Testing Weak $\sigma$-Convergence using HAR Inference*

Jianning Kong,† Peter C. B. Phillips‡ Donggyu Sul§

March 15, 2019

Abstract

Measurement of diminishing or divergent cross section dispersion in a panel plays an important role in the assessment of convergence or divergence over time in key economic indicators. Econometric methods, known as weak $\sigma$-convergence tests, have recently been developed (Kong et al., 2019) to evaluate such trends in dispersion in panel data using simple linear trend regressions. To achieve generality in applications, these tests rely on heteroskedastic and autocorrelation consistent (HAC) variance estimates. The present paper examines the behavior of these convergence tests when heteroskedastic and autocorrelation robust (HAR) variance estimates using fixed-$b$ methods are employed instead of HAC estimates. Asymptotic theory for both HAC and HAR convergence tests is derived and numerical simulations are used to assess performance in null (no convergence) and alternative (convergence) cases. While the use of HAR statistics tends to reduce size distortion, as has been found in earlier analytic and numerical research, use of HAR estimates in nonparametric standardization leads to significant power differences asymptotically, which are reflected in finite sample performance in numerical exercises. The explanation is that weak $\sigma$-convergence tests rely on intentionally misspecified linear trend regression formulations of unknown trend decay functions that model convergence behavior rather than regressions with correctly specified trend decay functions. Some new results on the use of HAR inference with trending regressors are derived and an empirical application to assess diminishing variation in US State unemployment rates is included.

Keywords: HAR estimation, HAC estimation, Nonparametric studentization, Weak $\sigma$—convergence.

JEL Classification: C33

---

*Phillips acknowledges support from the Kelly Foundation at the University of Auckland.
†Shandong University, China
‡Yale University, USA; University of Auckland, New Zealand; Singapore Management University, Singapore; University of Southampton, UK.
§University of Texas at Dallas, USA
1 Introduction

Amongst the many issues for which panel data enable empirical investigation, questions of convergence and divergence over time have attracted high interest. Particularly, but by no means exclusively, in the study of cross country economic performance, research has focussed on examining evidence of diminishing dispersion in key indicator variables such as income or consumption levels, poverty, and unemployment rates. These indicators all figure of importance in politico-economic discourse at both public and professional levels.

The general idea of diminishing variance is well understood, as is the notion of catch-up effects in economic development. Empirical testing of these concepts is much more subtle and has enlisted various econometric techniques, ranging from simple trend regression (Bunzel and Vogelsang, 2005; Campbell et al., 2001) to modern methods of cluster analysis, convergence, and classification (Phillips and Sul, 2007a, 2007b, 2009; Bonhomme and Manresa, 2015; Su et al., 2016; Wang et al. 2019) partly founded on machine learning methodologies. The latter techniques draw heavily on the discriminatory power of partial cross section averaging which forms one of the many advantages of panel data which were collectively explored in the masterful treatise by Cheng Hsiao (2014) that is now in a third updated edition.

A central concept in much of the empirical analysis is $\sigma$-convergence, which examines whether cross sectional variation diminishes over time. Econometric detection of this type of convergence typically relies on the assessment of statistical significance in any observed reductions in dispersion toward some ultimate (asymptotic) level associated with an ergodic limit distribution. Trend regression may then be formulated in terms of trend functions that decay over time. Regressions that employ such evaporating trends, as they are sometimes called, may be analyzed asymptotically and limit theory has been developed (Phillips, 2007; Robinson, 1995) to aid inference. Like all trend regressions, however, empirical formulations typically lack explicit justifications from economic theory and may be assumed to be misspecified. In consequence, the regression residuals are inevitably serially dependent and heterogeneous making robust inferential methods essential in validating such regressions.

In recent work (Kong, Phillips and Sul, 2019; KPS henceforth), the present authors developed a weak version of the $\sigma$–convergence concept that accommodates various forms of diminishing variation in the data and developed a linear trend regression method for its detection in empirical data. The approach relies on a simple t-statistic and explicitly allows for the fact that this linear trend regression is misspecified under diminishing variation but it makes use the fact that the behavior of the test statistic has a recognizable asymp-
totic signature that can be used in practical work to identify $\sigma-$convergence. In order to achieve robustness, the formulation of the t-statistic makes use of a HAC standard error normalization.

Inferential robustness has received a great deal of attention in econometrics since the 1980s and many different forms of heteroskedastic and autocorrelation consistent (HAC) and closely related heteroskedastic and autocorrelation robust (HAR) estimators have been suggested. The current paper explores the asymptotic and sampling properties of several of the main alternative procedures in the context of t-tests for $\sigma-$convergence. An important aspect of this analysis is that the properties are studied under the trend regression misspecification that is a general feature of this approach to convergence testing. We note that this is an area of research of extending the domain of validity in statistical testing where other ongoing work is relevant, including attempts to achieve valid regression testing in non-stationary regressions that include both cointegrated and spurious regression formulations (Chen and Tu, 2019; Wang, Phillips and Tu, 2019).

The paper is organized is as follows. Section 2 provides some background discussion of recent work on methods of robust inference concerning trend in time series and panel regression. Section 3 overviews the main features of the trend decay model, the simple fitted linear trend regression model recommended for practical implementation, and the $\sigma-$convergence concept for cross section dispersion developed in KPS (2019). Section 4 examines alternative robust methods of testing $\sigma-$convergence, including the ‘fixed-b’ lag truncation rule (Kiefer, Vogelsang and Bunzel, 2000; Kiefer and Vogelsang, 2002a, 2000b; Hwang and Sun, 2018), extending the asymptotic theory of KPS to those test procedures. A simulation experiment to assess the finite sample perforance of the various tests is reported in Section 5, together with an empirical application to assess convergence among unemployment rates in the 48 contiguous states of the USA. Section 6 concludes. Proofs of the main results and other technical derivations are given in the Appendix.

2 Preliminaries on Robust Inference concerning Trend

Methods to control for the effects of serial dependence and heterogeneity in regression errors play a key role in achieving robustness in inference. While conventional HAC methods have good asymptotic performance they are susceptible to large size distortions in practical work. Several alternatives have been proposed in the recent literature to improve finite
sample performance. Among these, the ‘fixed-b’ lag truncation rule (Kiefer, Vogelsang and Bunzel, 2000; Kiefer and Vogelsang, 2002a, 2000b) has attracted considerable interest. The method uses a truncation lag \( M \) that is proportional to the sample size \( T \) (i.e., \( M \sim bT \) for some fixed \( b \in (0, 1) \)) and sacrifices consistent estimation in the interest of achieving improved performance in statistical testing by mirroring finite sample characteristics of test statistics in the asymptotic theory. The formation of t-ratio and Wald statistics based on HAC estimators without truncation belongs to a general class of HAR test statistics\(^1\). There are known analytic advantages to the fixed-b approach, primarily related to controlling size distortion. In particular, research by Jansson (2004) and Sun et al (2008) has shown evidence from Edgeworth expansions of enhanced higher order asymptotic size control in the use of these tests. Recently, Müller (2014), Lazarus, Lewis, Stock and Watson (2018), and Sun (2018) have surveyed work in this literature and provided some further suggestions and recommendations for practical implementation.

One area where methods of achieving valid statistical inference has proved especially important in practice are regressions that involve trending variables, cointegration and possible spurious relationships. Spurious regressions misleadingly produce asymptotically divergent test statistics when there is no meaningful relationship (Phillips, 1986). In studying this phenomena more carefully, Phillips (1998) showed that the use of HAC methods attenuated the misleading divergence rate (under the null hypothesis of no association) by the extent to which the truncation lag \( M \rightarrow \infty \). In particular, the divergence rate of the \( t \) statistic in a spurious regression involving independent \( I(1) \) variables is \( O_p \left( \sqrt{T/M} \right) \) rather than \( O_p \left( \sqrt{T} \right) \). Concordant with this finding, Sun (2004) showed that the use of fixed-b methods (where \( M = bT \rightarrow \infty \) at the same rate as the sample size) in spurious regressions produces \( t \) statistics of order \( O_p(1) \) with convergent limit distributions. These discoveries revealed that prudent use of HAR techniques in regression testing can widen the range of valid inference to include spurious regression.

In the same spirit as Sun (2004, 2014), Phillips, Zhang and Wang (2012; PZW henceforth) considered possible advantages in using HAR test statistics in the context of simple trend regressions of the form

\[
x_t = at + z_t, \tag{1}
\]

\(^1\)Kiefer and Vogelsang (2002a, 2000b) introduced the fixed-b approach to heteroscedastic and autocorrelation robust construction of test statistics. The HAR terminology was used by Phillips (2005a) in an article concerned with the development of automated mechanisms of valid robust inference in econometrics.
where \( z_t \) is \( I(1) \) as well as similar trend regressions on orthonormal polynomials and independent random walks. For trend assessment in fitted models of the type (1) it is of interest to test the null hypothesis \( \mathcal{H}_0 : a = 0 \) of the absence of a deterministic trend in (1). PZW (2012) show that, upon least squares estimation of (1) with \( \hat{a} = \sum_{t=1}^{T} x_t t / \sum_{t=1}^{T} t^2 \), the conventional \( t \)-statistic.

\[
t_a = \frac{\hat{a}}{\left\{ T^{-1} \sum_{t=1}^{T} \hat{z}_t^2 \left( \sum_{t=1}^{T} t^2 \right)^{-1} \right\}^{1/2}} = O_p \left( \sqrt{T} \right),
\]

is divergent under the null, as is the \( t \)-ratio formed with a HAC estimator in sandwich form for which

\[
t_a^{HAC} = \frac{\hat{a}}{\left\{ \left( \sum_{t=1}^{T} t^2 \right)^{-1} \left[ T \hat{\Omega}_{HAC} \right] \left( \sum_{t=1}^{T} t^2 \right)^{-1} \right\}^{1/2}} = O_p \left( \sqrt{\frac{T}{L}} \right),
\]

where \( \hat{\Omega}_{HAC} = \frac{1}{T} \sum_{t=1}^{T} \hat{\omega}_t^2 + \frac{2}{T} \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{T-1} \right) \hat{\omega}_t \hat{\omega}_{t+\ell} \), with \( \hat{\omega}_t = \hat{z}_t t \) and \( \hat{z}_t = x_t - \hat{a} t \), \( L = \lfloor T^\kappa \rfloor \) for \( \kappa \in (0, 1) \). In contrast the \( t \)-ratio formed with a HAR estimator in sandwich form as

\[
t_a^{HAR} = \frac{\hat{a}}{\left\{ \left( \sum_{t=1}^{T} t^2 \right)^{-1} \left[ T \hat{\Omega}_{HAR} \right] \left( \sum_{t=1}^{T} t^2 \right)^{-1} \right\}^{1/2}} = O_p \left( 1 \right),
\]

where \( \hat{\Omega}_{HAR} = \frac{1}{T} \sum_{t=1}^{T} \hat{\omega}_t^2 + \frac{2}{T} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M-1} \right) \hat{\omega}_t \hat{\omega}_{t+\ell} \), has a nuisance parameter free limit distribution when \( M = [bT] \) for some \( b \in (0, 1) \). The intuition is clear: As the extent of the serial dependence in the regression error \( z_t \) rises, use of longer lag lengths to control this dependence help in controlling the size of the test statistic in both finite samples and in the limit theory. When the error becomes nonstationary, the infinite lag length in the limit when it is reproduced to match the rate at which \( T \to \infty \) leads to a \( t \)-ratio with a well defined pivotal limit distribution and \( t_a^{HAR} = O_p \left( 1 \right) \).

### 3 Testing Convergence

The present paper pursues these ideas on robust inference in the context of empirical work on convergence. We are motivated by a similar goal – to investigate whether HAR modifications to conventional testing have the capacity to improve tests for \( \sigma \)-convergence, examining whether cross sectional variation diminishes over time. It is widely understood that trend
specifications in applied econometric work are almost always inadequate approximations to the underlying trend mechanism. This limitation applies equally well to trend decay specifications in modeling convergence behavior in cross sectional variation. Our work involves the use of simple linear trend regressions of the form (1) but with intentional misspecification of the model to assess trend effects that enable testing for weak σ-convergence using the approach developed recently in KPS (2019). The advantage of this methodology for applications is that linear trend regression is simple to use in empirical work and its capacity to detect trend decay is unaffected by the deliberate misspecification of the fitted regression. The fitted model is just a device to determine whether there is evidence in the data to support trend decay and convergence.

KPS (2019) consider a data generating process for a (trend decay) panel \( y_{it} \) which can be written in terms of the general factor augmented system

\[
y_{it} = \theta_i' F_t + x_{it}, \quad x_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it} t^{-\beta},
\]

where \( F_t \) is a vector of common factors, \( \theta_i \) is a vector of factor loadings, \( x_{it} \) has a possible deterministic trend decay function \( t^{-\alpha} \) when \( \alpha > 0 \), and the error process \( \epsilon_{it} t^{-\beta} \) has unconditional variance decay \( \sigma^2 \epsilon_t t^{-2\beta} \) where \( \sigma^2 \) is the variance of \( \epsilon_{it} \) and \( \beta > 0 \). In this context, the convergence behavior of \( x_{it} \) is of primary interest. To simplify the presentation of the main effects of HAR inference here, we consider only the case where \( \alpha = \mu_i = 0 \) and \( \beta > 0 \) (This case is designated as model M2 in KPS). As shown in KPS (2019){2} these and other regularity conditions ensure that after fitting the common factor component of (5) the residual \( \hat{x}_{it} = y_{it} - \hat{\theta}_i' \hat{F}_t = x_{it} + O_p (C_{nT}^{-1}) \) where \( C_{nT} = \min \left[ \sqrt{n}, \sqrt{T} \right] \) and asymptotic analysis of the convergence tests is unaffected by working with \( x_{it} \) in place of \( \hat{x}_{it} \).

Using the notation \( \tilde{\epsilon}_{it} = \epsilon_{it} - n^{-1} \sum_{i=1}^{n} \epsilon_{it} \) for deviations from cross section means of \( \epsilon_{it} \) and similar notation for other variables, KPS (2019) show that the cross sectional variance \( K_{nt} := \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{it}^2 \) of \( x_{it} \) can be decomposed as follows

\[
K_{nt} = \frac{1}{n} \sum_{i=1}^{n} \left( x_{it} - \frac{1}{n} \sum_{i=1}^{n} x_{it} \right)^2 = \sigma^2_{a,n} + \eta_{n,t} + \varepsilon_{n,t},
\]

where \( \sigma^2_{a,n} = n^{-1} \sum_{i=1}^{n} \tilde{a}_{it}^2, \eta_{n,t} = \sigma^2_{\varepsilon,nT} t^{-2\beta} \) is the finite sample trend decay function, and

\[
\varepsilon_{n,t} = 2n^{-1} \sum_{i=1}^{n} \tilde{a}_{i} \tilde{\epsilon}_{it} t^{-\beta} + \left( \sigma^2_{\varepsilon,nT} - \sigma^2_{\varepsilon,nT} \right) t^{-2\beta}.
\]

\(^{2}\)See footnote 9 in KPS for more discussion of this issue involving the prior removal of a factor component from the data.
with \( \sigma^2_{e,nt} = n^{-1} \sum_{t=1}^{n} \tilde{e}_{it}^2 \), \( \sigma^2_{e,nT} = T^{-1} \sum_{t=1}^{T} \sigma^2_{e,nt} \). Since the coefficient on the time decay function \( \eta_{n,t} \) in (6) is random, the following representation of the decomposition is useful in the asymptotic development

\[
\begin{align*}
    K_{nt} &= \sigma^2_{a,n} + \eta_t + \varepsilon_{n,t} + \xi_{n,t},
\end{align*}
\]

where \( \eta_t = \sigma^2_{e} t^{-\lambda} \), \( \lambda = 2\beta \), \( \xi_{n,t} = \eta_{n,t} - \eta_t = \left( \sigma^2_{e,nT} - \sigma^2_{e} \right) t^{-\lambda} \) and \( \sigma^2_{e} \) is the variance of \( \epsilon_{it} \). It is easy to show that \( \xi_{n,t} = O_p \left( n^{-1/2} \right) \) uniformly in \( t \) for all \( \lambda \geq 0 \).

To test weak \( \sigma \)-convergence KPS (2019) propose the following simple linear trend regression fitted with \( T \) time series observations

\[
K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT} t + \hat{u}_{nt},
\]

using a robust \( t \)-statistic on the least squares regression coefficient \( \hat{\phi}_{nT} \). It is convenient to decompose \( \hat{\phi}_{nT} \) into component form as follows

\[
\hat{\phi}_{nT} = \sum_{t=1}^{T} a_{tT} \tilde{\eta}_t + \sum_{t=1}^{T} a_{tT} \tilde{\xi}_{n,t} + \sum_{t=1}^{T} a_{tT} \tilde{\varepsilon}_{n,t} =: I_A + I_B + I_C,
\]

where \( a_{tT} = \tilde{t} / \left( \sum_{s=1}^{T} \tilde{t}^2 \right) \), \( \tilde{\eta}_t = \sigma_{e}^2 t^{-\lambda} \), \( \tilde{t}^{-\lambda} = t^{-\lambda} - T^{-1} \sum_{t=1}^{T} t^{-\lambda} \), \( \tilde{\xi}_{n,t} = \xi_{n,t} - T^{-1} \sum_{t=1}^{T} \xi_{n,t} \) and \( \tilde{\varepsilon}_{n,t} = \varepsilon_{n,t} - T^{-1} \sum_{t=1}^{T} \varepsilon_{nt} \). The separate components \( I_A, I_B \) and \( I_C \) are useful in the proofs and are analyzed in full in KPS (2019). In view of the form of \( \eta_t = \sigma^2_{e} t^{-\lambda} \) and the presence of \( \eta_{n,t} \) in (6), the linear trend regression (9) is evidently misspecified unless \( \lambda = 2\beta = -1 \), in which case there is a specific form of divergence over time rather than convergence, or unless \( \lambda = 2\beta = 0 \), in which case there is neither convergence nor divergence over time. Weak \( \sigma \)-convergence of \( K_{nt} \) is formally defined in equation (2) of KPS (2019) and essentially requires that \( \hat{K}_t = \text{plim}_{n \to \infty} K_{nt} \) exists and decays over time to some constant value \( c \in [0, \infty) \). In what follows we refer to this concept as \( \sigma \)-convergence or more simply as convergence.

As indicated above, the model specification (9) is intentionally simple and inappropriate except for the special cases \( \lambda = 0, -1 \) where there is no cross section convergence. In particular, the linear trend specification would seem to be a poor proxy for capturing evaporating trend decay in variation which is inherently nonlinear because of the zero lower bound on variation. Nonetheless, while any particular parametric trend specification is likely to be misspecified and (9) itself is most likely quite wrong in practical work, intuition from spurious trend regression theory (Phillips, 1986; Durlauf and Phillips, 1988) suggests that a simple reduced form regression specification such as (9) is likely to reveal the presence of any trend effects that are manifest in the temporal evolution of \( K_{nt} \). Thus, in spite of misspecification,
the fitted trend regression (9) turns out to be revealing of both convergence and divergence in cross section dispersion.

KPS (2019) pursue this intuition by developing a formal test procedure with asymptotic theory that can be used to assess the presence of diminishing variation in $K_{nt}$. In mobilizing this test, some attention to serial dependence in the error is appropriate in view of the aggregated time series data and the simplicity and likely misspecification of the fitted model. The specific question that interests us in this paper is whether there is an advantage asymptotically or in finite samples in using fixed-$b$ HAR estimators rather than standard HAC estimators in the construction of the tests in this context of testing for $\sigma$-convergence.

To fix ideas on testing using the fitted model (9), consider the following $t$-ratio

$$t_1 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_1^2 \left( \sum_{t=1}^{T} \hat{i}_t^2 \right)^{-1}}},$$

(11)

where $\hat{\phi}_{nT}$ is the least squares estimate of the slope coefficient in (9), $\hat{i} = t - T^{-1} \sum_{t=1}^{T} t$, and $\hat{\Omega}_1^2$ is defined as

$$\hat{\Omega}_1^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \hat{u}_t \hat{u}_{t+\ell},$$

(12)

where $\hat{u}_t = \hat{K}_{nt} - \hat{\phi}_{nT} \hat{i}$ with $\hat{K}_{nt} = K_{nt} - T^{-1} \sum_{t=1}^{T} K_{nt}$, $L = \lfloor T^\kappa \rfloor$ for $\kappa \in (0,1)$, and more specifically $k = 1/3$, as in the Bartlett-Newey-West estimator.

Depending on the asymptotic behavior of the fitted coefficient $\hat{\phi}_{nT}$, the regression residuals $\hat{u}_t$ then bear the effects of a spurious imported trend from the regression, which influences the properties of long run variance estimators such as (12) that are used in the construction of the $t$-statistic. These effects, in turn, influence the asymptotic behavior of the test statistic. KPS (2019) show that in spite of its misspecification the fitted regression (9) enables a satisfactory test of $\sigma-$convergence. Here we investigate whether or not a HAR type correction, instead of a HAC correction, improves the testing procedure proposed by KPS (2019).

4 Robust Testing

4.1 Null and alternative hypotheses

As in KPS (2019), the hypothesis of interest is $\sigma-$convergence of $K_{nt}$, which naturally corresponds to the case where $\lambda > 0$. The null hypothesis is no convergence and has the
composite form\textsuperscript{3}

\[ H_0 : \lambda \leq 0. \]

(13)

The directed alternative hypothesis \( H_A : \lambda > 0 \) implies that testing for convergence is one-sided. Critical values are then delivered by the limit distribution under the point null \( \lambda = 0 \).

Even though the null and alternative hypotheses are well defined in terms of the unobserved parameter \( \lambda \) in the parametric trend decay model, testing is accomplished using the fitted coefficient \( \hat{\phi}_{nT} \) in the regression (9). It is hard to write down a generally applicable hypothesis of convergence in terms of (9) because this regression is misspecified and there are many possible forms of misspecification, including nonparametric specifications involving both mean and variance. In fact, as KPS (2019) show, the least squares estimator \( \hat{\phi}_{nT} \) approaches zero as \( n, T \to \infty \) when \( \lambda > 0 \). Nonetheless, the \( t_1 \) statistic in (11) diverges to negative infinity if \( 0 < \lambda < 1 \). This limit theory shows that use of the fitted regression (9) leads to a consistent one-sided test of convergence in spite of misspecification of the regression model (9). The test is a left-sided test based on \( \hat{\phi}_{nT} \). In fact, even when the decay parameter value \( \lambda \to \infty \) and the decay in variation \( K_{nt} \) is infinitely fast (implying that only a finite number of observations are helpful in detecting convergence), the \( t_1 \) statistic still remains negative, converging in probability to \(-\sqrt{3} = -1.732 < -1.65\). So the test remains consistent at a 5\% nominal size level in a one-sided asymptotic normal test.

The \( t_1 \) statistic is discontinuous in the limit around \( \lambda = 0 \) which includes the null hypothesis of no convergence. In this case when \( \lambda = 0 \), the limiting distribution of the \( t_1 \) statistic is standard normal, as shown in Theorem 1 below. This limit theory provides a convenient left-sided critical value for the test for convergence, i.e., convergent variation in \( K_{nt} \) over time.

When \( \lambda < 0 \), the \( t_1 \) statistic diverges to positive infinity, showing that the test is powerful in detecting divergent variation in \( K_{nt} \) (using a right-sided test) as well as convergent variation in \( K_{nt} \) (using the left-sided test).

\textsuperscript{3}The null hypothesis given in (13) is for the case \( \mu_t = \alpha = 0 \) in (5). This model is designated model M2 in KPS (2019) and other models are considered there, to which readers are referred for greater generality in HAC standardized \( t \)-ratio testing.
4.2 Test statistics and alternative nonparametric studentization

The $t_1$ test statistic defined by (11) involves a standard HAC studentization formula. We now consider the following $t$-ratios constructed using alternative variance estimates to standardize the coefficient estimate $\hat{\phi}_{nT}$ in the fitted regression (9). The first statistic is analogous to $t_1$ but uses a fixed-$b$ variance estimate

$$ t_2 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_2^2 \left( \sum_{t=1}^{T} \tilde{e}^2 \right)^{-1}}, } $$

where $\hat{\Omega}_2^2$ is given below in (19). In this formula and that for $\hat{\Omega}_M^2$ below it is convenient to use the fixed-$b$ lag truncation notation $M = [bT]$ with $b \in (0, 1)$. The next two statistics use sandwich formulae in the self normalization. Let $\tilde{z}_t = \tilde{u}_t \tilde{e}$, and define

$$ t_{\text{HAR}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left( \sum_{t=1}^{T} \tilde{e}^2 \right)^{-1} T \hat{\Omega}_M^2 \left( \sum_{t=1}^{T} \tilde{e}^2 \right)^{-1}}, } $$

$$ t_{\text{HAC}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left( \sum_{t=1}^{T} \tilde{e}^2 \right)^{-1} T \hat{\Omega}_L^2 \left( \sum_{t=1}^{T} \tilde{e}^2 \right)^{-1}}, } $$

where $M = [bT]$ for some $b \in (0, 1)$ and $L = [T^\kappa]$ for some $\kappa \in (0, \bar{\kappa})$ with $\bar{\kappa} < 1$. Next define

$$ \hat{\Omega}_M^2 = \frac{1}{T} \left[ \sum_{t=1}^{T} \tilde{z}_t^2 + 2 \frac{M}{T} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) \tilde{z}_t \tilde{z}_{t+\ell} \right], $$

$$ \hat{\Omega}_L^2 = \frac{1}{T} \left[ \sum_{t=1}^{T} \tilde{z}_t^2 + 2 \frac{L}{T} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \tilde{z}_t \tilde{z}_{t+\ell} \right], $$

$$ \hat{\Omega}_2^2 = \frac{1}{T} \left[ \sum_{t=1}^{T} \tilde{u}_t^2 + 2 \frac{M}{T} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) \tilde{u}_t \tilde{u}_{t+\ell} \right]. $$

With the formulation (19), the $t_2$ statistic has the same form as the $t_1$ statistic but uses a HAR variance estimator $\hat{\Omega}_2^2$ (with fixed-$b$ coefficient in lag truncation $M = [bT]$) in place of a HAC estimator. The two statistics $t_{\text{HAC}}$ and $t_{\text{HAR}}$ use sandwich formulae for the construction of the variance, with the HAR estimate $\hat{\Omega}_M^2$ in $t_{\text{HAR}}$ and the HAC estimate $\hat{\Omega}_L^2$ in $t_{\text{HAC}}$. Under the null hypothesis, the asymptotic behavior of the $t_{\text{HAC}}$ statistic is the same as that of the original $t_1$ statistic used in KPS (2019), as might be expected because both statistics use consistent estimates of the long run variance. The asymptotic properties under the null of
the $t_{\text{HAR}}$ and $t_2$ statistics differ from that $t_{\text{HAC}}$ and $t_1$, again as might be expected from standard limit theory for HAR testing. Versions of (17) - (19) with other kernels than the Bartlett kernel are possible and are considered in the Appendix.

4.3 Limit theory under the null

We now derive the limit theory of the statistics $\{t_1, t_2, t_{\text{HAC}}, t_{\text{HAR}}\}$. First, we consider asymptotic behavior under the null hypothesis of no convergence. When $\lambda = 0$, it is easy to see that $I_C$ is the only constituent part in (10). Under the null the OLS coefficient estimate in (9) can therefore be written in the simple form\(^4\)

$$\hat{\phi}_{nT} = \sum_{t=1}^{T} a_{tT} \tilde{\varepsilon}_{n,t} = I_C,$$

and $\hat{\phi}_{nT}$ is asymptotically normal. However, the proof is not immediate and, nor is the asymptotic variance formula, because of the complexity of the component variates $\tilde{\varepsilon}_{n,t} = \varepsilon_{n,t} - T^{-1} \sum_{t=1}^{T} \varepsilon_{n,t}$ and $\varepsilon_{n,t}$ in (7). In particular, the element $\varepsilon_{n,t}$ involves first and second sample moments of the original variates $\epsilon_{it}$ and these sample moments induce the presence of higher order moments in the limit theory, as shown in the following result whose proof is given in the Appendix together with the assumptions used in the derivations. In the statement and proof of the theorem the notation $\rightsquigarrow$ denotes both convergence of random sequences in distribution and weak convergence of random elements in the associated function space.

**Theorem 1 (Asymptotics under the Null)**

Under the null hypothesis $\lambda = 0$ and under Assumption A in the Appendix the coefficient estimate $\hat{\phi}_{nT}$ and associated t-ratio statistics have the following asymptotic behavior as $T, n \to \infty$.

(i) $\sqrt{n}T^{3/2} \hat{\phi}_{nT} \rightsquigarrow \mathcal{N} \left( 0, 12 \Omega_\phi^2 \right)$, where $\Omega_\phi^2 = 4 \sigma^2 \Omega_\tau^2 + \Omega_\tau^2$, $\Omega_\tau^2$ is the long run variance of $\epsilon_t$, $\Omega_\tau^2$ is the long run variance of $\epsilon_t^2$, and $\sigma^2 = \mathbb{E} (a_i - a)^2$.

(ii) $t_1 \rightsquigarrow \mathcal{N} (0, 1)$.

(iii) $t_2 = \frac{\sqrt{\hat{\phi}_{nT}^2 \left( \sum_{t=1}^{T} \tilde{t}_t \right)^{-2}}}{\left\{ \int_0^1 \int_0^1 \left( 1 - \frac{|r - s|}{b} \right) dW^r \left( t \right) dW^s \left( t \right) \right\}^{1/2}} \right\}^{1/2}.$

\(^4\)When $\lambda = 0$, $\tilde{t}_t^0 = t^0 - T^{-1} \sum_{t=1}^{T} t^0 = 0$ so that $\hat{\mu}_t = \sigma_t^2 \tilde{t}_t^0 = 0$. Also note that $\xi_{n,t} = n_{n,t} - \eta_t = (\sigma^2_{\epsilon,nT} - \sigma^2_{\epsilon}) t^{-\lambda} = \sigma^2_{\epsilon,nT} - \sigma^2_{\epsilon}$, and $\xi_{n,t} = \xi_{n,t} - T^{-1} \sum_{t=1}^{T} \xi_{nt} = 0$. 

11
(iv) $t_{\text{HAR}} = \frac{\hat{\phi}_n T}{\sqrt{\left(\sum_{t=1}^T \tilde{r}^2\right)^{-1} T\hat{\Omega}_M^2 \left(\sum_{t=1}^T \tilde{e}^2\right)^{-1}}} \sim Z \left\{ \int_0^1 \int_0^1 \left(1 - \frac{|r - s|}{b}\right) \tilde{r} \tilde{s} dW^r (r) dW^s (s) \right\}^{1/2}$.

(v) $t_{\text{HAC}} = \frac{\hat{\phi}_n T}{\sqrt{\left(\sum_{t=1}^T \tilde{r}^2\right)^{-1} T\hat{\Omega}_L^2 \left(\sum_{t=1}^T \tilde{e}^2\right)^{-1}}} \sim \mathcal{N}(0, 1)$.

In (iv) $\tilde{r} = r - \int_0^1 sds$. In (iii) and (iv) the stochastic process $W^r (\cdot)$ is the generalized standard Brownian Bridge $W^r (r) = W (r) - \alpha_B - \beta_B r$, which is the linearly $L_2 [0, 1]$ detrended version of the standard Brownian motion $W (r)$.

The coefficients $(\alpha_W, \beta_W)$ are the solution of the $L_2 [0, 1]$ optimization problem (Phillips, 1988; Park and Phillips, 1988, 1989;)

$$\begin{bmatrix} \alpha_B \\ \beta_B \end{bmatrix} = \arg \min_{(a,b)} \int_0^1 \left\{ W (r) - a - br \right\}^2 dr.$$ 

As discussed in the proof, two long run variance components $(\Omega_\epsilon^2, \Omega_\zeta^2)$ appear in the asymptotic variance $\hat{\Omega}_\phi^2$ of $\sqrt{nT^{3/2}} \hat{\phi}_n T_r$. This complication arises because the residual term $\varepsilon_{n,t}$ in the expression for the dependent variable $K_n$, involves the two second order moment quantities $2n^{-1} \sum_{i=1}^n \bar{a}_t \bar{\varepsilon}_{it}$ and $n^{-1} \sum_{i=1}^n \left( \tilde{e}_{it}^2 - \sigma_\epsilon^2 \right)$ that contribute to long run variation through the variable $\zeta_{it} = 2\bar{a}_t \bar{\varepsilon}_{it} + \left( \tilde{e}_{it}^2 - \sigma_\epsilon^2 \right)$. The quantity $\Omega_\phi^2 = 4\sigma_\zeta^2 \Omega_\epsilon^2 + \Omega_\zeta^2$ is the long run variance of $\zeta_{it}$.

The $t_1$ statistic employed by KPS (2019) utilizes the fact that the linear trend is deterministic and independent of the regression error. Standard theories of HAC and HAR estimation can therefore be applied directly and we might expect from earlier research that both $t_1$ and $t_{\text{HAC}}$ suffer from finite sample size distortion relative to asymptotic nominal size when the usual HAC formula with lag truncation parameter $L = \lfloor T^\kappa \rfloor$ and $\kappa = 1/3$ is employed. Also, as is apparent in (iii) and (iv), the limit theory for $t_2$ and $t_{\text{HAR}}$ is non-normal but still pivotal. This finding is consonant with standard fixed-$b$ limit theory, although the limit theory is of a different form due to the presence of the linear trend in the fitted regression. Both $t_2$ and $t_{\text{HAR}}$ typically show departures from normality in finite samples, and especially as $b$ approaches unity.
4.4 Limit theory under the alternative of convergence

We first discuss the sign of the coefficient \( \hat{\phi}_{nT} \). As shown in KPS (2019), the deterministic term \( I_A \) always dominates \( I_B \) and \( I_C \) under the alternative. In what follows we use the notation \( a_{nT} \sim c_{nT} \) if \( a_{nT}/c_{nT} \rightarrow_p 1 \) as \( n, T \rightarrow \infty \). Then

\[
\hat{\phi}_{nT} \sim a I_A = \sum_{t=1}^{T} a_{tT} \tilde{r}_t = \{1 + o_p(1)\} \begin{cases} -\sigma^2 T^{1-\lambda} & \text{if } \lambda < 1, \\ -6\sigma^2 T^{-2} \ln T & \text{if } \lambda = 1, \\ -6\sigma^2 Z(\lambda) T^{-2} & \text{if } \lambda > 1. \end{cases}
\]

where \( L_\lambda = 6\lambda[(2 - \lambda)(1 - \lambda)]^{-1} \) and \( Z(\lambda) = \sum_{t=1}^{\infty} t^{-\lambda} \) is the Riemann zeta function. Since \( \sigma^2 > 0 \), \( \hat{\phi}_{nT} \) is always negative when \( \lambda > 0 \) and the sign of the \( t \)-statistic is the same as \( \hat{\phi}_{nT} \). The denominators of the \( t \)-ratios become functions of \( \sigma^2 \) and the smoothing parameters \( \kappa \) and \( b \) in variance estimation. The precise form of this functional dependence affects on the individual statistic. For example, the long run variance estimate used in the \( t_1 \) statistic in (12) is a function of \( \tilde{u}_t \) and \( \kappa \). As shown in KPS (2019), when \( n, T \rightarrow \infty \), the dominant term of \( \tilde{u}_t \) becomes

\[
\hat{u}_t = \tilde{n}_{n,t} - \hat{\phi}_{nT} \tilde{t} + \tilde{\xi}_{nT} \sim a \tilde{n}_{n,t} - \hat{\phi}_{nT} \tilde{t},
\]

but

\[
\tilde{n}_{n,t} - \hat{\phi}_{nT} \tilde{t} = \tilde{n}_t - I_A \tilde{t} + \tilde{\xi}_{n,t} - \tilde{t} (I_B + I_C) \sim a \tilde{n}_t - I_A \tilde{t}.
\]

Hence the residual can be approximated as

\[
\hat{u}_t \sim a \tilde{n}_t - \left[ \sum_{t=1}^{T} a_{tT} \tilde{n}_t \right] \tilde{t},
\]

which is a linearly detrended form of \( \tilde{n}_t \). From the definitions following (10), \( \tilde{n}_t = \sigma^2 \tilde{t}^{-\lambda} \) and is a linear function of \( \sigma^2 \). So, in conjunction with (20), it is apparent that the \( t_1 \) test statistic is free from the scale nuisance parameter of \( \sigma^2 \).

We now study the asymptotic properties of the other three \( t \)-statistics under \( \sigma \)-convergence. For the \( t_1 \) and \( t_{HAC} \) statistics the alternative hypothesis of convergence in which \( \lambda > 0 \) can be written in terms of left-sided critical values and lead to a rejection of the null hypothesis of non convergence at the 5% level when \( t_1 < -1.65 \). The same is true for the HAR test statistics except that the left sided critical values depend on the pivotal limit theory given in Theorem 1 (iii) and (iv). These critical values can be obtained by simulation and this is done in the numerical exercises reported later. The following theorem provides details of the asymptotic behavior of these statistics.
Theorem 2 (Asymptotics under the Alternative)

Under weak $\sigma$-convergence with $\lambda > 0$ and under the regularity conditions given in Assumption A and B in the Appendix and Theorem 1 of KPS (2019), the $t$-ratio statistics have the following asymptotic behavior as $n, T \to \infty$:

$$t_1 = \begin{cases} 
O_p \left( T^{1/2-\kappa/2} \right) & \text{if } 0 < \lambda < 1/2, \\
O_p \left( T^{1/2-\kappa/2} \left( \ln T \right)^{-1/2} \right) & \text{if } \lambda = 1/2, \\
O_p \left( T^{1-\lambda-\kappa/2} \right) & \text{if } 1/2 < \lambda < 1/(1+\kappa), \\
O_p \left( T^{(1-\lambda)(1-\kappa)/2} \right) & \text{if } 1/(1+\kappa) \leq \lambda < 1, 
\end{cases} \quad (22)$$

and $\text{plim}_{n,T \to \infty} t_1 = \begin{cases} 
-\sqrt{6/\kappa^2} & \text{if } \lambda = 1, \\
-\mathcal{Z}(\lambda) \sqrt{3} & \text{if } 1 < \lambda < \infty, \\
-\sqrt{3} & \text{if } \lambda \to \infty. 
\end{cases} \quad (23)$

where $\kappa > 0$ is defined by the lag truncation parameter $L = \lfloor T^{\kappa} \rfloor$ in the Bartlett-Newey-West long run variance estimator. The function $\mathcal{Z}(\lambda) := \mathcal{Z}(\lambda) \left( \sum_{t=1}^{\infty} t^{-\lambda} \mathcal{Z}(\lambda, t) \right)^{-1/2} > 1$ for all $\lambda > 1$, where $\mathcal{Z}(\lambda, t) = \sum_{s=1}^{\infty} (s + t)^{-\lambda}$ is the Hurwitz zeta function.

$$t_2 = O_p(1) \quad \text{if } \lambda > 0 \quad (24)$$

$$t_{\text{HAR}} = O_p(1) \quad \text{if } \lambda > 0 \quad (25)$$

$$t_{\text{HAC}} = O_p \left( T^{1/2-\kappa/2} \right) \quad \text{if } \lambda > 0 \quad (26)$$

As $(n, T) \to \infty$ the statistics $t_1, t_2, t_{\text{HAR}}$ and $t_{\text{HAC}}$ are all negative under convergence since their signs are determined by the trend regression coefficient $\hat{\phi}_{it}$ which is always negative when $\lambda > 0$. Except for a few specific cases shown in the statement of the theorem, the asymptotic orders of the $t$-ratios that rely on HAR estimates can be obtained by replacing $\kappa$ by unity in (22). This accords with the understanding that the HAR statistics rely on the use of lag truncation parameters proportional to the sample size.

Theorem 2 provides the order of magnitude of each $t$ statistic under the alternative hypothesis of convergence, which is important for determining whether the associated test is consistent. The HAR test statistics $t_{\text{HAR}}$ and $t_2$ are $O_p(1)$ and their large sample behavior can be illustrated graphically. Figure 1a shows the empirical distribution of the $t_2$ statistic under the following data generating process: $y_{it} = \epsilon_{it} t^{-0.25}$ where $\epsilon_{it} \sim iid \mathcal{N}(0, 1)$. Evidently
as \( n \) increases, the distribution of the \( t_2 \) statistic concentrates around a value close to -3.8. When \( T \) increases as well, the distribution collapses more rapidly and to a slightly smaller value close to -3.4, as shown in Figure 1b. The limit behavior of the \( t_2 \) and \( t_{\text{HAR}} \) statistics depends on the value of \( \lambda \) and the smoothing parameter \( b \). This behavior is explored in the numerical simulations that follow.

Figure 1a: Densities of the \( t_2 \) statistic

\[ T = 100, \beta = 0.25 \text{ and } b = 0.3 \]

Figure 1b: Densities of the \( t_2 \) statistic

\[ \beta = 0.25 \text{ and } b = 0.3 \]

The orders of magnitude of the \( t_{\text{HAC}} \) and \( t_{\text{HAR}} \) statistics given in (25) bear a similar relation to those of the \( t_{a}^{\text{HAC}} \) and \( t_{a}^{\text{HAR}} \) statistics given earlier in (3) and (4) for testing the significance of the slope coefficient in the linear trend regression (1). Since under HAC estimation \( L = [T^\lambda] \), the order of magnitude of the \( t_{\text{HAC}} \) statistic is \( O_p \left( \sqrt{T/L} \right) = O_p \left( T^{1-\lambda}/2 \right) \) just as in (3); and \( t_{\text{HAR}} = O_p \left( 1 \right) \) as in the case of \( t_{a}^{\text{HAR}} \). Note that the asymptotic properties of both \( t_{a}^{\text{HAC}} \) and \( t_{a}^{\text{HAR}} \) were obtained under the null hypothesis of \( a = 0 \) in (1), but the linear trend regression in that model was misspecified (just as in the present case) and the error was \( I(1) \), so the data trends were stochastic rather deterministic in that case. The asymptotic properties of the \( t_{\text{HAC}} \) and the \( t_{\text{HAR}} \) test statistics for convergence are here driven under the alternative hypothesis (rather than the null) where \( \lambda > 0 \). But when \( \lambda > 0 \), the linear trend regression (9) is also misspecified and the regression residual in (21) has deterministic terms because \( \eta_t = \sigma^2 t^{-\lambda} \) and the regression weight \( a_{tT} \) is deterministic. Hence the residual \( \tilde{u}_t \) has persistent time series behavior. As will be apparent in the empirical example considered later, and as accords with earlier analyses of misspecification, the fitted AR(1) coefficient tends to be close to unity in this case. These properties lead to the asymptotic results for \( t_{\text{HAR}} \) and \( t_{\text{HAC}} \) given in (25) and to asymptotic behavior analogous to the \( t_{a}^{\text{HAC}} \) and \( t_{a}^{\text{HAR}} \) statistics in (1).
It is worth noting that the asymptotic behavior of the $t_1$ test differs from that of the $t_{HAC}$ test when $\lambda > 1/2$. As Phillips and Park (1988; theorem 3.1) and Park and Phillips (1988, pp. 486-487) show, in general trend regression where regressors have stochastic or deterministic trends but the regression errors are stationary, the commonly used sandwich variance matrix form can be further reduced to the form given in the $t_1$ statistic involving a suitable long run variance estimate. This result suggests that a similar equivalence in behavior (irrespective of the method of variance estimate construction) might be expected in the present context. However, as the trend decay parameter $\lambda$ increases, the fitted linear trend regression equation (9) reverts to the null specification as the effective regressor $\eta_t = \sigma^2_t t^{-\lambda}$ in the true model has negligible deterministic trend properties when $\lambda$ increases and the trend decay component is ultimately zero as $\lambda \to \infty$. Thus, as $\lambda$ increases we may expect differences to arise in finite sample and asymptotic behavior between the $t_1$ and $t_{HAC}$ statistics. We investigate this difference more carefully by means of numerical simulations in what follows (see, in particular, Figures 2 and 4 below).

To conduct simulations we first note that the data $K_{nt} = n^{-1} \sum_{i=1}^{n} (x_{it} - \frac{1}{n} \sum_{i=1}^{n} x_{it})^2$ become nonstochastic as $n \to \infty$. In particular, from (6) we have $K_{nt} = \sigma^2_{a,n} + \eta_{n,t} + \varepsilon_{n,t}$ and it is easy to see that $\sigma^2_{a,n} \to_p \sigma^2_a$, $\eta_{n,t} \to_p \sigma^2_t t^{-\lambda}$, and $\varepsilon_{n,t} \to_p 0$ as $n \to \infty$. Hence, for large $n$ we have the deterministic representation $K_{nt} \sim \sigma^2_a + \sigma^2_t t^{-\lambda}$. This large $n$ setting for $K_{nt}$ provides a convenient mechanism for studying the behavior of the trend coefficient estimate $\hat{\theta}_{nT}$ and the corresponding $t$-ratios in the fitted trend regression model (9). We use this device in the numerical exercises that follow, which should therefore be interpreted as simulations for very large $n$.

Figure 2 plots all four $t$-ratios over various $\lambda$ values with $\kappa = b = 1/3$ and $T = 1,000$. Evidently, all four $t$-ratios are discontinuous at $\lambda = 0$, where the model passes through the null hypothesis from convergence to divergence. Further, as $\lambda$ increases all the $t$-ratios seem to converge to a certain point. The $t_1$ and $t_{HAC}$ statistics converge to the same point $-\sqrt{3}$ and the $t_1$ statistic evidently has greater discriminatory power than $t_{HAC}$ over the full range of $\lambda$. It is also apparent in Figure 2 that the $t_2$ and $t_{HAR}$ statistics both converge to the same point $-\sqrt{3}$ as $\lambda \to \infty$ and this asymptotic behavior seems to be independent of the value of $b$. The magnitude (in absolute value) of the test statistics is as follows: $|t_1| \geq |t_{HAC}| \geq |t_2| \geq |t_{HAR}|$ for all values of $\lambda$ with $b = 1/3$. Only for very large $\lambda$ does equality hold. Subject to size correction, this outcome suggests that the $t_1$ statistic provides the most powerful test.

We next vary the setting of the time series sample size $T$ to examine the behavior of
the $t_2$ and $t_{\text{HAR}}$ statistics with big changes in $T$ and as $\lambda$ changes. We maintain the setting $\kappa = b = 1/3$ for comparability. Figure 3 shows the $t_2$ and $t_{\text{HAR}}$ statistic values as $\lambda$ changes with different values of $T$ ($T = 1,000$ v.s. $T = 5,000$). It is apparent that there are only very minor, virtually undetectable differences in test behavior between these large sample sizes. This finding corroborates the result in Theorem 2 that these two tests have $O_p(1)$ order under the alternative and are not dependent on $T$ as $T \to \infty$. Moreover, since $K_{nt} \sim \sigma^2_a + \sigma^2_\varepsilon t^{-\lambda}$ under the alternative, as $\lambda$ increases, we can expect the discriminatory power for detecting convergence to dissipate rapidly for large $t$ because the impact on $K_{nt}$ of $\sigma^2_\varepsilon t^{-\lambda}$ is small when $t$ is large and $\lambda$ is not small. The gain in moving from $T = 1,000$ to $T = 5,000$ can be expected to be negligible in this case, as evidenced in Figure 3 and in the companion Figure 4. Figure 4 displays the $t_1$ and $t_{\text{HAC}}$ statistics for the same two values of $T$. As Theorem 1 predicts, both $t_1$ and $t_{\text{HAC}}$ are time series sample size dependent as $T \to \infty$. As $T$ increases, both test statistics become noticeably larger in absolute value when $\lambda$ is moderately small. With a large $\lambda$, the effects of rising $T$ are attenuated for the reason explained above and both statistics converge to $-\sqrt{3}$ as $\lambda$ increases indefinitely.

Figure 5 explains how the smoothing parameter $b$ influences the $t_{\text{HAR}}$ statistic across various $\lambda$ values. At a given $\lambda$, the absolute value of the $t_{\text{HAR}}$ statistic decreases initially as $b$ increases, but then reaches a minimum and begins to increase slowly as $b$ increases further towards unity. This functional dependence of $t_{\text{HAR}}$ on $b$ is highly nonlinear. The behavior is somewhat expected since it is known that as $b$ approaches to zero, the $t_{\text{HAR}}$ statistic approaches the $t_{\text{HAC}}$ statistic whose asymptotic behavior and dependence on $T$ is very different from the fixed-$b$ HAR statistic, as indeed is indicated in Theorem 2.
Figure 2: Test statistic numerical calculations ($b = \kappa = 1/3$, $T = 1,000$)

Figure 3: Impact of large $T$ on tests $t_2$ and $t_{HAR}$ ($b = \kappa = 1/3$)
Figure 4: Time varying behavior of $t_1$ and $t_{HAC}$ ($b = \kappa = 1/3$)

Figure 5: Limits of $t_{HAR}$ for various $b$ and $\lambda$ ($T = 1,000$).
5 Monte Carlo Simulations and An Empirical Example

First we report results of a simulation experiment designed to assess the finite sample performance of the convergence tests based on the various HAC and HAR estimator normalizations that were studied in the previous section. We also report an empirical application of the methods to study convergence behavior in unemployment rates among the 48 contiguous states of the USA.

5.1 Monte Carlo Simulations

We use the same data generating process considered by KPS (2019), viz.,

\[ y_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it} t^{-\beta}, \]  

where \( \epsilon_{it} = \rho \epsilon_{it-1} + v_{it}, a_i \sim iidN(0, 1), \mu_i \sim iidN(0, 1), v_{it} \sim iidN(0, 1), \) and \( \rho = 0.5. \)

We evaluate the size properties based on the restrictions \( \mu_i = \alpha = \beta = 0 \) under which the model is simply \( y_{it} = a_i + \epsilon_{it} \) and there is no convergence. Given the non-normal limit theory of the HAR test statistics, we use simulations to obtain the asymptotic critical values for these statistics. To do so we set \( n = T = 500 \) and \( \rho = 0.9 \) with \( y_{it} = a_i + \epsilon_{it} \). We run the fitted trend regression and from 50,000 replications compute the empirical distributions of \( t_2 \) and \( t_{HAR} \) with \( b = 0.1, 0.2 \) and 0.3. Table 1 reports the asymptotic critical values at the 5% level obtained in this manner. Evidently, as \( b \) increases, the critical values also increase in absolute value, which corroborates the limit theory that indicates greater departures from normality as \( b \) departs from zero and approaches unity.

Table 1: Simulated 5% critical values for \( t_2 \) and \( t_{HAR} \)

<table>
<thead>
<tr>
<th>( b )</th>
<th>( t_2 )</th>
<th>( t_{HAR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2.155</td>
<td>2.341</td>
</tr>
<tr>
<td>0.2</td>
<td>2.499</td>
<td>2.746</td>
</tr>
<tr>
<td>0.3</td>
<td>2.938</td>
<td>3.118</td>
</tr>
</tbody>
</table>

Table 2 shows test size based on a nominal asymptotic 5% rejection rate. Evidently, the statistics \( t_2 \) and \( t_{HAR} \) that are based on HAR corrections exhibit much milder size distortion compared with the \( t_1 \) and \( t_{HAC} \) statistics. The size distortion for all statistics diminishes as \( T \) increases and the number of the cross sectional units \( n \) has little influence on test size,
which is as expected since the tests all focus on trend behavior. Since we use asymptotic critical values calculated for each smoothing parameter \( b \) (from Table 1), the size distortions in the HAR statistics are little affected by the value of \( b \). Interestingly, size distortions in the \( t_1 \) test are uniformly smaller than those of \( t_{\text{HAC}} \), showing that the sandwich correction in the latter is a source of greater distortion.

Table 3 shows size adjusted test powers under the model (27) with \( \alpha = 0 \) and \( \beta = 0.05 \) (so that \( \lambda = 2\beta = 0.1 \)), corresponding to model M2 in KPS (2019). As \( n \) and \( T \) increase, test power goes to unity in all cases. Generally, as \( b \) increases for a given \( n \) and \( T \), the test powers of \( t_2 \) and \( t_{\text{HAR}} \) decline, showing that the use of fixed-\( b \) methods typically diminishes power as the lag truncation parameter rises. Most importantly, with few exceptions, the power of the \( t_1 \) test uniformly exceeds that of the other tests. For example, when \( T = n = 100 \), the power of the test based on \( t_1 \) equals 0.636 and the next most powerful test is \( t_2 \) with \( b = 0.1 \) (0.604). These simulation results corroborate the limit theory in Theorem 2 and the numerical findings shown in Figure 3 and 4.

Table 4 reports size adjusted test powers for the model (27) with \( \alpha = 0.05 \) (so that \( \lambda = 0.05 \)) and \( \beta = 0 \), corresponding to model M1 in KPS (2019). The results mirror the simulation findings for model M2 reported in Table 3 and, as for that model, the \( t_1 \) test is evidently the most powerful of the four tests. Again, it is noticeable that powers of the tests \( t_2 \) and \( t_{\text{HAR}} \) that use HAR corrections decrease as the value of \( b \) increases.

5.2 Empirical Example: State Unemployment Rates

Here we revisit one of the empirical examples used in KPS (2019) concerning potential convergence among unemployment rates in the 48 contiguous States of the USA from 2009:M8 to 2016:M7. Panel A of Figure 7 in KPS (2019) shows that the t-ratio \( t_\phi \) is \(-21.95\). Figure 6 below shows the time path of the sample cross section variance among the 48 unemployment rates, wherein there is clear visual evidence of diminishing variation over this historical period. The fitted values in (9) are as follows

\[
K_{nt} = 4.433 - 0.047 \times t + \hat{u}_t, \quad \hat{u}_t = 1.005 \times \hat{u}_{t-1} + \hat{v}_t.
\]

In terms of these estimated coefficients, this case is similar to the one that Phillips et al (2012) consider. But as we will show it is not realistic to assume that the unemployment rate cross section variation measure \( K_{nt} \) follows a trend regression with a nonstationary regression error. Nevertheless there is vivid evidence for weak \( \sigma \)-convergence in the data.
The cross sectional variance seems to stabilize around unity towards the end of the sample period over 2015-2016. If that is so, then one potential inference from these data is that it has taken around 5-6 years for the economy to adjust to the shock on the labor market of the subprime mortgage crisis. The stabilization period is part of the adjustment process and one that is not accounted for directly in a linear trend regression.

Suppose that the true DGP for $K_{nt}$ can be represented in the form

$$K_{nt} = a + gt^{-\lambda} + e_t.$$  \hfill (28)

By a rough calculation, assuming a 5 year adjustment, the trend decay rate parameter is approximately 0.3 if $a = 0$ and $g = 1$ in (28) since we have

$$\lambda = - \left[ \frac{1 - \ln 4.5}{\ln 5} \right] \approx 0.313.$$

This crude estimate of $\lambda$ gives an approximate idea of how the test statistics should behave, based on the limit theory and numerical calculations reported above. Since this calculated value of $\lambda$ is less than $1/2$, the large $n$–asymptotic limits of both the $t_{\text{HAR}}$ and $t_2$ tests approach negative constants.

![Figure 6: Cross section variance of unemployment rates over the 48 US contiguous States](image)

Table 5 reports the empirical findings for the $t$-ratios including various choices of the fixed-$b$ smoothing parameter. As noted in Figures 2 and 3, the $t_1$ statistic is typically larger than $t_{\text{HAC}}$ in absolute value, and also $|t_2| < |t_1|$ and $|t_{\text{HAR}}| < |t_{\text{HAC}}|$. The empirical values of
the test statistics all satisfy these inequalities. Specific values of $t_2$ and $t_{\text{HAR}}$ change along with the different values of $b$, but clearly the values in the table are far distant from the asymptotic critical values given in Table 1 for the HAR test statistics. And the empirical outcomes of the $t_1$ and $t_{\text{HAC}}$ statistics a far distant from the nominal 5% critical value -1.65. It follows that the generally supportive evidence for convergence does not change and is little influenced by the choice of the long run variance (LRV) estimate used in standardizing the various $t$-ratios, including the choice of smoothing parameter $b$ in HAR standardizations.

Table 5: Effectiveness of Various HAC/HAR Tests

<table>
<thead>
<tr>
<th>$k$</th>
<th>$t$-ratio</th>
<th>$k$</th>
<th>$t$-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>4</td>
<td>-21.948</td>
<td>$t_{\text{HAC}}$</td>
</tr>
<tr>
<td>$t_2$</td>
<td>$b = 0.1$</td>
<td>9</td>
<td>-16.860</td>
</tr>
<tr>
<td></td>
<td>$b = 0.2$</td>
<td>17</td>
<td>-14.446</td>
</tr>
<tr>
<td></td>
<td>$b = 0.3$</td>
<td>30</td>
<td>-14.254</td>
</tr>
</tbody>
</table>

To highlight the differences among the various LRV estimators, we consider a longer time series trajectory. Figure 7 plots the first and second central cross section moments of the unemployment rates – giving the cross section mean and variance – from 1976.M1 to 2018.M8. The trajectory of the cross section mean or national average does not appear to have an overall positive or negative trend behavior since the twin (recession associated) peaks (around 1982 and 2009) in the series are roughly comparable in magnitude. On the other hand, the cross section variation does show evidence of a decline over time, subject to the interlude of a rapid rise of the unemployment rate in 2009 and noting that the peak rate in 2009 is noticeably lower than the peak rate in 1982.
Figure 7: Time Varying Patterns of Cross Section Mean and Variance in US State Unemployment Rates over 42 years

Figure 8 shows recursive $t$-ratios computed for the trend regression with the four different LRV estimators giving time series for the trajectories of the test statistics $t_1$, $t_2$, $t_{HAC}$ and $t_{HAR}$. The starting point in the recursions is fixed and the end point in the sample moves over time. Around 1992, all of the four $t$-ratios pass through zero, indicating a move towards convergence. After this point, as the recursive calculations continue the $t$-ratios $t_1$ and $t_{HAC}$ show very similar values. Both are much smaller than the $t_2$ and $t_{HAR}$ statistics with smoothing parameter $b = 0.1$, a pattern that reflects that of Figure 2. Higher values of $b$ do not lead to major differences in these time paths for $t_2$ or $t_{HAR}$. When the end point in the recursion is around 2006, the total number of observations is around 360, so that we can include around 36 lags in the computation. Adding more lags in the LRV calculation makes little difference from this point. However, irrespective of the sample end point in the recursion the $t_1$ and $t_{HAC}$ trajectories are always below the 5% critical value of $-1.65$ from around 1995 forwards. On the other hand, both $t_2$ and $t_{HAR}$ trajectories exceed the normal critical value $-1.65$ (and therefore conservative critical value, given the HAR critical values in Table 1) for several years in the aftermath of the subprime mortgage crisis before falling below this critical value again around 2017. Thus, all four series show some evidence of $\sigma$-convergence but the evidence is stronger and more sustained over the full sample period in the $t_1$ and $t_{HAC}$ trajectories than for $t_2$ and $t_{HAR}$. This outcome squares with the analytic and simulation evidence that the $t_1$ and $t_{HAC}$ tests tend to have the greater discriminatory
6 Concluding Remarks

We investigate the use of various HAC, HAR and sandwich-type long run variance estimators in testing weak $\sigma$-convergence. These tests are particularly useful in assessing evidence for sustained diminution in cross section dispersion over time. The approach is easy to use in practice and is based on a simple linear trend regression with cross sectional variance as the dependent variable, as suggested in recent research by Kong et al. (2019). Under the null of no convergence, the trend regression is generally well specified since the trend slope coefficient is zero and the regression errors are taken as stationary time series. So standard normal limit theory for nonparametric studentization (with consistent variance estimates) of $t$-tests continues to hold under the null. Non-normal limit theory applies to fixed-$b$ HAR studentized $t$-tests, which in the present case needs to account for the complexities induced by the trend regression with cross section averaged data. Under convergence, the empirical linear trend regression is generally misspecified because it does not correctly capture trend decay formulations of diminishing variation. Conventional asymptotic theory for nonparametrically studentized $t$-tests fails to apply in such regressions and the tests have different asymptotic behavior in the presence of $\sigma$-convergence depending on the specific nature of the studentization.

The limit theory in the paper shows that $t$-ratios formed using traditional HAC variance estimates have better asymptotic behavioral characteristics in terms of discriminatory power.
for distinguishing convergence than those based on fixed-\(b\) HAR variance estimates. These asymptotics are supported by numerical explorations of the finite sample power properties of the various tests. There is also evidence in the simulations that HAR tests have less size distortion than HAC tests in finite samples, supporting earlier findings from both limit theory and simulations in traditional location model and GMM settings (Jansson, 2004; Sun et al, 2008). The results from simulations and limit theory further suggest that simple HAC standardizations outperform sandwich formula standardizations in terms of discriminatory power.

Application of these methods to assess diminishing variation in US State unemployment rates is largely confirmatory of the diminution over time, particularly in the latter period following the 2008 financial crisis. The empirical findings also corroborate the differences and power orderings in the HAC and HAR test behavior noted in the limit theory and simulations.
## Table 2: Sizes of Various Tests (Nominal Size: 5%)

<table>
<thead>
<tr>
<th>$T$</th>
<th>$n$</th>
<th>$t_1$</th>
<th>$t_{HAC}$</th>
<th>$t_2$</th>
<th>$t_{HAR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$b=$0.1</td>
<td>$b=$0.2</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.131</td>
<td>0.154</td>
<td>0.089</td>
<td>0.076</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.131</td>
<td>0.149</td>
<td>0.079</td>
<td>0.070</td>
</tr>
<tr>
<td>25</td>
<td>100</td>
<td>0.146</td>
<td>0.165</td>
<td>0.085</td>
<td>0.073</td>
</tr>
<tr>
<td>25</td>
<td>200</td>
<td>0.135</td>
<td>0.155</td>
<td>0.088</td>
<td>0.074</td>
</tr>
<tr>
<td>25</td>
<td>500</td>
<td>0.132</td>
<td>0.152</td>
<td>0.081</td>
<td>0.073</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.114</td>
<td>0.128</td>
<td>0.061</td>
<td>0.060</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.100</td>
<td>0.117</td>
<td>0.057</td>
<td>0.055</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.104</td>
<td>0.119</td>
<td>0.056</td>
<td>0.053</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>0.110</td>
<td>0.128</td>
<td>0.059</td>
<td>0.056</td>
</tr>
<tr>
<td>50</td>
<td>500</td>
<td>0.105</td>
<td>0.121</td>
<td>0.062</td>
<td>0.057</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.089</td>
<td>0.101</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.096</td>
<td>0.107</td>
<td>0.048</td>
<td>0.050</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.092</td>
<td>0.102</td>
<td>0.048</td>
<td>0.047</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.083</td>
<td>0.097</td>
<td>0.048</td>
<td>0.050</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>0.092</td>
<td>0.101</td>
<td>0.046</td>
<td>0.046</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.076</td>
<td>0.083</td>
<td>0.035</td>
<td>0.038</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.076</td>
<td>0.081</td>
<td>0.040</td>
<td>0.043</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.086</td>
<td>0.093</td>
<td>0.041</td>
<td>0.043</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.080</td>
<td>0.086</td>
<td>0.041</td>
<td>0.041</td>
</tr>
<tr>
<td>200</td>
<td>500</td>
<td>0.078</td>
<td>0.084</td>
<td>0.043</td>
<td>0.046</td>
</tr>
</tbody>
</table>
Table 3: Power Comparison under M2 (Size Adjusted): $\lambda = 2\beta = 0.1$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$n$</th>
<th>$t_1$</th>
<th>$t_{HAC}$</th>
<th>$t_2$ $b=0.1$</th>
<th>$t_2$ $b=0.2$</th>
<th>$t_2$ $b=0.3$</th>
<th>$t_{HAR}$ $b=0.1$</th>
<th>$t_{HAR}$ $b=0.2$</th>
<th>$t_{HAR}$ $b=0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.126</td>
<td>0.124</td>
<td>0.126</td>
<td>0.122</td>
<td>0.119</td>
<td>0.124</td>
<td>0.118</td>
<td>0.116</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.182</td>
<td>0.181</td>
<td>0.182</td>
<td>0.173</td>
<td>0.165</td>
<td>0.181</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>25</td>
<td>100</td>
<td>0.244</td>
<td>0.230</td>
<td>0.244</td>
<td>0.230</td>
<td>0.223</td>
<td>0.230</td>
<td>0.224</td>
<td>0.212</td>
</tr>
<tr>
<td>25</td>
<td>200</td>
<td>0.398</td>
<td>0.377</td>
<td>0.398</td>
<td>0.380</td>
<td>0.366</td>
<td>0.377</td>
<td>0.367</td>
<td>0.363</td>
</tr>
<tr>
<td>25</td>
<td>500</td>
<td>0.649</td>
<td>0.598</td>
<td>0.649</td>
<td>0.609</td>
<td>0.569</td>
<td>0.598</td>
<td>0.565</td>
<td>0.561</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.187</td>
<td>0.161</td>
<td>0.178</td>
<td>0.174</td>
<td>0.164</td>
<td>0.161</td>
<td>0.158</td>
<td>0.159</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.267</td>
<td>0.247</td>
<td>0.261</td>
<td>0.243</td>
<td>0.231</td>
<td>0.240</td>
<td>0.239</td>
<td>0.232</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.418</td>
<td>0.385</td>
<td>0.409</td>
<td>0.368</td>
<td>0.336</td>
<td>0.379</td>
<td>0.354</td>
<td>0.347</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>0.608</td>
<td>0.558</td>
<td>0.595</td>
<td>0.544</td>
<td>0.503</td>
<td>0.550</td>
<td>0.511</td>
<td>0.491</td>
</tr>
<tr>
<td>50</td>
<td>500</td>
<td>0.901</td>
<td>0.835</td>
<td>0.889</td>
<td>0.810</td>
<td>0.747</td>
<td>0.823</td>
<td>0.740</td>
<td>0.705</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.277</td>
<td>0.256</td>
<td>0.261</td>
<td>0.238</td>
<td>0.227</td>
<td>0.249</td>
<td>0.227</td>
<td>0.221</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.430</td>
<td>0.399</td>
<td>0.404</td>
<td>0.374</td>
<td>0.350</td>
<td>0.368</td>
<td>0.342</td>
<td>0.332</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.636</td>
<td>0.616</td>
<td>0.604</td>
<td>0.542</td>
<td>0.508</td>
<td>0.571</td>
<td>0.518</td>
<td>0.495</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.878</td>
<td>0.829</td>
<td>0.848</td>
<td>0.769</td>
<td>0.683</td>
<td>0.771</td>
<td>0.705</td>
<td>0.663</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>0.997</td>
<td>0.988</td>
<td>0.990</td>
<td>0.960</td>
<td>0.921</td>
<td>0.973</td>
<td>0.912</td>
<td>0.871</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.410</td>
<td>0.402</td>
<td>0.398</td>
<td>0.366</td>
<td>0.337</td>
<td>0.362</td>
<td>0.339</td>
<td>0.328</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.657</td>
<td>0.626</td>
<td>0.609</td>
<td>0.546</td>
<td>0.509</td>
<td>0.571</td>
<td>0.503</td>
<td>0.484</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.880</td>
<td>0.854</td>
<td>0.839</td>
<td>0.763</td>
<td>0.708</td>
<td>0.771</td>
<td>0.686</td>
<td>0.651</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.991</td>
<td>0.982</td>
<td>0.977</td>
<td>0.938</td>
<td>0.873</td>
<td>0.936</td>
<td>0.866</td>
<td>0.814</td>
</tr>
<tr>
<td>200</td>
<td>500</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.982</td>
<td>0.997</td>
<td>0.979</td>
<td>0.947</td>
</tr>
</tbody>
</table>
Table 4: Power Comparison under M1 (Size Adjusted): $\lambda = \alpha = 0.05$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$n$</th>
<th>$t_1$</th>
<th>$t_{HAC}$</th>
<th>$b = 0.1$</th>
<th>$b = 0.2$</th>
<th>$b = 0.3$</th>
<th>$t_{HAR}$ $b = 0.1$</th>
<th>$b = 0.2$</th>
<th>$b = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.081</td>
<td>0.086</td>
<td>0.081</td>
<td>0.080</td>
<td>0.082</td>
<td>0.086</td>
<td>0.084</td>
<td>0.083</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.111</td>
<td>0.111</td>
<td>0.111</td>
<td>0.108</td>
<td>0.102</td>
<td>0.111</td>
<td>0.109</td>
<td>0.108</td>
</tr>
<tr>
<td>25</td>
<td>100</td>
<td>0.135</td>
<td>0.128</td>
<td>0.135</td>
<td>0.128</td>
<td>0.124</td>
<td>0.128</td>
<td>0.127</td>
<td>0.124</td>
</tr>
<tr>
<td>25</td>
<td>200</td>
<td>0.182</td>
<td>0.183</td>
<td>0.182</td>
<td>0.176</td>
<td>0.176</td>
<td>0.183</td>
<td>0.177</td>
<td>0.176</td>
</tr>
<tr>
<td>25</td>
<td>500</td>
<td>0.306</td>
<td>0.303</td>
<td>0.306</td>
<td>0.293</td>
<td>0.284</td>
<td>0.303</td>
<td>0.285</td>
<td>0.290</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.102</td>
<td>0.098</td>
<td>0.100</td>
<td>0.103</td>
<td>0.100</td>
<td>0.099</td>
<td>0.097</td>
<td>0.094</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.135</td>
<td>0.126</td>
<td>0.131</td>
<td>0.123</td>
<td>0.120</td>
<td>0.125</td>
<td>0.120</td>
<td>0.118</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.187</td>
<td>0.185</td>
<td>0.185</td>
<td>0.168</td>
<td>0.159</td>
<td>0.183</td>
<td>0.170</td>
<td>0.171</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>0.262</td>
<td>0.253</td>
<td>0.256</td>
<td>0.243</td>
<td>0.235</td>
<td>0.250</td>
<td>0.239</td>
<td>0.234</td>
</tr>
<tr>
<td>50</td>
<td>500</td>
<td>0.472</td>
<td>0.446</td>
<td>0.462</td>
<td>0.417</td>
<td>0.390</td>
<td>0.440</td>
<td>0.405</td>
<td>0.391</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.129</td>
<td>0.127</td>
<td>0.123</td>
<td>0.119</td>
<td>0.112</td>
<td>0.122</td>
<td>0.113</td>
<td>0.112</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.185</td>
<td>0.189</td>
<td>0.179</td>
<td>0.172</td>
<td>0.158</td>
<td>0.179</td>
<td>0.170</td>
<td>0.165</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.262</td>
<td>0.266</td>
<td>0.252</td>
<td>0.238</td>
<td>0.236</td>
<td>0.255</td>
<td>0.237</td>
<td>0.234</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.407</td>
<td>0.395</td>
<td>0.390</td>
<td>0.353</td>
<td>0.325</td>
<td>0.365</td>
<td>0.348</td>
<td>0.337</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>0.732</td>
<td>0.688</td>
<td>0.694</td>
<td>0.627</td>
<td>0.592</td>
<td>0.658</td>
<td>0.592</td>
<td>0.563</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.167</td>
<td>0.173</td>
<td>0.165</td>
<td>0.158</td>
<td>0.153</td>
<td>0.163</td>
<td>0.158</td>
<td>0.154</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.252</td>
<td>0.247</td>
<td>0.240</td>
<td>0.227</td>
<td>0.217</td>
<td>0.236</td>
<td>0.219</td>
<td>0.217</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.399</td>
<td>0.390</td>
<td>0.366</td>
<td>0.337</td>
<td>0.319</td>
<td>0.343</td>
<td>0.314</td>
<td>0.310</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.624</td>
<td>0.596</td>
<td>0.586</td>
<td>0.528</td>
<td>0.490</td>
<td>0.526</td>
<td>0.495</td>
<td>0.470</td>
</tr>
<tr>
<td>200</td>
<td>500</td>
<td>0.918</td>
<td>0.884</td>
<td>0.871</td>
<td>0.792</td>
<td>0.729</td>
<td>0.806</td>
<td>0.716</td>
<td>0.672</td>
</tr>
</tbody>
</table>
References


Appendix

Assumptions

We base our conditions in Assumption A below on those employed in KPS (2019), augmented by conditions on the existence of the long run variance of $\varepsilon_{it}^2$. The kernel conditions given in Assumption B are similar to those employed by Sun et al (2008) and Sun (2014) and are satisfied by the Bartlett-Newey-West kernel used in the body of the paper.

Assumption A:

(i) The model error term $\varepsilon_{it} \sim iid (0, \sigma_i^2)$ over $i$ with finite fourth moment $\mathbb{E}(\varepsilon_{it}^4) < \infty$ and is strictly stationary over $t$. The autocovariance sequence $\gamma_{\varepsilon}(h) = \mathbb{E}(\varepsilon_{it}\varepsilon_{it+h})$ of $\varepsilon_{it}$ satisfies the summability condition $\sum_{h=1}^{\infty} h |\gamma_{\varepsilon}(h)| < \infty$ and $\varepsilon_{it}$ has long run variance $\Omega_{\varepsilon}^2 = \sum_{h=\infty}^{h=-\infty} \gamma_{\varepsilon}(h) > 0$. The autocovariance sequence $\gamma_{\varepsilon^2}(h) = \mathbb{E}\{(\varepsilon_{it}^2 - \sigma_i^2)(\varepsilon_{it+h}^2 - \sigma_i^2)\}$ of $\varepsilon_{it}^2$ satisfies the summability condition $\sum_{h=1}^{\infty} h |\gamma_{\varepsilon^2}(h)| < \infty$ and $\varepsilon_{it}^2$ has long run variance $\Omega_{\varepsilon^2}^2 = \sum_{h=\infty}^{h=-\infty} \gamma_{\varepsilon^2}(h) > 0$.

(ii) The coefficients $a_i \sim iid (a, \sigma_a^2)$ and are independent of $\varepsilon_{js}$ for all $\{i, j, s, t\}$.

Assumption B:

(i) $k(x) : \mathbb{R} \rightarrow [0, 1]$ is symmetric, piecewise smooth with $k(x) = 0$ for $|x| > 1$, $k(0) = 1$, and $\int_{-1}^{1} k(x) \, dx = 1$.

(ii) The Parzen characteristic exponent defined by

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+, \, g_{q_0} = \lim_{x \to 0} \frac{1 - k(x)}{|x|^{q_0}} < \infty\}$$

is greater than or equal to 1.

(iii) $k(x)$ is positive semidefinite, i.e., for any square integrable function $f(x)$, $\int_0^{\infty} \int_0^{\infty} k(s-t) f(s) f(t) ds dt \geq 0$.

The identical distribution assumption in A(i) and A(ii) is convenient in what follows but can no doubt be relaxed under stronger uniform moment conditions that assure the validity of laws of large numbers and central limit theory. Some changes in the formulae for the variances and long run variances in those cases would be needed. Assumption B is similar
to the kernel conditions in Sun et al (2008) and Sun (2014), assures a nonnegative long run variance estimator, and is sufficient to validate first order and some higher order asymptotics, although the latter are not considered here. The conditions in Assumption B are satisfied by the Bartlett-Newey-West estimator employed in the long run variance estimators (17)- (19) used in the text.

Generating mechanism under the null hypothesis

In view of (6) and (8), the data on $K_{nt}$ are obtained by cross section aggregation as follows

$$K_{nt} = \frac{1}{n} \sum_{i=1}^{n} \left( x_{it} - \frac{1}{n} \sum_{i=1}^{n} x_{it} \right)^2 = \sigma_{a,n}^2 + \eta_{n,t} + \varepsilon_{n,t},$$

(30)

where $\sigma_{a,n}^2 = n^{-1} \sum_{i=1}^{n} \tilde{\alpha}_i^2$, $\eta_{n,t} = \sigma_{\varepsilon,T}^2 t^{-2\beta}$ is the finite sample trend decay function, and

$$\varepsilon_{n,t} = 2n^{-1} \sum_{i=1}^{n} \tilde{\alpha}_i \tilde{\varepsilon}_{it} t^{-\beta} + \left( \sigma_{\varepsilon,T}^2 - \sigma_{\varepsilon,T}^2 \right) t^{-2\beta},$$

(31)

with $\sigma_{\varepsilon,T}^2 = n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_{it}^2$, $\sigma_{\varepsilon,T}^2 = T^{-1} \sum_{t=1}^{T} \sigma_{\varepsilon,nT}^2$, and where the notation $\tilde{\varepsilon}_{it} = \varepsilon_{it} - n^{-1} \sum_{j=1}^{n} \varepsilon_{jt}$ is used for deviations from cross section means. It follows that

$$K_{nt} = \sigma_{a,n}^2 + \eta_t + \varepsilon_{n,t} + \xi_{n,t},$$

where $\eta_t = \sigma_{\varepsilon}^2 t^{-\lambda}$, $\lambda = 2\beta$, $\xi_{n,t} = \eta_{n,t} - \eta_t = \left( \sigma_{\varepsilon,T}^2 - \sigma_{\varepsilon}^2 \right) t^{-\lambda}$ and $\sigma_{\varepsilon}^2$ is the variance of $\varepsilon_{it}$.

Since we assume $a_i \sim iid (a, \sigma_a^2)$, we have $\sigma_{a,n}^2 = \sigma_a^2 + O_p \left( \frac{1}{\sqrt{n}} \right)$, $n^{-1} \sum_{i=1}^{n} \tilde{\alpha}_i \tilde{\varepsilon}_{it} = O_p \left( \frac{1}{\sqrt{n}} \right)$, $\sigma_{\varepsilon,T}^2 = n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_{it}^2 = \sigma_{\varepsilon}^2 + O_p \left( \frac{1}{\sqrt{n}} \right)$ uniformly in $t \leq T$, and $\sigma_{\varepsilon,nT}^2 = T^{-1} \sum_{t=1}^{T} \sigma_{\varepsilon,nT}^2 = \sigma_{\varepsilon}^2 + O_p \left( \frac{1}{\sqrt{\sqrt{T}}} \right)$. It follows that

$$\sigma_{\varepsilon,T}^2 - \sigma_{\varepsilon,nT}^2 = n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_{it}^2 - T^{-1} \sum_{t=1}^{T} \sigma_{\varepsilon,nT}^2$$

$$= n^{-1} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_{\varepsilon}^2 \right) - \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_{\varepsilon}^2 \right)$$

$$= n^{-1} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_{\varepsilon}^2 \right) + O_p \left( \frac{1}{\sqrt{nT}} \right),$$

and then, when $\lambda = 2\beta = 0$, we deduce that

$$K_{nt} = \sigma_{a,n}^2 + \sigma_{\varepsilon}^2 + 2n^{-1} \sum_{i=1}^{n} \tilde{\alpha}_i \tilde{\varepsilon}_{it} + \left( \sigma_{\varepsilon,T}^2 - \sigma_{\varepsilon,nT}^2 \right) + \left( \sigma_{\varepsilon,nT}^2 - \sigma_{\varepsilon}^2 \right)$$

$$= \sigma_{a,n}^2 + 2n^{-1} \sum_{i=1}^{n} \tilde{\alpha}_i \tilde{\varepsilon}_{it} + n^{-1} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_{\varepsilon}^2 \right) + O_p \left( \frac{1}{\sqrt{nT}} \right),$$

(32)
with $\sigma^2_n = \sigma^2_{a,n} + \sigma^2_{e} + \left(\sigma^2_{e,nT} - \sigma^2_{e}\right)$. We can write (32) as

$$K_{nt} = \sigma^2_n + u_{nt}, \quad \text{with } u_{nt} = \frac{1}{\sqrt{n}} \zeta_{nt} + O_p\left(\frac{1}{\sqrt{nT}}\right),$$

(33)

and

$$\zeta_{nt} = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i \tilde{\epsilon}_{it} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\tilde{\epsilon}_{it}^2 - \sigma^2_{e}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_{it},$$

(34)

$$\zeta_{it} = 2\tilde{a}_i \tilde{\epsilon}_{it} + \left(\tilde{\epsilon}_{it}^2 - \sigma^2_{e}\right).$$

(35)

The generating mechanism (33) is a panel location model in which the error term $u_{nt} \sim \frac{1}{\sqrt{n}} \zeta_{nt} = O_p\left(\frac{1}{\sqrt{n}}\right)$, which is a consequence of the cross section aggregation involved in the definition of $K_{nt}$. Upon taking deviations from time series means, we have

$$\tilde{K}_{nt} = K_{nt} - T^{-1} \sum_{t=1}^{T} K_{nt} = u_{nt} - T^{-1} \sum_{t=1}^{T} u_{nt} = \tilde{u}_{nt} = u_{nt} + O_p\left(\frac{1}{\sqrt{nT}}\right),$$

(36)

since

$$\frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \tilde{a}_i \tilde{\epsilon}_{it} = O_p\left(\frac{1}{\sqrt{nT}}\right) \quad \text{and} \quad \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \left(\tilde{\epsilon}_{it}^2 - \sigma^2_{e}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right).$$

Thus, the error term in the effective model under the null stems from equations (33) and (34), viz.,

$$K_{nt} = \sigma^2_n + u_{nt} = \sigma^2_n + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_{it} + O_p\left(\frac{1}{\sqrt{nT}}\right),$$

(37)

which is a simple location model in which the effective error $u_{nt}$ involves $\zeta_{it}$ and hence both first and second centred sample moments of $\epsilon_{it}$ in view of (35).

**Proof of Theorem 1**

**Proof of (i) The fitted trend regression of $K_{nt}$ on a linear time trend is**

$$K_{nt} = \tilde{a}_{nT} + \hat{\phi}_{nT} t + \hat{u}_{t}.$$

(38)

The regression coefficient $\hat{\phi}_{nT}$ has the explicit form

$$\hat{\phi}_{nT} = \sum_{t=1}^{T} c_{tT} \tilde{K}_{nt} = \sum_{t=1}^{T} c_{tT} u_{nt}, \quad \text{with } c_{tT} = \tilde{t} / \left(\sum_{s=1}^{T} \tilde{s}^2\right).$$


From (33) and (36) we have \( K_{nt} = \sigma_n^2 + u_{nt} \), with \( u_{nt} = \frac{1}{\sqrt{n}}\varepsilon_{nt} + O_p\left(\frac{1}{\sqrt{nT}}\right) \), and so

\[
\hat{\phi}_{nT} = \sum_{t=1}^{T} c_{iT} \widetilde{K}_{nt} = \sum_{t=1}^{T} c_{iT} u_{nt}
\]

\[
= \sum_{t=1}^{T} c_{iT} \left[ 2n^{-1} \sum_{i=1}^{n} \tilde{a}_i \tilde{\varepsilon}_{it} + n^{-1} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right) + O_p\left(\frac{1}{nT}\right) \right]
\]

\[
= \frac{1}{\sqrt{n}} \sum_{t=1}^{T} c_{iT} \left[ 2 \sqrt{n} \sum_{i=1}^{n} \tilde{a}_i \tilde{\varepsilon}_{it} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right) + O_p\left(\frac{1}{\sqrt{T}}\right) \right].
\]

Then,

\[
\sqrt{nT^{3/2}} \hat{\phi}_{nT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{t}_{t}}{T^3/2} \left[ 2 \sqrt{n} \sum_{i=1}^{n} \tilde{a}_i \tilde{\varepsilon}_{it} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right) + O_p\left(\frac{1}{\sqrt{T}}\right) \right]
\]

\[
= \frac{12}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{t}_{t}}{T} \left( 2 \sqrt{n} \sum_{i=1}^{n} \tilde{a}_i \tilde{\varepsilon}_{it} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right) \right) + o_p(1)
\]

\[\overset{\sim}{\sim}_{(n,T \to \infty)} \mathcal{N}\left(0, 12\Omega_\phi^2\right). \tag{39}\]

We proceed to prove (39). In this expression for the limit theory the variance \( \Omega_\phi^2 \) depends on the two components \( \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i \tilde{\varepsilon}_{it} \) and \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right) \). We can use sequential asymptotics to simplify the derivation and appeal to joint asymptotics using the limit theory of Phillips and Moon (1999) which applies under Assumption A. Note that \( \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i \tilde{\varepsilon}_{it} \) and \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right) \) are asymptotically normal and independent in view of the independence of \( a_i \) and \( \varepsilon_{it} \) as

\[
\mathbb{E} \left\{ (a_i - a) \varepsilon_{it} (\varepsilon_{is}^2 - \sigma_\varepsilon^2) \right\} = \mathbb{E} (a_i - a) \mathbb{E} \left\{ \varepsilon_{it} (\varepsilon_{is}^2 - \sigma_\varepsilon^2) \right\} = 0, \text{ for all } t, s
\]

Using \( \tilde{\varepsilon}_{it} = \varepsilon_{it} - n^{-1} \sum_{j=1}^{n} \varepsilon_{jt} = \varepsilon_{it} + O_p\left(\frac{n^{-1/2}}{}\right) \) and \( \tilde{a}_i = a_i - a + O_p\left(\frac{n^{-1/2}}{}\right) \), under Assumption A the limit theory of the two components as \( T \to \infty \) is therefore given by

\[
\left[ \begin{array}{c}
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{t}_{t}}{T} \tilde{a}_i \tilde{\varepsilon}_{it} \\
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{t}_{t}}{T} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right)
\end{array} \right] \overset{\sim}{\sim}_{(T \to \infty)} \mathcal{N}\left(0, \frac{1}{12} \begin{bmatrix}
\sigma_a^2 \Omega_\varepsilon^2 & 0 \\
0 & \Omega_\varepsilon^2
\end{bmatrix}\right). \tag{40}
\]

In the limit distribution (40) \( \Omega_\varepsilon^2 \) is the long run variance of \( \varepsilon_{it} \), and \( \Omega_\varepsilon^2 \) is the long run variance of \( \varepsilon_{it}^2 \). Then, as in Phillips and Moon (1999), we deduce the joint limit theory

\[
\left[ \begin{array}{c}
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\tilde{t}_{t}}{T} \tilde{a}_i \tilde{\varepsilon}_{it} \\
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\tilde{t}_{t}}{T} \left( \tilde{\varepsilon}_{it}^2 - \sigma_\varepsilon^2 \right)
\end{array} \right] \overset{\sim}{\sim}_{(n,T \to \infty)} \mathcal{N}\left(0, \frac{1}{12} \begin{bmatrix}
\sigma_a^2 \Omega_\varepsilon^2 & 0 \\
0 & \Omega_\varepsilon^2
\end{bmatrix}\right), \tag{41}
\]

37
where \( \sigma_a^2 = \mathbb{E} (a_i - \bar{a})^2 \). The long run variance of \( \zeta_{it} = \{ 2\bar{a}_i \hat{e}_{it} + (\hat{e}_{it}^2 - \sigma_e^2) \} \) is \( \Omega_\phi^2 = 4\sigma_a^2 \Omega_e^2 + \Omega_e^2 \) and it follows from (41) that

\[
\frac{1}{nT} \sum_{t=1}^n \sum_{i=1}^T \zeta_{it} \sim \frac{1}{\sqrt{nT}} \sum_{t=1}^n \sum_{i=1}^T \{ 2\hat{a}_i \hat{e}_{it} + (\hat{e}_{it}^2 - \sigma_e^2) \} \sim_{(n,T \rightarrow \infty)} \mathcal{N} \left( 0, \Omega_\phi^2 \right) . \tag{42}
\]

We deduce that

\[
\sqrt{nT^{3/2}} \hat{\phi}_{nT} = \frac{12}{\sqrt{T}} \sum_{t=1}^T \frac{\hat{t}}{T} \left[ \frac{2}{\sqrt{n}} \sum_{i=1}^n \hat{a}_i \hat{e}_{it} + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{e}_{it}^2 - \sigma_e^2) \right] + o_p(1)
\]

\[
\sim_{(n,T \rightarrow \infty)} Z_\phi = \mathcal{N} \left( 0, 12\Omega_\phi^2 \right) , \quad \text{with } \Omega_\phi^2 = 4\sigma_a^2 \Omega_e^2 + \Omega_e^2 . \tag{43}
\]

The formula for the limiting long run variance, \( \Omega_\phi^2 = 4\sigma_a^2 \Omega_e^2 + \Omega_e^2 \), is the sum of two components: One arises from the sample covariance term \( \sum_{i=1}^n \sum_{t=1}^T \hat{t} \hat{a}_i \hat{e}_{it} \), which is linear in \( \epsilon_{it} \); and the other from the term \( \sum_{i=1}^n \sum_{t=1}^T (\hat{e}_{it}^2 - \sigma_e^2) \), which is quadratic in \( \epsilon_{it} \).

In (43) the dependence of the asymptotic variance \( \Omega_\phi^2 \) of \( \sqrt{nT^{3/2}} \hat{\phi}_{nT} \) on the two long run variance components \( (\Omega_e^2, \Omega_e^2) \) is to be expected. This is because the dependent variable in the fitted trend regression \( K_{nt} \) given by (30) involves second order quantities that measure variation. Correspondingly, the ‘error’ term \( \varepsilon_{n,t} \) given in (31) involves both a first order (cross product) sample moment and a second order moment. These appear in (32) as the components \( 2n^{-1} \sum_{i=1}^n \bar{a}_i \hat{e}_{it} \) and \( n^{-1} \sum_{i=1}^n (\hat{e}_{it}^2 - \sigma_e^2) \), which in turn lead to the more complex limiting variance matrix \( \Omega_\phi^2 \), which depends on fourth moments of \( \epsilon_{it} \).

**Proof of (ii)** The \( t_1 \) statistic can be written as

\[
t_1 = \frac{\hat{\phi}_{nT}}{\sqrt{\Omega_1^2 (\sum_{t=1}^T \hat{t}^2)^{-1}}} = \frac{\sqrt{nT^{3/2}} \hat{\phi}_{nT}}{\sqrt{n\hat{\Omega}_1^2 (T^{-3} \sum_{t=1}^T \hat{t}^2)^{-1}}} ,
\]

where \( \hat{\Omega}_1^2 \) is the standard HAC estimator (with Bartlett-Newey-West kernel)

\[
\hat{\Omega}_1^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 + \frac{2}{T} \sum_{\ell=1}^L \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \hat{u}_t \hat{u}_{t+\ell} .
\]

We first derive asymptotics of the HAC estimator \( \hat{\Omega}_1^2 \). We start with an analysis of the residuals in the fitted regression (38). Under the null, the data \( K_{nt} \) has the location model form \( K_{nt} = \sigma_n^2 + u_{nt} = \sigma_n^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{it} + O_p \left( \frac{1}{\sqrt{nT}} \right) \) given in (37). The intercept in the
regression (38) therefore satisfies
\[
\hat{a}_{nT} = \frac{1}{T} \sum_{t=1}^{T} \left( K_{nt} - \hat{\phi}_{nT} t \right) + n^{2} + T^{-1} \sum_{t=1}^{T} u_{nt} - \frac{1}{2} \sqrt{nT^{3/2}} \hat{\phi}_{nT} \frac{(T + 1)}{\sqrt{nT^{3/2}}},
\]
\[
= \sigma_n^2 + O_p \left( \frac{1}{\sqrt{nT}} \right),
\]
using (36) and (39). Hence, the residual in (38), \( \hat{u}_t \), becomes
\[
\hat{u}_t = K_{nt} - \hat{a}_{nT} - \hat{\phi}_{nT} t = \tilde{u}_n - \hat{\phi}_{nT} \tilde{t},
\]
\[
= \tilde{u}_n - \left( \sum_{t=1}^{T} c_{T\tilde{u}_n} \right) \tilde{t},
\]
(44)
where \( \tilde{u}_t \) is simply linearly detrended \( u_{nt} \). Since \( u_{nt} = \frac{1}{\sqrt{n}} \zeta_{nt} \), it follows that \( \tilde{u}_t = n^{-1/2} \zeta_{nt} \), where \( \zeta_{nt} = \tilde{\zeta}_{nt} - \left( \sum_{t=1}^{T} c_{T\tilde{u}_n} \right) \tilde{t} \) is linearly detrended \( \tilde{\zeta}_{nt} \). We are effectively demeaning and detrending the errors \( u_{nt} \) by the trend regression which, from (44), gives the residuals
\[
\hat{u}_t = \tilde{u}_n - \left( \sum_{t=1}^{T} c_{T\tilde{u}_n} \right) \tilde{t} = \tilde{u}_n - \frac{\tilde{t}}{T} \times O_p \left( \frac{1}{\sqrt{nT}} \right).
\]
so that, since \( u_{nt} = n^{-1/2} \zeta_{nt} \), we have
\[
\sqrt{n} \hat{u}_t = \zeta_{nt} - \frac{t}{T} \times O_p \left( \frac{1}{\sqrt{T}} \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \zeta_{it} + o_p(1) \sim_{n \rightarrow \infty} \zeta_t^0.
\]
Thus, the sample second moments of the residuals \( \hat{u}_t \) are asymptotically, as \( n \rightarrow \infty \), equivalent to those of \( \zeta_t^0 \). It follows that the sample second moments and autocovariances of \( \hat{u}_t \) have the following limit behavior after scaling by \( n \)
\[
n^2 \hat{\gamma}(0) = \frac{n}{T} \sum_{t=1}^{T} \hat{u}_t^2 = \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_{it} + o_p(1) \right\}^2 \rightarrow_p \mathbb{E} \left( \zeta_{it}^2 \right)
\]
\[
= 4 \mathbb{E} \left\{ (a_i - a)^2 \sigma_t^2 \right\} + \mathbb{E} \left( \epsilon_t^2 - \sigma_t^2 \right)^2 = 4 \sigma_a^2 \sigma_t^2 + \sigma_t^2.
\]

Similarly, setting \( \hat{\gamma}(j) := \frac{n}{T} \sum_{1 \leq t \leq j \leq T} \sum_{1 \leq t \leq j \leq T} \zeta_{it} \zeta_{jt}^* \), we have
\[
n^2 \hat{\gamma}(j) = \frac{n}{T} \sum_{1 \leq t \leq j \leq T} \sum_{1 \leq t \leq j \leq T} \zeta_{it} \zeta_{jt}^* \rightarrow_p \mathbb{E} \left( \zeta_{it} \zeta_{jt} \right) = 4 \mathbb{E} \left\{ (a_i - a)^2 \mathbb{E} \left( \epsilon_t \epsilon_t \right) + \mathbb{E} \left( \epsilon_t^2 - \sigma_t^2 \right) \left( \epsilon_t^2 - \sigma_t^2 \right) \right\} = 4 \sigma_a^2 \Gamma_e (j) + \Gamma_{\epsilon^2} (j),
\]
where we use the notation \( \Gamma_e (j) = \mathbb{E} \left\{ (v_t - \mathbb{E} v_t) (v_t - \mathbb{E} v_t) \right\} \). We deduce that the HAC estimator has the following asymptotic behavior as \( T \rightarrow \infty \) with \( L = \lfloor T^{1/3} \rfloor \) as in the Bartlett-Newey-West estimator
\[
n \hat{\Omega}^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{t=1}^{L} \sum_{t=1}^{T-t} \left( 1 - \frac{\ell}{L+1} \right) \hat{u}_t \hat{u}_{t+\ell}
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} \zeta_t^0 \right)^2 + \frac{2}{T} \sum_{t=1}^{L} \sum_{t=1}^{T-t} \left( 1 - \frac{\ell}{L+1} \right) \zeta_t^0 \zeta_{t+\ell}^0 + o_p(1)
\]
\[
\rightarrow_p \sum_{j=-\infty}^{\infty} \left\{ 4 \sigma_a^2 \Gamma_e (j) + \Gamma_{\epsilon^2} (j) \right\} = \Omega_\zeta = \Omega_{\zeta}^2.
\]

The final result for the limit distribution of the \( t_1 \) statistic
\[
t_1 = \frac{\sqrt{n} T^{3/2} \hat{\phi}_{nT}}{\sqrt{n \hat{\Omega}_2^2 \left( T^{-3} \sum_{t=1}^{T} \hat{I}^2 \right)^{-1}}} \sim \mathcal{N} \left( 0, \frac{12 \Omega_{\zeta}^2}{\sqrt{12 \Omega_{\zeta}^2}} \right) \equiv \mathcal{N} \left( 0, 1 \right),
\]
now follows.

**Proof of (iii)** We next consider the analogous \( t \) statistic
\[
t_2 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_2^2 \left( \sum_{t=1}^{T} \hat{I}^2 \right)^{-1}}} = \frac{\sqrt{n} T^{3/2} \hat{\phi}_{nT}}{\sqrt{n \hat{\Omega}_2^2 \left( T^{-3} \sum_{t=1}^{T} \hat{I}^2 \right)^{-1}}}.
\]
in which the fixed-\( b \) HAR long run variance estimate

\[ \hat{\Omega}_2^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) \hat{u}_t \hat{u}_{t+\ell}, \]

from (19) is employed. In the analysis that follows it is convenient to use a general kernel function \( k(x) \) satisfying Assumption B.

We start by considering the scaled full sample (with fixed-\( b \) parameter \( b = 1 \) and \( M = T - 1 \)) HAR estimate

\[ n\hat{\Omega}_2^2 = \sum_{j=1-T}^{T-1} k \left( \frac{j}{T} \right) \gamma_u (j) = \frac{n}{T} \sum_{j=1-T}^{T-1} k \left( \frac{j}{T} \right) \sum_{1 \leq t,j \leq T} \hat{u}_t \hat{u}_{t-j} \]

\[ = \frac{n}{T} \sum_{t=1}^{T} \sum_{p=1}^{T} k \left( \frac{t-p}{T} \right) \hat{u}_t \hat{u}_p = \frac{1}{T} \sum_{t=1}^{T} \sum_{p=1}^{T} k \left( \frac{t-p}{T} \right) \zeta_{nt}^{\tau} \zeta_{np}^{\tau}. \]

In view of the functional laws (46) and using the same arguments as in Kiefer and Vogelsang (2002, 2005), Sun et al. (2008) and Sun (2014) we find that

\[
\begin{align*}
\sqrt{n} \hat{\Omega}_2^2 & = \sum_{t=1}^{T} \sum_{p=1}^{T} k \left( \frac{t-p}{T} \right) \frac{\zeta_{nt}^{\tau}}{\sqrt{T}} \frac{\zeta_{np}^{\tau}}{\sqrt{T}} \overset{\overset{\sim}{\rightarrow}}{\overset{(n,T \to \infty)}{}} \int_0^1 \int_0^1 k (r-s) dB^{\tau}_{\Omega_2} (r) d\sigma^{\tau}_{\Omega_2} (s) .
\end{align*}
\]

Using \( \Omega_2^{\tau} = \Omega_\phi^{\tau} \), the HAR \( t \) statistic for testing the significance of the linear trend in the empirical regression (38) therefore has the following limit theory as \( (n, T \to \infty) \)

\[ t_2 = \frac{\sqrt{nT^{3/2} \hat{\phi}_n \phi}}{\sqrt{n\hat{\Omega}_2 \left( T^{-3} \sum_{t=1}^{T} \hat{u}_t^2 \right)^{-1}}} \overset{\overset{\sim}{\rightarrow}}{\overset{(n,T \to \infty)}{}} \frac{Z_{\phi}}{\sqrt{12 \int_0^1 \int_0^1 k (r-s) dB_{\Omega_2}^{\tau} (r) d\sigma_{\Omega_2}^{\tau} (s)}} \]

\[
\begin{align*}
& \overset{\overset{\sim}{\rightarrow}}{\overset{(n,T \to \infty)}{}} \frac{Z}{\sqrt{12 \Omega_\phi \int_0^1 \int_0^1 k (r-s) dW^{\tau} (r) dW^{\tau} (s)}} \mathbb{I} \equiv \frac{Z}{\sqrt{12 \Omega_\phi \int_0^1 \int_0^1 k (r-s) D^{\tau} (r) dW^{\tau} (s)}} \overset{\overset{\sim}{\rightarrow}}{\overset{(n,T \to \infty)}{}} \frac{Z}{\sqrt{12 \Omega_\phi \int_0^1 \int_0^1 k (r-s) D^{\tau} (r) dW^{\tau} (s)}} \end{align*}
\]

where \( Z_\phi \equiv \mathcal{N} \left( 0, 12 \Omega_\phi^{2} \right), Z = \mathcal{N} (0, 1) \), and \( B_{\Omega_2}^{\tau} (r) = \Omega_\xi W^{\tau} (r) \) where \( W^{\tau} (r) \) is detrended standard brownian motion \( W (r) = BM (1) \). Since \( Z \) and \( W^{\tau} (r) \) are independent, the final expression for the limit theory given in (52) is therefore free of nuisance parameters.

When the fixed-\( b \) kernel smoothing parameter \( b \) satisfies \( b \in (0, 1) \) scale adjustments in the derivations show that the corresponding limit theory is given by

\[ t_{\text{HAR}} \overset{\overset{\sim}{\rightarrow}}{\overset{(n,T \to \infty)}{}} \frac{Z}{\sqrt{\left\{ \int_0^1 \int_0^1 k \left( \frac{r-s}{b} \right) dW^{\tau} (r) dW^{\tau} (s) \right\}^{1/2}}}, \]

41

**Proof of (iv)** In this case we use the variance estimate in sandwich form with $t$ ratio

$$t_{\text{HAR}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left(\sum_{t=1}^{T} \tilde{\ell}_{t}^{2}\right)^{-1} T \hat{\Omega}_{M}^{2} \left(\sum_{t=1}^{T} \tilde{\ell}_{t}^{2}\right)^{-1}}}$$

using $\hat{\Omega}_{M}^{2} = \frac{1}{T} \sum_{t=1}^{T} \tilde{z}_{t}^{2} + \frac{2}{T} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left(1 - \frac{\ell}{M+1}\right) \tilde{z}_{t} \tilde{z}_{t+\ell}$. In what follows, it is again convenient to use a general kernel $k(x)$ satisfying Assumption B.

In $\hat{\Omega}_{M}^{2}$, the variate $\tilde{z}_{t} = \hat{u}_{t} \tilde{t} = \frac{1}{\sqrt{n}} \tilde{\zeta}_{nt} \tilde{t}$. Setting $\hat{\gamma}_{\tilde{x}}(j) := \frac{n}{T} \sum_{1 \leq t \leq j \leq T} \tilde{z}_{t} \tilde{z}_{t-j}$, and proceeding as in (47) - (48), we have

$$\frac{n}{T^{2}} \hat{\gamma}_{\tilde{x}}(0) = \frac{n}{T} \sum_{1 \leq t \leq T} \hat{u}_{t}^{2} \left(\frac{t}{T}\right) = \frac{1}{T} \sum_{1 \leq t \leq T} \left(\tilde{\zeta}_{nt}\right)^{2} \left(\frac{t}{T}\right)^{2}$$

$$\sim_{p} 12 \mathbb{E}\left\{\left(\hat{\zeta}_{nt}^{0}\right)^{2}\right\} = 12 \left[4 \sigma_{2}^{2} \Gamma_{1}^{0}(0) + \Gamma_{1}^{2}(0)\right],$$

and similarly for all $j$ such that $\frac{j}{T} \rightarrow 0$

$$\frac{n}{T^{2}} \hat{\gamma}_{\tilde{x}}(j) = \frac{n}{T} \sum_{1 \leq t \leq j \leq T} \hat{u}_{t} \hat{u}_{t-j} \left(\frac{t-j}{T}\right) = \frac{1}{T} \sum_{1 \leq t \leq j \leq T} \tilde{\zeta}_{nt}^{\tau} \tilde{\zeta}_{nt-j}^{\tau} \left(\frac{t-j}{T}\right)$$

$$\sim_{p} 12 \mathbb{E}\left\{\tilde{\zeta}_{nt}^{0} \tilde{\zeta}_{nt-j}^{0}\right\} = 12 \left[4 \mathbb{E} \sigma_{2}^{2} \mathbb{E}_{t} \left(\epsilon_{it} \epsilon_{it-j}\right) + \mathbb{E} \left(\epsilon_{it}^{2} - \sigma_{t}^{2}\right) \left(\epsilon_{it-j}^{2} - \sigma_{t}^{2}\right)\right];$$

$$= 12 \left[4 \sigma_{2}^{2} \Gamma_{1}(j) + \Gamma_{1}^{2}(j)\right].$$

We may now proceed as in (50). Using the functional laws (46) and the arguments of Kiefer and Vogelsang (2002, 2005), Sun et al. (2008) and Sun (2014) we find that for the full sample ($b = 1$) estimator

$$\frac{n}{T^{2}} \hat{\Omega}_{M}^{b=1} = \frac{n}{T^{2}} \sum_{j=1-T}^{T-1} k\left(\frac{j}{T}\right) \hat{\gamma}_{\tilde{x}}(j) = \sum_{t=1}^{T} \sum_{p=1}^{T} k\left(\frac{t-p}{T}\right) \tilde{\zeta}_{nt}^{\tau} \tilde{\zeta}_{tp}^{\tau} \left(\frac{p}{T}\right)$$

$$\sim \left(\eta, T \rightarrow \infty\right) \int_{0}^{1} \int_{0}^{1} k\left(r-s\right) \tilde{r} \tilde{s} \mathbb{E} \left(\tilde{B}_{\Omega_{2}}^{\kappa}(r) \right) \mathbb{E} \left(\tilde{B}_{\Omega_{2}}^{\kappa}(s)\right),$$

where $\tilde{r} = r - \frac{1}{2}$ and $\tilde{s} = s - \frac{1}{2}$. In the same way, we have for the fixed-$b$ estimator with $b \in (0, 1)$
\[
\frac{n}{T^2} \hat{\Omega}_L^n := \frac{n}{T^2} \sum_{j=1-T}^{T-1} k_b \left( \frac{j}{T} \right) \hat{\zeta}_j (j) = \frac{n}{T^2} \sum_{j=1-T}^{T-1} k \left( \frac{j}{bT} \right) \frac{1}{T} \sum_{1 \leq t, t-j \leq T} \hat{z}_t \hat{z}_{t-j}
\]

\[
= \sum_{t=1}^{T} \sum_{p=1}^{T} k \left( \frac{t-p}{bT} \right) \frac{\zeta_{nt}^r}{\sqrt{T}} \frac{\zeta_{np}^r}{\sqrt{T}} \frac{\tilde{t}}{T} \left( \frac{\tilde{p}}{T} \right)
\]

\[\sim_{(n,T \to \infty)} \int_0^1 \int_0^1 k \left( \frac{r-s}{b} \right) \tilde{r} \tilde{s} dB^*_t \left( r \right) dB^*_s \left( s \right) = 12 \left\{ \int_0^1 \int_0^1 k \left( \frac{r-s}{b} \right) \tilde{r} \tilde{s} dW^r \left( r \right) dW^s \left( s \right) \right\}
\]

with \( k_b(r) = k(r/b) \). See equation (14) of Sun et al. (2008) for a comparable result in the simple time series location model. It follows that the \( t \) ratio test under sandwich HAR variance estimation has the limit theory

\[
t_{\text{HAR},b} \sim_{(n,T \to \infty)} \frac{Z_\phi}{\sqrt{12 \Omega_L^2 \int_0^1 \int_0^1 k \left( \frac{t-s}{b} \right) \tilde{r} \tilde{s} dW^r \left( r \right) dW^s \left( s \right)}}^{1/2}
\]

as stated.

**Proof of (v)** Finally, we consider the \( t \) ratio based on the sandwich form

\[
t_{\text{HAC}} = \frac{\hat{\phi}_{nT}}{\sqrt{\sum_{t=1}^{T} \hat{\zeta}_t^2})^{-1} \frac{T \hat{\Omega}_L^2 \left( \sum_{t=1}^{T} \hat{\zeta}_t^2 \right)^{-1}}{\sqrt{\left( T^{-3} \sum_{t=1}^{T} \hat{\zeta}_t^2 \right)^{-1} \frac{1}{nT^{3} \hat{\phi}_{nT}}}}
\]

with HAC estimate \( \hat{\Omega}_L^2 \) as given in (18). The limit behavior of \( \hat{\Omega}_L^2 \) as \( (T, n) \to \infty \) with lag truncation \( L = \lfloor T^{1/3} \rfloor \) is deduced as follows,

\[
\frac{n}{T^2} \hat{\Omega}_L^2 = \frac{n}{T} \sum_{t=1}^{T} \hat{\zeta}_t^2 + \frac{2n}{T} \sum_{t=1}^{L} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \hat{z}_t \hat{z}_{t+\ell}
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\hat{t}}{T} \right)^2 \mathbb{E} (\hat{\zeta}_t^2) + \frac{2}{T} \sum_{t=1}^{L} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \hat{z}_t \hat{z}_{t+\ell} \left( \frac{\hat{t}+\ell}{T} \right) \mathbb{E} (\zeta_{nt}^r \zeta_{nt+\ell}^r) + o_p(1)
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\hat{t}}{T} \right)^2 \left\{ \mathbb{E} (\zeta_t^2) + \frac{2}{T} \sum_{t=1}^{L} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \hat{z}_t \hat{z}_{t+\ell} \mathbb{E} (\zeta_{nt}^r \zeta_{nt+\ell}^r) \right\} + o_p(1)
\]

\[
= -\frac{1}{12} \sum_{j=-\infty}^{\infty} \mathbb{E} \zeta_t \zeta_{t+j} = \frac{1}{12} \sum_{j=-\infty}^{\infty} \{ 4 \sigma_n^2 \Gamma_{t} (j) + \Gamma_{t}^2 (j) \} = \frac{1}{12} \Omega_n^2 = \frac{1}{12} \Omega_{\phi}^2.
\]
Similar results on consistency of the HAC estimator apply with other kernels satisfying Assumption B. It follows that

\[
t_{\text{HAC}} = \frac{\sqrt{nT^3} \hat{\phi}_{nT}}{\left(T^{-3} \sum_{t=1}^{T} \tilde{i}^2\right)^{-1} nT^{-2} \hat{\Omega}^2_L \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{i}^2\right)^{-1} \sim_{(n,T \to \infty)} Z} \quad \text{with} \quad Z_{\phi} = \frac{Z}{\sqrt{12\Omega^2_{\phi}}},
\]

giving the stated result.

**Proof of Theorem 2**

**Proof of (25)**

We consider the exact orders of \(\hat{\Omega}^2_L\) and \(\hat{\Omega}^2_M\) first. Use \(\vartheta_{\ell L} = 1 - \ell / (1 + L)\) to denote the Bartlett lag kernel. The residuals from the trend regression are

\[
\hat{u}_t = \hat{K}_{nt} - \hat{\phi}_{nT} \tilde{f} = \left(\hat{\eta}_{n,t} - \hat{\phi}_{nT} \tilde{f}\right) + \hat{\varepsilon}_{nt} =: \hat{M}_{nt} + \hat{\varepsilon}_{nt}.
\]

We decompose \(\hat{\Omega}^2_L\), the long run variance estimate with lag truncation parameter \(L\) and \(\vartheta_{\ell L}\) as follows.

\[
\hat{\Omega}^2_L = \frac{1}{T} \sum_{t=1}^{T} \tilde{i}^2 \hat{u}_t^2 + \frac{2}{T} \sum_{\ell=1}^{L} \vartheta_{\ell L} \sum_{t=1}^{T-\ell} \hat{i}(t+\ell) \hat{u}_t \hat{u}_{t+\ell} = \frac{1}{T} \sum_{t=1}^{T} \tilde{i}^2 \hat{M}_{nt}^2 + 2 \frac{1}{T} \sum_{\ell=1}^{L} \vartheta_{\ell L} \sum_{t=1}^{T-\ell} \hat{i}(t+\ell) \hat{M}_{nt} \hat{M}_{nt+\ell}
\]

\[
+ \frac{1}{T} \sum_{t=1}^{T} \tilde{i}^2 \hat{\varepsilon}_{nt}^2 + 2 \frac{1}{T} \sum_{\ell=1}^{L} \vartheta_{\ell L} \sum_{t=1}^{T-\ell} \hat{i}(t+\ell) \hat{\varepsilon}_{nt} \hat{\varepsilon}_{nt+\ell}
\]

\[
+ 2 \frac{1}{T} \sum_{t=1}^{T} \tilde{i}^2 \hat{\varepsilon}_{nt} \hat{\varepsilon}_{nt} + 2 \frac{1}{T} \sum_{\ell=1}^{L} \vartheta_{\ell L} \sum_{t=1}^{T-\ell} \hat{i}(t+\ell) \hat{\varepsilon}_{nt} \hat{\varepsilon}_{nt+\ell}
\]

\[
=: \hat{\Omega}^2_{\ell} + \hat{\Omega}^2_{\varepsilon} + 2\hat{\Omega}_{\ell \varepsilon},
\]

where

\[
\hat{\Omega}^2_{\ell} = \frac{1}{T} \sum_{t=1}^{T} \tilde{i}^2 \hat{M}_{nt}^2 + 2 \frac{1}{T} \sum_{\ell=1}^{L} \vartheta_{\ell L} \tilde{i}(t+\ell) \hat{\varepsilon}_{nt} \hat{\varepsilon}_{nt+\ell}.
\]
It has been shown in KPS (2019) that the dominating term in $\hat{\Omega}_M^2$ is $\hat{\Omega}_I^2$. Note that

\[
\hat{\Omega}_I^2 = \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 \tilde{m}_t^2 + \frac{2}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \partial_{tL} \tilde{t}(t) \tilde{M}_{nt} \tilde{M}_{nt+\ell}
\]

\[
= \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 (\tilde{m}_t + R_{nt})^2 + \frac{2}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \partial_{tL} \tilde{t}(t) (\tilde{m}_t + R_{nt}) (\tilde{m}_{nt+\ell} + R_{nt+\ell})
\]

\[
= \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 \tilde{m}_t^2 + \frac{2}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \partial_{tL} \tilde{t}(t) \tilde{m}_t \tilde{m}_{t+\ell}
\]

\[
+ \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 R_{nt}^2 + \frac{2}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \partial_{tL} \tilde{t}(t) R_{nt} R_{nt+\ell}
\]

\[
+ 2 \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 \tilde{m}_t R_{nt} + \frac{2}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \partial_{tL} \tilde{t}(t) (\tilde{m}_t R_{nt+\ell} + R_{nt} \tilde{m}_{t+\ell}).
\]

Let

\[
\Omega_i^2 = \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 \tilde{m}_t^2 + \frac{2}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \partial_{tL} \tilde{t}(t) \tilde{m}_t \tilde{m}_{t+\ell}.
\]

We have shown that $R_{nt} = o_p(\tilde{m}_t)$ uniformly in $t \leq T$, from which it follows that the dominating term in $\hat{\Omega}_I^2$ is $\Omega_i^2$, which is represented by $\hat{\Omega}_I^2 \sim \Omega_i^2$. The decomposition of $\hat{\Omega}_I^2$ is similar to that of $\hat{\Omega}_M^2$. We denote the dominating term in $\hat{\Omega}_M^2$ by $\Omega_m^2$.

Let $\tilde{p}_t = \tilde{m}_t \tilde{t}$. Note that as $T \to \infty$, we have

\[
\frac{1}{T} \sum_{t=1}^T \tilde{p}_t^2 = \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 \tilde{m}_t^2
\]

\[
= \frac{1}{T} \sum_{t=1}^T \tilde{t}^2 \left[ \tilde{t}^{-\lambda} - \tilde{t} \left( \sum_{t=1}^T \tilde{t}^{-\lambda} \right) \left( \sum_{t=1}^T \tilde{t}^2 \right)^{-1} \right]^2
\]

\[
= \begin{cases} 
O(T^{2-2\lambda}) & \text{if } \lambda < 1 \\
O(1) & \text{if } \lambda = 1 \\
O(T^{2-2\lambda}) & \text{if } 1 < \lambda < 3/2 \\
O(T^{-1}\ln T) & \text{if } \lambda = 3/2 \\
O(T^{-1}) & \text{if } \lambda > 3/2
\end{cases}
\]

Next, let

\[
P_L(T,\lambda) = \frac{1}{T} \sum_{t=1}^L \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{L+1} \right) \tilde{p}_t \tilde{p}_{t+\ell},
\]

and

\[
P_M(T,\lambda) = \frac{1}{T} \sum_{t=1}^M \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) \tilde{p}_t \tilde{p}_{t+\ell}.
\]
We expand $P_M$ as

$$
\frac{1}{T} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) \tilde{p}_t \tilde{p}_{t+\ell} 
= \frac{1}{T} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) t^{1-\lambda}(t+\ell)^{1-\lambda} 
- T_T(1, \lambda) T^{-4} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) t^{1-\lambda}(t+\ell)^2 
- T_T(1, \lambda) T^{-4} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) t^2(t+\ell)^{1-\lambda} 
+ (T_T(1, \lambda))^2 T^{-7} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) (t+\ell)^2 \tilde{t}^2 
:= \Psi_M^M - \Psi_M^M - \Psi_M^M + \Psi_M^M,
$$

where $T_T(1, \alpha) = \sum_{t=1}^{T} \tilde{t}^{\tilde{t}-\alpha}$, which is defined in Lemma 4 in KPS (2019). Further note that

$$
\Psi_M^M = \frac{1}{T} \sum_{\ell=1}^{M} \left( 1 - \frac{\ell}{M+1} \right) \sum_{t=1}^{T-\ell} \tilde{t}^{1-\lambda}(t+\ell)^{1-\lambda} 
= \frac{1}{T} \sum_{\ell=1}^{M} \left( 1 - \frac{\ell}{M+1} \right) \sum_{t=1}^{T-\ell} (t^2 + t\ell)^{1-\lambda} 
- \frac{1}{T} \sum_{\ell=1}^{M} \left( 1 - \frac{\ell}{M+1} \right) \left( \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} t^{1-\lambda} \right) \sum_{t=1}^{T-\ell} (t+\ell)^{1-\lambda} 
:= \Psi_M^{M} - \Psi_M^{M},
$$

$$
\Psi_M^M = T_T(1, \lambda) T^{-4} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) t^{1-\lambda}(t+\ell)^2 
= T_T(1, \lambda) T^{-4} \sum_{\ell=1}^{M} \left( 1 - \frac{\ell}{M+1} \right) \sum_{t=1}^{T-\ell} t^{1-\lambda}(t+\ell)^2 
- T_T(1, \lambda) T^{-4} \sum_{\ell=1}^{M} \left( 1 - \frac{\ell}{M+1} \right) \frac{1}{T-\ell} \left( \sum_{t=1}^{T-\ell} t^{1-\lambda} \right) \sum_{t=1}^{T-\ell} (t+\ell)^2 
:= \Psi_M^{M} - \Psi_M^{M},
$$
\[
\Psi_3^M = T_T (1, \lambda) T^{-4} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) \tilde{t}^2 (t + \ell)^{1-\lambda}
\]
\[
= T_T (1, \lambda) T^{-4} \sum_{\ell=1}^{M} \sum_{t=1}^{T-\ell} \left( 1 - \frac{\ell}{M+1} \right) t^2 (t + \ell)^{1-\lambda}
\]
\[
- T_T (1, \lambda) T^{-4} \sum_{\ell=1}^{M} \left( 1 - \frac{\ell}{M+1} \right) \frac{1}{T - \ell} \left( \sum_{t=1}^{T-\ell} t^2 \right) \sum_{t=1}^{T-\ell} (t + \ell)^{1-\lambda}
\]
\[
\therefore = \Psi_3^M - \Psi_3^M.
\]

Direct calculation gives the order of each term. Rather than record all the derivations we show here how to get the exact order of \(\Psi_{21}^L\) and \(\Psi_{21}^M\). Note that
\[
\Psi_{21}^L = T_T (1, \lambda) T^{-4} \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \sum_{t=1}^{T-\ell} t^{1-\lambda} (t + \ell)^2
\]
\[
= T_T (1, \lambda) T^{-4} \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \sum_{t=1}^{T-\ell} \left( t^{3-\lambda} + t^{1-\lambda} \ell^2 + 2t^{2-\lambda} \ell \right).
\]

Consider each term.
\[
\sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \sum_{t=1}^{T-\ell} t^{3-\lambda} = \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \left\{ \begin{array}{ll}
\frac{1}{\ell} (T - \ell)^{4-\lambda} & \text{if } \lambda < 4 \\
\ln (T - \ell) & \text{if } \lambda = 4 \\
\zeta (\lambda - 3) & \text{if } \lambda > 4
\end{array} \right.
\]
\[
= \begin{cases}
O \left( T^{4-\lambda} L \right) & \text{if } \lambda < 4 \\
O \left( L \ln T \right) & \text{if } \lambda = 4 \\
O \left( L \right) & \text{if } \lambda > 4
\end{cases},
\]
\[
\sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \sum_{t=1}^{T-\ell} t^{1-\lambda} \ell^2 = \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \ell^2 \sum_{t=1}^{T-\ell} t^{1-\lambda}
\]
\[
= \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \ell^2 \left\{ \begin{array}{ll}
\frac{1}{\ell-\lambda} (T - \ell)^{2-\lambda} & \text{if } \lambda < 2 \\
\ln (T - \ell) & \text{if } \lambda = 2 \\
\zeta (\lambda - 1) & \text{if } \lambda > 2
\end{array} \right.
\]
\[
= \begin{cases}
O \left( T^{4-\lambda} L \right) & \text{if } \lambda < 2 \\
O \left( T^2 L \ln T \right) & \text{if } \lambda = 2 \\
O \left( T^2 L \right) & \text{if } \lambda > 2
\end{cases}.
\]
and
\[ \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \sum_{t=1}^{T-\ell} 2 t^{2-\lambda} \ell = 2 \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \ell \sum_{t=1}^{T-\ell} t^{2-\lambda} \]
\[ = 2 \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \ell \begin{cases} \frac{1}{3-\lambda} (T-\ell)^{3-\lambda} & \text{if } \lambda < 3 \\ \ln (T-\ell) & \text{if } \lambda = 3 \\ \zeta (\lambda - 2) & \text{if } \lambda > 3 \end{cases} \]
\[ = \begin{cases} O (T^{4-\lambda} L) & \text{if } \lambda < 3 \\ O (T L \ln T) & \text{if } \lambda = 3 \\ O (T L) & \text{if } \lambda > 3 \end{cases} . \]

Combining all terms yields
\[ \Psi_{21}^{L} = T_{T} (1, \lambda) T^{-4} \sum_{\ell=1}^{L} \left( 1 - \frac{\ell}{L+1} \right) \sum_{t=1}^{T-\ell} t^{1-\lambda} (t + \ell)^{2} \]
\[ = T_{T} (1, \lambda) T^{-4} \begin{cases} O (T^{4-\lambda} L) & \text{if } \lambda < 4 \\ O (L \ln T) & \text{if } \lambda = 4 + T_{T} (1, \lambda) T^{-4} \begin{cases} O (T^{4-\lambda} L) & \text{if } \lambda < 3 \\ O (T L \ln T) & \text{if } \lambda = 3 \\ O (T L) & \text{if } \lambda > 3 \end{cases} \\ O (T L) & \text{if } \lambda > 4 \end{cases} \]
\[ = T_{T} (1, \lambda) T^{-4} \begin{cases} O (T^{4-\lambda} L) & \text{if } \lambda < 2 \\ O (T^{2} L \ln T) & \text{if } \lambda = 2 \\ O (T^{2} L) & \text{if } \lambda > 2 \end{cases} . \]

Next, replacing \( L \) by \( T^{\kappa} \) leads to
\[ \Psi_{21}^{L} = T_{T} (1, \lambda) \begin{cases} O (T^{\lambda+\kappa}) & \text{if } \lambda < 2 \\ O (T^{2+\kappa} \ln T) & \text{if } \lambda = 2 \\ O (T^{2+\kappa}) & \text{if } \lambda > 2 \end{cases} \]
\[ = \begin{cases} T^{2-\lambda} & \text{if } \lambda < 1 \\ T \ln T & \text{if } \lambda = 1 \times \begin{cases} O (T^{\lambda+\kappa}) & \text{if } \lambda < 2 \\ O (T^{2+\kappa} \ln T) & \text{if } \lambda = 2 \\ O (T^{2+\kappa}) & \text{if } \lambda > 2 \end{cases} \\ \zeta (\lambda) T & \text{if } \lambda > 1 \end{cases} . \]

48
To calculate the order of \( \Psi_{21}^M \), we replace \( M \) by \( bT \). That is,

\[
\Psi_{21}^M = T_T (1, \lambda) T^{-4} \begin{cases}
O (T^{4-\lambda} M) & \text{if } \lambda < 2 \\
O (T^2 M \ln T) & \text{if } \lambda = 2 \\
O (T^2 M) & \text{if } \lambda > 2
\end{cases}
\]

\[
= T_T (1, \lambda) \begin{cases}
O (T^{5-\lambda}) & \text{if } \lambda < 2 \\
O (T^3 \ln T) & \text{if } \lambda = 2 \\
O (T^3) & \text{if } \lambda > 2
\end{cases}
\]

\[
= \begin{cases}
T^{2-\lambda} & \text{if } \lambda < 1 \\
T \ln T & \text{if } \lambda = 1 \\
\zeta (\lambda) T & \text{if } \lambda > 1
\end{cases}
\]

\[
\times \begin{cases}
O (T^{5-\lambda}) & \text{if } \lambda < 2 \\
O (T^3 \ln T) & \text{if } \lambda = 2 \\
O (T^3) & \text{if } \lambda > 2
\end{cases}
\]

\[
= \begin{cases}
O (T^{3-2\lambda}) & \text{if } \lambda < 1 \\
O (T^{2-\lambda} \ln T) & \text{if } \lambda = 1 \\
O (T^{2-\lambda}) & \text{if } 1 < \lambda < 2 \\
O (\ln T) & \text{if } \lambda = 2 \\
O (1) & \text{if } \lambda > 2
\end{cases}
\]

In the expressions below we provide the final order of each term.

\[
\Psi_{11}^M = \begin{cases}
O (T^{3-2\lambda}) & \text{if } \lambda < 3/2 \\
O (\ln T) & \text{if } \lambda = 3/2 \\
O (1) & \text{if } \lambda > 3/2
\end{cases}
\]

\[
\Psi_{12}^L = \begin{cases}
O (T^{3-2\lambda}) & \text{if } \lambda < 2 \\
O (T^{-1} \ln^2 T) & \text{if } \lambda = 2 \\
O (T^{-1}) & \text{if } \lambda > 2
\end{cases}
\]

\[
\Psi_{12}^L = \begin{cases}
O (T^{2-2\lambda+\lambda}) & \text{if } \lambda < 3/2 \\
O (T^{1+\lambda} \ln T) & \text{if } \lambda = 3/2 \\
O (T^{1+\lambda}) & \text{if } \lambda > 3/2
\end{cases}
\]

Hence

\[
\Psi_{1}^M = \begin{cases}
O (T^{3-2\lambda}) & \text{if } \lambda < 3/2 \\
O (\ln T) & \text{if } \lambda = 3/2 \\
O (1) & \text{if } \lambda > 3/2
\end{cases}
\]

\[
\Psi_{1}^L = \begin{cases}
O (T^{2-2\lambda+\lambda}) & \text{if } \lambda < 3/2 \\
O (T^{1+\lambda} \ln T) & \text{if } \lambda = 3/2 \\
O (T^{1+\lambda}) & \text{if } \lambda > 3/2
\end{cases}
\]
Next

\[ \Psi_{21}^M = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1 \\
O(T^{2-\lambda} \ln T) & \text{if } \lambda = 1 \\
O(T^{2-\lambda}) & \text{if } 1 < \lambda < 2 \\
O(\ln T) & \text{if } \lambda = 2 \\
O(1) & \text{if } \lambda > 2 
\end{cases} \quad \Psi_{21}^L = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^{1-\lambda+\kappa} \ln T) & \text{if } \lambda = 1 \\
O(T^{1-\lambda+\kappa}) & \text{if } 1 < \lambda < 2 \\
O(T^{-1+\kappa} \ln T) & \text{if } \lambda = 2 \\
O(T^{-1+\kappa}) & \text{if } \lambda > 2 
\end{cases} \]

\[ \Psi_{22}^M = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1 \\
O(T^{2-\lambda} \ln T) & \text{if } \lambda = 1 \\
O(T^{2-\lambda}) & \text{if } 1 < \lambda < 2 \\
O(\ln T) & \text{if } \lambda = 2 \\
O(1) & \text{if } \lambda > 2 
\end{cases} \quad \Psi_{22}^L = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^{1-\lambda+\kappa} \ln T) & \text{if } \lambda = 1 \\
O(T^{1-\lambda+\kappa}) & \text{if } 1 < \lambda < 2 \\
O(T^{-1+\kappa} \ln T) & \text{if } \lambda = 2 \\
O(T^{-1+\kappa}) & \text{if } \lambda > 2 
\end{cases} \]

Combining these two leads to

\[ \Psi_2^L = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1 \\
O(T^{2-\lambda} \ln T) & \text{if } \lambda = 1 \\
O(T^{2-\lambda}) & \text{if } 1 < \lambda < 2 \\
O(\ln T) & \text{if } \lambda = 2 \\
O(1) & \text{if } \lambda > 2 
\end{cases} \quad \Psi_2^M = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^{1-\lambda+\kappa} \ln T) & \text{if } \lambda = 1 \\
O(T^{1-\lambda+\kappa}) & \text{if } 1 < \lambda < 2 \\
O(T^{-1+\kappa} \ln T) & \text{if } \lambda = 2 \\
O(T^{-1+\kappa}) & \text{if } \lambda > 2 
\end{cases} \]

The third term becomes

\[ \Psi_{31}^M = \begin{cases} 
O(T^{2-2\lambda}) & \text{if } \lambda < 1 \\
O(T^{1-\lambda} \ln T) & \text{if } \lambda = 1 \\
O(T^{1-\lambda}) & \text{if } 1 < \lambda < 2 \\
O(\ln T) & \text{if } \lambda = 2 \\
O(1) & \text{if } \lambda > 2 
\end{cases} \quad \Psi_{31}^L = \begin{cases} 
O(T^{(5-\lambda)\kappa-2-\lambda}) & \text{if } \lambda < 1, \\
O(T^{(5-\lambda)\kappa-3} \ln T) & \text{if } \lambda = 1, \\
O(T^{(5-\lambda)\kappa-3}) & \text{if } 1 < \lambda < 2, \\
O(T^{3\kappa-3} \ln T) & \text{if } \lambda = 2, \\
O(T^{3\kappa-3}) & \text{if } \lambda > 2, 
\end{cases} \]

\[ \Psi_{32}^M = \begin{cases} 
O(T^{2-2\lambda}) & \text{if } \lambda < 1 \\
O(T^{2-\lambda} \ln T) & \text{if } \lambda = 1 \\
O(T^{2-\lambda}) & \text{if } 1 < \lambda < 2 \\
O(\ln T) & \text{if } \lambda = 2 \\
O(1) & \text{if } \lambda > 2 
\end{cases} \quad \Psi_{32}^L = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^{1-\lambda+\kappa} \ln T) & \text{if } \lambda = 1 \\
O(T^{1-\lambda+\kappa}) & \text{if } 1 < \lambda < 2 \\
O(T^{-1+\kappa} \ln T) & \text{if } \lambda = 2 \\
O(T^{-1+\kappa}) & \text{if } \lambda > 2 
\end{cases} \]
Combining these two terms yields

\[
\Psi_3^M = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1, \\
O(T^{2-\lambda \ln T}) & \text{if } \lambda = 1, \\
O(T^{2-\lambda}) & \text{if } 1 < \lambda < 2, \\
O(\ln T) & \text{if } \lambda = 2, \\
O(1) & \text{if } \lambda > 2,
\end{cases} \quad \Psi_3^L = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1, \\
O(T^{1-\lambda+\kappa} \ln T) & \text{if } \lambda = 1, \\
O(T^{1-\lambda+\kappa}) & \text{if } 1 < \lambda < 2, \\
O(T^{1+\kappa} \ln T) & \text{if } \lambda = 2, \\
O(T^{1+\kappa}) & \text{if } \lambda > 2,
\end{cases}
\]

Last, the fourth term becomes

\[
\Psi_4^M = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1, \\
O(T \ln^2 T) & \text{if } \lambda = 1, \\
O(T) & \text{if } \lambda > 1
\end{cases} \quad \Psi_4^L = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^\kappa \ln^2 T) & \text{if } \lambda = 1 \\
O(T^\kappa) & \text{if } \lambda > 1
\end{cases}
\]

After combining all terms, we have

\[
P_M(T, \lambda) = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1 \\
O(T \ln^2 T) & \text{if } \lambda = 1, \\
O(T) & \text{if } \lambda > 1
\end{cases} \quad P_L(T, \lambda) = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^\kappa \ln^2 T) & \text{if } \lambda = 1 \\
O(T^\kappa) & \text{if } \lambda > 1
\end{cases}
\]

Finally,

\[
\hat{\Omega}_M^2 \sim \frac{1}{T} \sum_{t=1}^{T} \tilde{p}_t^2 + 2P_M(T, \lambda) = \begin{cases} 
O(T^{3-2\lambda}) & \text{if } \lambda < 1 \\
O(T \ln^2 T) & \text{if } \lambda = 1 \\
O(T) & \text{if } \lambda > 1
\end{cases}
\]

\[
\hat{\Omega}_L^2 \sim \frac{1}{T} \sum_{t=1}^{T} \tilde{p}_t^2 + 2P_L(T, \lambda) = \begin{cases} 
O(T^{2-2\lambda+\kappa}) & \text{if } \lambda < 1 \\
O(T^\kappa \ln^2 T) & \text{if } \lambda = 1 \\
O(T^\kappa) & \text{if } \lambda > 1
\end{cases}
\]

Therefore we have

\[
t_{\text{HAR}} \sim a \begin{cases} 
O_p \left( \frac{T^{-1-\lambda} \ln T}{T^{1/2}} \right) = O_p(1) & \text{if } \lambda < 1 \\
O_p \left( \frac{T^{-2} \ln^2 T}{T^{1/2}} \right) = O_p(1) & \text{if } \lambda = 1, \\
O_p \left( \frac{T^{-2} \ln T}{T^{1/2}} \right) = O_p(1) & \text{if } \lambda > 1
\end{cases}
\]

\[
t_{\text{HAC}} \sim a \begin{cases} 
O_p \left( \frac{T^{1-\lambda} \ln T}{T^{1/2}} \right) = O_p(T^{1-\kappa}/2) & \text{if } \lambda < 1 \\
O_p \left( \frac{T^{-2} \ln^2 T}{T^{1/2} \ln^2 T} T^{1/2} \right) = O_p(T^{1-\kappa}/2) & \text{if } \lambda = 1 \\
O_p \left( \frac{T^{-2} \ln^2 T}{T^{1/2} \ln^2 T} \right) = O_p(T^{1-\kappa}/2) & \text{if } \lambda > 1
\end{cases}
\]

51
using representations in terms of the dominant deterministic order.

**Proof of (24)**

$P_L(T, \lambda)$ has been calculated in KPS(2019) as

$$
P_L(T, \lambda) = \begin{cases} 
O(T^{-2\lambda + \kappa}) & \text{if } \lambda < 1/2 \\
O(T^{\kappa - 1} \ln T) & \text{if } \lambda = 1/2 \\
O(T^{\kappa - 1}) & \text{if } 1/2 < \lambda < 1/(1 + \kappa) \\
O(T^{-\lambda + \kappa - \lambda \kappa}) & \text{if } 1/(1 + \kappa) \leq \lambda < 1 \\
O(T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O(T^{-1}) & \text{if } \lambda > 1
\end{cases}
$$

As shown in the previous proof, the order of the t-ratio based on the HAR estimator can be directly obtained by replacing $\kappa$ by 1, except when $\lambda > 1/2$. Below we show the main difference only. When $L = [T^\kappa]$, and $0 < \kappa < 1$, we have

$$
\frac{1}{T} \sum_{t=1}^L \left( 1 - \frac{\ell}{L + 1} \right) \sum_{t=1}^{T-\ell} (t^2 + t\ell)^{-\lambda} < \begin{cases} 
\min [O(T^{-2\lambda + \kappa}), O(T^{-\lambda + \kappa - \lambda \kappa})] = O(T^{-2\lambda + \kappa}) & \text{if } \lambda < 1/2 \\
\min [O(T^{\kappa - 1} \ln T), O(T^{\kappa - 1/2 - \kappa/2})] = O(T^{\kappa - 1} \ln T) & \text{if } \lambda = 1/2 \\
\min [O(T^{\kappa - 1}), O(T^{-\lambda + \kappa - \lambda \kappa})] = O(T^{\kappa - 1}) & \text{if } 1/2 < \lambda < 1/(1 + \kappa) \\
\min [O(T^{\kappa - 1}), O(T^{-\lambda + \kappa - \lambda \kappa})] = O(T^{-\lambda + \kappa - \lambda \kappa}) & \text{if } 1/(1 + \kappa) \leq \lambda < 1
\end{cases}
$$

Details can be found in KPS(2019). But when $M = [bT]$, which corresponds to $\kappa = 1$,

$$
\frac{1}{T} \sum_{t=1}^M \left( 1 - \frac{\ell}{M + 1} \right) \sum_{t=1}^{T-\ell} (t^2 + t\ell)^{-\lambda} < \begin{cases} 
\min [O(T^{-2\lambda + 1}), O(T^{-2\lambda + 1})] = O(T^{-2\lambda + 1}) & \text{if } \lambda < 1/2 \\
\min [O(\ln T), O(T^{-2\lambda + 1})] = O(1) & \text{if } \lambda = 1/2 \\
\min [O(1), O(T^{-2\lambda + 1})] = O(T^{-2\lambda + 1}) & \text{if } \lambda > 1/2
\end{cases}
$$

This difference resulted in the difference in the order of $\Psi^M_{11}$ and $\Psi^L_{11}$. 

52
Below we only show the details of the decomposition for \( P_M (T, \lambda) \).

\[
\begin{align*}
\Psi_{11}^M &= \begin{cases} 
O (T^{-2\lambda+1}) & \text{if } \lambda < 1 \\
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases} \\
\Psi_{12}^M &= \begin{cases} 
O (T^{1-2\lambda}) & \text{if } \lambda < 1 \\
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases} \\
\Psi_{21}^M &= \begin{cases} 
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases} \\
\Psi_{22}^M &= \begin{cases} 
O (T^{1-2\lambda}) & \text{if } \lambda < 1 \\
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases} \\
\Psi_{31}^M &= \begin{cases} 
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases} \\
\Psi_{32}^M &= \begin{cases} 
O (T^{1-2\lambda}) & \text{if } \lambda < 1 \\
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases} \\
\Psi_{4}^M &= \begin{cases} 
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases}
\end{align*}
\]

Hence,

\[
P_M (T, \lambda) = \Psi_1 - \Psi_2 - \Psi_3 + \Psi_4
\]

\[
\sim a \begin{cases} 
O (T^{-2\lambda+1}) & \text{if } \lambda < 1 \\
O (T^{-1} \ln^2 T) & \text{if } \lambda = 1 \\
O (T^{-1}) & \text{if } \lambda > 1
\end{cases}
\]

Then

\[
t_2 \sim a \frac{\hat{\phi}_{nT} \left( \sum T^2 \right)^{1/2}}{\sqrt{\Omega_m^2}}
\]

\[
= \begin{cases} 
O_p \left( \frac{T^{-1-\lambda} \ln T^{3/2}}{T^{1/2-\lambda}} \right) = O_p (1) & \text{if } \lambda < 1 \\
O_p \left( \frac{T^{-2} \ln T^{3/2}}{T^{1/2} \ln T T^{3/2}} \right) = O_p (1) & \text{if } \lambda = 1 \\
O_p \left( \frac{T^{-2} \ln T^{3/2}}{T^{1/2} \ln T T^{3/2}} \right) = O_p (1) & \text{if } \lambda > 1
\end{cases}
\]

where again the representations are given in terms of the dominating deterministic order.