/2}$	100.00	100.00	100.00	100.00

REFERENCES

[1] BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31(3), 307–327.

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

[3] DAVYDOV, Y. A. (1968): "Convergence of distributions generated by stationary stochastic processes", *Theory of Probability & Its Applications*, 13(4), 691–696.

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

[5] DENDRAMIS, Y., L. GIRAITIS, AND G. KAPETANIOS (2021): “Estimation of time-varying covariance matrices for large datasets,” *Econometric Theory*, *published online*.

[6] DING, Z., C.W. GRANGER, AND R.F. ENGLE (1993): *A long memory property of stock market returns and a new model*. *Journal of Empirical Finance*, 1(1), 83–106.

[7] GIRAITIS, L., H. KOUL, AND D. SURGAILIS (2012): *Large Sample Inference for Long memory Processes*. *Imperial College Press, London*

[8] GLOSTEN, LR., R. JAGANNATHAN, AND D.E. RUNKLE (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks", *The Journal of Finance*, 542–547.

[9] RUIZ, E. (1994): "Quasi-maximum likelihood estimation of stochastic volatility models", *Journal of Econometrics*, 63, 289–306.

[10] TRAPANI, L. (2016): "Testing for (in) finite moments", *Journal of Econometrics*, 191, 57–68.