Adaptive Long Memory Testing under Heteroskedasticity

David Harris* and Hsein Kew
Department of Econometrics and Business Statistics
Monash University
May 29, 2015

Abstract
This paper considers adaptive hypothesis testing for the fractional differencing parameter in a parametric ARFIMA model with unconditional heteroskedasticity of unknown form. A weighted score test based on a non-parametric variance estimator is proposed and shown to be asymptotically equivalent, under the null and local alternatives, to the Neyman-Rao effective score test constructed under Gaussianity and known variance process. The proposed test is therefore asymptotically efficient under Gaussianity. The finite sample properties of the test are investigated in a Monte Carlo experiment and shown to provide potentially large power gains over the usual unweighted long memory test.

1 Introduction
There is a large literature on statistical inference for the fractional differencing parameter in a stationary ARFIMA model. Of particular note is Robinson (1994), who derived asymptotically efficient score-based tests; see also Tanaka (1999), Nielsen (2004) and Johansen and Nielsen (2010). Regression based LM tests of fractional integration have been developed by Robinson (1991), Agiakloglou and Newbold (1994), Breitung and Hassler (2002), Demetrescu, Kuzin and Hassler (2008) and Hassler, Rodrigues and Rubia (2009); and Wald version of these regression based tests have been proposed by Dolado, Gonzalo and Mayoral (2002) and Lobato and Velasco (2006, 2007). All of this literature maintains an assumption of unconditional homoskedasticity. That is, while the disturbances of the model may be permitted to follow a martingale difference structure that allows for some degree of conditional heteroskedasticity, this literature does not allow for changes in the unconditional variance.

There is, however, abundant empirical evidence that macroeconomic and financial time series exhibit unconditional heteroskedasticity; see for example Pagan and Schwert (1990), Loretan and Phillips (1994), Watson (1999), McConnell and Perez-Quiros (2000), van Dijk, Osborn and Sensier (2002), Sensier and Van Dijk (2004), and Stârîcă and Granger (2005). Kew and Harris (2009) and Cavaliere, Nielsen and Taylor (2015, hereafter CNT, 2014) derived some implications for the size of long memory tests in the presence of such heteroskedasticity and constructed a heteroskedasticity-robust test, but they did not pursue the possibility of adapting the test to recover the power losses that unmodelled heteroskedasticity can incur. This paper takes up this point and derives a test that non-parametrically adapts to unconditional heteroskedasticity of unknown form. In particular, we first derive the infeasible asymptotically efficient score test for

*We are grateful to the co-editor and two referees for very helpful suggestions and comments that led to significant improvement on an earlier version of this paper. This research is supported by the Australian Research Council Discovery Grant DP1094010. Correspondence to: David Harris, E-mail: david.harris3@monash.edu
a known variance process and then prove the asymptotic equivalence of a feasible version of this test that estimates the variance process using a kernel-based non-parametric regression on the squares of the residuals of the model. This approach closely follows that taken by Xu and Phillips (2008) for an AR model and extends it to long memory testing in ARFIMA models.


In a closely related and important literature that deals with conditional heteroskedasticity, Baillie, Chung and Tieslau (1996), Ling and Li (1997), Li, Ling and McAleer (2002) and Ling (2003) consider efficient Maximum Likelihood estimation of an ARFIMA model in the presence of parametric GARCH models under Gaussianity. They however maintain the unconditional homoskedasticity assumption. More recently, Cavaliere, Nielsen and Taylor (2014) extend the consistency and asymptotic normality properties of the conditional sum-of-square estimators proposed by Hualde and Robinson (2011) to include both conditional and unconditional heteroskedasticity of a very general and unknown form.

The paper is structured as follows. In section 2 we introduce the ARFIMA model and the general model of heteroskedasticity for the disturbances, and derive score tests for the fractional differencing parameter. The score test based on a Gaussian likelihood with known variance process is shown to be asymptotically efficient. A robust score test based on a quasi likelihood that imposes a constant variance is derived and shown to be asymptotically inefficient. In section 3 we provide the main result of the paper, which is that a feasible test, based on re-weighting using a non-parametric variance estimator, is asymptotically equivalent to the efficient score test. This new test is shown to have superior asymptotic local power properties to the robust test, and hence to the robust tests of Kew and Harris (2009) and CNT. These properties are evaluated in finite samples in section 4, where it is shown that the new re-weighted test can achieve substantial power gains over robust tests for certain patterns of heteroskedasticity. In the following, $\overset{P}{\rightarrow}$ denotes (weak) convergence in probability and $\rightarrowD$ denotes convergence in distribution.
2 Infeasible tests

Suppose the observed time series $z_t$ satisfies
\[ \Delta^d z_t = y_t, \]
where $d$ is a known differencing parameter of any value, $\Delta^d$ is the Type II fractional differencing operator\(^1\)
\[ \Delta^d = \sum_{j=0}^{t-1} \Gamma (j - d) / (\Gamma (j + 1) \Gamma (-d)) L^j, \]
and $y_t$ follows an ARFIMA process of the form
\[ a(L; \psi_0) \Delta^{\theta_0} y_t = e_t, \]
where
\[ a(L; \psi) = \sum_{j=0}^{\infty} a_j(\psi) L^j = \frac{\phi(L)}{\eta(L)} \]
is a rational lag polynomial defined in terms of an autoregressive component $\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$ and a moving average component $\eta(L) = 1 - \sum_{j=1}^{q} \eta_j L^j$ of known fixed orders $p$ and $q$ respectively. In (2) the parameter vector $\psi$ is $\psi = (\phi_1, \ldots, \phi_p, \eta_1, \ldots, \eta_q)^t$, and $\theta_0$ and $\psi_0$ in (1) denote the true values of the parameters. Define the full parameter vector $\gamma = (\theta, \psi')^t$ on a parameter space satisfying the following assumption, which is the same one made in CNT.

Assumption R
The true values $(\theta_0, \psi_0)$ lie in the interior of a convex, compact parameter space $\Gamma = \Theta \times \Psi$, such that for all $\psi \in \Psi$, the polynomial functions $\phi(L)$ and $\eta(L)$ have no common roots and all their roots lie strictly outside the unit circle.

As in Robinson (1991, 1994) and Tanaka (1999), we wish to test
\[ H_0 : \theta_0 = 0, \]
against
\[ H^L_1 : \theta_0 < 0 \quad \text{or} \quad H^U_1 : \theta_0 > 0, \]
which is equivalent to testing the null hypothesis that $z_t$ is $I(d)$ for the known value of $d$. Henceforth we discuss the testing problem in terms of the observable time series $y_t$. It can equivalently be considered as a specification test of the choice of $d$ for the original time series $z_t$.

The disturbance term $e_t$ in (1) is assumed to have the heteroskedastic specification
\[ e_t = \sigma_t \varepsilon_t, \quad t = 1, 2, \ldots \]
where $\sigma_t^2$ is the unconditional variance, with $e_t = 0$ for $t \leq 0$. We do not assume a specific parametric functional form for $\sigma_t^2$. In this section $\sigma_t$ will be treated as known, with a feasible non-parametric estimator of $\sigma_t$ given in the next section. For the purposes of the likelihood-based efficiency theory in this section it will be assumed that
\[ \varepsilon_t \sim \text{i.i.d.} N(0, 1), \]
although this can be weakened for some subsequent asymptotic results, see Assumption E below.

For simplicity, the model in (1) ignores any nonstochastic variables ($x_t$) such as an unknown mean and trend terms. CNT Remark 2.3 provides a detailed discussion about how $x_t$ can be taken into account; see also Robinson (1994), Tanaka (1999) and Nielsen (2004).

\(^1\)See equation (4) of Tanaka (1999) for the computation of this operator.
2.1 Scores

The log-likelihood under (4) is

$$L (\gamma) = \text{constant} + \sum_{t=1}^{T} l_t (\gamma) ,$$

where

$$l_t (\gamma) = - \frac{1}{2} \left( \frac{e_t (\gamma)}{\sigma_t} \right)^2 , \quad e_t (\gamma) = a (L; \psi) \Delta^0 y_t .$$

Denote the score vector as $s_t (\gamma) = \left( s_{\theta,t} (\gamma) , s_{\psi,t} (\gamma) \right)'$ where

$$s_{\theta,t} (\gamma) = \frac{\partial l_t (\gamma)}{\partial \theta} = - \frac{a (L; \psi) \Delta^0 (\ln \Delta) y_t}{\sigma_t} \cdot \frac{e_t (\gamma)}{\sigma_t} ,$$

$$s_{\psi,t} (\gamma) = \frac{\partial l_t (\gamma)}{\partial \psi} = - \frac{a (L; \psi) \Delta^0 y_t}{\sigma_t} \cdot \frac{e_t (\gamma)}{\sigma_t} ,$$

in which $\ln \Delta = - \sum_{j=1}^{t-1} j^{-1} L^j$ and $a (L; \psi) = \sum_{j=1}^{\infty} a_{\psi,j} (\psi) L^j$ with $a_{\psi,j} (\psi) = \partial a_j (\psi) / \partial \psi$. It is shown in section 2.2 that an asymptotically efficient test of $H_0$ against $H_1^U$ or $H_1^L$ is based on these scores.

For comparison purposes, define the quasi log-likelihood function

$$K (\gamma) = \sum_{t=1}^{T} k_t (\gamma) , \quad k_t (\gamma) = - \frac{1}{2} e_t (\gamma)^2 ,$$

which includes no weights to allow for heteroskedasticity. The “quasi-score” vector is similarly denoted by $r_t (\gamma) = \left( r_{\theta,t} (\gamma) , r_{\psi,t} (\gamma) \right)'$ where $r_{\theta,t} (\gamma) = \partial k_t (\gamma) / \partial \theta$ and $r_{\psi,t} (\gamma) = \partial k_t (\gamma) / \partial \psi$. These unweighted scores provide the basis for the test statistics of Robinson (1994) and Tanaka (1999) derived under homoskedastic errors. CNT Theorem 1 shows that these homoskedastic score tests suffer from asymptotic size distortions in the presence of both conditional and unconditional heteroskedasticity. To resolve this problem, CNT propose a wild bootstrap method for these score tests and show that their testing procedure is robust to both conditional and unconditional heteroskedasticity of a very general and unknown form.

Our asymptotic distribution theory follows from a Central Limit Theorem for the scores, which will be shown to hold under the following assumptions on the components of $e_t$.

Assumption E

$\{ \varepsilon_t \}$ is a martingale difference sequence that satisfies: (i) $E (\varepsilon_t^2) = 1$; (ii) $\tau_{r,s} = E (\varepsilon_t^2 \varepsilon_{t-r} \varepsilon_{t-s})$ is uniformly bounded for all $t \geq 1, r \geq 0, s \geq 0$, where also $\tau_{r,0} > 0$ for all $r \geq 0$; (iii) for all integers $q$ such that $3 \leq q \leq 8$ and for all integers $r_1, \ldots, r_{q-2} \geq 1$, the $q$’th order cumulants $k_q (t, t - r_1, \ldots, t - r_{q-2})$ of $(\varepsilon_t, \varepsilon_{t-r_1}, \ldots, \varepsilon_{t-r_{q-2}})$ satisfy

$$\sup_{t} \sum_{r_1, \ldots, r_{q-2} = 1}^{\infty} |k_q (t, t - r_1, \ldots, t - r_{q-2})| < \infty;$$

(iv) $E (\varepsilon_t^2 | F_{t-1}) = 1$; and (v) $E (\varepsilon_t^4 | F_{t-1}) = \tau_{0,0}$ where $F_t$ is the $\sigma$-field of events generated by $\varepsilon_s, s \leq t$.

Assumption S

$\sigma_t$ satisfies $\sigma_t = \sigma (t/T)$, where $\sigma (.)$ is a non-stochastic function with at most a finite number
of points of discontinuity; moreover \( \sigma (.) \) is a measurable function on the interval \((0, 1]\) such that 
\[ 0 < \inf_{r \in (0, 1]} \sigma (r) \leq \sup_{r \in (0, 1]} \sigma (r) < \infty, \]
and \( \sigma (r) \) satisfies a (uniform) first-order Lipschitz condition except at the points of discontinuity.

Assumption E (i) to (iii) are the same as CNT Assumption V (b), while the remaining assumptions are made in Phillips and Xu (2006) and Xu and Phillips (2008) in the context of autoregressive models, and Hualde and Robinson (2011) and Nielsen (2014) in the context of fractionally integrated models. Assumption S, which was first introduced by Cavaliere (2004a), allows for a single structural break or multiple breaks in the volatility of the observed series \( z_t \). It also allows for smooth transition instead of abrupt variance breaks as well as linear or non-linear trending variances.

Define \( \gamma_0 = (0, \psi_0')' \) to be the parameter vector under \( H_0 \), and define the lag polynomial 
\[ b (L; \psi) = a_\psi (L; \psi) / a (L; \psi_0) = \sum_{j=1}^{\infty} b_j (\psi) L^j. \]
The following Lemma gives a joint Central Limit Theorem for \( s_t (\gamma_0) \) and \( r_t (\gamma_0) \).

**Lemma 1** Under \( H_0 \) and Assumptions E and S

\[
T^{-1/2} \sum_{t=1}^{T} \left( \begin{array}{c} s_t (\gamma_0) \\ r_t (\gamma_0) \end{array} \right) \overset{d}{\to} N \left( \begin{array}{c} 0 \\ 0 \end{array} \right), 
\left( \begin{array}{cc} V & V \int_0^1 \sigma^2 (s) ds \\ V \int_0^1 \sigma^4 (s) ds & V \int_0^1 \sigma^2 (s) ds \end{array} \right) \right)
\]

where

\[
V = \left( \begin{array}{cc} V_{\theta \theta} & V_{\theta \psi} \\ V_{\psi \theta} & V_{\psi \psi} \end{array} \right) = \left( \begin{array}{cc} \pi^2 / 6 & \sum_{j=1}^{\infty} j^{-1} b_j (\psi_0) \\ \sum_{j=1}^{\infty} j^{-1} b_j (\psi_0) & \sum_{j=1}^{\infty} j^{-1} b_j (\psi_0) b_j (\psi_0)' \end{array} \right).
\]

This lemma is fundamental to our subsequent asymptotic theory and the form of \( V \) is also required for the definitions of the effective score tests that now follow.

### 2.2 Effective score tests

Choi, Hall and Schick (1996, hereafter CHS) provide a general optimality theory of hypothesis testing in likelihood-based models with unknown nuisance parameters. We follow their approach in deriving an infeasible test as a function of the nuisance parameter, but defining asymptotic efficiency in a manner that anticipates the estimation of the nuisance parameter in section 3. CHS show that an asymptotically efficient test against a one-sided alternative uses the effective score test statistic, which in our case is

\[
\xi_T = \frac{T^{-1/2} \sum_{t=1}^{T} s_{\theta; \psi, t} (\gamma_0)}{V_{\theta \theta}^{1/2}},
\]

where

\[
s_{\theta; \psi, t} (\gamma) = s_{\theta, t} (\gamma) - s_{\psi, t} (\gamma)' V_{\psi \psi}^{-1} V_{\psi \theta}
\]
is the effective score (as defined by CHS and previously by Hall and Mathiason (1990)) and

\[
V_{\theta \theta} = V_{\theta \theta} - V_{\theta \psi} V_{\psi \psi}^{-1} V_{\psi \theta}
\]
is its asymptotic variance. CHS prove that if the log-likelihood has the LAN (Locally Asymptotically Normal) property then an asymptotically efficient test is based on \( \xi_T \). The effective quasi-score test statistic can be defined as

\[
\zeta_T = \frac{T^{-1/2} \sum_{t=1}^{T} r_{\theta; \psi, t} (\gamma_0)}{\sqrt{V_{\theta \theta} \int_0^1 \sigma^4 (s) ds}},
\]

5
where
\[ r_{\psi,t}(\gamma_0) = r_{\psi,t}(\gamma_0) - r_{\psi,t}(\gamma_0) V_{\psi}^{-1} V_{\psi}. \]

The lack of weighting for heteroskedasticity in this statistic results in an inefficient test relative to \( \xi_T \). This is all formalised in the following theorem.

**Theorem 2** Define the local sequence \( \gamma_T = \gamma_0 + T^{-1/2} g \) for a fixed finite vector \( g = (g_0, g_0)' \).

(a) Under \( \gamma_0 \) and Assumptions R, E, S, the log-likelihood \( L(\gamma) \) admits the LAN representation
\[ L(\gamma_T) - L(\gamma_0) = g' T^{-1/2} \sum_{t=1}^{T} s_t(\gamma_0) + \frac{1}{2} g' V g + o_p(1), \]
and the effective score and quasi-score statistics satisfy
\[ \xi_T, \zeta_T \sim N(0,1). \]

The score tests reject \( H_0 \) against \( H_1^U \) for \( \xi_T, \zeta_T > z_\alpha \), where \( z_\alpha \) is the 100 \( (1 - \alpha) \% \) percentile of the standard normal distribution, and similarly for the lower tailed tests.

(b) Under \( \gamma_T \) and Assumptions R, S and \( \epsilon_I \sim i.i.d. N(0,1) \), the statistics satisfy
\[ \xi_T \sim N\left( g_0' V_{\theta_0}^{1/2}, 1 \right), \]
\[ \zeta_T \sim N\left( g_0' V_{\theta_0}^{1/2} \nu, 1 \right) \]
where \( \nu = \int_0^1 \sigma(s)^2 ds / \sqrt{\int_0^1 \sigma(s)^4 ds} \), and the \( \xi_T \) test is asymptotically efficient.

Theorem 2 (a) extends the LAN property for ARFIMA models in Proposition 1 of Hallin, Taniguchi, Serroukh and Choy (1998) to allow for heteroskedasticity of a known form. Similarly, the rest of the Theorem extends the asymptotic efficiency results of Robinson (1994) and Tanaka (1999) for the score test to include heteroskedasticity of a known form. The asymptotic efficiency property requires Gaussianity, and the study of efficient tests under non-Gaussianity is left to future research. The asymptotic distributions under local alternatives in (12) and (13) are given under \( \epsilon_I \sim i.i.d. N(0,1) \) but those results could be generalised, as in CNT, to allow \( \epsilon_I \) to satisfy Assumption E.

It follows from (13) that the effective quasi-score test \( \zeta_T \) is asymptotically inefficient, with loss of power relative to \( \xi_T \) determined by the constant \( v \leq 1 \), with \( v = 1 \) under homoskedasticity. The smaller the value of \( v \), for a given value of \( g_0 \), the \( \zeta_T \) test suffers the larger power loss and this is illustrated in Kew and Harris (2009) Corollary 2 and Figure 2, CNT Figures 2 and 3, and CNT Remark 3.1. Xu and Phillips (2008) derives explicit expressions for \( v \) under a single shift variance model (see their Example 1) and a trending variance model (Example 2). Under a single downward shift variance model, it is clear that the quasi-score \( \zeta_T \) test suffers from substantial asymptotic power loss relative to the efficient score \( \xi_T \) test when this downward shift occurs early in the sample. Finite-sample power loss due to this variance model is also reflected in our Monte Carlo simulation results presented in Section 4.

### 3 Feasible Tests

In this section we propose feasible versions of the \( \xi_T \) and \( \zeta_T \) tests. For the quasi-score \( \zeta_T \) test in (10), \( \psi_0 \) is unknown and so we define the quasi-MLE \( \hat{\psi} \) under the null as
\[ \hat{\psi} = \arg \max_{\psi \in \Psi, \theta = 0} K(\gamma), \]
and \( \hat{\gamma} = (0, \hat{\psi})' \). That is, \( \hat{\psi} \) is the standard ARMA coefficient estimator assuming a constant variance, shown in CNT Lemma A.1 to be consistent under the null. The first-order condition for \( \hat{\psi} \) is \( \sum_{t=1}^{T} r_{\psi,t}(\hat{\gamma}) = 0 \) and hence \( T^{-1/2} \sum_{t=1}^{T} r_{\psi,t}(\hat{\gamma}) = T^{-1/2} \sum_{t=1}^{T} r_{\theta,t}(\hat{\gamma}) \) when substituting \( \hat{\gamma} \) for \( \gamma_0 \) in the numerator of \( \zeta_T \). A feasible denominator for \( \zeta_T \) is found by defining the estimated variance matrix

\[
\hat{W} = \left( \begin{array}{cc} \hat{W}_{\theta \theta} & \hat{W}_{\theta \psi} \\ \hat{W}_{\psi \theta} & \hat{W}_{\psi \psi} \end{array} \right) = T^{-1} \sum_{t=1}^{T} r_t(\hat{\gamma}) r_t(\hat{\gamma})',
\]

and \( \hat{W}_{\theta|\psi} = \hat{W}_{\theta \theta} - \hat{W}_{\theta \psi} \hat{W}_{\psi \psi}^{-1} \hat{W}_{\psi \theta} \). The feasible quasi-score statistic is then

\[
\hat{\zeta}_T = \frac{T^{-1/2} \sum_{t=1}^{T} r_{\theta,t}(\hat{\gamma})}{\hat{W}_{\theta|\psi}^{1/2}}.
\]

Theorem 3 below establishes that \( \hat{\zeta}_T \) is asymptotically equivalent to its infeasible counterpart.

Turning to the efficient score \( \xi_T \) test in (9), both \( \sigma_2^2 \) and \( \psi_0 \) are unknown. Following the approach of Xu and Phillips (2008) we estimate \( \sigma_2^2 \) non-parametrically and then adaptively estimate \( \psi_0 \). Xu and Phillips (2008) deal with an unconditionally heteroskedastic AR model and propose an Adaptive Least Squares estimator that has the same asymptotic distribution as the infeasible Generalised Least Squares estimator. We show that their method can be extended to our ARFIMA testing framework. Specifically we construct, under \( H_0 \), \( \hat{\gamma} = a \left( L; \hat{\psi} \right) y_t \) and define the non-parametric variance estimator as

\[
\hat{\sigma}_t^2 = \sum_{i=1}^{T} w_{ti} e_t^2, \tag{14}
\]

where \( w_{ti} = \left( \sum_{i=1}^{T} K_{ti} \right)^{-1} K_{ti} \) and \( K_{ti} = K \left( \frac{t-i}{b} \right) \), with \( K (\cdot) \) a bounded nonnegative continuous kernel function defined on the real line such that \( \int_{-\infty}^{\infty} K (z) \, dz = 1 \), and \( b \) is a bandwidth parameter. Following Xu and Phillips (2008), we define \( K_{ti} = 0 \), leaving out the \( t \)th observation of \( e_t^2 \) when estimating \( \hat{\sigma}_t^2 \). We use the cross validation method to select \( b \); i.e. we calculate \( CV (b) = T^{-1} \sum_{t=1}^{T} (\epsilon_t^2 - \hat{\sigma}_t^2)^2 \) for a range of values of \( b \) and select \( b^* \) such that \( CV (b) \) is minimised.

The feasible log-likelihood is then defined by replacing \( \sigma_t \) in (5) with \( \hat{\sigma}_t \) to give

\[
\hat{L} (\gamma) = \text{constant} + \sum_{t=1}^{T} \hat{e}_t (\gamma), \quad \hat{e}_t (\gamma) = -\frac{1}{2} \left( \frac{e_t (\gamma)}{\hat{\sigma}_t} \right)^2.
\]

Similarly the score vector \( \hat{s}_t (\gamma) = \left( \hat{s}_{\theta,t} (\gamma), \hat{s}_{\psi,t} (\gamma) \right)' \) is defined by replacing \( \sigma_t \) in (6) and (7) with \( \hat{\sigma}_t \). Define the feasible MLE \( \hat{\psi} \) under the null as

\[
\hat{\psi} = \arg \max_{\psi \in \Psi, \theta = 0} \hat{L} (\gamma),
\]

giving \( \hat{\gamma} = \left( 0, \hat{\psi} \right)' \). A feasible version of \( \xi_T \) is constructed similarly to \( \hat{\zeta}_T \), exploiting \( \sum_{t=1}^{T} \hat{s}_{\psi,t} (\gamma) = 0 \) in the numerator of the statistic. A variance matrix estimator may be defined as

\[
\hat{V} = \left( \begin{array}{cc} \hat{V}_{\theta \theta} & \hat{V}_{\theta \psi} \\ \hat{V}_{\psi \theta} & \hat{V}_{\psi \psi} \end{array} \right) = T^{-1} \sum_{t=1}^{T} \frac{\partial e_t (\gamma)}{\partial \gamma} \bigg|_{\gamma = \hat{\gamma}} \frac{\partial e_t (\gamma)}{\partial \gamma'} \bigg|_{\gamma = \hat{\gamma}}',
\]
based on the information equality holding once the likelihood has been weighted appropriately (asymptotically). An “outer product of gradients” estimator $\tilde{V} = T^{-1} \sum_{t=1}^{T} \tilde{s}_t (\gamma) \tilde{s}_t (\gamma)'$ can also be shown to be consistent. In either case we define $\hat{V}_{\theta \theta | \psi} = \hat{V}_{\theta \theta} - \hat{V}_{\theta \psi} \hat{V}_{\psi \psi}^{-1} \hat{V}_{\psi \theta}$, and the feasible score statistic is

$$\hat{\xi}_T = \frac{T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\theta t} (\gamma)}{\hat{V}_{\theta \theta | \psi}^{1/2}}.$$ 

To establish the asymptotic equivalence of $\hat{\xi}_T$ with its infeasible counterpart, we require the following Assumption B, which is from Xu and Phillips (2008).

**Assumption B**

As $T \to \infty, b + 1/Tb^2 \to 0$.

The following theorem gives the main result of our paper.

**Theorem 3** (a) Under $H_0$ and Assumptions S, R, and E, and B,

$$\hat{\xi}_T - \xi_T = o_p (1) \quad \text{and} \quad \hat{\xi}_T - \xi_T = o_p (1).$$

(b) These asymptotic equivalences also hold under $\gamma_T$ and Assumptions S, R, and B and $\varepsilon_t \sim \text{i.i.d.} N (0, 1)$.

The implication of this theorem is that the feasible tests $\hat{\xi}_T$ and $\tilde{\xi}_T$ inherit the asymptotic properties of $\xi_T$ and $\xi_T$ respectively. In particular, the non-parametrically variance-weighted test $\hat{\xi}_T$ is asymptotically efficient in the Gaussian model, and retains the same asymptotic properties as the correctly weighted test when $\varepsilon_t$ is not Gaussian.

The asymptotic efficiency of $\hat{\xi}_T$ that has been shown for unconditional heteroskedasticity is not expected to hold under conditional heteroskedasticity (that is if Assumption E(iv) were relaxed), and adaptation to the latter remains an open question in this context. Also CNT have shown that the wild bootstrap provides robust inference on long memory in the presence of conditional heteroskedasticity and the combination of their bootstrap with the kernel re-weighting developed here could be a productive topic for future research.

### 4 Simulation Evidence

This section compares the finite sample size and power properties of the various tests described in Theorem 3 when $\sigma_t^2$ follows a one-time structural break model with $\sigma_t^2 = \beta_1^2$ for $t \leq \lfloor \tau T \rfloor$ and $\sigma_t^2 = \beta_2^2$ for $t > \lfloor \tau T \rfloor$ for some $\tau \in (0, 1)$. We set, without loss of generality, $\beta_1 = 1$. Let $\delta = \beta_2 / \beta_1$ measure the size of the shift and, following Cavaliere (2004) and Cavaliere and Taylor (2007), we set $\delta = 1/3$ (downward variance shift) and $\tau = 0.2$ (early shift) and $\tau = 0.8$ (late shift). Simulation results for $\delta = 3$ (upward variance shift) are omitted since they are quite similar. For comparison purposes, we also give results for the homoskedastic case where $\delta = 1$. The innovation $\varepsilon_t$ is generated using the `rndn` routine in Gauss. The sample sizes $T = 100, 400$ and the number of replications is 50000.

Following Tanaka (1999), the data generating process for $y_t$ is $(1 - \phi_0 L) \Delta^{b_0} y_t = \varepsilon_t$. We test $H_0 : \theta_0 = 0$ vs $H_1^L : \theta_0 < 0$ or $H_1^U : \theta_0 > 0$ and we report the null rejection percentages based on a 5% nominal level. We follow Tanaka (1999) and set the values for the AR coefficient $\phi_0 = 0, 0.6$ and $-0.8$. If $\phi_0 = 0$ we let $\theta_0$ range between $-0.2$ and $0.2$ in steps of 0.05 and if $\phi_0 \not= 0$ we let $\theta_0$ range between $-0.4$ and $0.4$ in steps of 0.1.

In the simulations, the feasible quasi $\tilde{\xi}_T$ and efficient $\hat{\xi}_T$ test are computed assuming that the true orders $p$ and $q$ are known. CNT Remark 2.5 suggests that in practice the orders $p$ and $q$
can be selected by employing the usual Schwarz information criterion. As for the $\hat{\xi}_T$ test, the estimator $\hat{\sigma}_T^2$ in (14) is computed using the Gaussian kernel and the cross validation method to select $b$ and the estimator $\hat{V}$ is computed via the “outer product of gradients” method. \(^2\)

Table 1 reports the case when no autocorrelation is present (i.e. $\phi_0 = 0$). It shows that when the errors are homoskedastic ($\delta = 1$) the $\hat{\zeta}_T$ and $\hat{\xi}_T$ tests display acceptable size properties. The efficient $\hat{\xi}_T$ test does not yield any power gains over the quasi $\hat{\zeta}_T$ test and this is expected since $\nu$ in Theorem 2 is equal to 1. Also as expected, the empirical power of each test increases as $T$ increases for a given $\theta_0$, and the power increases as $|\theta_0|$ becomes large for a given $T$.

When $\sigma_t$ is not constant because of an early downward variance shift with $\tau = 0.2$, both our proposed $\hat{\zeta}_T$ and $\hat{\xi}_T$ tests display relatively good size properties. In all cases, the powers of the efficient score $\hat{\xi}_T$ test clearly exceed those of the quasi-score $\hat{\zeta}_T$ test and these observed power gains are expected since, by Xu and Phillips (2008) Example 1, $\nu = 0.63$, which is far less than 1. By comparison, we consider a late variance shift with $\tau = 0.8$. The efficient score $\hat{\xi}_T$ test no longer yields significant power gains over the quasi-score $\hat{\zeta}_T$ test and this too is expected since now $\nu = 0.92$, which is close to 1.

Table 2 reports the case when first order autocorrelation is present. Again $\hat{\zeta}_T$ continues to yield substantial power gains over $\hat{\zeta}_T$ in the presence of heteroskedasticity. Under homoskedasticity there are very small differences in terms of size and power between $\hat{\xi}_T$ and $\hat{\zeta}_T$. Results for $\tau = 0.8$ are not reported since, like the previous $\phi_0 = 0$ case in Table 1 and as expected, there are very small differences in size and power between the two tests.

\begin{table}[h]
\centering
\begin{tabular}{ccc|cccc|cccc}
\hline
 & & $H_1 : \theta_0 < 0$ & & $H_1 : \theta_0 > 0$ & \\
\hline
 & $T \theta_0$ & 0 & $-0.05$ & $-0.10$ & $-0.15$ & $-0.20$ & 0 & $0.05$ & $0.10$ & $0.15$ & $0.20$ \\
\hline
$\delta = 1$ & $\hat{\zeta}_T$ & 100 & 5.60 & 14.95 & 31.48 & 52.90 & 74.26 & 3.88 & 14.40 & 33.92 & 56.41 & 75.96 \\
 & & 400 & 5.47 & 34.30 & 79.86 & 98.27 & 99.96 & 4.43 & 35.02 & 80.09 & 97.49 & 99.90 \\
$\hat{\xi}_T$ & 100 & 5.55 & 14.73 & 31.16 & 52.42 & 73.75 & 3.97 & 14.26 & 33.21 & 55.39 & 74.78 \\
 & & 400 & 5.48 & 34.21 & 79.56 & 98.21 & 99.95 & 4.42 & 34.81 & 79.75 & 97.44 & 99.89 \\
$\tau = 0.2$ & $\hat{\zeta}_T$ & 100 & 5.62 & 11.56 & 19.63 & 30.47 & 43.38 & 3.57 & 10.23 & 22.01 & 38.97 & 56.88 \\
 & & 400 & 5.56 & 20.75 & 46.84 & 74.95 & 92.20 & 4.00 & 19.43 & 50.79 & 81.12 & 95.83 \\
$\hat{\xi}_T$ & 100 & 5.57 & 14.47 & 30.07 & 50.05 & 70.57 & 3.96 & 14.60 & 33.94 & 56.68 & 75.55 \\
 & & 400 & 5.44 & 33.31 & 77.69 & 97.58 & 99.92 & 4.47 & 34.58 & 79.06 & 97.15 & 99.86 \\
$\tau = 0.8$ & $\hat{\zeta}_T$ & 100 & 5.58 & 14.03 & 28.56 & 47.81 & 68.14 & 3.84 & 13.71 & 31.46 & 53.15 & 72.60 \\
 & & 400 & 5.53 & 30.99 & 73.74 & 96.37 & 99.84 & 4.27 & 31.61 & 74.90 & 95.86 & 99.68 \\
$\hat{\xi}_T$ & 100 & 5.50 & 14.34 & 29.62 & 49.63 & 70.30 & 4.02 & 14.31 & 33.04 & 55.08 & 74.15 \\
 & & 400 & 5.48 & 32.81 & 76.98 & 97.52 & 99.91 & 4.32 & 34.22 & 78.40 & 96.96 & 99.84 \\
\hline
\end{tabular}
\caption{Empirical size and power of tests when $\phi_0 = 0$}
\end{table}

\[^2\]We do not report results for the homoskedastic $S^T_{\tau 1}$ test in Tanaka (1999) because CNT demonstrate that, under a single downward variance shift model, this $S^T_{\tau 1}$ test, as expected, is severely over-sized even when the sample size increases.
Table 2: Empirical size and power of tests when $\phi_0 \neq 0$

<table>
<thead>
<tr>
<th>$T/\theta_0$</th>
<th>$\phi_0 = 0.6$</th>
<th>$H_1: \theta_0 &lt; 0$</th>
<th>$-0.1$</th>
<th>$-0.2$</th>
<th>$-0.3$</th>
<th>$-0.4$</th>
<th>$0$</th>
<th>$0.1$</th>
<th>$0.2$</th>
<th>$0.3$</th>
<th>$0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta = 1$</td>
<td>$\hat{c}_T$</td>
<td>100</td>
<td>7.16</td>
<td>11.14</td>
<td>17.01</td>
<td>25.34</td>
<td>36.96</td>
<td>3.55</td>
<td>7.27</td>
<td>10.88</td>
<td>16.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>6.25</td>
<td>18.40</td>
<td>41.62</td>
<td>70.72</td>
<td>90.14</td>
<td>3.58</td>
<td>14.55</td>
<td>28.62</td>
<td>39.69</td>
</tr>
<tr>
<td>$\xi_T$</td>
<td>100</td>
<td>7.23</td>
<td>11.21</td>
<td>16.93</td>
<td>25.22</td>
<td>36.54</td>
<td>3.59</td>
<td>7.26</td>
<td>10.72</td>
<td>15.55</td>
<td>24.16</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>6.19</td>
<td>18.37</td>
<td>41.50</td>
<td>70.57</td>
<td>90.16</td>
<td>3.60</td>
<td>14.56</td>
<td>28.55</td>
<td>39.34</td>
<td>54.09</td>
</tr>
<tr>
<td>$\tau = 0.2$</td>
<td>$\hat{c}_T$</td>
<td>100</td>
<td>5.04</td>
<td>8.70</td>
<td>13.92</td>
<td>20.31</td>
<td>26.92</td>
<td>4.17</td>
<td>10.29</td>
<td>14.09</td>
<td>16.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>5.53</td>
<td>14.89</td>
<td>25.85</td>
<td>40.28</td>
<td>57.28</td>
<td>3.78</td>
<td>13.28</td>
<td>22.21</td>
<td>27.75</td>
</tr>
<tr>
<td>$\xi_T$</td>
<td>100</td>
<td>8.16</td>
<td>12.44</td>
<td>17.82</td>
<td>24.58</td>
<td>33.16</td>
<td>3.91</td>
<td>7.95</td>
<td>10.74</td>
<td>13.25</td>
<td>18.37</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>6.37</td>
<td>18.54</td>
<td>40.15</td>
<td>66.81</td>
<td>87.66</td>
<td>3.67</td>
<td>16.35</td>
<td>30.61</td>
<td>36.14</td>
<td>43.52</td>
</tr>
<tr>
<td>$\phi_0 = -0.8$</td>
<td>$\hat{c}_T$</td>
<td>100</td>
<td>6.41</td>
<td>31.44</td>
<td>71.61</td>
<td>94.67</td>
<td>99.41</td>
<td>3.48</td>
<td>28.73</td>
<td>68.93</td>
<td>91.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>5.84</td>
<td>76.51</td>
<td>99.89</td>
<td>100.00</td>
<td>100.00</td>
<td>4.16</td>
<td>74.13</td>
<td>99.63</td>
<td>100.00</td>
</tr>
<tr>
<td>$\xi_T$</td>
<td>100</td>
<td>6.51</td>
<td>31.11</td>
<td>71.11</td>
<td>94.38</td>
<td>99.37</td>
<td>3.54</td>
<td>28.40</td>
<td>68.07</td>
<td>90.74</td>
<td>98.01</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>5.82</td>
<td>76.36</td>
<td>99.89</td>
<td>100.00</td>
<td>100.00</td>
<td>4.16</td>
<td>73.84</td>
<td>99.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$\tau = 0.2$</td>
<td>$\hat{c}_T$</td>
<td>100</td>
<td>7.33</td>
<td>23.43</td>
<td>46.33</td>
<td>68.55</td>
<td>84.16</td>
<td>3.68</td>
<td>20.50</td>
<td>51.60</td>
<td>78.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>6.44</td>
<td>46.80</td>
<td>90.16</td>
<td>99.55</td>
<td>99.99</td>
<td>3.70</td>
<td>45.42</td>
<td>92.75</td>
<td>99.85</td>
</tr>
<tr>
<td>$\xi_T$</td>
<td>100</td>
<td>6.62</td>
<td>30.00</td>
<td>67.85</td>
<td>92.19</td>
<td>98.85</td>
<td>3.74</td>
<td>30.00</td>
<td>69.97</td>
<td>91.65</td>
<td>98.19</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>5.76</td>
<td>74.42</td>
<td>99.82</td>
<td>100.00</td>
<td>100.00</td>
<td>4.23</td>
<td>73.49</td>
<td>99.54</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

References


A Proofs of main results

We first define some lag polynomials. Recalling \( \psi = (\phi_1, \ldots, \phi_p, \eta_1, \ldots, \eta_q)' \), and the short run lag polynomial \( a(L; \psi) = \phi(L)/\eta(L) \), the first derivative vector is the \( p + q \) dimensional vector

\[
a_{\psi}(L; \psi) = \frac{\partial a(L; \psi)}{\partial \psi'} = \left( \frac{\partial a(L; \psi)}{\partial \phi'} \frac{\partial a(L; \psi)}{\partial \eta'} \right)'
\]

where \( \phi = (\phi_1, \ldots, \phi_p)' \) and \( \eta = (\eta_1, \ldots, \eta_q)' \) and

\[
\frac{\partial a(L; \psi)}{\partial \phi_k} = \frac{-1}{1 - \sum_{j=1}^q \eta_j L^j} \cdot L^k, \quad \frac{\partial a(L; \psi)}{\partial \eta_k} = \frac{1 - \sum_{j=1}^p \phi_j L^j}{(1 - \sum_{j=1}^q \eta_j L^j)^2} \cdot L^k.
\]

The second derivative is the \((p + q) \times (p + q)\) matrix

\[
a_{\psi\psi}(L; \psi) = \frac{\partial^2 a(L; \psi)}{\partial \psi \partial \psi'} = \left( \begin{array}{cc} \frac{\partial^2 a(L; \psi)}{\partial \phi \partial \phi'} & \frac{\partial^2 a(L; \psi)}{\partial \phi \partial \eta'} \\ \frac{\partial^2 a(L; \psi)}{\partial \eta \partial \phi'} & \frac{\partial^2 a(L; \psi)}{\partial \eta \partial \eta'} \end{array} \right)
\]

in which

\[
\frac{\partial^2 a(L; \psi)}{\partial \phi_k \partial \phi_h} = 0, \quad \frac{\partial^2 a(L; \psi)}{\partial \phi_k \partial \eta_h} = \frac{-1}{(1 - \sum_{j=1}^q \eta_j L^j)^2} \cdot L^{k+h}, \quad \frac{\partial^2 a(L; \psi)}{\partial \eta_k \partial \eta_h} = \frac{2 \left(1 - \sum_{j=1}^p \phi_j L^j\right)}{(1 - \sum_{j=1}^q \eta_j L^j)^3} \cdot L^{k+h}.
\]

Then we can define

\[
c_0(L; \gamma) = \frac{a(L; \psi)}{a(L; \psi_0)} \Delta^\theta
\]

\[
c_1(L; \gamma) = \frac{\partial c_0(L; \gamma)}{\partial \gamma} = \left( \frac{a(L; \psi)}{a(L; \psi_0)} \Delta^\theta \ln \Delta \right)
\]

\[
c_2(L; \gamma) = \frac{\partial^2 c_0(L; \gamma)}{\partial \gamma \partial \gamma'} = \left( \frac{a(L; \psi)}{a(L; \psi_0)} \Delta^\theta (\ln \Delta)^2 \right) \frac{a_{\psi}(L; \psi)}{a(L; \psi_0)} \Delta^\theta \frac{a_{\psi \psi}(L; \psi)}{a(L; \psi_0)} \Delta^\theta \ln \Delta
\]

**Proof of Lemma 1**

Under \( H_0 \) we have \( y_t = a(L; \psi_0)^{-1} \epsilon_t \) and hence \( e_t(\gamma) = c_0(L; \gamma) \epsilon_t \). Similarly \( \partial e_t(\gamma) / \partial \gamma = c_1(L; \gamma) \epsilon_t \), giving

\[
s_t(\gamma) = -\frac{1}{\sigma_t^2} \frac{\partial e_t(\gamma)}{\partial \gamma} \epsilon_t(\gamma) = -\frac{c_1(L; \gamma) \epsilon_t}{\sigma_t}, \quad c_0(L; \gamma) \epsilon_t
\]

14
The Central Limit Theorem (CLT) in (A.11) of Lemma A.2 of CNT applies directly to
and (ii) to prove that 
Assumption S and similar arguments to Phillips and Xu (2006) Lemma A gives
in which, as required,
but not to the weighted version \( s_t^\# (\gamma_0) = \sigma_t^{-1} c_1 (L; \gamma_0) e_t \cdot \varepsilon_t \). However we can define
so that the reasoning leading to CNT’s result (A.11) can be immediately applied jointly to
\( (s_t^\# (\gamma_0), r_t (\gamma_0)) \). It then remains (i) to check the form of the asymptotic variance in the CLT
and (ii) to prove that
\[
T^{-1/2} \sum_{t=1}^{T} (s_t^\# (\gamma_0) - s_t (\gamma_0)) \xrightarrow{p} 0. \tag{A.1}
\]

(i) To derive the form of \( V \), use
\[
E \left[ \begin{pmatrix} s_t^\# (\gamma_0) \\ r_t (\gamma_0) \end{pmatrix} \begin{pmatrix} s_t^\# (\gamma_0) \\ r_t (\gamma_0) \end{pmatrix}' \right] = \begin{pmatrix} \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' & \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \sigma_{t-j} \sigma_t \\ \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' & \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \sigma_{t-j} \sigma_t^2 \end{pmatrix} \cdot \int_0^1 \sigma (s)^2 \, ds
\]
Assumption S and similar arguments to Phillips and Xu (2006) Lemma A gives
\[
T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \rightarrow \sum_{j=1}^{\infty} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)', \\
T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \sigma_{t-j} \sigma_t \rightarrow \sum_{j=1}^{\infty} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \cdot \int_0^1 \sigma (s)^2 \, ds \\
T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \sigma_{t-j}^2 \sigma_t^2 \rightarrow \sum_{j=1}^{\infty} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' \cdot \int_0^1 \sigma (s)^4 \, ds
\]
in which, as required,
\[
V = \sum_{j=1}^{\infty} c_{1,j} (\gamma_0) c_{1,j} (\gamma_0)' = \sum_{j=1}^{\infty} \begin{pmatrix} j^{-1} \\ b_j (\psi_0) \end{pmatrix}' \begin{pmatrix} j^{-1} \\ b_j (\psi_0) \end{pmatrix} \cdot \int_0^1 \sigma (s)^2 \, ds
\]
(ii) To show (A.1), we write
\[
T^{-1/2} \sum_{t=1}^{T} (s_t^\# (\gamma_0) - s_t (\gamma_0)) = T^{-1/2} \sum_{t=2}^{T} \left( \sum_{j=1}^{t-1} c_{1,j} (\gamma_0) \left( \frac{\sigma_{t-j}}{\sigma_t} - 1 \right) \varepsilon_{t-j} \right)
\]
which is shown to satisfy (A.1) if we prove the generic result
\[
r_{s,T} = T^{-1/2} \sum_{t=2}^{T} \left( \varepsilon_{t} \sum_{j=1}^{t-1} c_{j} \left( \frac{\sigma_{t-j}}{\sigma_t} - 1 \right) \varepsilon_{t-j} \right) \xrightarrow{p} 0
\]
for any coefficients $c_j$ satisfying $\sum_{j=1}^{\infty} c_j^2 < \infty$. Using that $\varepsilon_t \sim \text{i.i.d.} \ (0, 1)$, it follows that

$$E (r_{s,t}^2) = T^{-1} \sum_{t=2}^{T} \sum_{j=1}^{t-1} c_j^2 \left( \frac{\sigma_j - \sigma_t}{\sigma_t} - 1 \right)^2$$

$$\leq \frac{1}{\inf_r \sigma (r)^2} \sum_{j=1}^{T-1} c_j^2 T^{-1} \sum_{t=j+1}^{T} \left( \sigma \left( \frac{t}{T} \right) - \sigma \left( \frac{t-j}{T} \right) \right)^2$$

$$= \frac{1}{\inf_r \sigma (r)^2} \sum_{j=1}^{T-1} c_j^2 T^{-1} \sum_{t=j+1}^{T} \left( \sum_{i=t-j}^{t} \sigma \left( \frac{i+1}{T} \right) - \sigma \left( \frac{i}{T} \right) \right)^2.$$

The proof that this disappears under Assumption S is given allowing for a single discontinuity in $\sigma (.)$ to illustrate, with extension to a finite number of discontinuities following identically. Suppose there is a single discontinuity at $\tau \in (0, 1)$ such that $\lim_{\tau \uparrow \tau} \sigma (r) - \sigma (\tau) = \delta$, $0 < \delta < \infty$. It follows that $\limsup_{\tau \uparrow \tau} \left| \frac{\sigma (\tau T + 1)}{T} - \sigma (\frac{\tau T}{T}) \right| = \delta$, while for $i \neq \tau$ the Lipschitz condition imposed in Assumption S implies that $|\sigma (\frac{i+1}{T}) - \sigma (\frac{i}{T})| \leq \frac{\ell}{T}$ for some $\ell < \infty$. Thus

$$\sum_{t=j+1}^{T} \left( \sum_{i=t-j}^{t} \sigma \left( \frac{i+1}{T} \right) - \sigma \left( \frac{i}{T} \right) \right)^2$$

$$= \sum_{t=j+1}^{T} \left( \sum_{i=t-j}^{t} \left( \sigma \left( \frac{i+1}{T} \right) - \sigma \left( \frac{i}{T} \right) \right) + 1 \right)^2 \left( \sigma \left( \frac{\tau T + 1}{T} \right) - \sigma \left( \frac{\tau T}{T} \right) \right)^2$$

$$\leq 2 \sum_{t=j+1}^{T} \left( \sum_{i=t-j}^{t} \left( \sigma \left( \frac{i+1}{T} \right) - \sigma \left( \frac{i}{T} \right) \right) \right)^2 + 2 \left( \sigma \left( \frac{\tau T + 1}{T} \right) - \sigma \left( \frac{\tau T}{T} \right) \right)^2 \sum_{t=j+1}^{T} 1_{t-j \leq \tau T} \leq 2 \sum_{t=j+1}^{T} \left( \frac{\ell T}{T} \right)^2 + 2 j \left( \sigma \left( \frac{\tau T + 1}{T} \right) - \sigma \left( \frac{\tau T}{T} \right) \right)^2.$$

Using this bound in (A.2) gives

$$E (r_{s,T}^2) \leq \frac{2}{\inf_r \sigma (r)^2} \sum_{j=1}^{T-1} c_j^2 T^{-1} \left( \sum_{t=j+1}^{T} \left( \frac{\ell T}{T} \right)^2 + j \left( \sigma \left( \frac{\tau T + 1}{T} \right) - \sigma \left( \frac{\tau T}{T} \right) \right)^2 \right)$$

$$\leq 2 \left( T^{-1} \sum_{j=1}^{T-1} j c_j^2 \right) \left( \ell^2 + \left( \sigma \left( \frac{\tau T + 1}{T} \right) - \sigma \left( \frac{\tau T}{T} \right) \right)^2 \right) \to 0,$$

since $\sum_{j=1}^{\infty} c_j^2 < \infty$ implies the Cesaro sum $T^{-1} \sum_{j=1}^{T-1} j c_j^2 \to 0$.

**Proof of Theorem 2**

(a) The LAN representation is based on the standard mean value expansion

$$\lambda_T (g) = L (\gamma_T) - L (\gamma_0) = g' T^{-1/2} \sum_{t=1}^{T} s_t (\gamma_0) + \frac{1}{2} g' T^{-1} \sum_{t=1}^{T} h_t (\gamma_t^\perp) g$$

16
where \( \gamma_T \) is a convex combination of \( \gamma_T \) and \( \gamma_0 \) and
\[
 h_t(\gamma) = \frac{\partial^2 l_t(\gamma)}{\partial \gamma \partial \gamma'} = -\frac{1}{\sigma_t^2} \left( e_t(\gamma) \frac{\partial^2 e_t(\gamma)}{\partial \gamma \partial \gamma'} + \frac{\partial e_t(\gamma)}{\partial \gamma} \frac{\partial e_t(\gamma)}{\partial \gamma'} \right).
\]
Given Lemma 1, it remains to show that
\[
 T^{-1} \sum_{t=1}^{\infty} h_t(\gamma_T^*) \stackrel{P}{\rightarrow} -V.
\]
Under \( H_0 \) we write \( \frac{\partial e_t(\gamma)}{\partial \gamma} = c_2(L; \gamma) e_t \), and hence
\[
 h_t(\gamma) = -\frac{1}{\sigma_t^2} \left( c_0(L; \gamma) e_t \cdot c_2(L; \gamma) e_t + (c_1(L; \gamma) e_t) (c_1(L; \gamma) e_t) \right).
\]
We define
\[
 h_t^\#(\gamma) = -\left( c_0(L; \gamma) \varepsilon_t \cdot c_2(L; \gamma) \varepsilon_t + (c_1(L; \gamma) \varepsilon_t) (c_1(L; \gamma) \varepsilon_t) \right)
\]
and show
\[
 T^{-1} \sum_{t=1}^{\infty} \left( h_t^\#(\gamma_T) - h_t(\gamma_T) \right) \stackrel{P}{\rightarrow} 0, 
\]
so that (A.12) of CNT applies to conclude the required convergence
\[
 T^{-1} \sum_{t=1}^{\infty} h_t^\#(\gamma_T) \stackrel{P}{\rightarrow} -V.
\]
To show (A.3), we show that
\[
 T^{-1} \sum_{t=1}^{\infty} \left( h_t^\#(\gamma) - h_t(\gamma) \right) = -T^{-1} \sum_{t=1}^{\infty} \left( \sum_{j=0}^{t-1} c_{0,j}(\gamma) \varepsilon_{t-j} \cdot \sum_{j=0}^{t-1} c_{2,j}(\gamma) \varepsilon_{t-j} \right)
\]
\[
 -T^{-1} \sum_{t=1}^{\infty} \left( \sum_{j=0}^{t-1} c_{1,j}(\gamma) \varepsilon_{t-j} \cdot \sum_{j=0}^{t-1} c_{1,j}(\gamma) \varepsilon_{t-j} \right)' \]
\[
 -T^{-1} \sum_{t=1}^{\infty} \left( \sum_{j=0}^{t-1} c_{1,j}(\gamma) \frac{\sigma_{t-j}}{\sigma_t} \varepsilon_{t-j} \cdot \sum_{j=0}^{t-1} c_{1,j}(\gamma) \frac{\sigma_{t-j}}{\sigma_t} \varepsilon_{t-j} \right)' \]
\[
 \quad -T^{-1} \sum_{t=1}^{\infty} \left( \sum_{j=0}^{t-1} c_{1,j}(\gamma) \frac{\sigma_{t-j}}{\sigma_t} \varepsilon_{t-j} \cdot \sum_{j=0}^{t-1} c_{1,j}(\gamma) \frac{\sigma_{t-j}}{\sigma_t} \varepsilon_{t-j} \right) \quad \text{P. (A.4)}
\]
uniformly on \( \Gamma_h = \Theta_h \times \Psi \), where \( \Theta_h = [-1, 1/2 - \varepsilon] \) for any \( \varepsilon > 0 \), and \( \Psi \) satisfies Assumption R. This parameter space is large enough to accommodate \( \gamma_T \) as required in (A.3), at least for large enough \( T \). We note that each element of the coefficients of \( c_0(L; \gamma) \), \( c_1(L; \gamma) \) and \( c_2(L; \gamma) \) are square summable uniformly on \( \Gamma_h \). Therefore each element of each term in (A.4) and (A.5) will be shown to satisfy the general convergence
\[
 T^{-1} \sum_{t=1}^{\infty} \left[ \sum_{j=0}^{t-1} c_{1,j}(\gamma) \frac{\sigma_{t-j}}{\sigma_t} \varepsilon_{t-j} \right] \left( \sum_{j=0}^{t-1} c_{2,j}(\gamma) \frac{\sigma_{t-j}}{\sigma_t} \varepsilon_{t-j} \right) - \left( \sum_{j=0}^{t-1} c_{1,j} \varepsilon_{t-j} \right) \left( \sum_{j=0}^{t-1} c_{2,j} \varepsilon_{t-j} \right)' \quad \text{P. (A.5)}
\]

where \( c_{0,j}(\gamma) \) and the individual elements of \( c_{1,j}(\gamma) \) and \( c_{2,j}(\gamma) \) are represented as generic scalar coefficients \( c_{1,j} \) and \( c_{2,j} \) that satisfy \( \sum_{j=0}^{\infty} c_{1,j}^2 < \infty \) and \( \sum_{j=0}^{\infty} c_{2,j}^2 < \infty \) (the \( \gamma \) can be dropped from the generic notation because of the uniform square summability of the coefficients on \( \Gamma_h \)). This is sufficient for (A.5) and hence (A.3).

In (A.6) the convergence in probability of \( T^{-1} \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} c_{1,j} \varepsilon_{t-j} \right) \left( \sum_{j=0}^{t-1} c_{2,j} \varepsilon_{t-j} \right) \) to some limit is standard, while that of \( T^{-1} \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} c_{1,j} \left( \sigma_{t-j} / \sigma_t \right) \varepsilon_{t-j} \right) \left( \sum_{j=0}^{t-1} c_{2,j} \left( \sigma_{t-j} / \sigma_t \right) \varepsilon_{t-j} \right) \) follows by similar arguments while exploiting \( 0 < \inf_r \sigma(r) \) and \( \sup_r \sigma(r) < \infty \). The proof of (A.6) therefore consists of verifying that both terms have the same probability limit. This follows by using

\[
E \left( \sum_{j=0}^{t-1} c_{1,j} \left( \sigma_{t-j} / \sigma_t \right) \varepsilon_{t-j} \right) \left( \sum_{j=0}^{t-1} c_{2,j} \left( \sigma_{t-j} / \sigma_t \right) \varepsilon_{t-j} \right) = \sum_{j=0}^{t-1} c_{1,j} c_{2,j} \left( \sigma_{t-j} / \sigma_t \right)^2 ,
\]

and arguing that the average difference between these converges to zero:

\[
T^{-1} \sum_{t=1}^{T} \frac{1}{\sigma_t^2} \sum_{j=0}^{t-1} c_{1,j} c_{2,j} \left( \sigma_{t-j}^2 / \sigma_t^2 \right) \leq \frac{1}{\inf_r \sigma(r)^2} T^{-1} \sum_{j=0}^{t-1} |c_{1,j}| \left| c_{2,j} \right| \sum_{t=j+1}^{T} \left| \left( \sigma \left( \frac{t}{T} \right)^2 - \sigma \left( \frac{t-j}{T} \right)^2 \right) \right| \leq \frac{2 \sup_r \sigma(r)}{\inf_r \sigma(r)^2} T^{-1} \sum_{j=0}^{t-1} |c_{1,j}| \left| c_{2,j} \right| \sum_{t=j+1}^{T} \left| \left( \sigma \left( \frac{i+1}{T} \right) - \sigma \left( \frac{i}{T} \right) \right) \right| \leq \frac{2 \sup_r \sigma(r)}{\inf_r \sigma(r)^2} \left( T^{-1} \sum_{j=0}^{t-1} \frac{1}{2} \right)^{1/2} \left( T^{-1} \sum_{j=0}^{t-1} \frac{1}{2} \right)^{1/2} \left( \left( \ell + \left| \sigma \left( \frac{\tau T}{T} + 1 \right) - \sigma \left( \frac{\tau T}{T} \right) \right) \right) \to 0.
\]

(b) Define the shorthand notation \( \omega^2 \psi = V_{\theta \psi \psi} \), \( v_2 = \int_0^1 \sigma(s)^2 ds \) and \( v_4 = \int_0^1 \sigma(s)^4 ds \). The definitions of \( \xi_T \) and \( \zeta_T \) and the LAN representation in (a) give

\[
\begin{pmatrix} \xi_T \\ \zeta_T \\ \chi_T(g) - \frac{1}{2} g' V g \end{pmatrix} = \begin{pmatrix} \frac{1}{\omega \xi} - \frac{V_{\theta \psi V^{-1}_{\psi \psi}}}{\omega \xi} & 0 & 0 \\ 0 & \frac{1}{\omega \xi v_4^{1/2}} & \frac{V_{\theta \psi V^{-1}_{\psi \psi}}}{\omega \xi v_4^{1/2}} \\ g_\theta & g_\psi & 0 \end{pmatrix} T^{-1/2} \sum_{t=1}^{T} \begin{pmatrix} s_{\theta,t}(\gamma_0) \\ s_{\psi,t}(\gamma_0) \\ r_{\theta,t}(\gamma_0) \\ r_{\psi,t}(\gamma_0) \end{pmatrix} + o_p(1)
\]

\[
\sim N \left( \begin{pmatrix} \frac{1}{\omega \xi} - \frac{V_{\theta \psi V^{-1}_{\psi \psi}}}{\omega \xi} & 0 & 0 \\ 0 & \frac{1}{\omega \xi v_4^{1/2}} & \frac{V_{\theta \psi V^{-1}_{\psi \psi}}}{\omega \xi v_4^{1/2}} \\ g_\theta & g_\psi & 0 \end{pmatrix} \right) \begin{pmatrix} v_2/v_4^{1/2} \\ v_2/v_4^{1/2} \\ g_\theta \omega \xi \\ g_\psi \omega \xi \\ g_\theta \omega \xi \\ g_\psi \omega \xi \\ g_\theta \omega \xi \\ g_\psi \omega \xi \\ g_\theta \omega \xi \\ g_\psi \omega \xi \end{pmatrix}
\]
by Lemma 1. The null distributions of $\xi_T$ and $\zeta_T$ follow immediately. The distributions of $\xi_T$ and $\zeta_T$ under $\gamma_T$ then follow from Le Cam’s third lemma. The asymptotic efficiency of the test based on $\xi_T$ follows from Theorem 1 of CHS.

**Proof of Theorem 3.**

(a) The proof for $\hat{\zeta}_T$ is essentially a special case (with $\hat{\sigma}_T^2 = 1$) of that for $\hat{\xi}_T$, so we focus on the efficient test. It is shown in Lemma B that $\hat{\psi} \overset{D}{=} \hat{\psi}_0$. For clarity we write $(\theta, \psi)$ for $\gamma$ in the rest of the proof of this Theorem. Define the Hessian

$$\hat{h}_t(\theta, \psi) = \begin{pmatrix} \hat{h}_{\theta\theta,t}(\theta, \psi) & \hat{h}_{\theta\psi,t}(\theta, \psi) \\ \hat{h}_{\psi\theta,t}(\theta, \psi) & \hat{h}_{\psi\psi,t}(\theta, \psi) \end{pmatrix}$$

as the partitioned matrix of second derivatives of $\hat{l}_t(\theta, \psi)$. The mean value equality

$$\hat{s}_{\psi,t} \left( 0, \hat{\psi} \right) = \hat{s}_{\psi,t} \left( 0, \psi_0 \right) + \hat{h}_{\psi\psi,t} \left( 0, \psi^* \right) \left( \hat{\psi} - \psi_0 \right)$$

in the first order conditions $\sum_{t=1}^{T} \hat{s}_{\psi,t} \left( 0, \hat{\psi} \right) = 0$ gives

$$\sqrt{T} \left( \hat{\psi} - \psi_0 \right) = - \left( T^{-1} \sum_{t=1}^{T} \hat{h}_{\psi\psi,t} \left( 0, \psi^* \right) \right)^{-1} T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\psi,t} \left( 0, \psi_0 \right).$$

Another mean value expansion in $\hat{s}_{\theta,t} \left( 0, \hat{\psi} \right)$ gives

$$\begin{align*}
T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\theta,t} \left( 0, \hat{\psi} \right) &= T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\theta,t} \left( 0, \psi_0 \right) + T^{-1} \sum_{t=1}^{T} \hat{h}_{\theta\psi,t} \left( 0, \psi^* \right) \cdot \sqrt{T} \left( \hat{\psi} - \psi_0 \right) \\
&= T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\theta,t} \left( 0, \psi_0 \right) - T^{-1} \sum_{t=1}^{T} \hat{h}_{\theta\psi,t} \left( 0, \psi^* \right) \left( T^{-1} \sum_{t=1}^{T} \hat{h}_{\psi\psi,t} \left( 0, \psi^* \right) \right)^{-1} T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\psi,t} \left( 0, \psi_0 \right).
\end{align*}$$

In Lemma B we show that

$$T^{-1/2} \sum_{t=1}^{T} \left( \hat{s}_t \left( 0, \psi_0 \right) - s_t \left( 0, \psi_0 \right) \right) \overset{D}{=} 0, \quad \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left( \hat{h}_t \left( 0, \psi \right) - h_t \left( 0, \psi \right) \right) \overset{D}{=} 0. \quad (A.7)$$

Also the Hessian is stochastically equicontinuous. To see this, for any $\hat{\psi} \overset{D}{=} \psi_0$ we have

$$\begin{align*}
T^{-1} \sum_{t=1}^{T} \left( h_t \left( 0, \hat{\psi} \right) - h_t \left( 0, \psi_0 \right) \right) &= T^{-1} \sum_{t=1}^{T} \left( h_t \left( 0, \hat{\psi} \right) - h_t^\# \left( 0, \hat{\psi} \right) \right) \\
&\quad + T^{-1} \sum_{t=1}^{T} \left( h_t^\# \left( 0, \hat{\psi} \right) - h_t^\# \left( 0, \psi_0 \right) \right) - T^{-1} \sum_{t=1}^{T} \left( h_t \left( 0, \psi_0 \right) - h_t^\# \left( 0, \psi_0 \right) \right).
\end{align*}$$

19
The first and third terms $\frac{p}{T} \to 0$ follow from equation (A.5) and the second term $\frac{p}{T} \to 0$ follows from CNT equation A.12. These, together with $\psi^*, \psi^{**} \frac{p}{T} \to \psi_0$, are sufficient to conclude that

$$T^{-1/2} \sum_{t=1}^{T} \hat{s}_{\theta,t}(0, \tilde{\psi}) = T^{-1/2} \sum_{t=1}^{T} s_{\theta,t}(0, \psi_0)$$

$$- T^{-1} \sum_{t=1}^{T} h_{\theta,\psi,t}(0, \psi_0) \left( T^{-1} \sum_{t=1}^{T} h_{\psi,\psi,t}(0, \psi_0) \right)^{-1} T^{-1/2} \sum_{t=1}^{T} s_{\psi,t}(0, \psi_0) + o_p(1)$$

$$= T^{-1/2} \sum_{t=1}^{T} s_{\theta|\psi,t}(0, \psi_0) + o_p(1).$$

The consistency of $\tilde{V}$ is implied by the arguments that lead to (A.7), see below from (B.28) onwards. The “outer product of gradients” estimator is consistent by similar lengthy algebra. This proves asymptotic equivalence of the feasible and infeasible statistics under $H_0$.

(b) Le Cam’s third lemma implies equivalence under $\gamma_T$ for both $\hat{\zeta}_T$ and $\hat{\xi}_T$. That is, $\hat{\zeta}_T - \zeta_T \frac{p}{T} \to 0$ and $\hat{\xi}_T - \xi_T \frac{p}{T} \to 0$ imply that the joint distributions of $\left(\hat{\xi}_T, \hat{\zeta}_T, \lambda_T(g)\right)$ and $\left(\xi_T, \zeta_T, \lambda_T(g)\right)$ are asymptotically equivalent, so the conclusions of Theorem 2 apply to $\left(\hat{\xi}_T, \hat{\zeta}_T, \lambda_T(g)\right)$. 
B Supplementary Results and Proofs

In this section, we give additional results which are required for proving the theorems. Let $C$ denote a generic positive constant.

**Lemma A.** Let $\hat{\sigma}_t^2 = \sum_{i=1}^{T} w_{ti} \sigma_i^2$ and $\hat{\sigma}_t^2 = \sum_{i=1}^{T} w_{ti} e_i^2$. Under Assumptions R, S, E and B,

(a) $\hat{\psi} \xrightarrow{P} \psi_0$;
(b) $\sqrt{T} \left( \hat{\psi} - \psi_0 \right) \rightsquigarrow N \left( 0, \nu^{-2} \Sigma_{\psi\psi}^{-1} \right)$;
(c) let $t = [Tr]$, for any fixed $r \in (0, 1]$, $\frac{1}{Tb} \sum_{i=1}^{T} K_{ti} \rightarrow \int_{-\infty}^{\infty} K(z) dz = 1$;
(d) $\max_{t,i} w_{ti} = O \left( \frac{1}{Tb} \right)$;
(e) $\min_{1 \leq t \leq T} \hat{\sigma}_t^2 \geq C > 0$;
(f) $\max_{1 \leq t \leq T} E \left[ |\hat{\sigma}_t^2 - \sigma_t^2|^4 \right] = O \left( \frac{1}{(Tb)^2} \right)$;
(g) $\max_t |\hat{\sigma}_t^2 - \sigma_t^2| = O_p \left( T^{-1/4} b^{-1/2} \right)$;
(h) $\left( \min_{1 \leq t \leq T} \hat{\sigma}_t^2 \right)^{-1} = O_p (1)$;
(i) $\max_{1 \leq t \leq T} |\hat{\sigma}_t^2 - \sigma_t^2| = O_p \left( \frac{1}{\sqrt{T}} \right)$;
(j) $\left( \min_{1 \leq t \leq T} \hat{\sigma}_t^2 \right)^{-1} = O_p (1)$;
(k) $\sum_{t=1}^{T} (\hat{\sigma}_t^2 - \sigma_t^2)^2 = O_p \left( 1/\sqrt{T} \right)$;
(l) $T^{-1} \sum_{t=1}^{T} |\hat{\sigma}_t^2 - \sigma_t^2| = o(1)$.

**Lemma B.** Under Assumptions R, S, E and B,

(a) $T^{-1/2} \sum_{t=1}^{T} \left( \hat{s}_t \left( 0, \psi_0 \right) - s_t \left( 0, \psi_0 \right) \right) \xrightarrow{P} 0$,

where $\hat{s}_t \left( 0, \psi_0 \right)$ and $s_t \left( 0, \psi_0 \right)$ represent, respectively, the score vectors $\hat{s}_t \left( \gamma \right)$ and $s_t \left( \gamma \right)$ but evaluated at $\theta = 0$ and $\psi = \psi_0$;

(b) $\sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left( \hat{h}_t \left( 0, \psi \right) - h_t \left( 0, \psi \right) \right) \xrightarrow{P} 0$,

where $\hat{h}_t \left( 0, \psi \right)$ and $h_t \left( 0, \psi \right)$ represent, respectively, the matrix of second derivatives of $\hat{L}_t \left( \gamma \right)$ and $L_t \left( \gamma \right)$ but evaluated at $\theta = 0$;

(c) Let $\bar{\psi} = \arg \max_{\psi \in \Psi, \theta = 0} L \left( \gamma \right)$. Then

$\bar{\psi} \xrightarrow{P} \psi_0$;

(d) $\bar{\psi} \xrightarrow{P} \psi_0$. 

1
Part (a) The proof is given in CNT Lemma A.1.

Part (b) This result follows from Cavaliere, Nielsen and Taylor (2014) Lemma C.3, CNT Lemma A.2 and the standard Mean Value Theorem.

Parts (c) to (h) follows directly from Xu and Phillips (2008) Lemma A(c) to Lemma A(h).

Part (i) For \( \tilde{\psi}, \psi_0 \in \Psi \) and for some \( \psi^* \) in between \( \tilde{\psi} \) and \( \psi_0 \),

\[
\hat{e}_i = a(L; \tilde{\psi}) y_i 
= a(L; \psi_0) y_i + (a_\psi(L; \psi^*) y_i)' (\tilde{\psi} - \psi_0) 
= e_i + (a_\psi(L; \psi^*) y_i)' (\tilde{\psi} - \psi_0) 
\]

and so

\[
\hat{e}_i^2 = e_i^2 + 2e_i (a_\psi(L; \psi^*) y_i)' (\tilde{\psi} - \psi_0) + (\tilde{\psi} - \psi_0)' (a_\psi(L; \psi^*) y_i) (a_\psi(L; \psi^*) y_i)' (\tilde{\psi} - \psi_0).
\]

Denote \( \hat{\psi} = (\hat{\psi}_1, \ldots, \hat{\psi}_{p+q})' \), \( \psi_0 = (\psi_{0,1}, \ldots, \psi_{0,p+q})' \) and \( \psi^* = (\psi^*_1, \ldots, \psi^*_{p+q})' \). We have \( a_\psi(L; \psi) y_i = (z_{1,i}(\psi), \ldots, z_{p+q,i}(\psi))' \) where \( z_{k,i}(\psi) = \sum_{j=1}^{s-1} \varphi_{k,j}(\psi) e_{i-j} \) and \( \varphi_{k,j}(\psi) \) decays exponentially in \( j \) for all \( \psi \in \Psi \) under Assumption R and for \( k = 1, \ldots, p+q \).

We write

\[
\hat{\sigma}_t^2 - \tilde{\sigma}_t^2 = \sum_{i=1}^{T} w_{ti} (\hat{e}_i^2 - e_i^2) = A_{1,t} + A_{2,t}, \tag{B.1}
\]

where

\[
A_{1,t} = 2 \sum_{i=1}^{T} w_{ti} e_i (a_\psi(L; \psi^*) y_i)' (\hat{\psi} - \psi_0) = 2 \sum_{j=1}^{p+q} (\hat{\psi}_j - \psi_{0,j}) \sum_{i=1}^{T} w_{ti} e_i z_{j,i}(\psi^*),
\]

and

\[
A_{2,t} = \sum_{i=1}^{T} w_{ti} (\hat{\psi} - \psi_0)' (a_\psi(L; \psi^*) y_i) (a_\psi(L; \psi^*) y_i)' (\hat{\psi} - \psi_0)
= \sum_{k=1}^{p+q} \sum_{l=1}^{p+q} (\hat{\psi}_l - \psi_{0,l}) \sum_{i=1}^{T} w_{ti} z_{k,i}(\psi^*) z_{l,i}(\psi^*).
\]

Now the required results that

\[
\max_{1 \leq t \leq T} |\hat{\sigma}_t^2 - \tilde{\sigma}_t^2| \leq \max_{1 \leq t \leq T} |A_{1,t}| + \max_{1 \leq t \leq T} |A_{2,t}| = O_p \left( \frac{1}{\sqrt{Tb}} \right)
\]

will follow from

\[
\max_{1 \leq t \leq T} |A_{1,t}| = O_p \left( \frac{1}{\sqrt{Tb}} \right) \tag{B.2}
\]

and

\[
\max_{1 \leq t \leq T} |A_{2,t}| = O_p \left( \frac{1}{Tb} \right), \tag{B.3}
\]

as we will now show.
First note that
\[ \sup_{\psi \in \Psi} T^{-1} \sum_{i=1}^{T} |z_{k,i} (\psi) z_{l,i} (\psi)| = O_p (1) \quad \text{for } k, l = 1, \ldots, p + q \quad (B.4) \]

and
\[ \sup_{\psi \in \Psi} T^{-1} \sum_{i=1}^{T} |e_i z_{j,i} (\psi)| = O_p (1) \quad \text{for } j = 1, \ldots, p + q. \quad (B.5) \]

We will only show (B.4) since (B.5) follows as a special case. To show (B.4) we will follow precisely the same lines as the proof given in Cavaliere, Nielsen and Taylor (2014) Lemma C.3 with \( u_1 = u_2 = 0 \) in their \( Q_T (u_1, u_2, \psi) \) notation. Following them, we write
\[
T^{-1} \sum_{i=1}^{T} |z_{k,i} (\psi) z_{l,i} (\psi)| \leq T^{-1} \sum_{i=1}^{T} \sum_{m=1}^{i-1} \sum_{n=1}^{i-1} |\varphi_{k,m} (\psi)| |\varphi_{l,n} (\psi)| |e_{i-m} e_{i-n}| \\
= \sum_{m=1}^{T-1} \sum_{n=1}^{T-1} |\varphi_{k,m} (\psi)| |\varphi_{l,n} (\psi)| T^{-1} \sum_{i=\max(m,n)+1}^{T} |e_{i-m} e_{i-n}|. \quad (B.6)
\]

Since \( T^{-1} \sum_{i=\max(m,n)+1}^{T} |e_{i-m} e_{i-n}| = O_p (1) \) uniformly in \( m, n \) because \( E |e_{i-m} e_{i-n}| \leq \sup_r \sigma (r)^2 < \infty \) it thus follows that (see the treatment of Cavaliere, Nielsen and Taylor (2014) equation (20))
\[
\sup_{\psi \in \Psi} (B.6) = O_p \left( \sup_{\psi \in \Psi} \sum_{m=1}^{T-1} \sum_{n=1}^{T-1} |\varphi_{k,m} (\psi)| |\varphi_{l,n} (\psi)| \right) = O_p (1),
\]
because \( \sum_{j=1}^{\infty} |\varphi_{k,j} (\psi)| < \infty \) uniformly in \( \psi \in \Psi \) under Assumption R.

To show (B.2), for \( j = 1, \ldots, p + q \) and since \( p \) and \( q \) are known fixed orders, we have
\[
\max_{t} \left| 2 \left( \hat{\psi}_j - \psi_{0,j} \right) \sum_{i=1}^{T} w_{ti} e_i z_{j,i} (\psi^*) \right| \leq \max_{t} \left| \hat{\psi}_j - \psi_{0,j} \right| \sum_{i=1}^{T} w_{ti} |e_i z_{j,i} (\psi^*)| \\
\leq 2T \left| \hat{\psi}_j - \psi_{0,j} \right| \left( \max_{t,i} w_{ti} \right) \sup_{\psi \in \Psi} T^{-1} \sum_{i=1}^{T} |e_i z_{j,i} (\psi)| \\
= O_p (T^{1/2}) O \left( \frac{1}{T^b} \right) O_p (1) = O_p \left( \frac{1}{\sqrt{T^b}} \right),
\]
by Lemma A(b,d) and (B.5).

To show (B.3), for \( k, l = 1, \ldots, p + q \) and given that \( p \) and \( q \) are known fixed orders, we have
\[
\max_{t} \left| \hat{\psi}_k - \psi_{0,k} \right| \left| \hat{\psi}_l - \psi_{0,l} \right| \sum_{i=1}^{T} w_{ti} |z_{k,i} (\psi^*) z_{l,i} (\psi^*)| \\
\leq T \left| \hat{\psi}_k - \psi_{0,k} \right| \left| \hat{\psi}_l - \psi_{0,l} \right| \left( \max_{t,i} w_{ti} \right) \sup_{\psi \in \Psi} T^{-1} \sum_{i=1}^{T} |z_{k,i} (\psi) z_{l,i} (\psi)| \\
= O_p \left( \frac{1}{T^b} \right),
\]
by (B.4). This concludes the proof for this part (i).

Part (j) This follows from Xu and Phillips (2008) Lemma A(j) and our Lemma A(i) above.
\textbf{Part (k)} Recall equation (B.1), we have
\[
\sum_{t=1}^{T} (\hat{\sigma}^2_t - \tilde{\sigma}^2_t)^2 = \sum_{t=1}^{T} (A_{1,t} + A_{2,t})^2 = \sum_{t=1}^{T} A^2_{1,t} + \sum_{t=1}^{T} A^2_{2,t} + 2 \sum_{t=1}^{T} A_{1,t}A_{2,t}
\]
and so the required result that \(\sum_{t=1}^{T} (\hat{\sigma}^2_t - \tilde{\sigma}^2_t)^2 = O_p \left(1/\sqrt{T}\right)\) will follow if we show that:
\[
\sum_{t=1}^{T} A^2_{1,t} = O_p \left(1/\sqrt{T}\right); \quad (B.7)
\]
\[
\sum_{t=1}^{T} A^2_{2,t} = O_p \left(1/T\right); \quad (B.8)
\]
\[
\sum_{t=1}^{T} A_{1,t}A_{2,t} = O_p \left(1/T^{3/4}\right). \quad (B.9)
\]
To show (B.7), we write
\[
\sum_{t=1}^{T} A^2_{1,t} = 4 \sum_{t=1}^{T} \left( \sum_{k=1}^{p+q} \left( \hat{\psi}_k - \psi_{0,k} \right) \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi^* \right) \right)^2 \\
\leq 4 \sum_{t=1}^{T} \left( \sum_{k=1}^{p+q} \left( \hat{\psi}_k - \psi_{0,k} \right)^2 \sum_{k=1}^{p+q} \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi^* \right) \right)^2 \\
= 4 \left( \sum_{k=1}^{p+q} T \left( \hat{\psi}_k - \psi_{0,k} \right)^2 \right) \sum_{k=1}^{p+q} \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi^* \right)^2.
\]
Since \(p\) and \(q\) are known fixed orders and \(\left( \hat{\psi}_k - \psi_{0,k} \right) = O_p \left(T^{-1/2}\right)\) by Lemma A(b), equation (B.7) follows if, for \(k = 1, \ldots, p + q\), we show that
\[
T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi^* \right) \right)^2 - T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi_0 \right) \right)^2 = O_p \left(1/\sqrt{T}\right) \quad (B.10)
\]
and
\[
T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi_0 \right) \right)^2 = O_p \left(1/T^b\right). \quad (B.11)
\]
To show (B.10), for some \(\tilde{\psi}\) in between \(\hat{\psi}\) and \(\psi_0\),
\[
T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi^* \right) \right)^2 - T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti}e_{i}z_{k,i} \left( \psi_0 \right) \right)^2 \\
\leq \sum_{l=1}^{p+q} \left| T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{ti}w_{tj}e_{i}e_{j} \left( \frac{\partial}{\partial \psi_l} z_{k,i} \left( \hat{\psi} \right) z_{k,j} \left( \hat{\psi} \right) \right) \left| \psi_l^* - \psi_{0,l} \right| \right| \\
\leq \sum_{l=1}^{p+q} \sup_{\psi \in \Psi} \left| T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{ti}w_{tj}e_{i}e_{j} \left( \frac{\partial}{\partial \psi_l} z_{k,i} \left( \psi \right) z_{k,j} \left( \psi \right) \right) \left| \psi_l^* - \psi_{0,l} \right| \right|.
Recall that $\psi^*$ lies in between $\hat{\psi}$ and $\psi_0$ (see part (i)) and Lemma A(b) implies that $|\psi^*_l - \psi_{0,l}| = O_p\left(T^{-1/2}\right)$ then (B.10) follows if, for $m, n = 0, 1$, we show that

$$
\sup_{\psi \in \Psi} \left| T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{ti} w_{tj} \left( e_{i,j} \frac{\partial^n}{\partial \psi^m} z_{k,i} (\psi) \right) \left( e_{j} \frac{\partial^n}{\partial \psi^m} z_{l,j} (\psi) \right) \right| = O_p\left(1\right). \tag{B.12}
$$

We show only when $m = n = 0$ since the remaining cases may be shown in a similar fashion given that the coefficients of the linear processes resulting from the derivative wrt $\psi$ still decay exponentially under Assumption R. To show (B.12), it suffices to show that, for $k, l = 1, \ldots, p + q$,

$$
\sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti} \left| z_{k,i} (\psi) \right| z_{l,i} (\psi) \right)^2 = O_p\left(1\right), \tag{B.13}
$$

since (B.12) is a special case. To show (B.13), we will follow the same lines as those in the proof of Cavaliere, Nielsen and Taylor (2014) Lemma C.3. Following them, we write

$$
T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti} \left| z_{k,i} (\psi) \right| z_{l,i} (\psi) \right)^2 \leq T^{-1} \sum_{t=1}^{T-1} \sum_{i=1}^{T-1} \sum_{a=1}^{T-1} \sum_{b=1}^{T-1} \left| \varphi_{k,m} (\psi) \right| \left| \varphi_{l,n} (\psi) \right| \left| \varphi_{k,a} (\psi) \right| \left| \varphi_{l,b} (\psi) \right| \times T^{-1} \sum_{t=1}^{T-1} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} w_{ti} w_{tj} \left| e_{i-m} e_{j-n} e_{l-a} e_{j-b} \right|. \tag{B.15}
$$

In (B.16), $T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} w_{ti} w_{tj} \left| e_{i-m} e_{j-n} e_{l-a} e_{j-b} \right| = O_p\left(1\right)$ uniformly in $m, n, a, b$ because

$$
\begin{align*}
E \left| T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} w_{ti} w_{tj} \left| e_{i-m} e_{j-n} e_{l-a} e_{j-b} \right| \right| & \leq T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} w_{ti} w_{tj} E \left| e_{i-m} e_{j-n} e_{l-a} e_{j-b} \right| \\
& \leq T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} w_{ti} w_{tj} \left( E \left( e_{i-m}^4 \right) E \left( e_{j-n}^4 \right) E \left( e_{l-a}^4 \right) E \left( e_{j-b}^4 \right) \right)^{1/4} \\
& \leq \tau_{0.0} \left( \sup_{r} \sigma (r)^2 \right)^{2} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{ti} w_{tj} = \tau_{0.0} \left( \sup_{r} \sigma (r)^2 \right)^{2} < \infty.
\end{align*}
$$

In (B.15), since $\sum_{n=1}^{\infty} \left| \varphi_{k,m} (\psi) \right| < \infty$ uniformly for all $\psi \in \Psi$ under assumption R, it thus follows that (see the treatment of Cavaliere, Nielsen and Taylor (2014) equation (20))

$$
\sup_{\psi \in \Psi} (B.14) = O_p \left( \left( \sum_{m=1}^{T-1} \left| \varphi_{k,m} (\psi) \right| \right)^{2} \left( \sum_{n=1}^{T-1} \left| \varphi_{l,n} (\psi) \right| \right)^{2} \right) = O_p\left(1\right).
$$
To show (B.11), we write

\[
T^{-1} \sum_{t=1}^{T} E \left( \sum_{i=1}^{T} w_{ti} e_{i, k, i} \left( \psi_{0} \right) \right)^{2} = T^{-1} \sum_{t=1}^{T} E \left( \sum_{i=2}^{T} w_{ti} e_{i} \sum_{m=1}^{i-1} \varphi_{k,m} \left( \psi_{0} \right) e_{i-m} \right)^{2} \\
\leq T^{-1} \sum_{t=1}^{T} \sum_{i=2}^{T} \sum_{j=2}^{T} w_{ti} w_{tj} \sum_{m=1}^{i-1} | \varphi_{k,m} \left( \psi_{0} \right) | \sum_{n=1}^{j-1} | \varphi_{l,n} \left( \psi_{0} \right) | | E \left( e_{i-m} e_{i-n} e_{i-j} \right) | \\
= T^{-1} \sum_{t=1}^{T} \sum_{i=2}^{T} \sum_{m=1}^{i-1} | \varphi_{k,m} \left( \psi_{0} \right) | \sum_{n=1}^{i-1} | \varphi_{l,n} \left( \psi_{0} \right) | | E \left( e_{i-m} e_{i-n} E \left( e_{i-j}^2 | F_{i-1} \right) \right) | \\
\leq \left( \sup_{r} \sigma \left( r \right) \right)^{2} T^{-1} \sum_{t=1}^{T} \sum_{i=2}^{T} w_{ti}^{2} \sum_{m=1}^{i-1} | \varphi_{k,m} \left( \psi_{0} \right) | \sum_{n=1}^{i-1} | \varphi_{l,n} \left( \psi_{0} \right) | | E \left( e_{i-m} e_{i-n} \right) | \\
= \left( \sup_{r} \sigma \left( r \right) \right)^{2} T^{-1} \sum_{t=1}^{T} \sum_{i=2}^{T} w_{ti}^{2} \sum_{m=1}^{i-1} \varphi_{k,m} \left( \psi_{0} \right)^{2} E \left( e_{i-m}^2 \right) \\
\leq \left( \sup_{r} \sigma \left( r \right) \right)^{2} \left( \max_{t,i} w_{ti} \right) \sum_{m=1}^{\infty} \varphi_{k,m} \left( \psi_{0} \right)^{2} T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti} \right) \\
\leq \left( \sup_{r} \sigma \left( r \right) \right)^{2} \left( \max_{t,i} w_{ti} \right) \sum_{m=1}^{\infty} \varphi_{k,m} \left( \psi_{0} \right)^{2} = O \left( \frac{1}{T^{b}} \right).
\]

To show (B.8), for \( k, l = 1, \ldots, p + q \) and since \( p \) and \( q \) are known fixed orders, we write

\[
\left( \hat{\psi}_{k} - \psi_{0,k} \right) \left( \hat{\psi}_{l} - \psi_{0,l} \right) T \left( \hat{\psi}_{l} - \psi_{0,l} \right)^{2} T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{T} w_{ti} z_{k,i} \left( \psi_{0} \right) z_{l,i} \left( \psi_{0} \right) \right)^{2} \\
\leq \left( \hat{\psi}_{k} - \psi_{0,k} \right) \left( \hat{\psi}_{l} - \psi_{0,l} \right) T \left( \hat{\psi}_{l} - \psi_{0,l} \right)^{2} \sup_{\psi \in \Psi} T^{-1} \sum_{i=1}^{T} \left( \sum_{i=1}^{T} w_{ti} z_{k,i} \left( \psi \right) z_{l,i} \left( \psi \right) \right)^{2} \\
= O_{p} \left( T^{-1} \right)
\]

by Lemma A(b) and equation (B.13).

Finally (B.9) follows immediately from (B.7), (B.8) and Cauchy-Schwarz inequality.


**Proof of Lemma B**

**Part (a)** We have

\[
T^{-1/2} \sum_{t=1}^{T} \left( \delta_{0,t} \left( 0, \psi_{0} \right) - s_{0,t} \left( 0, \psi_{0} \right) \right) = T^{-1/2} \sum_{t=1}^{T} \left( \sigma_{t}^{-2} - \sigma_{0}^{-2} \right) \sum_{i=1}^{l-1} \epsilon_{t-i} \epsilon_{t}, \quad (B.17)
\]

and

\[
T^{-1/2} \sum_{t=1}^{T} \left( \delta_{0,t} \left( 0, \psi_{0} \right) - s_{0,t} \left( 0, \psi_{0} \right) \right) = T^{-1/2} \sum_{t=1}^{T} \left( \sigma_{t}^{-2} - \sigma_{0}^{-2} \right) \sum_{j=1}^{l-1} b_{j} \left( \psi_{0} \right) e_{t-j} \epsilon_{t}, \quad (B.18)
\]
where \(b_i(\psi_0)\) decays exponentially under Assumption R. We will only show \((B.17) \xrightarrow{p} 0\) since \((B.18) \xrightarrow{p} 0\) follows immediately from Xu and Phillips (2008). Let \(\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_T)'\) and define

\[
B(\sigma) = T^{-1/2} \sum_{t=1}^{T} \sigma_t^{-2} \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t.
\]

Following from Robinson (1987) and Xu and Phillips (2008), \((B.17) \xrightarrow{p} 0\) follows if we show the following three results:

\[
\begin{align*}
B(\hat{\sigma}) - B(\bar{\sigma}) &\xrightarrow{p} 0, \quad (B.19) \\
B(\bar{\sigma}) - B(\bar{\sigma}) &\xrightarrow{p} 0, \quad (B.20) \\
B(\bar{\sigma}) - B(\bar{\sigma}) &\xrightarrow{p} 0. \quad (B.21)
\end{align*}
\]

Recall that

\[
\hat{\sigma}_t^2 = \sum_{i=1}^{T} w_i e_i^2 \quad \text{and} \quad \bar{\sigma}_t^2 = \sum_{i=1}^{T} w_i \bar{\sigma}^2.
\]

To show \((B.19)\), first note that

\[
T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right)^2 \xrightarrow{p} O_p(1) \quad (B.22)
\]

because

\[
E \left( \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right)^2 \leq \sum_{i=1}^{t-1} \sum_{j=1}^{i-1} \frac{1}{i} E(e_{t-i} e_{t-j} E(e_t^2|F_{t-1}))
\]

\[
= \sup_r \sigma(r)^2 \sum_{i=1}^{t-1} \frac{1}{i^2} E(e_{t-i}^2) \leq \left( \sup_r \sigma(r)^2 \right)^2 \frac{\pi^2}{6} < \infty \quad \text{uniformly in } t.
\]

Then,

\[
\begin{align*}
|B(\hat{\sigma}) - B(\bar{\sigma})| &= \left| T^{-1/2} \sum_{t=1}^{T} \left( \hat{\sigma}_t^{-2} - \bar{\sigma}_t^{-2} \right) \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right| \\
&= \left| T^{-1/2} \sum_{t=1}^{T} \frac{(\hat{\sigma}_t^2 - \bar{\sigma}_t^2)^2}{\hat{\sigma}_t^2 \bar{\sigma}_t^2} \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right| \\
&\leq \left( \frac{1}{\min \hat{\sigma}_t^2} \right) \left( \frac{1}{\min \bar{\sigma}_t^2} \right) \left( \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \bar{\sigma}_t^2)^2 \right)^{1/2} \left( T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right)^2 \right)^{1/2} \\
&= O_p \left( \frac{1}{T^{1/2}} \right) = o_p(1),
\end{align*}
\]

where the second last equality uses Lemma A(h,j,k) and (B.22).

To show \((B.20)\), we have

\[
T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^{-2} - \bar{\sigma}_t^{-2}) \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t
\]

\[
= T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^{-2} - \bar{\sigma}_t^{-2}) \hat{\sigma}_t^{-4} \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t + T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^{-2} - \bar{\sigma}_t^{-2})^2 \bar{\sigma}_t^{-2} \sum_{t=1}^{T} \frac{1}{i} e_{t-i} e_t, \quad (B.23)
\]

7
where the last equality uses
\[ p^{-1} - q^{-1} = (q - p) q^{-2} + (q - p)^2 p^{-1} q^{-2} \]  \hspace{1cm} (B.24)
for any non-zero real numbers \( p \) and \( q \) (see Xu and Phillips (2008) equation (21) page 276).

We now show that the first term in (B.23) \( T \rightarrow 0 \). To see this, let
\[
x_{t-1} = \sum_{i=1}^{t-1} i^{-1} e_{t-i}.
\]
Then \( E \left( \left( \tilde{\sigma}_t^2 - \tilde{\sigma}_t^2 \right) \tilde{\sigma}_t^{-4} x_{t-1} e_t \right) = 0 \) follows the same arguments as in Xu and Phillips equation (22). Now
\[
E \left( T^{-1/2} \sum_{t=1}^{T} (\tilde{\sigma}_t^2 - \tilde{\sigma}_t^2) \tilde{\sigma}_t^{-4} x_{t-1} e_t \right)^2
= T^{-1} \sum_{t=1}^{T} E \left( (\tilde{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 \tilde{\sigma}_t^{-8} x_{t-1}^2 e_t^2 \right)
+ T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} E \left( \tilde{\sigma}_t^{-2} - \tilde{\sigma}_s^{-2} \right) \left( \tilde{\sigma}_t^2 - \tilde{\sigma}_s^2 \right) \tilde{\sigma}_t^{-4} \tilde{\sigma}_s^{-4} x_{t-s-1} x_{t-1} e_{t-s} e_t. \hspace{1cm} (B.25)
\]
By Lemma A(e,f), the first term \( \rightarrow 0 \) since
\[
T^{-1} \sum_{t=1}^{T} E \left( (\tilde{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 \tilde{\sigma}_t^{-8} x_{t-1}^2 e_t^2 \right)
= T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{t-1} \sum_{j=1}^{l-1} \frac{1}{i j} E \left( (\tilde{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 \tilde{\sigma}_t^{-8} e_{t-l-j} e_t^2 \right)
\leq \left( \frac{1}{\min \tilde{\sigma}_t^2} \right)^4 T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{t-1} \sum_{j=1}^{l-1} \frac{1}{i j} \left( E \left( \tilde{\sigma}_t^2 - \tilde{\sigma}_t^2 \right)^4 E \left( e_{t-l-j}^2 e_t^2 \right) \right)^{1/2}
\leq \left( \frac{1}{\min \tilde{\sigma}_t^2} \right)^4 \left( \max_l E \left( \tilde{\sigma}_t^2 - \tilde{\sigma}_t^2 \right)^4 \right)^{1/2} T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{t-1} \sum_{j=1}^{l-1} \frac{1}{i j} \left( E \left( e_{t-l-j}^2 \right) E \left( e_t^8 \right) \right)^{1/4}
\leq \left( \frac{1}{\min \tilde{\sigma}_t^2} \right)^4 \left( \max_l E \left( \tilde{\sigma}_t^2 - \tilde{\sigma}_t^2 \right)^4 \right)^{1/2} T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{t-1} \sum_{j=1}^{l-1} \frac{1}{i j} \left( E \left( e_{t-l-j}^8 \right) E \left( e_t^8 \right) \right)^{1/4} E \left( e_t^8 \right)^{1/4}
\leq \left( \frac{1}{\min \tilde{\sigma}_t^2} \right)^4 \left( \max_l E \left( \tilde{\sigma}_t^2 - \tilde{\sigma}_t^2 \right)^4 \right)^{1/2} T^{-1} \sum_{t=1}^{T} \left( \sum_{i=1}^{t-1} \frac{1}{i} \right)^2
= O \left( \frac{(\log T)^2}{T b} \right).
As for (B.25), we have, by Lemma A(e),

\[ T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \sum_{t}^{T} \left( \hat{\sigma}_{t-s}^{2} - \hat{\sigma}_{t}^{2} \right) \left( \hat{\sigma}_{t-s}^{2} \left( E T_{s} - 1 \right) x_{t-s} x_{t-s} e_{t} e_{t} \right) \right) \]

\[ = T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \left( \hat{\sigma}_{t-s}^{2} - \hat{\sigma}_{t}^{2} \right) \left( \hat{\sigma}_{t-s}^{2} \left( E T_{s} - 1 \right) x_{t-s} x_{t-s} e_{t} e_{t} \right) \right) \]

\[ \leq C T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \left( \hat{\sigma}_{t-s}^{2} - \hat{\sigma}_{t}^{2} \right) \left( \hat{\sigma}_{t-s}^{2} \left( E T_{s} - 1 \right) x_{t-s} x_{t-s} e_{t} e_{t} \right) \right) \]

\[ = C T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \hat{\sigma}_{t-s}^{2} x_{t-s} x_{t-s} e_{t} e_{t} \right) \]  

(B.26)

We will only show that (B.26) is zero since the remaining terms follow immediately. First, when \( i = j \) we write (B.26) as

\[ T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \hat{\sigma}_{t-s}^{2} x_{t-s} x_{t-s} e_{t} e_{t} \right) \]

(B.27)

The first term is zero since

\[ E \left( \hat{\sigma}_{t-s}^{4} x_{t-s} x_{t-s} e_{t} e_{t} | \mathcal{F}_{t-1} \right) = \hat{\sigma}_{t-s}^{4} x_{t-s} x_{t-s} e_{t} e_{t} (\mathcal{F}_{i} | \mathcal{F}_{t-1}) = 0, \]

the second term is zero since \( w_{tt} = 0 \) and finally the third term is zero because

\[ T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \hat{\sigma}_{t-s}^{2} x_{t-s} x_{t-s} e_{t} e_{t} | \mathcal{F}_{t-1} \right) \]

\[ = T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \hat{\sigma}_{t-s}^{2} x_{t-s} x_{t-s} e_{t} e_{t} | \mathcal{F}_{t-1} \right) \]

\[ = T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{t-s} w_{t} E \left( \hat{\sigma}_{t-s}^{2} x_{t-s} x_{t-s} e_{t} e_{t} | \mathcal{F}_{t-1} \right) = 0. \]
Now when \( i \neq j \) in (B.26), the arguments follows similarly from (B.27). For example if \( i > j \) and both \( i \) and \( j \) greater than \( t \) then

\[
E \left( e_t^2 x_{t-s-1} x_{t-1} e_{t-s} e_t \right) = E \left( E \left( E \left( e_t^2 x_{t-s-1} x_{t-1} e_{t-s} e_t | F_{t-1} \right) | F_{t-1} \right) \right) = 0.
\]

For the second term in (B.23), we have, by Lemma A(e,g,h) and (B.22),

\[
\left| T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \overline{\sigma}_t^2)^2 \overline{\sigma}_t^{-2} \overline{\sigma}_t^{-4} \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right| \\
\leq \left( \frac{1}{\min \hat{\sigma}_t^2} \right)^2 \left( \frac{1}{\min \overline{\sigma}_t^2} \right)^2 \left| T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \overline{\sigma}_t^2)^2 \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right| \\
\leq \left( \frac{1}{\min \hat{\sigma}_t^2} \right)^2 \left( \frac{1}{\min \overline{\sigma}_t^2} \right)^2 \left( \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \overline{\sigma}_t^2)^4 \right)^{1/2} \left( T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right)^{1/2} \\
= O_p \left( \frac{1}{T^{1/2}} \right) = o_p (1).
\]

To show (B.21), we have

\[
B (\hat{\sigma}) - B (\sigma) = T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \sigma_t^2) \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t.
\]

Note that

\[
E \left( \left( \hat{\sigma}_t^2 - \overline{\sigma}_t^2 \right) \sum_{i=1}^{t-1} i^{-1} e_{t-i} e_t | F_{t-1} \right) = (\hat{\sigma}_t^2 - \sigma_t^2) \sum_{i=1}^{t-1} i^{-1} e_{t-i} E (e_t | F_{t-1}) = 0
\]

so that \((\hat{\sigma}_t^2 - \sigma_t^2) \sum_{i=1}^{t-1} i^{-1} e_{t-i} e_t\) is a martingale difference sequence with respect to \( F_{t-1} \). Also note that

\[
\sigma_t^2 (\hat{\sigma}_t^2 - \sigma_t^2)^2 \leq \hat{\sigma}_t^{-4} \sigma_t^{-2} |\sigma_t^2 + \sigma_t^2| \cdot |\sigma_t^2 - \sigma_t^2| \leq C |\sigma_t^2 - \sigma_t^2|
\]

from Xu and Phillips (2008) equation (23), page 277. Then

\[
E \left( T^{-1/2} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \sigma_t^2) \sum_{i=1}^{t-1} \frac{1}{i} e_{t-i} e_t \right)^2 \\
= T^{-1} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \sigma_t^2)^2 \sum_{i=1}^{t-1} \frac{1}{i} \sum_{j=1}^{t-1} E \left( e_{t-i} e_{t-j} E \left( e_t^2 | F_{t-1} \right) \right) \\
= T^{-1} \sum_{t=1}^{T} \sigma_t^2 (\hat{\sigma}_t^2 - \sigma_t^2)^2 \sum_{i=1}^{t-1} \frac{1}{i} E \left( e_{t-i}^2 \right) \\
\leq C \sup_r \sigma (r)^2 T^{-1} \sum_{t=1}^{T} \frac{1}{t^2} \sum_{i=1}^{t-1} \frac{1}{i^2} \\
\leq C \sup_r \sigma (r)^2 \frac{\pi^2}{6} T^{-1} \sum_{t=1}^{T} |\hat{\sigma}_t^2 - \sigma_t^2| \to 0,
\]

by Lemma A(l).
Part (b) Recall $h_t(\gamma)$ given just above equation (A.3) and define its partitioned matrix as

$$
\begin{align*}
h_t(\gamma) &= -\frac{1}{\sigma_t^2} (c_0 (L; \gamma) e_t \cdot c_2 (L; \gamma) e_t + (c_1 (L; \gamma) e_t) (c_1 (L; \gamma) e_t)^T) \\
&= \begin{pmatrix} h_{\theta \theta, t}(\gamma) & h_{\theta \psi, t}(\gamma) \\ h_{\psi \theta, t}(\gamma) & h_{\psi \psi, t}(\gamma) \end{pmatrix}.
\end{align*}
$$

We will only deal with the term $h_{\theta \theta, t}(\gamma)$ as the remaining terms are dealt with in the same fashion. Note that $h_{\theta \theta, t}(\gamma) = -\frac{1}{\sigma_t^2} (X_{0,t}(\gamma) X_{2,t-2}(\gamma) + X_{1,t-1}(\gamma)^2)$

where $X_{0,t}(\gamma) = c_0 (L; \gamma) e_t$, $X_{2,t-2}(\gamma)$ is the $(1, 1)$-element of the matrix $c_2 (L; \gamma) e_t$ and $X_{1,t-1}(\gamma)$ is the first element of the vector $c_1 (L; \gamma) e_t$. Also $h_{\theta \theta, t}(\gamma)$ is similarly defined by replacing $\sigma_t^2$ in $h_{\theta \theta, t}(\gamma)$ with $\hat{\sigma}_t^2$.

For clarity we write $(\theta, \psi)$ for $\gamma$ and we will show that

$$
\sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left( h_{\theta \theta, t}(0, \psi) - h_{\theta \theta, t}(0, \psi) \right) \xrightarrow{p} 0. \tag{B.28}
$$

To show (B.28), we have

$$
T^{-1} \sum_{t=1}^{T} \left( h_{\theta \theta, t}(0, \psi) - h_{\theta \theta, t}(0, \psi) \right) \cong T^{-1} \sum_{t=1}^{T} (\hat{\sigma}_t^{-2} - \sigma_t^{-2}) \left( X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right).
$$

First we will show that

$$
\sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right| = O_p(1). \tag{B.29}
$$

To show (B.29) we write

$$
\begin{align*}
\sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right| &\leq \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) \right| + \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} X_{1,t-1}(0, \psi)^2 \\
&\leq \left( \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} X_{0,t}(0, \psi)^2 \right)^{1/2} + \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} X_{1,t-1}(0, \psi)^2.
\end{align*}
$$

Now from Cavaliere, Nielsen and Taylor (2014) Lemma C.3,

$$
\sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} X_{0,t}(0, \psi)^2 = O_p(1)
$$

follows from setting $k = l = u_1 = u_2 = 0$ in their Lemma C.3. Similarly for $X_{2,t-2}(0, \psi)^2$ with $k = l = 2$ and $u_1 = u_2 = 0$ and lastly for $X_{1,t-1}(0, \psi)^2$ with $k = l = 1$ and $u_1 = u_2 = 0$.

Second we define

$$
A(\sigma, \psi) = T^{-1} \sum_{t=1}^{T} \sigma_t^{-2} \left( X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right).
$$
Now, following Xu and Phillips (2008), (B.28) follows from showing the following results:

\[ \sup_{\psi \in \Psi} |A(\hat{\sigma}, \psi) - A(\bar{\sigma}, \psi)| \xrightarrow{p} 0; \]  
\[ (B.30) \]

\[ \sup_{\psi \in \Psi} |A(\hat{\sigma}, \psi) - A(\bar{\sigma}, \psi)| \xrightarrow{p} 0; \]  
\[ (B.31) \]

\[ \sup_{\psi \in \Psi} |A(\hat{\sigma}, \psi) - A(\sigma, \psi)| \xrightarrow{p} 0. \]  
\[ (B.32) \]

To show (B.30), by Lemma A(h,i,j) and equation (B.29) above, we write

\[ \sup_{\psi \in \Psi} |A(\hat{\sigma}, \psi) - A(\bar{\sigma}, \psi)| = \sup_{\psi \in \Psi} \left| T^{-1} \sum_{t=1}^{T} \left( \frac{\hat{\sigma}_t - \bar{\sigma}_t}{\bar{\sigma}_t^2} \right) \left( X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right) \right| \]

\leq \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| \frac{\hat{\sigma}_t^2 - \bar{\sigma}_t^2}{\bar{\sigma}_t^2 \hat{\sigma}_t} \right| \left| X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right|

\leq \frac{\max_t \left| \frac{\hat{\sigma}_t^2 - \bar{\sigma}_t^2}{\bar{\sigma}_t^2 \hat{\sigma}_t} \right|}{\left( \min_t \hat{\sigma}_t \right)^2} \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right|

= O_p \left( \frac{1}{\sqrt{Tb}} \right). \]

To show (B.31), using the same arguments as (B.30) but now using Lemma A(e,g,h) and (B.29), we have (B.31) = \( O_p \left( T^{-1/4} b^{-1/2} \right) \).

We now show (B.32). To simplify notation let \( X_t(\psi) = X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \). We write

\[ \sup_{\psi \in \Psi} |A(\hat{\sigma}, \psi) - A(\sigma, \psi)| = \sup_{\psi \in \Psi} \left| T^{-1} \sum_{t=1}^{T} \left( \frac{\hat{\sigma}_t^2 - \bar{\sigma}_t^2}{\bar{\sigma}_t^2 \hat{\sigma}_t} \right) X_t(\psi) \right| \]

\leq \left( \frac{1}{\inf_t \bar{\sigma}(r)^2} \right) \left( \frac{1}{\min_t \bar{\sigma}_t^2} \right) \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| \hat{\sigma}_t^2 - \bar{\sigma}_t^2 \right| \left| X_t(\psi) \right|

Then, given Lemma A(e), (B.32) follows if we show

\[ \sup_{\psi \in \Psi} T^{-1} \sum_{t=1}^{T} \left| \hat{\sigma}_t^2 - \bar{\sigma}_t^2 \right| \left| X_t(\psi) \right| = o_p(1). \]  
\[ (B.33) \]

To show (B.33), we first show pointwise convergence in \( \psi \); that is

\[ T^{-1} \sum_{t=1}^{T} \left| \hat{\sigma}_t^2 - \bar{\sigma}_t^2 \right| \left| X_t(\psi) \right| \xrightarrow{p} 0. \]  
\[ (B.34) \]

To show this, we first note that

\[ E \left| X_t(\psi) \right| = E \left| X_{0,t}(0, \psi) X_{2,t-2}(0, \psi) + X_{1,t-1}(0, \psi)^2 \right| \]

\leq \left( E \left( X_{0,t}(0, \psi) \right)^2 \right)^{1/2} \left( E \left( X_{2,t-2}(0, \psi) \right)^2 \right)^{1/2} + E \left( X_{1,t-1}(0, \psi)^2 \right)^{1/2} \]

\leq C < \infty \text{ uniformly in } t, \]  
\[ (B.35) \]
following the proof of CNT Lemma A.1, the required result that
and because

\[ \sum_{j=0}^{t-1} \varphi_j(\psi) \leq \sup_r \sigma^2(r) \sum_{j=0}^{\infty} \varphi_j(\psi)^2 < C < \infty. \]

Similarly for \( \sup_{1 \leq t \leq T} E(X_{2t-2}(0, \psi)^2 < \infty \) and \( \sup_{1 \leq t \leq T} E(X_{1t-1}(0, \psi)^2 < \infty \) because the
coefficients of these linear processes are square summable. Now to show (B.34), we have

\[ E(T^{-1} \sum_{t=1}^{T} |\tilde{\sigma}_t^2 - \sigma_t^2| |X_t(\psi)|) \leq T^{-1} \sum_{t=1}^{T} |\tilde{\sigma}_t^2 - \sigma_t^2| E|X_t(\psi)| \leq CT^{-1} \sum_{t=1}^{T} |\tilde{\sigma}_t^2 - \sigma_t^2| = o(1) \]

by Lemma A(l) and (B.35).

Finally, given (B.34), (B.33) follows if we show that \( T^{-1} \sum_{t=1}^{T} (\tilde{\sigma}_t^2 - \sigma_t^2) \sigma_t^{-2} \tilde{\sigma}_t^{-2} X_t(\psi) \) is
stochastically equicontinuous. By the same argument that shows (B.29) and Cavaliere, Nielsen and Taylor (2014) Lemma C.3, the stochastic equicontinuity holds since \( (\tilde{\sigma}_t^2 - \sigma_t^2) \sigma_t^{-2} \tilde{\sigma}_t^{-2} \) is a
non-stochastic sequence and does not depend on \( \psi \) and hence does not change the steps of their
proof for Lemma C.3.

**Part (c)** For clarity we write \((\theta, \psi) \) for \( \gamma \) in what follows. Let

\[ r(\psi) = \lim_{t \to \infty} T^{-1} \sum_{t=1}^{T} E \left( \frac{e_t(0, \psi)}{\sigma_t} \right)^2 = \lim_{t \to \infty} T^{-1} \sum_{t=1}^{T} E \left( \frac{a(L; \psi_a(L; \psi_0)^{-1} e_t}{\sigma_t} \right)^2 \]

and

\[ r_T(\psi) = T^{-1} \sum_{t=1}^{T} \left( \frac{e_t(0, \psi)}{\sigma_t} \right)^2 = T^{-1} \sum_{t=1}^{T} \left( \frac{a(L; \psi_a(L; \psi_0)^{-1} e_t}{\sigma_t} \right)^2. \]

Following the proof of CNT Lemma A.1, the required result that \( r \Rightarrow \psi \)
holds if we show

\[ \sup_{\psi \in \Phi} |r_T(\psi) - r(\psi)| \overset{P}{\to} 0, \quad (B.36) \]

and

\[ \inf_{\psi \in \Phi \cap \{\psi \colon \|\psi - \psi_0\| \geq \epsilon\}} r(\psi) > r(\psi_0) \quad \text{for all} \ \epsilon > 0, \quad (B.37) \]

where \( \|x\| \) denote the usual Euclidean norm for any vector \( x \).

We first define

\[ r^\#(\psi) = \lim_{t \to \infty} T^{-1} \sum_{t=1}^{T} E \left( a(L; \psi_a(L; \psi_0)^{-1} e_t \right)^2 \]

and

\[ r_T^\#(\psi) = T^{-1} \sum_{t=1}^{T} \left( a(L; \psi_a(L; \psi_0)^{-1} e_t \right)^2. \]

To show (B.36), we write

\[ \sup_{\psi \in \Phi} |r_T(\psi) - r(\psi)| = \sup_{\psi \in \Phi} |r_T(\psi) - r_T^\#(\psi) + r_T^\#(\psi) - r^\#(\psi) + r^\#(\psi) - r(\psi)| \leq \sup_{\psi \in \Phi} |r_T(\psi) - r_T^\#(\psi)| + \sup_{\psi \in \Phi} |r_T^\#(\psi) - r^\#(\psi)| + \sup_{\psi \in \Phi} |r^\#(\psi) - r(\psi)| \]

(B.38)
For the first term in (B.38) we have

\[
\sup_{\psi \in \Psi} \left| r_T (\psi) - r_T^\# (\psi) \right| = \sup_{\psi \in \Psi} \left| T^{-1} \sum_{t=1}^{T} \left( \left( \sum_{j=0}^{t-1} \varphi_j (\psi) \left( \frac{\sigma_{t-j}}{\sigma_t} \right) \varepsilon_{t-j} \right)^2 - \left( \sum_{j=0}^{t-1} \varphi_j (\psi) \varepsilon_{t-j} \right)^2 \right) \right|. \tag{B.39}
\]

Since \( \varphi_i (\psi) \) decays exponentially in \( i \) uniformly over all \( \psi \in \Psi \) under Assumption R, (B.39) \( \xrightarrow{p} 0 \) follows by the same arguments leading to (A.6). The second term in (B.38) \( \xrightarrow{p} 0 \) follows from CNT Lemma A.1 equation (A.1). The third term \( \xrightarrow{p} 0 \) follows from the proof of (A.6).

Finally, equation (B.37) holds because the third term in (B.38) \( \xrightarrow{p} 0 \) and CNT Lemma A.1 equation (A.2).

**Part (d).** Define

\[
\tilde{r}_T (\psi) = T^{-1} \sum_{t=1}^{T} \left( \frac{e_t (0, \psi)}{\sigma_t} \right)^2.
\]

Recall \( r_T (\psi) \) and \( r (\psi) \) given in part (c) above. The required result that \( \tilde{\psi} \xrightarrow{p} \psi_0 \) holds if we show

\[
\sup_{\psi \in \Psi} |\tilde{r}_T (\psi) - r (\psi)| \xrightarrow{p} 0.
\]

To see this, we write

\[
\sup_{\psi \in \Psi} |\tilde{r}_T (\psi) - r_T (\psi) + r_T (\psi) - r (\psi)| \leq \sup_{\psi \in \Psi} |\tilde{r}_T (\psi) - r_T (\psi)| + \sup_{\psi \in \Psi} |r_T (\psi) - r (\psi)|.
\]

The second term \( \xrightarrow{p} 0 \) because of (B.36). As for the first term,

\[
\sup_{\psi \in \Psi} |\tilde{r}_T (\psi) - r_T (\psi)| = \sup_{\psi \in \Psi} \left| T^{-1} \sum_{t=1}^{T} \left( \left( \frac{e_t (0, \psi)}{\sigma_t} \right)^2 - \left( \frac{e_t (0, \psi)}{\sigma_t} \right)^2 \right) \right| \xrightarrow{p} 0,
\]

by the same arguments that give (B.28) with \( e_t (0, \psi) = X_{0,t} (\psi) \).