KING'S College LONDON

Choosing between persistent and stationary volatility

Ilias Chronopoulos*† Liudas Giraitis ‡ George Kapetanios §

February 14, 2021

Abstract

In this paper, we analyse persistent and possibly non-stationary processes that have the potential to characterise volatility better than stationary alternatives. We discuss in detail both the conditions needed for their consistent estimation and conditions that enable the use of standard ARCH tests to detect presence of stationary volatility after persistent volatility is taken into account. We provide Monte Carlo evidence that supports our testing strategy in small samples and present extensive empirical evidence clearly supporting the persistent volatility paradigm, suggesting that stationary time–varying conditional volatility is less pronounced than previously thought. Finally, results from an out-of-sample forecasting exercise are presented, that support our proposed persistent volatility paradigm.

Key words: Time-varying coefficient models, random coefficient models, non-parametric estimation, kernel estimation, persistence, volatility.

JEL classification: C10, C12, C14.

*Chronopoulos's research is supported by the ESRC grant ES/P000703/1

†King's Business School, King's College London, Email: Ilias.Chronopoulos@kcl.ac.uk

‡ Queen Mary, University of London, Email: L.Giraitis@qmul.ac.uk

§ King's Business School, King's College London, Email: George.Kapetanios@kcl.ac.uk

1 Introduction

Two important issues widely discussed in empirical econometric analysis for macroeconomics and finance, over the last 25 years, are structural change and volatility modelling. Starting with the seminal work of Engle (1982), volatility modelling has developed into a large topic of study - perhaps the major preoccupation of financial econometrics. Most work has produced models that are stationary but, crucially, allow for time variation in conditional variances. There are two important groups of parametric models used to model volatility. Firstly, the generalised $ARCH$ class, where specific models contain a single innovation process and, secondly, Stochastic Volatility (SV) models, where the conditional variance is treated as a latent variable and more than one innovation processes enter the model.

Empirical work though, has repeatedly concluded that variation in volatility can be extremely persistent. Such a finding is not easily accommodated within the above stationary model classes. The challenge is revealed in observed parameter estimates that are close to the boundary of stationarity. This *integrated GARCH effect*, see e.g. Mikosch and Stărică (2004), can be caused by structural change in the unconditional variance, where it changes either smoothly or abruptly 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 processes that allow for gradual change in the unconditional variance. First, we succinctly summarise the main ways this is addressed in the literature, and then present our main contributions.

The first line of research, following recent work on structural change, has focused on paradigms coming from the statistical literature, such as the work of Priestley (1965) and Dahlhaus (2000), where processes are smooth deterministic functions of time. Dahlhaus, Rao, et al. (2006) proposed the locally stationary time-varying $ARCH$ model, where stationarity is assumed locally, but the process is globally nonstationary. Along the same lines Van Bellegem and Von Sachs (2004) proposed another smooth deterministic model, where they assumed that volatility is multiplicatively decomposed into a stationary and nonstationary part. The assumption that there is a nonstationary part that could at least partially drive the volatility, has recently permeated in standard $ARCH$ (stationary) models as well. Specifically, Engle and Rangel (2008), used exponential splines to specify the nonstationary part. Brownlees and Gallo (2009) followed closely using a different version of splines. Mazur and Pipień (2012), used the Fourier Flexible form of Gallant (1984), and finally, in a series of papers Amado and Teräsvirta (2013) and Amado and Teräsvirta (2017), suggested the use of a linear combination of logistic functions and their generalisations.

While the above characterisations provide an avenue to describe and estimate nonstationary processes for the volatility process, either with kernel type methods or semiparametrically, there is no clear way to separate the two kinds. Further, the above characterisations are tied to either parametric forms for the nonstationary component or are subject to be smooth deterministic functions of time, which disallows the use of stochastic components.

It is in this context that this paper makes a number of contributions. First, we discuss possible setups for processes that are persistent and can potentially account for the persistence of volatility observed in data. Then, we use ideas from recent work in the structural change literature to discuss how persistence, perhaps surprisingly, allows estimation of the unobserved volatility process without strong parametric assumptions and the requirement that they are smooth and deterministic functions of time[1]. While such a focus on smooth deterministic functions of time allows estimation using kernel methods, it is not satisfactory as it does not allow for stochastic elements that can provide a richer representation.

Recent work by Giraitis, Kapetanios, and Yates (2014) address this by showing that as long as a process satisfies a smoothness and boundedness (or moment) condition, then it can be stochastic but still estimable using kernel estimation. This opens up the possibility that such processes may adequately fit the observed behaviour of time-varying volatility, being clearly more persistent than stationary processes. In fact this persistence is their most distinctive characteristic. To appreciate this, it is important to focus on the estimation strategy considered by Giraitis, Kapetanios, and Yates (2018). They essentially ask the following question: Assuming a decomposition of the form $y_t = h_t u_t$, for some observed process y_t, what properties should h_t have, so that h_t^2 can be consistently estimated by, essentially, a rolling window type mean estimation of y_t^2, with an expanding window size? It follows that the answer cannot lie within the class of stationary processes whose squares satisfy a law of large numbers. As a result, the vast majority of stationary models do not qualify. Instead, the answer seems to be that a process has to change slowly, in the sense that $|h_t - h_s|$ has to be small. Further, and for obvious reasons, it has to be bounded, a condition that disallows the use of random walks, which would otherwise be of great relevance due to their persistence. However, a normalised, and therefore bounded, random walk provides a canonical example for the sort of processes we have in mind.

The ability to consistently estimate h_t if it is persistent but not otherwise, provides a clear avenue for a strategy to separate stationary from persistent processes, that is clearly missing from the literature. If the unobserved volatility of a process y_t is persistent then it can be estimated and then the rescaled series $y_t/\widehat{h}_t$ can be tested for the presence of time varying stationary volatility using existing standard $ARCH$ tests. If there is only persistent volatility, the tests will not reject. If there is only stationary volatility, normalisation by the estimate of h_t^2 will not remove it, since the estimate will simply converge to $E(h_t^2)$, and the tests will reject. Finally, the possibility exists that both the persistent and stationary volatility components co-exist. Estimation of the persistent part can proceed in the presence of the stationary part and then again the test will reject. In a second step, a stationary volatility model can be fitted and estimated. This case is not covered in this paper and we leave it for future work.

[1] see e.g. Priestley (1965), Dahlhaus (2000), Kapetanios and Yates (2008) and Van Bellegem and Von Sachs (2004), among others.

In this paper, we discuss in detail conditions needed for consistent estimation of persistent volatility scaling factor h_t 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 in small samples. We proceed and present extensive empirical evidence clearly supporting the persistent volatility paradigm, suggesting that stationary time varying volatility is less pronounced than previously thought and that further conditional second moments of asset returns change slowly. Finally, results from an out-of-sample recursive forecasting exercise are presented and found to support the use of persistent volatility modelling

The remainder of this paper is organised as follows. Section 2 presents our econometric procedure and theoretical results. Section 3 contains the Monte Carlo exercise. Section 4 presents the empirical results from implementing our testing strategy to the data, and from the forecasting exercise. Section 5 concludes. Proofs are relegated to the appendix.

2 Theoretical considerations

We consider the following white noise model:[2]

$$y_t = h_t u_t, \quad t = 1, \ldots, T, \tag{2.1}$$

where u_t is a stationary sequence of uncorrelated random variables with $Eu_t = 0$, $Eu_t^2 = 1$, and h_t is a persistent scale factor (stochastic or deterministic). We assume that sequences $\{u_t\}$ and $\{h_t\}$ are mutually independent. Then

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

Given observations $y_1, ..., y_T$, our objective is to test for the presence of conditional heteroscedasticity (ARCH effect) in u_t^2, i.e. to determine whether u_t is an i.i.d noise or dependent random variables which in addition to h_t contributes in (2.1) a stationary conditional e.g. GARCH type volatility $E[u_t^2 | u_{t-1}, u_{t-2}, ...]$. Since u_t is not observed, we will estimate h_t by an estimate $\widehat{h}_t$ and base testing for ARCH effects on residuals $\widehat{u}_t = \widehat{h}_t^{-1} y_t$. Such testing requires uniform consistency in estimation of u_t^2 by $\widehat{u}_t^2$ and thus, stronger conditions on (h_t, u_t) than in point estimation of h_t at time t. This is reflected in Assumptions M and H we make on u_t and h_t, in particular the assumption of mutual independence of $\{h_t^2\}$ and $\{u_t^2\}$ we impose. The latter clearly holds for a deterministic scaling factor h_t.

[2]This model abstracts from the general case of a model with a specified conditional mean and time varying volatility of the form $y_t = E(y_t | \mathcal{F}_{t-1}) + h_t u_t, \quad t = 1, \ldots, T$, where $\mathcal{F}_{t-1}$ is a sigma field: $\mathcal{F}_{t-1} = \sigma\{h_{t-1}, \ldots, h_0, u_{t-1}, \ldots, u_1\}$, by setting $E(y_t | \mathcal{F}_{t-1})$ to zero for simplicity. The general case can be handled straightforwardly.

Assumption M (α-mixing)
1. $\{u_t\}$ is a stationary ergodic sequence with $Eu_t = 0$, $Eu_t^2 = 1$, $Eu_t u_s = 0$ for $t \neq s$, and $E|u_1|^\theta < \infty$ for some $\theta > 6$.
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. Variables $h_1, ..., h_T$ satisfy the following smoothness condition. For some $\gamma \in (1/2, 1]$,

$$|h_t - h_j| \leq (|t-j|/T)^\gamma \xi_{tj}, \quad t, j = 1, ..., T \tag{2.2}$$

and for some $0 < \alpha \leq \infty$, $c > 0$

$$\max_{t,j\geq 1} E[\exp(c|\xi_{tj}|^\alpha] < \infty, \quad \max_{t,j\geq 1} E[\exp(c|h_t|^\alpha)] < \infty. \tag{2.3}$$

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

Remark 2.1. 1. Condition (2.2) in Assumption H implies that the volatility process h_t drifts slowly in time, which essentially rules out explosive behaviour. This is a widely used assumption in the literature. It allows the use of both deterministic varying process of the form $h_t = g(t/T)$, where $g(\cdot)$ is a Lipschitz smooth function with parameter $1/2 < \gamma \leq 1$, i.e. $|g(x) - g(y)| \leq C|x - y|^\gamma$, as well as a stochastic process h_t.

2. The deterministic specification $h_t = g(t/T)$, $t = 1, ..., T$ is a standard assumption in the work of Dahlhaus on locally stationary processes (see, e.g. Dahlhaus (2000) or Dahlhaus and Polonik (2006)). The stochastic time-variation of h_t was proposed by Giraitis et al. (2014, 2016, 2018), to allow for stochastic processes that can be presented as non-stationary random walks. The combinations between the two, satisfying (2.2) can be summarised as

$$h_t = |T^{-\gamma}(v_1 + ... + v_t) + g(t/T)| + a, \quad t = 1, ..., T, \tag{2.4}$$

where $a > 0$ and v_t is a stationary zero mean sequence, see Example 2.1.

3. 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. Chronopoulos, Kapetanios, and Petrova (2019).

Example 2.1. Let v_j be a stationary zero mean sequence of Gaussian $ARFIMA(p, d, q)$ variables with parameter $d \in (0, 1/2)$[3]. Set $\gamma = 1/2 + d$. If $|g(x) - g(y)| \leq C|x - y|^\nu$ for $x, y \in (0, 1)$ where $\gamma \leq \nu \leq 1$ then h_t^2 in (2.4) satisfies Assumption H with $\gamma = 1/2 + d$ and $\alpha = 2$. Indeed, for $t > s$,

$$\begin{aligned} |h_t - h_s| &= \left|T^{-\gamma} \textstyle\sum_{j=1}^t v_j + g(t/T)\right| - \left|T^{-\gamma} \textstyle\sum_{j=1}^s v_j + g(s/T)\right| \\ &\leq |T^{-\gamma} \textstyle\sum_{j=s+1}^t v_j| + |g(t/T) - g(s/T)| \leq (|t-s|/T)^\gamma |\xi_{ts}| + C(|t-s|/T)^\gamma \end{aligned}$$

[3]see Chapter 7 in Giraitis, Koul, and Surgailis (2012)

where $\xi_{tj} = (t-s)^{-\gamma}\sum_{j=s+1}^{t} v_j$ is a Gaussian r.v. Since $\mathrm{var}(\xi_{tj}) = \mathrm{var}(\xi_{t-j,0}) \to v_d^2$ as $t-j \to \infty$[4], ξ_{tj} satisfies (2.2) with $\gamma = 1/2+d$ and $\alpha = 2$. Recall that $\sum_{j=1}^{t} v_j$ is $ARFIMA(p, 1+d, q)$ process.

2.1 Volatility estimation

For the estimation of h_t, we follow Giraitis, Kapetanios, and Yates (2018). They considered estimation of h_t in the context of VAR(1) model. We show that in the model described in (2.1) and under Assumptions H and M, h_t^2 can be consistently estimated as

$$\widehat{h}_t^2 = K_t^{-1}\sum_{j=1}^{T} b_{|t-j|} y_j^2, \quad K_t = \sum_{j=1}^{T} b_{|t-j|}, \quad t = 1, ..., T, \tag{2.5}$$

where $b_{|t-j|} = K((t-j)/H)$ 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:

$$\begin{aligned} K(x) &= (1/2)I(|x| \le 1), \quad \text{flat kernel,} \\ K(x) &= (3/4)(1-x^2)I(|x| \le 1), \quad \text{Epanechnikov kernel,} \\ K(x) &= (1/\sqrt{2\pi})e^{-x^2/2}, \quad \text{Gaussian kernel.} \end{aligned}$$

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

$$K(x) \le C(1+x^\nu)^{-1}, \ \ |(d/dx)K(x)| \le C(1+x^\nu)^{-1}, \ \ x \ge 0 \ \text{ for some } \nu > 3, \ C > 0. \tag{2.6}$$

Under this setup, in Lemma 6.5 we show that

$$|\widehat{h}_t^2 - h_t^2| = O_p\left((H/T)^\gamma + H^{-1/2}\right), \tag{2.7}$$

and in Lemma 6.3 using Dendramis, Giraitis, and Kapetanios (2021) results we show uniform convergence

$$\max_{t=1,\dots,T} |\widehat{h}_t^2 - h_t^2| = o_P(1). \tag{2.8}$$

The uniform convergence result on the asymptotic consistency in the estimation of $\widehat{h}_t$, at a non-parametric rate, will prove useful in our testing procedure for distinguishing between persistent and stationary volatility that follows.

[4] see Proposition 3.3.1 in Giraitis, Koul, and Surgailis (2012)

2.2 Testing

In this subsection we consider how our strategy for discriminating between persistent and stationary volatility works. First, let us briefly summarise the most basic tests used in testing for the presence of $ARCH$ effects for a stationary sequence of uncorrelated variables $y_t = h_t u_t$ with $h_t = const.$

To test for ARCH effect in u_t^2, we use the test Lagrange Multiplier (LM) test by Engle (1982) which is similar to the LM test for autocorrelation. We fit to u_t^2 an AR(p), $p \geq 1$ model

$$u_t^2 = \beta_0 + \beta_1 u_{t-1}^2 + ... + \beta_p u_{t-p}^2 + \eta_t \tag{2.9}$$

where $\beta_0 > 0$ and test the following null hypothesis:
$H_0: \ \beta_1 = \beta_2 = \ldots = \beta_p = 0$

against the alternative
$H_1: \ \beta_j \neq 0$ for some $j = 1, ..., p$.

Engle (1982) derived a Lagrange Multiplier (LM) statistic for testing H_0, based on TR^2, where T is the sample size and R-squared is obtained from the auxiliary regression of u_t^2 on a constant and $u_{t-1}^2, \ldots, u_{t-p}^2$. Under H_0 ($u_t \sim$ i.i.d.), 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 tries to address two distinct problems: i) the misspecification of the conditional mean, see e.g. the discussion in Bera and Higgins (1993) and ii) the correct specification of the volatility process. Our work naturally falls in the second class, and in essence testing is useful to address whether persistent processes, defined as the ones that follow the smoothness condition (2.2) in Assumption H, provide a better specification for the volatility process.

Consider the model for y_t in (2.1) where u_t's are not observed, and let the standardised residuals be

$$\widehat{u}_t = \frac{y_t}{\widehat{h}_t}, \quad t = 1, \ldots, T, \tag{2.10}$$

where $\widehat{h}_t$ is defined as $\sqrt{\widehat{h}_t^2}$ of the kernel estimate of the persistent volatility using the estimator described in (2.5). Our aim is to show that asymptotically, it is equivalent to test for $ARCH$ effects using the standardised residuals, $\widehat{\boldsymbol{u}} = [\widehat{u}_1, \widehat{u}_2, \ldots, \widehat{u}_T]'$, instead of $\boldsymbol{u} = [u_1, u_2, \ldots, u_T]'$. Such an equivalence implies that if a process h_t, that follows Assumption H, drives the conditional second moment, then, under consistent estimation of h_t using the estimator $\widehat{h}_t$ in (2.5), the standardised residuals, $\widehat{u}_t = \widehat{h}_t^{-1} u_t$, should behave as white noise. In Theorem 2.4 we show that $\widehat{u}_t$ can be used to compute the correlogram of u_t and test for absence of correlation in u_t.

Testing for ARCH effects using regression (2.9) for squares $\widehat{u}_t^2$, note, that if stationary processes (co-)drive the volatility via u_t, then the normalisation by $\widehat{h}_t$ will not corrupt the properties of testing. In our setup and for $p \geq 1$, consider $TS(\boldsymbol{u}) = TR^2$, the test statistic based on $\boldsymbol{u}$ and $TS(\widehat{\boldsymbol{u}}) = TR^2$ based on the standardised residuals, $\widehat{\boldsymbol{u}}$, as described above.

The formulas of $S(\boldsymbol{u})$ and $S(\widehat{\boldsymbol{u}})$ are given in (6.1) and (6.2) of the proof. The following Theorem, that is proven in the Appendix, gives the sufficient condition for LM test for ARCH effects in squares u_t^2 to be asymptotically valid when applied to $\widehat{\boldsymbol{u}}$, instead of $\boldsymbol{u}$.

Denote $\gamma_k \equiv \gamma_{u^2,k} = \mathrm{cov}(u_k^2, u_0^2)$, $k \geq 0$ and set $\boldsymbol{\Gamma}_p = \left(\gamma_{|j-k|}\right)_{j,k=1,...,p}$ and $\boldsymbol{\gamma}_p = \left(\gamma_1, ..., \gamma_p\right)'$. Denote $\boldsymbol{\beta}_p = (\beta_1, ..., \beta_p)' = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p$, $\sigma_p^2 = \mathrm{var}(u_{p+1}^2 - \beta_1 u_p^2 - ... - \beta_p u_1^2)$. Recall notation γ of the smoothness parameter of h_t in (2.2). Notation $a_n << b_n$ means $a_n = o(b_n)$. Notice that $\boldsymbol{\Gamma}_p^{-1}$ exists[5].

Theorem 2.1. *(a) Let y_t follow (2.1) where h_t and u_t satisfy Assumptions H and M. Suppose that H satisfies*

$$T^{1/2} << H << T^{1-(1/4\gamma)}. \tag{2.11}$$

Then, for any $p \geq 1$, the test statistics $TS(\boldsymbol{u})$ and $TS(\widehat{\boldsymbol{u}})$ corresponding to regression (2.9) on squares u_t^2, have the following property

$$S(\widehat{\boldsymbol{u}}) = S(\boldsymbol{u}) + o_P(1) = \sigma_p^{-2}\boldsymbol{\beta}_p'\boldsymbol{\Gamma}_p\boldsymbol{\beta}_p + o_P(1). \tag{2.12}$$

(b) If $\{u_t\}$ is also an i.i.d. sequence, then for any $p \geq 1$, $\boldsymbol{\beta}_p = 0$, and

$$TS(\widehat{\boldsymbol{u}}) = TS(\boldsymbol{u}) + o_P(1) \to_D \chi_p^2. \tag{2.13}$$

This theorem shows, that the alternative $H_1 : \boldsymbol{\beta}_p \neq 0$, is equivalent to $\boldsymbol{\gamma}_p \neq 0$ and detected with the rate T. Indeed, $\boldsymbol{\beta}_p'\boldsymbol{\Gamma}_p\boldsymbol{\beta}_p = \boldsymbol{\gamma}_p'\boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p$, 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 $0 < \lambda_{min} \leq \lambda_{max} < \infty$ are the smallest and largest eigenvalues of $\boldsymbol{\Gamma}_p$ and $||\boldsymbol{\beta}_p||$ denotes the Euclidean norm of $\boldsymbol{\beta}_p$. It also shows that $\boldsymbol{\beta}_p = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p$ in the "true" value of $\boldsymbol{\beta}_p$ estimated by OLS in regression (2.9).

Recall that Assumption H is satisfied with $\gamma = 1$ by deterministic weights $h_t = g(t/T)$ where g is a continuous piecewise differentiable function with a bounded derivative. Then assumption (2.11) on bandwidth H becomes

$$T^{1/2} << H << T^{3/4}. \tag{2.14}$$

Test statistic $TS(\widehat{\boldsymbol{u}})$ for ARCH effects in u_t is based on regression (2.9) on squares u_t^2 and requires 6 finite moment of u_t, see Assumption M. The same test based on the regression

$$|u_t| = \beta_0 + \beta_1|u_{t-1}| + ... + \beta_p|u_{t-p}| + \eta_t \tag{2.15}$$

allows to reduce the number of moments to $E|u_t|^\theta < \infty$, $\theta > 3$. The following theorem shows that the results of Theorem 2.1 remain valid with obvious corrections: $TS(\boldsymbol{u})$ and $TS(\widehat{\boldsymbol{u}})$

[5]The existence of $\boldsymbol{\Gamma}_p^{-1}$ follows from Lemma 3.1(i) in Dalla, Giraitis, and Phillips (2020) because the stationary sequence $\{u_t^2\}$ has spectral density. The latter follows from the absolute sumability of the covariance function γ_k, see (6.52).

are computed using $|u_t|$ and $|\widehat{u}_t|$ instead of u_t^2 and $\widehat{u}_t^2$, and in (2.12) σ_p, $\boldsymbol{\beta}_p$, $\boldsymbol{\Gamma}_p$ are defined setting

$$\gamma_k = \mathrm{cov}(|u_k|, |u_0|), \quad \sigma_p^2 = \mathrm{var}(|u_{p+1}| - \beta_1|u_p| - ... - \beta_p|u_1|). \tag{2.16}$$

THEOREM 2.2. *The results of Theorem 2.2 remain valid for test statistics $TS(\boldsymbol{u})$ and $TS(\widehat{\boldsymbol{u}})$ obtained from regression (2.15) on $|u_t|$ with σ_p, $\boldsymbol{\beta}_p$, $\boldsymbol{\Gamma}_p$ in (2.12) defined using (2.16).*
In Assumption M, it suffices to assume $E|u_t|^\theta < \infty$, $\theta > 3$ instead of $\theta > 6$.

Similarly, testing for ARCH effects based on regression

$$u_t = \beta_0 + \beta_1 u_{t-1} + ... + \beta_p u_{t-p} + \eta_t \tag{2.17}$$

results in testing for the presence of correlation of u_t at lags $1, ..., p$. In such case, the statistics $TS(u)$ and $TS(\widehat{u})$ computed using u_t and $\widehat{u}_t$ instead of u_t^2 and $\widehat{u}_t^2$ satisfy the results of Theorem 2.1 with σ_p, $\boldsymbol{\beta}_p$, $\boldsymbol{\Gamma}_p$ in (2.12) defined correspondingly setting

$$\gamma_k = \mathrm{cov}(u_k, u_0), \quad \sigma_p^2 = \mathrm{var}(u_{p+1} - \beta_1 u_p - ... - \beta_p u_1). \tag{2.18}$$

THEOREM 2.3. *The results of Theorem 2.2 remain valid for test statistics $TS(\boldsymbol{u})$ and $TS(\widehat{\boldsymbol{u}})$ obtained from regression (2.17) on u_t with σ_p, $\boldsymbol{\beta}_p$, $\boldsymbol{\Gamma}_p$ in (2.12) defined using (2.18).*
In Assumption M, it suffices to assume $E|u_t|^\theta < \infty$, $\theta > 3$ instead of $\theta > 6$.

Next we show that we can can test for absence of correlation in u_t using the correlogram of $\widehat{u}_t$. This is an important step in data analysis where we do not know in advance whether the series $y_t = h_t u_t$ we observe is a sequence of uncorrelated random variables. In particular, before proceeding to testing for ARCH effects in u_t, we may wish first to test for absence of correlation in u_t. Next theorem shows that such testing can be based on $\widehat{u}_t$.

Denote for $k = 0, 1, ...$

$$\begin{aligned} \widehat{r}_{\widehat{u},k} &= T^{-1} \textstyle\sum_{t=k+1}^T (\widehat{u}_t - \overline{\widehat{u}})(\widehat{u}_{t-k} - \overline{\widehat{u}}), \\ \widetilde{r}_{u,k} &= T^{-1} \textstyle\sum_{t=k+1}^T (u_t - Eu_t)(u_{t-k} - Eu_{t-k}). \end{aligned} \tag{2.19}$$

THEOREM 2.4. *Suppose that assumptions of Theorem 2.1 are satisfied. Then*

$$\widehat{r}_{\widehat{u},k} = \widetilde{r}_{u,k} + o_P(1) = \mathrm{cov}(u_k, u_0) + o_P(1), \quad k \geq 0. \tag{2.20}$$

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

$$T^{1/2}\widehat{r}_{\widehat{u},k} = T^{1/2}\widetilde{r}_{u,k} + o_P(1) \to \mathcal{N}(0, (Eu_1^2)^2) \quad k \geq 1. \tag{2.21}$$

Moreover, in Assumption M it suffices to assume $E|u_t|^\theta < \infty$, $\theta > 3$ instead of $\theta > 6$.

Denote $\widehat{\rho}_{\widehat{u},k} = \widehat{r}_{\widehat{u},k}/\widehat{r}_{\widehat{u},0}$, $k = 0, 1, 2, ...$ If $\{u_t\}$ is an i.i.d. sequence then (2.21) of Theorem 2.4 implies that for any $m = 1, 2, ...$

$$T^{1/2}(\widehat{\rho}_{\widehat{u},1}, ..., \widehat{\rho}_{\widehat{u},m}) \to_D \mathcal{N}(0, I_m). \tag{2.22}$$

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

Similarly, it can be show that approximation $|\widehat{u}_t| \sim |u_t|$ and $\widehat{u}_t^2 \sim u_t^2$ allows to use $|\widehat{u}_t|$ and $\widehat{u}_t^2$ to test for absence of correlation in $|u_t|$ and u_t^2 as if $|u_t|$ and u_t^2 were observed. Testing for correlation in $|u_t|$ requires $E|u_t|^\theta < \infty$, $\theta > 3$ while testing for correlation in u_t^2 requires $E|u_t|^\theta < \infty$, $\theta > 6$.

Notice that no ARCH effects in u_t means no correlation in u_t and u_t^2. This is a slightly weaker assumption than the assumption of the i.i.d. property of u_t. The later lead to standard approximations (2.13) and (2.22), which are not guaranteed for non i.i.d. noises u_t.

3 Monte Carlo Analysis

In this section we present Monte Carlo findings on the finite sample performance of the test for the presence of stationary volatility (ARCH effects) in artificially generated uncorrelated data, given by

$$y_t = h_t u_t, \quad u_t = \sigma_t \varepsilon_t, \quad t = 1, ..., T. \tag{3.1}$$

We evaluate the rejection frequency (in %) of the test statistics $TS(\widehat{u})$ for the presence of a stationary non-constant volatility factor σ_t. We use sample sizes $T = 200, 400, 800, 1600$ and report testing results for the lag order, $p = 5$. Testing results for lags $p = 1, 10$ suggest similar patterns to those discussed below, and are available upon request. Each experiment involves 5000 replications. After estimating the persistent scale factor h_t of volatility, the test is applied on squares $\widehat{u}_t^2 = y_t^2/\widehat{h}_t^2$, y_t^2 and for comparison on u_t^2. We expect similar empirical size and power properties, when testing is based on either $\widehat{u}_t^2$ or unobserved u_t^2, while the presence of h_t in $y_t = h_t \varepsilon_t$ testing based on y_t^2 leads to detection of spurious ARCH effects. We report results for four kernel estimates $\widehat{h}_t^2$ computed with bandwidth parameters $H = T^{0.5}, T^{0.6}, T^{0.7}, T^{0.8}$.

Table 5.1 reports the size and power of the test for the ARCH model $y_t = u_t = \sigma_t \varepsilon_t$ where the persistent component h_t is absent. This model allows to verify whether estimation of h_t introduces any distortions to the empirical size and power of the standard test for ARCH effects. For σ_t, ARCH(1) and GARCH(1,1) models are used:

$$\sigma_t^2 = 1 + \beta u_{t-1}^2, \quad \beta = 0, 0.2, 0.4, \tag{3.2}$$

$$\sigma_t^2 = 1 + 0.2 u_{t-1}^2 + 0.7 \sigma_{t-1}^2. \tag{3.3}$$

Except for the case of no ARCH effects, $\beta = 0$, such y_t contains a stationary volatility component σ_t which should result in detection of ARCH effects in the data.

From the table, it is clear that for $\beta = 0$ testing for ARCH effects based on the residuals $\widehat{u}_t^2$ achieves the nominal size of $\alpha = 5\%$ as both T and H increase and H meets the requirement $T^{0.5} << H << T^{0.75}$ of (2.11), and power is good throughout.

Table 5.2 reports the size and power of the test for y_t as in (3.1) with a stochastic persistent factor h_t generated by a non-stationary ARFIMA$(0, d, 0)$, $d > 1$ process $I_t^{(d)}$ that is re-scaled and bounded away from zero:

$$h_t \;=\; T^{-(d-1/2)}|I_t^{(d)}| + 1, \quad d = 1.2, 1.4, 1.5, \tag{3.4}$$

and σ_t follows an ARCH(1) model (3.2) with $\beta = 0, 0.2, 0.4$. The case $\beta = 0$ corresponds to $y_t = h_t\varepsilon_t$ satisfying the null hypothesis of no ARCH effect, i.e. $\sigma_t = const$.

We assume that $\{h_t\}$ and $\{u_t\}$ are mutually independent. Such h_t satisfies Assumption H with $\gamma = d - 1/2 > 1/2$, see Example 2.1. Here the importance and novelty of our new testing procedure becomes clear. Specifically, we find that the standard ARCH test based on y_t^2, which *a priory* assumes a constant h_t, produces severe size distortions, under the null hypothesis on no ARCH effects ($\beta = 0$). On the contrary, testing using the residuals $\widehat{u}_t^2$ attains the nominal 5% size, as T increases and H satisfies the requirement $T^{0.5} << H << T^{1-(1/4\gamma)}$ for the range of H in (2.11). The test based on $\widehat{u}_t^2$ tracks well the power of the test based on u_t^2.

Table 5.3 reports testing results for y_t as in (3.1) with a deterministic persistent scale factor, given by

$$h_t \;=\; \sin(2\pi t/T) + 2. \tag{3.5}$$

σ_t follows either an ARCH(1) model (3.2) with $\beta = 0, 0.2, 0.4$ or a GARCH (1,1) model given by (3.3). As before, the case $\beta = 0$ corresponds to $y_t = h_t\varepsilon_t$ satisfying the null hypothesis of no ARCH effects which allows the evaluation of the size of the test. Such h_t satisfies Assumption H with $\gamma = 1$. This is the least favourable case for satisfactory performance of the standard $ARCH$ test. We find severe size distortions when such test is applied directly on y_t^2. The test based on $\widehat{u}_t^2$ approaches the nominal 5% size as long as H satisfies the range condition $T^{0.5} << H << T^{0.75}$ given in (2.11). Overall, the results of the test based on the residuals $\widehat{u}_t^2$ and y_t^2 are similar to those in Table 5.3.

The above simulations are a small part of a much more extensive set of simulations. These include considering tests for absence of ARCH effects based on untransformed residuals $\widehat{u}_t$ and absolute values of the residuals $|\widehat{u}_t|$, as well as considering a wide variety of deterministic and stochastic settings for the persistent factor h_t. All these results produce patterns that are similar to those presented above and are available upon request. They are also presented in detail in Chronopoulos (2021).

Some general conclusions follow from the simulation study. First we find that size and power of the test for ARCH effects are not distorted by the estimation of h_t, when h_t is constant. Second, the standard $ARCH$ test is severely affected by the presence of a persistent time varying scaling factor h_t; it may suffer severe size distortions and produce spurious power even for small sample size T. Overall, the new test on residuals $\widehat{u}_t^2$, produces adequate nominal size and power throughout our experiments, illustrating the usefulness and novelty of the approach.

4 Empirical Analysis

In this section we illustrate how the proposed testing methodology in Section 2 can be applied to real data, to explore the presence or absence of stationary volatility component σ_t, besides the persistent component, h_t, in data y_t presented as in (3.1). We examine the difference in proportions (in %) of S&P 500 stock returns that test positive for $ARCH$ effects when: i) the random scale factor h_t is not assumed constant, estimated by a kernel estimate $\widehat{h}_t$, and a scale robust series $\widehat{u}_t = \widehat{h}_t^{-1} y_t$ is used in testing, and ii) h_t is assumed constant (equal to 1), i.e. the implicit assumption in regular $ARCH$ testing. Further, we also explore one-step-ahead volatility forecasting of $\mathrm{var}(y_t|\mathcal{F}_{t-1}) = h_t^2\sigma_t^2$ in order to assess how, forecasting methods/models that that allow for the persistent non-stationary scaling factor h_t, compare to commonly used stationary alternatives for σ_t^2.

Our data is a panel of $T = 1434$ weekly observations on $N = 505$ S&P 500 stock prices over the period 08/Jan/1993-27/Dec/2019[6]. Following common practice we trim non–working days, keep stocks that do not have missing values, and convert stock prices into returns using log–differencing. After these operations, we retain $T = 1340$ observations of $N = 254$ stocks. Further, we consider fitting a conditional mean model for stock returns, and for that model we also require data on the *S&P* index[6], data on the three Fama and French (1993) factors and the risk free rate, available in French's website.[7]

4.1 Testing for $ARCH$ effect in empirical data

Following the literature, we assume that stock returns can be specified by the model

$$y_{it} = \mu_{it} + h_{it}u_{it}, \quad u_{it} = \sigma_{it}\varepsilon, \quad \mu_{it} = E[y_{it}|\mathcal{F}_{t-1}], \quad \mathrm{var}(y_{it}|\mathcal{F}_{t-1}) = h_{it}^2\sigma_{it}^2. \tag{4.1}$$

Our aim is to discriminate between persistent, h_{it}, and persistent and stationary volatility, $h_{it}\sigma_{it}$, by testing stock returns $y_{it} - \mu_{it}$ for $ARCH$ effects. Following testing setup in Section 2, for $p = 1, 5, 10$, we consider $TS\,(u_i)$, the test statistic based on u_{it}^2 and $TS\,(\widehat{u}_{it})$ based on the squares $\widehat{u}_{it}^2$ of the rescaled residuals, $\widehat{u}_{it} = \widehat{h}_{it}^{-1}(y_{it} - \widehat{\mu}_{it})$, for every stock i, where $\widehat{\mu}_{it}$ is an estimate of the conditional mean, and $\widehat{h}_{it}$ is the kernel estimate defined in (2.5), for a range of pre-selected bandwidths H. We consider the full dataset, from 28/Oct/1994 to 27/Dec/2019, and the following three subsamples: i) 28/Oct/1994 – 28/Dec/2007, ii) 7/Jan/2000 - 30/Dec/2011 and iii) 7/Jan/2011–27/Dec/2019. We use these subsamples to examine how the exclusion of highly volatility periods, like the 2008 financial crisis can affect

[6]Data are obtained from *Bloomberg*

[7]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

our testing results. Further, we examine how potential misspecification of the conditional mean, $\mu_{it} = E(y_{it}|F_{t-1}) \neq 0$ and the inclusion of the estimate $\widehat{\mu}_{it}$ of conditional mean build on covariates/factors that are known to proxy well for sensitivity to common risk factors and capture strong common variation in stock returns, affect our testing for ARCH effects results. The models we consider to estimate the conditional mean μ_t are the market factor model and the three factor Fama and French (1993) model[8]. We report percentage of stocks with ARCH effects across $N = 254$ stocks for the baseline model for conditional mean, $\mu_t = 0$ and the two estimates of the conditional mean, see footnote 7.

Our empirical testing results are presented in Table 5.4. We report the proportion of stocks with ARCH effects across the i stocks, for different bandwidths H, lags p and subsamples, for both the baseline model, $\mu_t = 0$, where we do not estimate the conditional mean μ_t and set it equal to 0, and where we use the market factor model and Fama-French model to estimate it. Below, we provide a discussion on testing in subsamples and an overall comment, for the baseline model, $\mu_t = 0$, and the three factor Fama French model for μ_t.

The 2007 (28/Oct/1994 – 28/Dec/2007) subsample was generally a moderate period, with the small outlier being the *dotcom* bubble. During this period we find that, in the case where we do not estimate the conditional mean, i.e. $\mu_t = 0$, the the proportion of stock returns with ARCH effects, detected in the rescaled squared residuals $\widehat{u}_t^2$ across stocks, is considerably smaller than in the squared errors $(y_t - \mu_t)^2$, i.e. a drop from 61.42% to 11.02% $-$ 38.19 for different H's.

The 2011–2019 (7/Jan/2011–27/Dec/2019) subsample, is the most moderate period we consider. During this period, we still find a drop in proportion of stocks with ARCH effects between the two tests, corresponding to the base line model, $\mu_t = 0$, and Fama French factor model for μ_t, from 33.86% to 1.57% $-$ 18.50% across H. It is evident from comparisons with other periods, that the proportion of stocks with $ARCH$ effects when we do not estimate the conditional mean is considerably smaller.

The 2000 -2011 (7/Jan/2000 - 30/Dec/2011) subsample, is more turbulent since it contains the 2008 financial crisis. During this period, we find that for the *baseline* model, $\mu_t = 0$, the majority (83.47%) of the stocks test positive for $ARCH$ effects in squares y_{it}^2. Using the test on the rescaled residuals, we again, find that the proportion of stocks with ARCH effects in squares $\widehat{u}_{it}^2$ falls from 83.47% to 18.11% $-$ 74.02 for different values of H. Similar drops in the proportion of returns with ARCH effects in squares $\widehat{u}_{it}^2$ is observed for the case where a market factor or the Fama French factors are used to model the conditional mean.

The sample starting on 28/Oct/1994 and ending 27/Dec/2019, is the longest and most turbulent period we have considered. We abstract from the impact of the CoViD-19 pan-

[8]The model with a conditional mean, can be written as $y_t = \mu_t + h_t u_t$ with $\mu_t = r_{f,t} + \beta_1(R_{m,t} - r_{f,t})$ for the market factor model, and $\mu_t = r_{f,t} + \beta_1(R_{m,t} - r_{f,t}) + \beta_2 SMB_t + \beta_3 HML_t + h_t\epsilon_t$, for the three factor Fama French model. $R_{m,t}$ is the market factor, $r_{f,t}$ is the risk free rate and SMB, HML are the firm size (small-minus-big) and book-to-market (high-minus-low) factors described in Fama and French (1993). h_t can be estimated using the errors $\widehat{y}_t = y_t - \mu_t$ of the two models, given estimates on $\beta_i, i = 1, 2, 3$.

demic, since it is still not clear how it will affect the economy. During this period we find for the *baseline* model, $\mu_t = 0$, almost all (93.31%) of the stocks test positive for $ARCH$ effects in y_{it}^2, and similarly as above, the proportion stock returns with ARCH effects in the squares $\widehat{u}_t^2$ of rescaled residuals drops to $25.98\% - 84.65\%$ for different H values. Similar pattern is observed for two our models for the conditional mean.

Across lags and subsamples, we find that the proportion of returns with ARCH effects generally falls when testing the squares of the rescaled residuals and increases for the non-rescaled ones. Further, accounting for conditional mean via the inclusion of covariates leads to a further decrease of this proportion across sub-samples and lags, which is not an unexpected result, since misspecification of the conditional mean, is known to affect $ARCH$ testing resulting is spurious ARCH effects. Overall, these empirical results are in line with our Monte Carlo experiments. We treat these findings as evidence that stationary volatility factor is overall considerably less pronounced in the data, than previously thought and that further, the second moments of asset returns change slowly and persistently over time.

4.2 Volatility Forecasting

4.2.1 Forecasting Setup

Our in-sample empirical results suggest that stationary volatility is considerably less pronounced in real data than what was previously thought. Specifically, we find that non-stationary processes seem to co-drive the overall latent volatility. This can naturally have implication for both the estimation and forecasting of volatility. In this section we explore how this finding affects volatility forecasting, and we leave the case of estimation for future work.

We assume that stock returns can be specified by the model

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

where h_t is a persistent scale factor that satisfies Assumption H and u_t is a stationary sequence of uncorrelated random variables with appropriate moments. We re-write u_t as

$$u_t = \sigma_t \varepsilon_t \tag{4.2}$$

to make explicit that σ_t^2 follows some general stationary volatility process. We proceed to produce volatility forecasts for a variety of model choices for σ_t^2. Specifically, we produce 1-step ahead forecasts of $v_t^2 = \mathbb{E}_t(y_{t+1}^2)$ via a recursive (expanding window) $y_1, ..., y_t$ and compare them with a volatility proxy r_t^2, discussed below.

In our forecasting exercise we consider three approaches to estimate and subsequently forecast volatility. First, we set $y_t = \sigma_t \varepsilon_t$, $h_t = 1$, and fit the following commonly used stationary models in the literature for σ_t^2. The models we consider are: $ARCH(1)$, $GARCH(1,1)$, the Glosten, Jagannathan, and Runkle (1993) $GARCH$, $GJR(1,1,1)$, and the random walk stochastic volatility ($RW - SV$) of Ruiz (1994)[9]. Next, we set $y_t = h_t \varepsilon_t$, $\sigma_t^2 = 1$, and use the estimator $\widehat{h}_t$ described in (2.5) to model volatility h_t^2 non-parametrically. Because it can be the case that both types of volatility can co-drive the overall latent volatility, we further use forecast combinations where at a first step we rescale the squared stock returns $y_s^2/\widehat{h}_s^2$, $1 \leq s \leq t$ using the non-parametric estimate of volatility, $\widehat{h}_s$, and subsequently fit to them a parametric stationary model σ_s^2 and produce forecasts $\widehat{\sigma}_t^2$ for σ_t^2. Finally we use the latest value of the non-parametric estimate of volatility, $\widehat{h}_t^2$, and $\widehat{\sigma}_t^2$ to produce the overall volatility forecasts, $\widehat{v}_t^2 = \widehat{h}_t^2 \widehat{\sigma}_t^2$, $\widehat{v}_t^2 = \widehat{\sigma}_t^2$ and $\widehat{v}_t^2 = \widehat{h}_t^2$ in the third, first and the second case.

Generally, we use the $ARCH(1)$ model for performance check, since it is well known that it does not give good forecasts especially for large T. Hence we also refrain from doing any forecast combination with it, while we employ combinations of $\widehat{h}_t^2$ with the rest of the models. We use the subsamples $y_1, ..., y_t$, $t \in [0.2T,\, T-1]$ to evaluate forecasting performance. Before we present our results there are two important matters worthy of discussion.

First, because volatility is unobserved, the choice of a well-behaved proxy, to compare with our forecast, is pertinent. Following Patton (2011), we use the squared demeaned returns because they have been shown to be an unbiased proxy for volatility. We acknowledge that while this is an unbiased proxy, it is noisy. This requires attention when the exercise is about ranking different models in terms of their performance. Because the results are subject to the proxy, this boils down in the choice of a "robust" loss function, that can be used in common tests, e.g. Diebold and Mariano (1995) or West (1996). Patton (2011) has illustrated that the MSE loss function (among others), is reasonably robust. We use the Diebold and Mariano (1995) test to rank the forecasting performance of the different models considered.

Second, the estimation period of the initial volatility model parameters, i.e. the origin $0.2T$ where the recursive forecasting exercise starts can potentially be important in the performance of the forecasting exercise, especially for the stationary models due to parameter instability. There are arguments both for and against the importance of parameter instability in forecasting models. Clements and Hendry (1996, 2000) among others, argue that ignoring parameter instability is one of the main sources of forecast breakdowns, if not modelled explicitly. On the other hand, Stock and Watson (1996) argue that there is very little benefit in modelling this and Kim and Swanson (2014) argue that forecasts estimated by recursive estimation performed as well or better than rolling windows forecasts, for a range of models estimated from a large panel of macroeconomic time series.

[9] For the *GJR-GARCH(1,1,1)* which is an extension to $GARCH(1,1)$, $\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \beta_2 I(u_{t-1} < 0) u_{t-1}^2 + \beta_3 u_{t-1}^2$. For the RW-SV model, $\sigma_t^2 = \exp(\kappa_t/2)$, where $\kappa_t = \omega + \kappa_{t-1} + \eta_t$, $\eta_t \sim N(0,1)$ and $\omega \in \mathbb{R}$.

4.2.2 Results

In Table 5.5 we present the results from our main forecasting exercise where we use the last decade of data as our sample, from 04/Jan/2008 to 27/Dec/2019. The table evaluates forecasting performance of $ARCH(1)$, $GARCH(1)$, $GJR(1,1,1)$, and $RW-SW$ models, kernel estimator $\widehat{h}_t^2$ (k-method) and some forecast combinations. These are denoted as $k-GARCH(1,1)$, $k-GJR(1,1,1)$ and $k-RWSV$, where we omit the order for the $ARCH$ type models to save space in the table. We consider four values of H. In table position (i,j) we report the proportion of stocks, where according to Diebold and Mariano (1995) test, model i is significantly better at forecasting volatility than model j[10].

Every j-th row relates to a particular volatility forecasting method j and elements of the row report the proportion of stocks for which the method j significantly outperforms the other methods according to the DM test. Similarly every j-th column reports the proportion of stocks for which method j is significantly outperformed by the other methods according to the DM test. Therefore, row and column averages are metrics for the forecasting performance of the methods. Well performing methods have small column and large row averages and vice versa.

The results from the forecasting exercise suggest that forecasts produced using the kernel estimator outperform forecasts obtained from the other stationary model we considered. Furthermore, using the kernel estimator to produce forecast combinations, in a two step fashion, ameliorates the forecasting performance of stationary models. We can see that across all models, $k-GARCH$ produces the best forecasts, then comes kernel and $k-GJR$ that are followed by stationary alternatives.

5 Conclusion

Volatility modelling is at the core of financial econometrics where the majority of the literature, so far, focus on models that are stationary. Time variation in the conditional variance comes from the use of either $ARCH$ or SV types of processes. However, parameter estimates near the boundary of stationarity suggest that $ARCH/SV$ types of models cannot easily accommodate the observed persistent variation in volatility estimates of macro/finance datasets.

In this paper we contribute to the literature in three ways. Firstly, we provide a possible setup for persistent processes, that can provide a superior approximation of the volatility process. Secondly, we discuss how persistence allows the consistent estimation of unobserved volatility, without strong parametric assumptions, and finally we provide a novel testing

[10] Specifically, if the Diebold and Mariano (2002) test statistic is below -1.96, we reject the null hypothesis of equal forecasting performance of forecasting models at 5% significance level in favour of the alternative that model i produces superior forecasts compared to model j. If test statistic above 1.96 we reject the null in favour of the alternative that model j is significantly better at forecasting volatility than model i.

strategy that enable standard $ARCH$ tests to separate persistent from stationary volatility and discuss the conditions needed for this.

We provide Monte Carlo evidence, illustrating that the existence of a persistent scale factor affects adversely standard ARCH tests, but not our newly proposed test.

We use our new testing scheme on U.S. stock returns and find, in line with our Monte Carlo results, extensive support for the persistent volatility paradigm, suggesting that the role of stationary time–varying volatility is not as outstanding in the data, as was previously thought and that further, the second moment of asset returns varies slowly over time. Further, our forecasting exercise illustrates that this strategy has merit also in volatility forecasting, either on its own or via forecasting combinations.

Concluding, our results, suggest a new avenue for testing for the presence of ARCH effects in practice. Without using our new testing scheme, a typical $ARCH$ test is applied to some non-rescaled residuals, and then a researcher will proceed by fitting an $ARCH(p)$ or $GARCH$ model, given a rejection of the null hypothesis. Using our proposed testing strategy, the researcher needs to further run a second $ARCH$ test on rescaled residuals. If the new test cannot reject the null, then the rejection from the first test is spurious, since it is the existence of a persistent scale factor h_t that drives the volatility and not some stationary process. In the case that the second test also rejects the null hypothesis of no ARCH effects, then we need to inspect the kernel estimate for the presence of persistent deterministic or random scale factor. If this appears constant across time, an $ARCH(p)$ or $GARCH$ model should be fitted, whereas if it varies considerably in time, then a two-step estimation is required. There are a number of avenues for future work, in particular, developing a stability test for the persistent volatility component h_t, similar to the work of Chen and Hong (2012).

$p = 5$						
T	H	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$Garch$
200	$T^{0.5}$	$\hat{u}_t$	12.98	19.04	48.82	7.22
	$T^{0.6}$		4.90	23.30	60.82	28.66
	$T^{0.7}$		3.72	30.14	69.60	52.42
	$T^{0.8}$		3.60	34.64	74.32	65.52
		u_t	4.34	40.30	78.94	77.12
		y_t	4.34	40.30	78.94	77.12
400	$T^{0.5}$	$\hat{u}_t$	13.18	42.80	89.26	33.86
	$T^{0.6}$		4.80	52.00	94.12	76.14
	$T^{0.7}$		3.62	60.04	96.08	90.56
	$T^{0.8}$		3.80	64.70	97.14	94.94
		u_t	4.68	69.06	97.60	97.22
		y_t	4.68	69.06	97.60	97.22
800	$T^{0.5}$	$\hat{u}_t$	15.32	80.58	99.92	89.00
	$T^{0.6}$		5.60	87.42	100.00	99.38
	$T^{0.7}$		4.38	91.38	100.00	99.88
	$T^{0.8}$		4.18	92.96	100.00	99.92
		u_t	4.74	94.14	100.00	99.98
		y_t	4.74	94.14	100.00	99.98
1600	$T^{0.5}$	$\hat{u}_t$	16.44	99.04	100.00	99.96
	$T^{0.6}$		6.20	99.72	100.00	100.00
	$T^{0.7}$		5.08	99.82	100.00	100.00
	$T^{0.8}$		5.16	99.84	100.00	100.00
		u_t	5.44	99.86	100.00	100.00
		y_t	5.44	99.86	100.00	100.00

Table 5.1: Test for ARCH effects. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). Model: $y_t = \sigma_t \varepsilon_t$ where $\sigma_t^2 = 1 + \beta u_{t-1}^2$; under $GARCH$ $\sigma_t^2 = 1 + 0.2u_{t-1}^2 + 0.7\sigma_{t-1}^2$; $\varepsilon_t \sim i.i.d.\mathcal{N}(0, 1)$.

$p = 5$											
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	$T^{0.5}$	$\hat{u}_t$	11.20	11.80	12.14	19.12	19.48	19.38	49.26	49.00	49.14
	$T^{0.6}$		3.90	4.06	4.30	23.40	23.30	23.36	62.06	61.40	61.52
	$T^{0.7}$		3.92	3.82	3.68	31.40	30.00	29.70	71.22	70.86	70.70
	$T^{0.8}$		6.32	5.04	4.78	38.68	37.14	36.54	77.26	76.54	76.38
		u_t	4.52	4.52	4.52	39.76	39.76	39.76	79.92	79.92	79.92
		y_t	22.58	20.24	19.36	56.08	54.18	53.76	84.12	83.46	83.34
400	$T^{0.5}$	$\hat{u}_t$	12.24	13.02	13.10	43.94	43.90	43.86	89.96	90.02	89.88
	$T^{0.6}$		4.08	4.58	4.68	54.08	53.34	53.50	94.64	94.54	94.42
	$T^{0.7}$		4.52	4.20	4.04	63.70	62.56	62.22	96.80	96.60	96.56
	$T^{0.8}$		8.34	6.62	6.00	71.02	69.42	68.80	97.92	97.70	97.68
		u_t	5.12	5.12	5.12	70.10	70.10	70.10	97.84	97.84	97.84
		y_t	36.40	32.48	30.64	83.80	81.96	81.70	98.78	98.68	98.52
800	$T^{0.5}$	$\hat{u}_t$	14.22	15.06	15.30	81.62	81.34	81.34	99.92	99.92	99.92
	$T^{0.6}$		4.86	5.22	5.38	88.68	88.30	88.22	99.98	99.98	99.98
	$T^{0.7}$		4.66	4.32	4.52	92.58	92.10	91.94	99.98	99.98	99.98
	$T^{0.8}$		9.46	6.28	5.74	95.22	94.10	93.82	99.98	99.98	99.98
		u_t	5.34	5.34	5.34	94.46	94.46	94.46	99.98	99.98	99.98
		y_t	51.20	45.04	43.28	98.02	97.68	97.46	99.98	99.98	99.98
1600	$T^{0.5}$	$\hat{u}_t$	15.44	15.94	16.10	99.28	99.26	99.24	100.00	100.00	100.00
	$T^{0.6}$		5.32	5.64	5.92	99.72	99.68	99.66	100.00	100.00	100.00
	$T^{0.7}$		5.76	5.14	5.06	99.86	99.84	99.80	100.00	100.00	100.00
	$T^{0.8}$		11.98	7.16	6.22	99.92	99.90	99.88	100.00	100.00	100.00
		u_t	5.28	5.28	5.28	99.88	99.88	99.88	100.00	100.00	100.00
		y_t	65.84	57.46	55.22	99.98	99.96	99.96	100.00	100.00	100.00

Table 5.2: Test for ARCH effects. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). Model: $y_t = h_t u_t$, $h_t = |I_t^{(d)}|/T^{d-1/2}+1$, $I_t^{(d)} \sim ARFIMA(0, d, 0)$, $u_t = \sigma_t \varepsilon_t$ where $\sigma_t^2 = 1 + \beta u_{t-1}^2$; $\varepsilon_t \sim i.i.d.\mathcal{N}(0, 1)$.

$p = 5$						
T	H	data	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$Garch$
200	$T^{0.5}$	$\hat{u}_t$	10.50	17.96	49.14	8.02
	$T^{0.6}$		3.20	24.42	62.84	34.96
	$T^{0.7}$		7.64	40.68	74.92	66.44
	$T^{0.8}$		25.10	60.12	84.08	83.34
		u_t	4.34	40.30	78.96	77.12
		y_t	73.66	86.78	93.62	95.22
400	$T^{0.5}$	$\hat{u}_t$	12.10	42.62	89.30	35.12
	$T^{0.6}$		3.58	53.36	94.34	79.00
	$T^{0.7}$		7.90	68.78	96.86	95.10
	$T^{0.8}$		40.86	86.58	98.64	98.60
		u_t	4.68	69.04	97.60	97.22
		y_t	95.40	99.14	99.80	99.92
800	$T^{0.5}$	$\hat{u}_t$	14.42	80.62	99.92	89.58
	$T^{0.6}$		4.74	88.16	100.00	99.52
	$T^{0.7}$		7.78	93.88	100.00	99.96
	$T^{0.8}$		60.80	98.98	100.00	100.00
		u_t	4.74	94.14	100.00	99.98
		y_t	99.98	100.00	100.00	100.00
1600	$T^{0.5}$	$\hat{u}_t$	15.96	99.08	100.00	99.96
	$T^{0.6}$		5.68	99.72	100.00	100.00
	$T^{0.7}$		7.06	99.84	100.00	100.00
	$T^{0.8}$		80.06	100.00	100.00	100.00
		u_t	5.44	99.86	100.00	100.00
		y_t	100.00	100.00	100.00	100.00

Table 5.3: Test for ARCH effects. Rejection frequencies (in %) at the 5% significance level ($\beta = 0$ size, $\beta > 0$ power). Model: $y_t = h_t u_t$, $h_t = \sin(2\pi t/T) + 2$, $u_t = \sigma_t \varepsilon_t$ where $\sigma_t^2 = 1 + \beta u_{t-1}^2$; under $GARCH$ $\sigma_t^2 = 1 + 0.2 u_{t-1}^2 + 0.7 \sigma_{t-1}^2$; $\varepsilon_t \sim i.i.d.\mathcal{N}(0, 1)$.

Testing for ARCH Effects Results

No conditional mean

H	data	$p=1$				$p=5$				$p=10$			
		2007	2000-2011	2011-2019	2019	2007	2000-2011	2011-2019	2019	2007	2000-2011	2011-2019	2019
$T^{0.5}$	$\hat{u}_t$	11.02	18.11	1.57	25.98	6.30	6.69	1.57	13.39	5.91	5.12	3.54	11.02
$T^{0.55}$		16.54	28.35	3.54	42.52	8.27	14.96	2.36	36.61	7.09	11.42	2.36	30.32
$T^{0.6}$		18.90	37.80	7.48	61.02	14.17	29.92	4.72	64.57	12.21	24.41	5.51	59.45
$T^{0.7}$		29.92	66.54	16.54	81.50	31.89	72.05	18.90	87.80	31.50	70.47	18.90	90.16
$T^{0.75}$		38.19	74.02	18.50	84.65	46.85	81.89	27.17	91.73	46.85	83.07	29.92	94.49
	y_t	61.42	83.47	33.86	93.31	79.13	90.16	55.51	96.46	79.53	92.13	56.30	98.43

Market factor model for conditional mean

H	data	$p=1$				$p=5$				$p=10$			
		2007	2000-2011	2011-2019	2019	2007	2000-2011	2011-2019	2019	2007	2000-2011	2011-2019	2019
$T^{0.5}$	$\hat{u}_t$	6.69	6.30	2.36	12.21	2.36	4.33	3.54	7.09	5.12	4.33	7.09	6.69
$T^{0.55}$		10.24	13.39	3.15	25.20	5.51	7.09	2.76	15.75	3.15	4.33	1.97	9.84
$T^{0.6}$		13.39	20.47	4.33	37.01	9.45	12.21	3.94	33.07	6.69	12.21	1.57	30.71
$T^{0.7}$		23.23	37.40	7.48	56.69	22.05	42.13	7.87	61.02	22.84	45.28	6.69	67.72
$T^{0.75}$		31.50	48.82	7.48	66.54	35.83	58.27	9.84	73.62	37.80	62.99	8.27	79.92
	y_t	60.63	74.02	21.65	84.25	77.95	83.07	19.69	93.70	79.13	88.98	22.84	94.09

Fama French model for conditional mean

H	data	$p=1$				$p=5$				$p=10$			
		2007	2000-2011	2011-2019	2019	2007	2000-2011	2011-2019	2019	2007	2000-2011	2011-2019	2019
$T^{0.5}$	$\hat{u}_t$	9.45	16.14	0.79	23.23	5.12	5.51	3.15	11.02	5.91	4.33	2.36	9.84
$T^{0.55}$		14.57	26.77	3.15	38.58	7.87	13.39	1.97	33.86	5.12	9.45	1.18	29.92
$T^{0.6}$		18.50	36.61	6.69	57.87	10.63	29.53	3.94	62.21	9.06	22.84	4.72	58.27
$T^{0.7}$		27.17	62.60	13.39	79.53	29.92	68.11	15.75	85.83	28.74	68.50	16.54	88.58
$T^{0.75}$		35.04	72.05	16.54	82.68	43.70	78.35	24.02	90.16	44.09	79.53	24.02	93.70
	y_t	59.45	80.71	31.89	92.91	79.53	88.98	49.21	96.06	79.13	92.13	52.76	98.82

Table 5.4: Proportion of stock returns with ARCH effects in squares across the i stocks, for different bandwidths H, lags p and subsamples defined in the main text. Testing at the 5% significance level. Presented results are for baseline model, $\mu_t = 0$, and for the cases where the conditional mean μ_t is estimated using the Market factor and Fama-French models.

Method	H	$ARCH$	$GARCH$	$k-GARCH$	$kernel$	GJR	$k-GJR$	SV	$k-SV$	$AVERAGE\,(row)$
$ARCH$	$T^{0.50}$	-	0.04	0.00	0.01	0.05	0.02	0.01	0.05	0.02
	$T^{0.55}$			0.00	0.01		0.01		0.02	0.02
	$T^{0.60}$			0.00	0.02		0.02		0.02	0.02
	$T^{0.65}$			0.00	0.03		0.02		0.02	0.02
	$T^{0.70}$			0.00	0.03		0.04		0.02	0.02
$GARCH$	$T^{0.50}$	0.54	-	0.02	0.04	0.09	0.02	0.10	0.14	0.12
	$T^{0.55}$			0.02	0.05		0.05		0.12	0.12
	$T^{0.60}$			0.02	0.07		0.07		0.11	0.13
	$T^{0.65}$			0.03	0.14		0.09		0.10	0.14
	$T^{0.70}$			0.04	0.22		0.15		0.10	0.15
$k-GARCH$	$T^{0.50}$	0.72	0.46	-	0.14	0.33	0.11	0.32	0.43	0.31
	$T^{0.55}$	0.71	0.47		0.19	0.33	0.15	0.34	0.34	0.32
	$T^{0.60}$	0.69	0.43		0.27	0.31	0.19	0.30	0.29	0.31
	$T^{0.65}$	0.69	0.36		0.36	0.25	0.23	0.23	0.23	0.29
	$T^{0.70}$	0.63	0.31		0.50	0.23	0.28	0.18	0.18	0.29
$Kernel$	$T^{0.50}$	0.67	0.37	0.00	-	0.30	0.04	0.30	0.37	0.26
	$T^{0.55}$	0.63	0.39	0.00		0.28	0.07	0.29	0.27	0.24
	$T^{0.60}$	0.57	0.31	0.00		0.25	0.08	0.23	0.18	0.20
	$T^{0.65}$	0.48	0.22	0.00		0.18	0.07	0.13	0.10	0.15
	$T^{0.70}$	0.37	0.19	0.01		0.16	0.07	0.07	0.07	0.12
GJR	$T^{0.50}$	0.53	0.17	0.03	0.04	-	0.05	0.13	0.17	0.14
	$T^{0.55}$			0.03	0.05		0.08		0.14	0.14
	$T^{0.60}$			0.04	0.09		0.08		0.13	0.15
	$T^{0.65}$			0.04	0.15		0.13		0.14	0.16
	$T^{0.70}$			0.04	0.25		0.17		0.14	0.18
$k-GJR$	$T^{0.50}$	0.59	0.32	0.00	0.04	0.24	-	0.21	0.26	0.21
	$T^{0.55}$	0.57	0.30	0.00	0.06	0.25		0.20	0.20	0.20
	$T^{0.60}$	0.53	0.25	0.00	0.09	0.18		0.17	0.14	0.17
	$T^{0.65}$	0.43	0.20	0.00	0.13	0.15		0.09	0.09	0.14
	$T^{0.70}$	0.35	0.17	0.02	0.15	0.12		0.07	0.06	0.12
SV	$T^{0.50}$	0.46	0.21	0.03	0.05	0.17	0.03	-	0.23	0.15
	$T^{0.55}$			0.03	0.05		0.04		0.18	0.14
	$T^{0.60}$			0.04	0.07		0.06		0.17	0.15
	$T^{0.65}$			0.05	0.13		0.09		0.17	0.16
	$T^{0.70}$			0.07	0.24		0.13		0.13	0.18
$k-SV$	$T^{0.50}$	0.46	0.22	0.02	0.04	0.18	0.02	0.10	-	0.13
	$T^{0.55}$	0.50	0.22	0.02	0.05	0.17	0.05	0.15		0.15
	$T^{0.60}$	0.50	0.21	0.04	0.09	0.17	0.07	0.14		0.15
	$T^{0.65}$	0.47	0.22	0.05	0.17	0.16	0.09	0.13		0.16
	$T^{0.70}$	0.46	0.21	0.06	0.28	0.15	0.14	0.11		0.18
$AVERAGE$ $(column)$	$T^{0.50}$	0.50	0.22	0.01	0.05	0.17	0.04	0.15	0.21	
	$T^{0.55}$	0.49	0.22	0.01	0.06	0.17	0.06	0.15	0.16	
	$T^{0.60}$	0.48	0.20	0.02	0.09	0.15	0.07	0.14	0.13	
	$T^{0.65}$	0.45	0.18	0.02	0.14	0.13	0.09	0.10	0.11	
	$T^{0.70}$	0.42	0.16	0.03	0.21	0.12	0.12	0.08	0.09	

Table 5.5: In table position (i, j) we report the proportion of stock returns, where Diebold and Mariano (1995) test rejects at the 5% significance level the equal performance of forecasting models i and j in favour of the alternative that the model i is significantly better at forecasting volatility than model j. Row and column averages are metrics for the forecasting performance of the models. The best performing model is the one with the smallest column and highest row average.

References

Amado, C., and T. Teräsvirta (2013): "Modelling volatility by variance decomposition," *Journal of Econometrics*, 175(2), 142–153.

——— (2017): "Specification and testing of multiplicative time-varying GARCH models with applications," *Econometric Reviews*, 36(4), 421–446.

Bera, A. K., and M. L. Higgins (1993): "ARCH models: properties, estimation and testing," *Journal of Economic Surveys*, 7(4), 305–366.

Brownlees, C. T., and G. M. Gallo (2009): "Comparison of volatility measures: a risk management perspective," *Journal of Financial Econometrics*, 8(1), 29–56.

Chen, B., and Y. Hong (2012): "Testing for smooth structural changes in time series models via nonparametric regression," *Econometrica*, 80(3), 1157–1183.

Chronopoulos, I. (2021): "Essays in Econometrics," *PhD thesis, in preparation.*

Chronopoulos, I., G. Kapetanios, and K. Petrova (2019): "Kernel-based volatility generalised least squares," *Econometrics and Statistics.*

Clements, M. P., and D. F. Hendry (1996): "Intercept corrections and structural change," *Journal of Applied Econometrics*, 11(5), 475–494.

——— (2000): *Forecasting Non-stationary Economic Time Series.* MIT Press.

Dahlhaus, R. (2000): "A likelihood approximation for locally stationary processes," *The Annals of Statistics*, 28(6), 1762–1794.

Dahlhaus, R., and W. Polonik (2006): "Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes," *The Annals of Statistics*, 34(6), 2790–2824.

Dahlhaus, R., S. S. Rao, et al. (2006): "Statistical inference for time-varying ARCH processes," *The Annals of Statistics*, 34(3), 1075–1114.

Dalla, V., L. Giraitis, and P. Phillips (2020): "Robust tests for white noise and cross-correlation," *Econometric Theory, online.*

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

Davydov, Y. (1968): "Convergence of distributions generated by stationary stochastic processes," *Theory of Probability and Its Applications*, 13, 691–696.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2021): "Estimation of Time-Varying Covariance Matrices for Large Datasets," *Econometric Theory, online.*

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3).

Engle, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and J. G. Rangel (2008): "The spline-GARCH model for low-frequency volatility and its global macroeconomic causes," *The Review of Financial Studies*, 21(3), 1187–1222.

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Gallant, A. R. (1984): "The Fourier flexible form," *American Journal of Agricultural Economics*, 66(2), 204–208.

Giraitis, L., G. Kapetanios, A. Wetherilt, and F. Žikeš (2016): "Estimating the dynamics and persistence of financial networks, with an application to the sterling money market," *Journal of Applied Econometrics*, 31(1), 58–84.

Giraitis, L., G. Kapetanios, and T. Yates (2014): "Inference on stochastic time-varying coefficient models," *Journal of Econometrics*, 179, 46–65.

——— (2018): "Inference on multivariate heteroscedastic time varying random coefficient models," *Journal of Time Series Analysis*, 39(2), 129–149.

Giraitis, L., H. Koul, and D. Surgailis (2012): *Large Sample Inference for Long memory Processes*. Imperial College Press, London.

Glosten, L. R., 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*, 48(5), 1779–1801.

Kapetanios, G., and T. Yates (2008): "An analysis of inflation persistence and the Great Moderation using a model of deterministic structural change," *Mimeo, Bank of England.*

Kim, H. H., and N. R. Swanson (2014): "Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence," *Journal of Econometrics*, 178, 352–367.

Mazur, B., and M. Pipień (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(2), 95–116.

Mikosch, T., and C. Stărică (2004): "Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects," *The Review of Economics and Statistics*, 86(1), 378–390.

Patton, A. J. (2011): "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, 160(1), 246–256.

Priestley, M. B. (1965): "Evolutionary spectra and non-stationary processes," *Journal of the Royal Statistical Society, Series B*, 27, 204–237.

Ruiz, E. (1994): "Quasi-maximum likelihood estimation of stochastic volatility models," *Journal of Econometrics*, 63(1), 289–306.

Stock, J. H., and M. W. Watson (1996): "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics*, 14(1), 11–30.

Van Bellegem, S., and R. Von Sachs (2004): "Forecasting economic time series with unconditional time-varying variance," *International Journal of Forecasting*, 20, 611–627.

West, K. D. (1996): "Asymptotic inference about predictive ability," *Econometrica: Journal of the Econometric Society*, pp. 1067–1084.

6 Appendix

6.1 Proof of Theorems 2.1, 2.2 and 2.4.

In the proof of Theorem 2.1 we use the claim that h_t^2 satisfies Assumption H which is verified the next proposition.

PROPOSITION 6.1. *If h_t, $t = 1, ..., T$ satisfies Assumption H with parameters γ and α then h_t^2, $t = 1, ..., T$ satisfies Assumption H with parameters γ and $\alpha/2$.*

Proof of Proposition 6.1 By assumption, h_t satisfies (2.2),

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

and ξ_{tj}, h_t satisfies (2.3). Therefore

$$|h_t^2 - h_j^2| \le |h_t - h_j|\,|h_t + h_j| \le (|t-j|/T)^\gamma \xi^*_{tj}, \quad \xi^*_{tj} = \xi_{tj}(|h_t| + |h_j|).$$

Condition (2.3) implies that ξ^*_{tj} and h_t^2 satisfy (2.3) with parameter $\alpha/2$. $\square$

Proof of Theorem 2.1. In Theorem 2.1 we analyse the Wald version of test for H_0 hypothesis for absence of ARCH effects u_t. First we recall definitions of $S(\boldsymbol{u})$ and $S(\widehat{\boldsymbol{u}})$. Given data $\boldsymbol{u} = [u_1, u_2, \ldots, u_T]$, we define the test statistic for testing H_0 as follows:

$$S(\boldsymbol{u}) = \widetilde{\sigma}^{-2}\widetilde{\boldsymbol{\beta}}_p'(X'X)\widetilde{\boldsymbol{\beta}}_p, \quad \widetilde{\boldsymbol{\beta}}_p = (X'X)^{-1}X'Y, \quad \widetilde{\sigma}_p^2 = (Y - X\widetilde{\boldsymbol{\beta}}_p)'(Y - X\widetilde{\boldsymbol{\beta}_p}) \qquad (6.1)$$

where $\overline{u^2} = T^{-1}\sum_{t=1}^T u_t^2$, Y is $(T-p)\times 1$ vector and X is $(T-p)\times 1$ design matrix:

$$Y = (u_{p+1}^2 - \overline{u^2}, ..., u_T^2 - \overline{u^2})',$$

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & ... & x_{1,p} \\ x_{2,1} & x_{2,2} & ... & x_{2,p} \\ ... & ... & ... & ... \\ x_{T-p,1} & x_{T-p,2} & ... & x_{T-p,p} \end{bmatrix} = \begin{bmatrix} u_p^2 - \overline{u^2} & u_{p-1}^2 - \overline{u^2} & ... & u_1^2 - \overline{u^2} \\ u_{p+1}^2 - \overline{u^2} & u_p^2 - \overline{u^2} & ... & u_2^2 - \overline{u^2} \\ ... & ... & ... & ... \\ u_{T-1}^2 - \overline{u^2} & u_{T-2}^2 - \overline{u^2} & ... & u_{T-p}^2 - \overline{u^2} \end{bmatrix}.$$

Here $\widetilde{\boldsymbol{\beta}}_p$ denotes the estimated regression coefficients of $u_{t-1}^2, \ldots, u_{t-p}^2$ in a regression of u_t^2 on a constant and $u_{t-1}^2, \ldots, u_{t-p}^2$. Similarly, let $\widehat{\boldsymbol{\beta}}_p$ denote the estimated regression coefficients of $\widehat{u}_{t-1}^2, \ldots, \widehat{u}_{t-p}^2$ in a regression of $\widehat{u}_t^2$ on a constant and $\widehat{u}_{t-1}^2, \ldots, \widehat{u}_{t-p}^2$. Then

$$S(\widehat{\boldsymbol{u}}) = \widehat{\sigma}_p^{-2}\widehat{\boldsymbol{\beta}}_p'(\widehat{X}'\widehat{X})\widehat{\boldsymbol{\beta}}_p, \quad \widehat{\boldsymbol{\beta}}_p = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'\widehat{Y}, \quad \widehat{\sigma}_p^2 = (\widehat{Y} - \widehat{X}\widehat{\boldsymbol{\beta}}_p)'(\widehat{Y} - \widehat{X}\widehat{\boldsymbol{\beta}}_p) \qquad (6.2)$$

where $\overline{\widehat{u}^2} = T^{-1}\sum_{t=1}^T \widehat{u}_t^2$,

$$\widehat{Y} = (\widehat{u}_{p+1}^2 - \overline{\widehat{u}^2}, ..., \widehat{u}_T^2 - \overline{\widehat{u}^2})',$$

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

Observe that we can write

$$T^{-1}(X'X) \;=\; (g_{ij})_{i,j=1,\dots,p}, \quad T^{-1}(X'Y)=(g_{0j})_{j=1,\dots,p}, \quad \text{where} \qquad (6.3)$$
$$g_{ij}=T^{-1}\textstyle\sum_{t=p+1}^{T}(u_{t-i}^2-\overline{u^2})(u_{t-j}^2-\overline{u^2}).$$

Similarly,

$$T^{-1}(\widehat{X}'\widehat{X}) \;=\; (\widehat{g}_{ij})_{i,j=1,\dots,p}, \quad T^{-1}(\widehat{X}'\widehat{Y})=(\widehat{g}_{0j})_{j=1,\dots,p}, \quad \text{where} \qquad (6.4)$$
$$\widehat{g}_{ij}=T^{-1}\textstyle\sum_{t=p+1}^{T}(\widehat{u}_{t-i}^2-\overline{\widehat{u}^2})(\widehat{u}_{t-j}^2-\overline{\widehat{u}^2}).$$

Proof of Theorem 2.1 is based on the Lemmas 6.1 and 6.2 below. Auxiliary results used to prove these lemmas are placed in Section 6.2. Denote

$$\widetilde{\gamma}_k \;=\; T^{-1}\sum_{t=k+1}^{T}\big(u_t^2-Eu_t^2\big)\big(u_{t-k}^2-Eu_{t-k}^2\big) \qquad (6.5)$$

the autocovariance function of squared residuals $\widehat{u}_t^2$ and of u_t^2, respectively, where $\overline{\widehat{u}^2}=T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2$. Denote $\widetilde{\mathbf{\Gamma}}_p=(\widetilde{\gamma}_{|i-j|})_{i,j=1,\dots,p}$, $\widetilde{\boldsymbol{\gamma}}_p=(\widetilde{\gamma}_1,\dots,\widetilde{\gamma}_p)'$, $\widetilde{\boldsymbol{\beta}}_p=\widetilde{\mathbf{\Gamma}}_p^{-1}\widetilde{\boldsymbol{\gamma}}_p$.

To prove the theorem we will derive the following results:

$$T^{-1}(\widehat{X}'\widehat{X})=\widetilde{\mathbf{\Gamma}}_p+o_P(1), \quad T^{-1}(\widehat{X}'\widehat{Y})=\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \qquad (6.6)$$
$$T^{-1}(X'X)=\widetilde{\mathbf{\Gamma}}_p+o_P(1), \quad T^{-1}(X'Y)=\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \qquad (6.7)$$
$$\widetilde{\mathbf{\Gamma}}_p\to_p\mathbf{\Gamma}_p, \quad \widetilde{\boldsymbol{\gamma}}_p\to_P\boldsymbol{\gamma}_p, \quad \widetilde{\boldsymbol{\beta}}_p\to\boldsymbol{\beta}_p, \qquad (6.8)$$
$$T^{-1}\widehat{\sigma}_p^2\to_P\sigma_p^2, \quad T^{-1}\widetilde{\sigma}_p^2\to_P\sigma_p^2, \qquad (6.9)$$

where $\boldsymbol{\beta}_p=\mathbf{\Gamma}_p^{-1}\boldsymbol{\gamma}_p$, $\mathbf{\Gamma}_p=\big(\gamma_{|j-k|}\big)_{j,k=1,\dots,p}$, $\boldsymbol{\gamma}_p=\big(\gamma_1,\dots,\gamma_p\big)'$, $\gamma_k=\mathrm{cov}(u_k^2,u_0^2)$, $k\ge 0$, and $\sigma_p^2=\mathrm{var}(u_{p+1}^2-\beta_1u_p^2-\dots-\beta_pu_1^2)$.

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

$$T^{-1/2}(\widehat{X}'\widehat{Y})=T^{-1/2}\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \qquad (6.10)$$
$$T^{-1/2}(X'Y)=T^{-1/2}\widetilde{\boldsymbol{\gamma}}_p+o_P(1), \qquad (6.11)$$
$$T^{-1/2}\widetilde{\boldsymbol{\gamma}}_p\to\mathcal{N}(0,\boldsymbol{I}_p\gamma_0), \quad \gamma_0=\mathrm{var}(u_1^2). \qquad (6.12)$$

In view of (6.3) and (6.4), relations (6.6), (6.7) and (6.8) follow from (6.13), (6.14) and (6.15) of Lemma 6.1, respectively, while (6.10) and (6.11) follow from (6.16) and (6.17) of Lemma 6.1. Convergence (6.12) for i.i.d. r.v. $u_t^2-Eu_t^2$ is a well known.

Convergence (6.9) follows from the definitions of $\widehat{\sigma}_p^2$ and $\widetilde{\sigma}_p^2$, (6.6)-(6.9), 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 $\sigma_p^2=\gamma_0-2\boldsymbol{\gamma}_p'\boldsymbol{\beta}_p+\boldsymbol{\beta}_p'\mathbf{\Gamma}_p\boldsymbol{\beta}_p$.

Applying in $S(\widehat{\boldsymbol{u}})$ and $S(\boldsymbol{u})$, given by (6.2) and (6.1), relations (6.6)-(6.9), we obtain

$$\widehat{\boldsymbol{\beta}}_p\to_P\boldsymbol{\beta}_p=\mathbf{\Gamma}_p^{-1}\boldsymbol{\gamma}_p, \quad \widetilde{\boldsymbol{\beta}}_p\to_P\boldsymbol{\beta}_p,$$

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

which proves the claim (2.12) of the theorem.

In addition, if $\{u_t\}$ is an i.i.d. sequence, then it holds $\boldsymbol{\Gamma}_p = \gamma_0 \boldsymbol{I}_p$ and $\sigma_p^2 = \gamma_0$. Then, using (6.10)–(6.12), 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 definitions of $S(\widehat{\boldsymbol{u}})$ and $S(\boldsymbol{u})$ implies

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

which proves (2.13). This completes the proof of the theorem. □

Proof of Theorems 2.2-2.3. It follows using the same arguments as in the proof of Theorem 2.1. □

Proof of Theorem 2.4. The claims (2.20) and (2.21) follow using the same argument as in the proof of (6.21) and (6.22) of Lemma 6.2 noting that convergence $T^{1/2}\widetilde{r}_{u,k} \to \mathcal{N}(0, (Eu_1^2)^2)$ is well known for i.i.d. r.v.'s. □

LEMMA 6.1. *(a) Suppose that (h_t, u_t) satisfy Assumptions M and H, and the bandwidth H satisfies (2.11). Then for $i, j = 0, 1, ..., p$,*

$$\widehat{g}_{ij} = \widetilde{\gamma}_{|i-j|} + o_P(1), \tag{6.13}$$
$$g_{ij} = \widetilde{\gamma}_{|i-j|} + o_P(1), \tag{6.14}$$
$$\widetilde{\gamma}_k \to_p \gamma_k, \quad k \geq 0. \tag{6.15}$$

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

$$T^{1/2}\widehat{g}_{0j} = T^{1/2}\widetilde{\gamma}_j + o_P(1), \tag{6.16}$$
$$T^{1/2}g_{0j} = T^{1/2}\widetilde{\gamma}_j + o_P(1). \tag{6.17}$$

Proof of Lemma 6.1.
Proof of (6.13). It suffices to verify (6.13) for $i \leq j$. Denote

$$\widehat{\gamma}_k = T^{-1}\sum_{t=k+1}^{T}\left(\widehat{u}_t^2 - \overline{\widehat{u}^2}\right)\left(\widehat{u}_{t-k}^2 - \overline{\widehat{u}^2}\right), \quad k = 0, 1, 2, ... \tag{6.18}$$

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

$$\widehat{g}_{ij} = T^{-1}\sum_{t=p+1-i}^{T-i}(\widehat{u}_t^2 - \overline{\widehat{u}^2})(\widehat{u}_{t-k}^2 - \overline{\widehat{u}^2}) = T^{-1}\sum_{t=k+1+(p-j)}^{T-i}(\widehat{u}_t^2 - \overline{\widehat{u}^2})(\widehat{u}_{t-k}^2 - \overline{\widehat{u}^2})$$

$$= \widehat{\gamma}_{|j-i|} - \delta_{ij}, \quad \delta_{ij} = T^{-1}[\sum_{t=k+1}^{k+p-j} + \sum_{t=T-i+1}^{T}](\widehat{u}_t^2 - \overline{\widehat{u}^2})(\widehat{u}_{t-k|}^2 - \overline{\widehat{u}^2}).$$

So,

$$\widehat{g}_{ij} = \widetilde{\gamma}_{|i-j|} + (\widehat{\gamma}_{|i-j|} - \widetilde{\gamma}_{|i-j|}) - \delta_{ij}. \tag{6.19}$$

By (6.21) of Lemma 6.2, $\widehat{\gamma}_{|i-j|} - \widetilde{\gamma}_{|i-j|} = o_P(1)$. On the other hand, straightforward use of (6.45) and (6.47) of Lemma 6.4 implies

$$\delta_{ij} = O_P(T^{-1}) \tag{6.20}$$

which together with (6.19) proves (6.13): $\widehat{g}_{ij} = \widetilde{\gamma}_{|i-j|} + o_P(1)$.
Proof of (6.14). Property (6.14) follows from the proof of (6.13) setting $\widehat{h}_t = h_t$ which implies $\widehat{u}_t^2 = u_t^2$.
Proof of (6.15). By assumption, u_t is a stationary ergodic sequence and $Eu_t^6 < \infty$. Then sequence $z_t = \big(u_t^2 - Eu_t^2\big)\big(u_{t-k}^2 - Eu_{t-k}^2\big)$ is stationary and ergodic with $E|z_t| < \infty$ which implies $\widetilde{\gamma}_k \to_P Ez_k = \mathrm{cov}(u_k^2, u_0^2) = \gamma_k$.
Proof of (6.16). By (6.19), (6.20) and (6.22), for $j = 1, ..., p$,

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

Proof of (6.17). Property (6.17) follows from (6.16) setting $\widehat{h}_t = h_t$. □

LEMMA 6.2. *(a) Suppose that (h_t, u_t) satisfy Assumptions M and H, and the bandwidth H satisfies (2.11). Then*

$$\widehat{\gamma}_k - \widetilde{\gamma}_k = o_P(1), \quad k \geq 0. \tag{6.21}$$

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

$$\widehat{\gamma}_k - \widetilde{\gamma}_k = o_P(T^{-1/2}), \quad k \geq 1. \tag{6.22}$$

Proof of Lemma 6.2. Denote

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

Then

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

Thus, to prove (6.21), it suffices to show

$$\widehat{\gamma}_k^* - \widetilde{\gamma}_k = o_P(1), \quad k \geq 1, \tag{6.24}$$
$$\widehat{\gamma}_k - \widehat{\gamma}_k^* = o_P(1), \quad k \geq 0, \tag{6.25}$$

$$\widehat{\gamma}_0^* - \widetilde{\gamma}_0 = o_P(1). \tag{6.26}$$

In turn, to prove (6.22), we show in addition that for an i.i.d. sequence $\{u_t\}$ it holds

$$\widehat{\gamma}_k^* - \widetilde{\gamma}_k = o_P(T^{-1/2}), \quad k \geq 1, \tag{6.27}$$
$$\widehat{\gamma}_k - \widehat{\gamma}_k^* = o_P(T^{-1/2}), \quad k \geq 1. \tag{6.28}$$

Proof of (6.24). Recall, that by assumption, $Eu_t^2 = 1$. We have,

$$\begin{aligned}
&(\widehat{u}_t^2 - Eu_t^2)(\widehat{u}_{t-k}^2 - Eu_{t-1}^2) - (u_t^2 - Eu_t^2)(u_{t-k}^2 - Eu_{t-k}^2)\\
&\quad = \{(\widehat{u}_t^2 - u_t^2) + (u_t^2 - 1)\}\{(\widehat{u}_{t-k}^2 - u_{t-k}^2) + (u_{t-k}^2 - 1)\} - (u_t^2 - 1)(u_{t-k}^2 - 1)\\
&\quad = (\widehat{u}_t^2 - u_t^2)(\widehat{u}_{t-k}^2 - u_{t-k}^2) + (\widehat{u}_t^2 - u_t^2)(u_{t-k}^2 - 1) + (u_t^2 - 1)(\widehat{u}_{t-k}^2 - u_{t-k}^2).
\end{aligned}$$

Hence,

$$\widehat{\gamma}_k^* - \widetilde{\gamma}_k = S_{T,1} + S_{T,2} + S_{T,3}, \tag{6.29}$$

where

$$\begin{aligned}
S_{T,1} &= T^{-1}\textstyle\sum_{t=k+1}^T(\widehat{u}_t^2 - u_t^2)(\widehat{u}_{t-k}^2 - u_{t-k}^2), \quad S_{T,2} = T^{-1}\sum_{t=k+1}^T(\widehat{u}_t^2 - u_t^2)(u_{t-k}^2 - 1),\\
S_{T,3} &= T^{-1}\textstyle\sum_{t=k+1}^T(u_t^2 - 1)(\widehat{u}_{t-k}^2 - u_{t-k}^2).
\end{aligned} \tag{6.30}$$

To prove (6.24), it remains to show that $S_{T,\ell}$ in (6.30) satisfy

$$S_{T,\ell} = o_P(1), \quad \ell = 1, 2, 3. \tag{6.31}$$

Notice that

$$\widehat{u}_t^2 - u_t^2 = (h_t^2\widehat{h}_t^{-2} - 1)u_t^2. \tag{6.32}$$

We have

$$\begin{aligned}
|S_{T,1}| &\leq T^{-1}|\sum_{t=k+1}^T |(h_t^2\widehat{h}_t^{-2} - 1)(h_{t-k}^2\widehat{h}_{t-k}^{-2} - 1)|u_t^2u_{t-k}^2\\
&\leq (\max_{t=1,\dots,T}|h_t^2/\widehat{h}_t^2 - 1|)^2\, s_T, \quad s_T = T^{-1}\sum_{t=k+1}^T u_t^2u_{t-k}^2.
\end{aligned}$$

By (6.39) of Lemma 6.3, $\max_{t=1,\dots,T}|h_t^2/\widehat{h}_t^2 - 1| = o_P(1)$. By Assumption M, $\{u_t\}$ is a stationary sequence, and $Es_T = E[u_1^2u_{1-k]}^2] \leq E[u_1^4] < \infty$. Therefore, $s_T = O_P(1)$. This implies

$$S_{T,1} = o_P(1).$$

The proof of (6.31) for $S_{T,2}$, $S_{T,3}$ is similar to that for $S_{T,1}$. This completes the proof of (6.24).

Proof of (6.25). Recall, that by assumption, $Eu_t^2 = 1$. We have,

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

Notice that

$$
\begin{aligned}
T^{-1}\textstyle\sum_{t=k+1}^T(\widehat{u}_t^2 - 1) &= (\overline{\widehat{u}^2} - 1) - T^{-1}\textstyle\sum_{t=1}^k(\widehat{u}_t^2 - 1), \\
T^{-1}\textstyle\sum_{t=k+1}^T(\widehat{u}_{t-k}^2 - 1) &= (\overline{\widehat{u}^2} - 1) - T^{-1}\textstyle\sum_{t=T-k+1}^T(\widehat{u}_t^2 - 1).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\widehat{\gamma}_k - \widehat{\gamma}_k^* &= T^{-1}\textstyle\sum_{t=k+1}^T\{(\overline{\widehat{u}^2} - 1)^2 - (\overline{\widehat{u}^2} - 1)(\widehat{u}_{t-k}^2 - 1) - (\widehat{u}_t^2 - 1)(\overline{\widehat{u}^2} - 1)\} \\
&= (\overline{\widehat{u}^2} - 1)^2\{T^{-1}(T-k) - 2) \\
&\quad + (\overline{\widehat{u}^2} - 1)\{T^{-1}\textstyle\sum_{t=1}^k(\widehat{u}_t^2 - 1) + T^{-1}\textstyle\sum_{t=T-k+1}^T(\widehat{u}_t^2 - 1)\}. \qquad (6.33)
\end{aligned}
$$

By Assumption M, $\{u_t\}$ is a stationary α-mixing sequence. Applying in (6.33) the bounds (6.45) and (6.46) of Lemma 6.4, we obtain (6.25):

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

Proof of (6.26) By Assumption M, $\{u_t\}$ is a stationary α-mixing sequence. Using equality $a^2 - b^2 = (a-b)^2 + (a-b)2b$ with $a = \widehat{u}_t^2 - 1$, $b = u_t^2 - 1$ we obtain

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

Recall that $\widehat{u}_t^2 - u_t^2 = (h_t^2/\widehat{h}_t^2 - 1)u_t^2$, and by (6.39) of Lemma 6.3, $i_T := \max_{t=1,\dots,T}|h_t^2/\widehat{h}_t^2 - 1| = o_P(1)$. Therefore,

$$
|\widehat{\gamma}_0^* - \widetilde{\gamma}_0| \leq i_T^2\big(T^{-1}\sum_{t=1}^T u_t^4\big) + 2i_T\big(T^{-1}\sum_{t=1}^T u_t^2|u_t^2 - 1|\big) = o_P(1)q_T, \quad q_T = T^{-1}\sum_{t=1}^T(u_t^4 + u_t^2).
$$

Since $Eq_T = Eu_1^4 + Eu_1^2 < \infty$ this implies $q_T = O_P(1)$ which proves (6.26): $\widehat{\gamma}_0^* - \widetilde{\gamma}_0 = o_P(1)$.

Proof of (6.27). By (6.29), $\widehat{\gamma}_k^* - \widetilde{\gamma}_k = S_{T,1} + S_{T,2} + S_{T,3}$. To prove (6.27), it remains to show that

$$
S_{T,\ell} = o_P(T^{-1/2}), \quad \ell = 1, 2, 3. \qquad (6.34)
$$

First we evaluate $S_{T,1}$. Using (6.32), we can bound

$$
|S_{T,1}| \;\leq\; T^{-1}|\sum_{t=k+1}^T|(h_t^2\widehat{h}_t^{-2} - 1)(h_{t-k}^2\widehat{h}_{t-k}^{-2} - 1)|u_t^2u_{t-k}^2
$$

$$\leq \ (\max_{t=1,\ldots,T} \widehat{h}_t^{-2})^2 S'_{T,1}, \quad S'_{T,1} = T^{-1} \sum_{t=k+1}^{T} \big|(h_t^2 - \widehat{h}_t^2)(h_{t-k}^2 - \widehat{h}_{t-k}^2)\big| u_t^2 u_{t-k}^2.$$

By Lemma 6.3, $\max_{t=1,\ldots,T} \widehat{h}_t^{-2} = O_p(1)$. Moreover, by (6.58) of Lemma 6.5,

$$\begin{aligned}
&E\big[\big|(h_t^2 - \widehat{h}_t^2)(h_{t-k}^2 - \widehat{h}_{t-k}^2)\big| u_t^2 u_{t-k}^2\big] \\
&\quad \leq (E[(h_t^2 - \widehat{h}_t^2)^2 u_t^2 u_{t-k}^2])^{1/2} (E[(h_{t-k}^2 - \widehat{h}_{t-k}^2)^2 u_t^2 u_{t-k}^2])^{1/2} \\
&\quad \leq C\big((H/T)^{2\gamma} + H^{-1}\big)
\end{aligned}$$

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

$$E S'_{T,1} \leq C\big((H/T)^{2\gamma} + H^{-1}\big) = o(T^{-1/2})$$

under assumption (2.11) on H. So, $S'_{T,1} = o_P(T^{-1/2})$. This proves (6.34) for $S_{T,1}$.

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

$$S_{T,2} = T^{-1} \sum_{t=k+1}^{T} (h_t^2 \widehat{h}_t^{-2} - 1)\zeta_t, \quad \zeta_t = u_t^2(u_{t-k}^2 - 1). \tag{6.35}$$

Write

$$\widehat{h}_t^2 = h_t^2 + (\widehat{h}_t^2 - h_t^2) = h_t^2(1 + x_t), \ x_t = \frac{\widehat{h}_t^2 - h_t^2}{h_t^2}.$$

Then

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

Then, by (6.35),

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

To prove (6.34) for $S_{T,2}$, we verify that

$$q_{T,\ell} = o_P(T^{-1/2}), \quad \ell = 1, 2. \tag{6.36}$$

In (6.56) of Lemma 6.5 it is shown $E|q_{T,1}| = O\big((H/T)^{2\gamma} + H^{-1}\big) = o(T^{-1/2})$ under assumption (2.11) which proves (6.39) for $q_{T,1}$.

Next, we bound

$$|q_{T,2}| \leq (\max_{t=1,\dots,T} h_t^{-2})(\max_{t=1,\dots,T} \widehat{h}_t^{-2})\, v_T \qquad v_T = T^{-1}\sum_{t=k+1}^{T}(\widehat{h}_t^2 - h_t^2)^2|\zeta_t|.$$

By Assumption H, $h_t \geq a > 0$ a.s., and by (6.38) of Lemma 6.3, $\max_{t=1,\dots,T} \widehat{h}_t^{-2} = O_P(1)$, whereas using (6.57) of Lemma 6.5, we obtain

$$Ev_T \leq T^{-1}\sum_{t=k+1}^{T} E[(\widehat{h}_t^2 - h_t^2)^2|\zeta_t|] = O\big((H/T)^{2\gamma} + H^{-1}\big).$$

This implies

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

under assumption (2.11) which proves (6.39). This verifies (6.34) for $S_{T,2}$. The proof of (6.34) for $S_{T,3}$ is similar to the proof for $S_{T,2}$. This completes the proof of (6.27).

Proof of (6.28) Using in (6.33) the bounds (6.48) and (6.46) of Lemma 6.4 we obtain

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

under assumption (2.11) on H. This proves (6.28) and completes the proof of the Lemma 6.2. □.

LEMMA 6.3. *Suppose that (h_t, u_t) satisfy Assumptions M and H, and the bandwidth H satisfies (2.11). Then there exists $\delta > 0$ such that*

$$\max_{t=1,\dots,T} |h_t^2 - \widehat{h}_t^2| = o_P(T^{-\delta}), \tag{6.37}$$

$$\max_{t=1,\dots,T} \widehat{h}_t^{-2} = O_P(1), \tag{6.38}$$

$$\max_{t=1,\dots,T} |h_t^2/\widehat{h}_t^2 - 1| = o_P(T^{-\delta}). \tag{6.39}$$

Proof of Lemma 6.3. *Proof of (6.37)*. We have

$$\begin{aligned} h_t^2 - \widehat{h}_t^2 &= K_t^{-1}\textstyle\sum_{j=1}^{T} b_{H,|t-j|}(h_t^2 - h_j^2 u_j^2) \\ &= K_t^{-1}\textstyle\sum_{k=1}^{T} b_{H,|t-j|}(h_t^2 - h_j^2)u_j^2 - K_t^{-1}h_t^2\sum_{j=1}^{T} b_{H,|t-j|}(u_j^2 - 1). \end{aligned} \tag{6.40}$$

Properties (2.6) of the kernel function K imply

$$\max_{t=1,\dots,T} K_t^{-1} \leq CH^{-1} \tag{6.41}$$

where C does not depend on T. By Assumption H and Proposition 6.1,

$$|h_t^2 - h_j^2| \leq (|t-j|/T)^{\gamma}\xi_{tj} = (H/T)^{\gamma}(|t-j|/H)^{\gamma}\xi_{tj}$$

where $\{\xi_{tj}\}$ and $\{h_t^2\}$ satisfy the condition (2.3) of finite exponential moment with parameter $\alpha/2 > 0$. From (6.40) it follows

$$\begin{aligned} |h_t^2 - h_t^2| \;\le\; & (H/T)^\gamma (\max_{j,t=1,\dots,T} |\xi_{tj}|)\{\max_{t=1,\dots,T} K_t^{-1} \textstyle\sum_{j=1}^T b_{H,|t-j|}(|t-j|/H)^\gamma u_j^2\} \\ & +(\max_{t=1,\dots,T} h_t^2)\{\max_{t=1,\dots,T} \big|\textstyle\sum_{j=1}^T b_{H,|t-j|}(u_j^2-1)\big|\}. \qquad (6.42) \end{aligned}$$

By (v) of Lemma C1 in the online supplement of Dendramis, Giraitis and Kapetanios (2021), under Assumption H h_t^2 and ξ_{tk} satisfy

$$\max_{1\le t\le T} h_t^2 = O_P\big((\log T)^{2/\alpha}\big), \quad \max_{1\le t,j\le T} |\xi_{tj}| = O_P\big((\log T)^{2/\alpha}\big). \qquad (6.43)$$

By Assumption M, $\{u_j\}$ is a stationary α-mixing sequence, and $E|u_1|^\theta < \infty$ for some $\theta > 6$. Then $\xi_j = u_j^2 - 1$ is also a stationary α-mixing sequence which satisfies α-mixing Assumption M, see Theorem 14.1 in Davidson (1994), and $E|\xi_1|^{\theta'} < \infty$ where $\theta' = \theta/2 > 3$. Under these assumptions, Corollary 6(b) of Dendramis, Giraitis and Kapetanios (2018) implies that for any $\varepsilon > 0$,

$$\max_{1\le t\le T} \big|H^{-1} \sum_{j=1}^T b_{H,|t-j|}(\xi_j - E\xi_j)\big| = O_P\big(H^{-1/2}\sqrt{\log T} + (HT)^{1/\theta'} H^{\varepsilon-1}\big). \qquad (6.44)$$

For $H \ge T^{1/2}$ and $\theta' > 3$, the r.h.s. of (6.44) is of order $o_P(T^{-\delta})$ for some $\delta > 0$ when ε is selected sufficiently small.
Setting $b^*_{H,|t-j|} = b_{H,|t-j|}(|t-j|/H)^\gamma$, we have

$$K_t^{-1} \textstyle\sum_{j=1}^T b^*_{H,|t-j|} u_j^2 = K_t^{-1} \sum_{j=1}^T b^*_{H,|t-j|}(u_j^2 - 1) + K_t^{-1} \sum_{j=1}^T b^*_{H,|t-j|},$$

where under assumption (2.6), $\max_{t=1,\dots,T} K_t^{-1} \sum_{j=1}^T b^*_{H,|t-j|} = O(1)$. Then, using similar argument as in the proof of (6.44), it follows that

$$\max_{t=1,\dots,T} K_t^{-1} \big|\textstyle\sum_{j=1}^T b^*_{H,|t-j|}(u_j^2-1)\big| = o_P(T^{-\delta}).$$

Applying these bound in (6.42) we obtain

$$\max_{t=1,\dots,T} |h_t^2 - \widehat{h}_t^2| = O_P\big((\log T)^{1/\alpha}\big)\{O_P\big((H/T)^\gamma\big) + o_P(T^{-\delta})\} = o_P(T^{-\delta'})$$

for some $\delta' > 0$ under assumption (2.11) on H. This proves (6.37).

Finally, by Assumption H, $h_t \ge a > 0$ a.s. Thus,

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

which proves (6.38. In turn, (6.38) and (6.37) imply (6.39):

$$\max_{t=1,\dots,T} |h_t^2/\widehat{h}_t^2 - 1| \le (\max_{t=1,\dots,T} \widehat{h}_t^{-2})(\max_{t=1,\dots,T} |h_t^2 - \widehat{h}_t^2|) = o_P(T^{-\delta}).$$

This completes the proof of the lemma. □

6.2 Auxiliary lemmas

This section contains auxiliary lemmas used on the proofs of Section 6.1.

LEMMA 6.4. *(a) Under assumptions of Lemma 6.2(a), for any fixed $k \geq 1$,*

$$\overline{\widehat{u}^2} - 1 = o_P\big(1\big), \tag{6.45}$$

$$\textstyle\sum_{t=1}^{k} |\widehat{u}_t^2 - 1| = O_p(1), \quad \sum_{t=T-k+1}^{T} |\widehat{u}_t^2 - 1| = O_p(1). \tag{6.46}$$

$$\textstyle\sum_{t=1}^{k} (\widehat{u}_t^2 + \widehat{u}_t^4) = O_p(1), \quad \sum_{t=T-k+1}^{T} (\widehat{u}_t^2 + \widehat{u}_t^4) = O_p(1). \tag{6.47}$$

(b) In addition, if u_t is an i.i.d. sequence, then

$$\overline{\widehat{u}^2} - 1 = O_P\big((H/T)^\gamma + H^{-1/2}\big). \tag{6.48}$$

Proof of Lemma 6.4. *Proof of (6.45).* We have

$$\begin{aligned} \overline{\widehat{u}^2} - 1 = T^{-1}\sum_{t=1}^{T}(\widehat{u}_t^2 - 1) &= T^{-1}\sum_{t=1}^{T}(\widehat{u}_t^2 - u_t^2) + T^{-1}\sum_{t=1}^{T}(u_t^2 - 1) \\ &=: Q_{1,T} + Q_{2,T}. \end{aligned} \tag{6.49}$$

We will show

$$Q_{1,T} = o_P(1), \qquad Q_{2,T} = O_P(T^{-1/2}) \tag{6.50}$$

which proves (6.45). We have,

$$\begin{aligned} |Q_{1,T}| &= T^{-1}\Big|\sum_{t=1}^{T}(h_t^2/\widehat{h}_t^2 - 1)u_t^2\Big| \leq \big(\max_{t=1,\ldots,T} |h_t^2/\widehat{h}_t^2 - 1|\big)\, T^{-1}\sum_{t=1}^{T} u_t^2 \\ &= o_P(1)\, T^{-1}\sum_{t=1}^{T} u_t^2 = o_P(1), \end{aligned} \tag{6.51}$$

by (6.39) of Lemma 6.3, noting that $E(T^{-1}\sum_{t=1}^{T} u_t^2) = Eu_1^2$ implies $T^{-1}\sum_{t=1}^{T} u_t^2 = O_P(1)$.

Under Assumption M, by Conclusion 2.2 in Davydov (1968) (for more details see (A.11) in Dendramis *et al.* (2021)), stationary α-mixing sequence $z_t = u_t^2 - 1$ has property

$$\sum_{k=0}^{\infty} |\mathrm{cov}(z_k, z_0)| < \infty. \tag{6.52}$$

Therefore,

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

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

which proves the second claim in (6.50). This completes the proof of (6.45).

Proof of (6.46). We have

$$\sum_{t=1}^{k} |\widehat{u}_t^2 - 1| = \sum_{t=1}^{k} |h_t^2/\widehat{h}_t^2 - 1| u_t^2 \leq (\max_{t=1,\dots,T} |h_t^2/\widehat{h}_t^2 - 1|) \sum_{t=1}^{k} u_t^2 = o_P(1)$$

by (6.39) of Lemma 6.3 noting that for any fixed k, $\sum_{t=1}^{k} u_t^2 = O_P(1)$. This proves the first claim in (6.46). The proof of the second claim is similar.

Proof of (6.47). We can bound

$$\widehat{u}_t^2 + \widehat{u}_t^4 = \{(\widehat{u}_t^2 - 1) + 1\} + \{(\widehat{u}_t^2 - 1) + 1\}^2 \leq |\widehat{u}_t^2 - 1| + 1 + 2((\widehat{u}_t^2 - 1)^2 + 1).$$

Then (6.47) follows by the same argument as in the proof of (6.46).

Proof of (6.48). In view of (6.49) and (6.50), it suffices to show

$$Q_{1,T} = O_p\big((H/T)^{\gamma} + H^{-1/2}\big). \tag{6.53}$$

By (6.51),

$$|Q_{1,T}| \;\leq\; (\max_{t=1,\dots,T} \widehat{h}_t^{-2}) d_T, \quad d_T = T^{-1} \sum_{t=1}^{T} |h_t^2 - \widehat{h}_t^2| u_t^2. \tag{6.54}$$

By Lemma 6.3, $\max_{t=1,\dots,T} \widehat{h}_t^{-2} = O_P(1)$. On the other hand, by (6.55) of Lemma 6.5, $E(h_t^2 - \widehat{h}_t^2)^2 \leq C((H/T)^{2\gamma} + H^{-1})$ where C does not depend on t, H, T, which implies

$$\begin{aligned} Ed_T \;&\leq\; T^{-1} \sum_{t=1}^{T} \big(E(h_t^2 - \widehat{h}_t^2)^2\big)^{1/2} (Eu_t^4)^{1/2} \\ &\leq\; C((H/T)^{2\gamma} + H^{-1})^{1/2} T^{-1} \sum_{t=1}^{T} (Eu_1^4)^{1/2} = C((H/T)^{2\gamma} + H^{-1})^{1/2}. \end{aligned}$$

Hence $d_T = O_P\big((H/T)^{\gamma} + H^{-1/2}\big)$ which together with (6.54) proves (6.53). This completes the proof of the lemma. □

LEMMA 6.5. *(a) Under assumptions of Lemma 6.2(a),*

$$E(\widehat{h}_t^2 - h_t^2)^2 \;\leq\; C\big((H/T)^{2\gamma} + H^{-1}\big). \tag{6.55}$$

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

(b) If in addition, u_t is a sequence of i.i.d. random variables, then for any $k \geq 1$,

$$E\Big|T^{-1}\sum_{t=k+1}^{T} h_t^{-2}(\widehat{h}_t^2 - h_t^2)\zeta_t\Big| = O\big((H/T)^{2\gamma} + H^{-1}\big), \tag{6.56}$$

$$E[(h_t^2 - \widehat{h}_t^2)^2|\zeta_t|] \leq C\big((H/T)^{2\gamma} + H^{-1}\big), \tag{6.57}$$

$$E[(h_{t-s}^2 - \widehat{h}_{t-s}^2)^2 u_t^2 u_{t-k}^2] \leq C\big((H/T)^{2\gamma} + H^{-1}\big), \ \textit{for } s = 0, k \tag{6.58}$$

where $\zeta_t = u_t^2(u_{t-k}^2 - 1)$ and C does not depend on t, H, T.

Proof of Lemma 6.5.
Proof of (6.55). Denote $\bar{h}_t^2 = K_t^{-1}\sum_{j=1}^{T} b_{|t-j|}h_j^2$. Then

$$\begin{aligned} h_t^2 - \widehat{h}_t^2 &= (h_t^2 - \bar{h}_t^2) + (\bar{h}_t^2 - \widehat{h}_t^2), \\ (h_t^2 - \widehat{h}_t^2)^2 &\leq 2(h_t^2 - \bar{h}_t^2)^2 + 2(\bar{h}_t^2 - \widehat{h}_t^2)^2. \end{aligned} \tag{6.59}$$

We will show that

$$E(h_t^2 - \bar{h}_t^2)^2 \leq C(H/T)^{2\gamma}, \tag{6.60}$$

$$E(\widehat{h}_t^2 - \bar{h}_t^2)^2 \leq CH^{-1}, \tag{6.61}$$

which together with (6.59) proves (6.55).
Proof of (6.60). We have

$$\begin{aligned} E(h_t^2 - \bar{h}_t^2)^2 &= E\big(K_t^{-1}\sum_{j=1}^{T} b_{|t-j|}(h_t^2 - h_j^2)\big)^2 \\ &\leq K_t^{-2}\sum_{j,k=1}^{T} b_{|t-j|}b_{|t-k|}E[(h_t^2 - h_j^2)(h_t^2 - h_k^2)] \\ &\leq CK_t^{-2}\sum_{j,k=1}^{T} b_{|t-j|}b_{|t-k|}\big(E(h_t^2 - h_j^2)^2 E(h_t^2 - h_k^2)^2\big)^{1/2}. \end{aligned}$$

By Assumption H and Proposition 6.1, $E(h_t^2 - h_j^2)^2 \leq C((t-j)/T)^{\gamma}$ where C does not depend on t, j, T. Properties (6.41) and (2.6) of K_t and b_t imply

$$\max_{t=1,\dots,T} K_t^{-1}\sum_{j=1}^{T} b_{|t-j|}(\frac{|t-j|}{H})^{\gamma} \leq C. \tag{6.62}$$

So,

$$E(h_t^2 - \bar{h}_t^2)^2 \leq C(H/T)^{2\gamma}\big(K_t^{-1}\sum_{j=1}^{T} b_{|t-j|}(|t-j|/H)^{\gamma}\big)^2 \leq C(H/T)^{2\gamma}$$

where C does not depend on t, H. This proves (6.60).
Proof of (6.61). By Assumption M, $\{h_t\}$ and $\{u_t\}$ are mutually independent sequences. Then

$$E(\widehat{h}_t^2 - \bar{h}_t^2)^2 = E\big(K_t^{-1}\sum_{j=1}^{T} b_{|t-j|}h_j^2(u_j^2 - 1)\big)^2$$

$$= \quad K_t^{-2} \sum_{j,k=1}^{T} b_{|t-j|} b_{|t-k|} E[h_j^2 h_k^2] \mathrm{cov}(u_j^2, u_k^2)].$$

By Assumption H, $E[h_j^2 h_k^2] \le (E[h_j^4][h_k^4])^{1/2} \le \max_j E[h_j^4] < \infty$, and by definition of b_j, $\max_j |b_{|j|}| \le C$. Using stationarity of $\{u_j^2\}$, (6.52) and (6.41), we obtain

$$\begin{aligned} E[(\widehat{h}_t^2 - \bar{h}_t^2)^2] &= CK_t^{-2} \sum_{j,k=1}^{T} b_{|t-j|} |\mathrm{cov}(u_j^2, u_k^2)| \\ &\le CK_t^{-1} (K_t^{-1} \sum_{j=1}^{T} b_{|t-j|}) \sum_{k=-\infty}^{\infty} |\mathrm{cov}(u_0^2, u_k^2)| \le CK_t^{-1} \le CH^{-1} \end{aligned}$$

which proves (6.61).
Proof of (6.56). Using (6.59), write

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

We will show that

$$E|Q_{\ell,T}| = O\big((H/T)^{2\gamma} + H^{-1}\big), \quad \ell = 1, 2 \tag{6.63}$$

which proves (6.56)

Notice that the sequence $\{h_t^{-2}(\bar{h}_t^2 - h_t^2)\}$ is independent of $\{\zeta_t\}$ while the i.i.d. property of $\{u_t\}$ implies $E\zeta_t\zeta_s = 0$ for $t \ne s$. Therefore,

$$\begin{aligned} EQ_{1,T}^2 &= T^{-2} \sum_{t,s=k+1}^{T} E[h_t^{-2}(\bar{h}_t^2 - h_t^2) h_s^{-2} (\bar{h}_s^2 - h_s^2)] E[\zeta_t \zeta_s] \\ &= T^{-2} \sum_{t=k+1}^{T} E[h_t^{-4} (\bar{h}_t^2 - h_t^2)^2] E\zeta_t^2 = E[\zeta_1^2] T^{-2} \sum_{t=k+1}^{T} E[h_t^{-4} (\bar{h}_t^2 - h_t^2)^2]. \end{aligned}$$

By Assumption M, $E\zeta_1^2 = Eu_1^4 E(u_{1-k}^2 - 1)^2 < \infty$, and by Assumption H, $\min_t h_t^2 \ge a > 0$ *a.s.* Hence,

$$E[h_t^{-4} (\bar{h}_t^2 - h_t^2))^2] \le a^{-4} E(h_t^2 - \bar{h}_t^2)^2 \le C(H/T)^{2\gamma} \tag{6.64}$$

by (6.60). Hence

$$E|Q_{1,T}| \le (EQ_{1,T}^2)^{1/2} \le C(H/T)^{\gamma} T^{-1/2} \le 2C\{(H/T)^{2\gamma} + T^{-1}\} \tag{6.65}$$

which proves (6.63) for $Q_{1,T}$.

Next we prove (6.63) for $Q_{2,T}$. Write

$$\begin{aligned}
\widehat{h}_t^2 - \bar{h}_t^2 &= K_t^{-1} \textstyle\sum_{j=1}^T b_{|t-j|} h_j^2 (u_j^2 - 1), \\
Q_{2,T} &= T^{-1} \sum_{t=k+1}^T h_t^{-2} (\widehat{h}_t^2 - \bar{h}_t^2) \zeta_t = T^{-1} K_t^{-1} \sum_{t=k+1}^T \sum_{j=1}^T b_{|t-j|} h_t^{-2} \zeta_t h_j^2 (u_j^2 - 1) \\
&= T^{-1} K_t^{-1} \sum_{t=k+1}^T \sum_{j=1:\, j>t+3k}^T [...] + T^{-1} K_t^{-1} \sum_{t=k+1}^T \sum_{j=1:\, j<t-3k}^T [...] + T^{-1} K_t^{-1} \sum_{t=k+1}^T \sum_{j=1:\, |j-t|\le 3k}^T [...] \\
&= R_{1,T} + R_{2,T} + R_{3,T}.
\end{aligned}$$

We will show

$$E|R_{\ell,T}| = O(H^{-1}), \quad \ell = 1,2,3 \tag{6.66}$$

which implies $E|Q_{2,T}| \le CH^{-1}$ which proves (6.63) for $Q_{2,T}$.

To evaluate $ER_{1,T}^2$, recall that by assumption $\{h_t\}$ and $\{u_t\}$ are mutually independent. Then, since u_t are i.i.d. random variables, for any $j > t + 3k$, $j' > t' + 3k$,

$$\begin{aligned}
&E[\{h_t^{-2} \zeta_t h_j^2 (u_j^2 - 1)\}\{h_{t'}^{-2} \zeta_{t'} h_{j'}^2 (u_{j'}^2 - 1)\}] \\
&\quad = E[h_t^{-2} h_j^2 h_{t'}^{-2} h_{j'}^2] E[\zeta_{t'} \zeta_t] E[(u_j^2 - 1)(u_{j'}^2 - 1)] = 0 \quad \text{if } t \ne t' \text{ or } j \ne j'; \\
&\quad = E[h_t^{-4} h_j^4] E[\zeta_t^2] E[(u_j^2 - 1)^2] \text{ if } \quad t = t',\ j = j'.
\end{aligned}$$

Hence,

$$\begin{aligned}
ER_{1,T}^2 &= T^{-2} K_t^{-2} \sum_{t=k+1}^T \sum_{j=1}^T b_{|t-j|}^2 E[h_t^{-4} h_j^4] E[\zeta_t^2] E[(u_j^2 - 1)^2] \\
&\le E[\zeta_1^2] E[(u_1^2 - 1)^2] (\max_{t,j} E[h_t^{-4} h_j^4]) T^{-2} K_t^{-2} \sum_{t=k+1}^T \sum_{j=1}^T b_{|t-j|}^2.
\end{aligned}$$

Under Assumption M, $E[\zeta_1^2] = E[u_1^4 (u_{1-k}^2 - 1)^2] = E[u_1^4] E[(u_{1-k}^2 - 1)^2] < \infty$, under Assumption H, $h_t^{-4} \le a^{-4} < \infty$ and $\max_j E[h_j^4] < \infty$, and $\max_j |b_{|j|}| < \infty$. Thus,

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

in view of (6.41). Then $E|R_{1,T}| \le (ER_{1,T}^2)^{-1/2} \le CH^{-1}$ which proves (6.66). The proof of (6.66) for $R_{2,T}$ is similar as the proof for $R_{1,T}$.
Finally, using similar arguments as above, we obtain

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

$$\leq \quad (\max_{t,k} E[h_t^{-2}h_j^2])E|\zeta_1(u_1^2-1)|(\max_j b_{|j|})K_t^{-1}(6k+1) \leq CH^{-1}$$

which proves (6.66) for $R_{3,T}$. This completes the proof of (6.56).
Proof of (6.57). Using (6.59), we can bound

$$E[(h_t^2-\widehat{h}_t^2)^2|\zeta_t|] \leq 2E[(h_t^2-\bar{h}_t^2)^2|\zeta_t|] + 2E[(\widehat{h}_t^2-\bar{h}_t^2)^2|\zeta_t|]. \tag{6.67}$$

From (6.60) and independence of $\{h_t\}$ and $\{u_t\}$ it follows

$$E[(h_t^2-\bar{h}_t^2)^2|\zeta_t|] = E[|\zeta_1|]E[(h_t^2-\bar{h}_t^2)^2] \leq C(H/T)^{2\gamma}$$

where C does not depend on t, H, T. In addition, we will show that

$$E[(\widehat{h}_t^2-\bar{h}_t^2)^2|\zeta_t|] \leq CH^{-1} \tag{6.68}$$

which together with (6.67) proves (6.57).

It remains to prove (6.68). Write

$$\widehat{h}_t^2-\bar{h}_t^2 = K_t^{-1}\sum_{j=1}^{T} b_{|t-j|}h_j^2(u_j^2-1) = v_t + K_t^{-1}\big(b_0h_t^2(u_t^2-1) + b_kh_{t-k}^2(u_{t-k}^2-1)\big),$$

where $v_t = K_t^{-1}\sum_{j=1:j\neq t,t-k}^{T} b_{|t-j|}h_j^2(u_j^2-1)$. Then,

$$(\widehat{h}_t^2-\bar{h}_t^2)^2 \quad \leq \quad 2v_t^2 + 4K_t^{-2}\big(b_0^2h_t^4(u_t^2-1)^2 + b_k^2h_{t-k}^4(u_{t-k}^2-1)^2\big).$$

By assumption $\{h_t\}$ and $\{u_t\}$ are mutually independent, and by construction, v_t is independent of ζ_t. Hence,

$$\begin{aligned} E[(\widehat{h}_t^2-\bar{h}_t^2)^2|\zeta_t|] \quad &\leq \quad 2E[v_t^2]E[|\zeta_t|] \\ &\qquad +4K_t^{-2}\big(b_0^2E[h_t^4]E[(u_t^2-1)^2|\zeta_t|] + b_k^2E[h_{t-k}^4]E[(u_{t-k}^2-1)^2|\zeta_t|]\big). \end{aligned}$$

Using similar argument as in the proof of (6.61) it follows that

$$Ev_t^2 \leq CH^{-1}$$

where C does not depend on t, H, T, and (6.41) implies $K_t^{-1} \leq CH^{-1}$. Under Assumption M, $Eu_t^6 < \infty$ which together with the i.i.d.property of variables u_t implies $E|\zeta_t| = E|\zeta_1| < \infty$, $E[(u_t^2-1)^2|\zeta_t|] = E[(u_1^2-1)^2|\zeta_1|] < \infty$, $E[(u_{t-k}^2-1)^2|\zeta_t|] = E[(u_{1-k}^2-1)^2|\zeta_1|] < \infty$. Hence,

$$E[(\widehat{h}_t^2-\bar{h}_t^2)^2|\zeta_t|] \quad \leq \quad CH^{-1}$$

which proves (6.68). This completes the proof of (6.57).
Proof of (6.58). This bound can be verified using the same arguments as in the proof of (6.57). □