Testing Convergence using HAR Inference*

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

April 30, 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 σ-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 σ-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

Convergence studies flourished in the cross-country growth literature during 1980s and 1990s. During that time much attention among empirical researchers was given to econometric tests of convergence. But the original discussions of convergence began earlier in the 1930s with a debate between Secrist (1933) and Hotelling (1933). That debate was revived in the work of Barro and Sala-i-Martin (1992) and Friedman (1992). One group argued that the convergence between rich and poor countries required much faster growth among initially poor countries than among those countries that were initially rich (Secrist, 1933; Barro and Sala-i-Martin, 1992). This type of the convergence was termed '$\beta-$convergence' as it focused on the relevant regression coefficient estimates. The other group was critical of β-convergence comparisons as a statistical artifice and instead suggested that the most appropriate concept for convergence should focus on consistent diminution of cross sectional variance (Hotelling, 1933; Friedman, 1992). Phillips and Sul (2007) and Kong, Phillips and Sul (2019) showed statistical problems with $\beta-$convergence tests and proposed new approaches to assessing convergence based on notions of relative and weak $\sigma-$convergence which analyze cross section variation directly for diminution. Sul (2019) provides a detailed discussion of these various convergence tests.

Amongst the many issues for which panel data enable empirical investigation, these questions of convergence and divergence over time have attracted high interest. In the study of cross country economic performance, research has focussed particularly on examining evidence of diminishing dispersion in key indicator variables such as income or consumption levels, poverty, and unemployment rates. These indicators figure 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 σ-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 asympotically 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 asymptotic 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 nonstationary 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, other technical derivations, and assumptions employed are given in the Appendix. The paper uses the same notation as KPS (2019) to assist in cross-referencing the derivations and results.

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 im-

[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.

portant 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 $M \to \infty$. 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 $M = bT \to \infty$ 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

$$x_t = at + z_t, \tag{1}$$

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.

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

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

$$t_a^{\mathrm{HAC}} = \frac{\hat{a}}{\left\{\left(\sum_{t=1}^{T}t^2\right)^{-1}\left[T\hat{\Omega}_{\mathrm{HAC}}\right]\left(\sum_{t=1}^{T}t^2\right)^{-1}\right\}^{1/2}} = O_p\left(\sqrt{\frac{T}{L}}\right), \tag{3}$$

where $\hat{\Omega}_{\mathrm{HAC}} = \frac{1}{T}\sum_{t=1}^{T}\varpi_t^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\varpi_t\varpi_{t+\ell}$, with $\varpi_t = \hat{z}_t t$ 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

$$t_a^{\mathrm{HAR}} = \frac{\hat{a}}{\left\{\left(\sum_{t=1}^{T} t^2\right)^{-1}\left[T\hat{\Omega}_{\mathrm{HAR}}\right]\left(\sum_{t=1}^{T} t^2\right)^{-1}\right\}^{1/2}} = O_p\left(1\right), \tag{4}$$

where $\hat{\Omega}_{\mathrm{HAR}} = \frac{1}{T}\sum_{t=1}^{T}\varpi_t^2 + \frac{2}{T}\sum_{\ell=1}^{M}\sum_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\varpi_t\varpi_{t+\ell}$, has a nuisance parameter free limit distribution when $M = \lfloor bT \rfloor$ 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 $t_a^{\mathrm{HAR}} = O_p\left(1\right)$.

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 σ-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

$$y_{it} = \theta_i' F_t + x_{it}, \quad x_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it} t^{-\beta}, \tag{5}$$

where F_t is a vector of common factors, θ_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 $\sigma_\epsilon^2 t^{-2\beta}$ where σ_ϵ^2 is the variance of ϵ_{it} and $\beta > 0$. For our following analysis we assume that ϵ_{it} is strictly stationary over t with 1-summable autocovariance sequence $(\gamma_i(h))$ and long run variance $\Omega_e^2 = \sum_{h=-\infty}^{\infty} \gamma_i(h) > 0$, but independently distributed over i with uniform fourth moments, and the slope coefficients a_i and μ_i are cross sectionally independent with finite second moments. Full details are given in Assumptions A and B in the Appendix.

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 $\hat{x}_{it} = y_{it} - \hat{\theta}_i' \hat{F}_t = x_{it} + O_p\left(C_{nT}^{-1}\right)$ 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 ϵ_{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

$$K_{nt} = \frac{1}{n}\sum_{i=1}^n \left(x_{it} - \frac{1}{n}\sum_{i=1}^n x_{it}\right)^2 = \sigma_{a,n}^2 + \eta_{n,t} + \varepsilon_{n,t}, \tag{6}$$

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

$$\varepsilon_{n,t} = 2n^{-1}\sum_{i=1}^n \tilde{a}_i \tilde{\epsilon}_{it} t^{-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) t^{-2\beta}, \tag{7}$$

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

$$K_{nt} = \sigma_{a,n}^2 + \eta_t + \varepsilon_{n,t} + \xi_{n,t}, \tag{8}$$

where $\eta_t = \sigma_\epsilon^2 t^{-\lambda}$, $\lambda = 2\beta$, $\xi_{n,t} = \eta_{n,t} - \eta_t = \left(\sigma_{\epsilon,nT}^2 - \sigma_\epsilon^2\right) t^{-\lambda}$ and σ_ϵ^2 is the variance of ϵ_{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 σ-convergence KPS (2019) propose the following simple linear trend regression fitted with T time series observations

$$K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT} t + \hat{u}_t, \tag{9}$$

[2] See footnote 9 in KPS for more discussion of this issue involving the prior removal of a factor component from the data.

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

$$\hat{\phi}_{nT} = \sum\nolimits_{t=1}^{T} a_{tT}\tilde{\eta}_t + \sum\nolimits_{t=1}^{T} a_{tT}\tilde{\xi}_{n,t} + \sum\nolimits_{t=1}^{T} a_{tT}\tilde{\varepsilon}_{n,t} =: I_A + I_B + I_C, \tag{10}$$

where $a_{tT} = \tilde{t}/\left(\sum_{s=1}^{T} \tilde{s}^2\right)$, $\tilde{\eta}_t = \sigma_\epsilon^2 \widetilde{t^{-\lambda}}$, $\widetilde{t^{-\lambda}} = t^{-\lambda} - T^{-1}\sum_{t=1}^{T} t^{-\lambda}$, $\tilde{\xi}_{n,t} = \xi_{n,t} - T^{-1}\sum_{t=1}^{T}\xi_{n,t}$ 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 $\eta_t = \sigma_\epsilon^2 t^{-\lambda}$ 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 σ-convergence of K_{nt} is formally defined in equation (2) of KPS (2019) and essentially requires that $\bar{K}_t = \text{plim}_{n\to\infty} K_{nt}$ exists and decays over time to some constant value $c \in [0, \infty)$. In what follows we refer to this concept as σ-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 σ-convergence.

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

$$t_1 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_1^2 \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1}}}, \tag{11}$$

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

$$\hat{\Omega}_1^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left(1 - \frac{\ell}{L+1}\right) \hat{u}_t \hat{u}_{t+\ell}, \tag{12}$$

where $\hat{u}_t = \widetilde{K_{nt}} - \hat{\phi}_{nT}\tilde{t}$ with $\widetilde{K_{nt}} = K_{nt} - T^{-1} \sum_{t=1}^{T} K_{nt}$, $L = \lfloor T^{\kappa} \rfloor$ 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[3]

$$\mathcal{H}_0 : \lambda \leq 0. \tag{13}$$

The directed alternative hypothesis $\mathcal{H}_A : \lambda > 0$ implies that testing for convergence is one-sided. Critical values are then delivered by the limit distribution under the point null $\lambda = 0$.

[3] The null hypothesis given in (13) is for the case $\mu_i = \alpha = 0$ 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.

Even though the null and alternative hypotheses are well defined in terms of the unobserved parameter λ 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 in this extreme situation. Theorem 2 in KPS (2019) provides further details and discussion. Obviously the test becomes inconsistent in this extreme situation where $\lambda \to \infty$ when a smaller nominal size is employed (e.g. for 4% nominal size the critical value is 1.75 and $-\sqrt{3} = -1.732 > -1.75$). The use of a 5% test is therefore relevant in determining asymptotic properties of the test in this extreme case where $\lambda \to \infty$.

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).

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

$$t_2 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_2^2 \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1}}}, \tag{14}$$

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 $M = \lfloor bT \rfloor$ with $b \in (0,1)$. The next two statistics use sandwich formulae in the self normalization. Let $\tilde{\varkappa}_t = \hat{u}_t \tilde{t}$, and define

$$t_{\text{HAR}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} T\hat{\Omega}_M^2 \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1}}}, \tag{15}$$

$$t_{\text{HAC}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} T\hat{\Omega}_L^2 \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1}}}, \tag{16}$$

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

$$\hat{\Omega}_M^2 = \frac{1}{T}\sum_{t=1}^{T} \tilde{\varkappa}_t^2 + \frac{2}{T}\sum_{\ell=1}^{M}\sum_{t=1}^{T-\ell}\left(1 - \frac{\ell}{M+1}\right)\tilde{\varkappa}_t \tilde{\varkappa}_{t+\ell}, \tag{17}$$

$$\hat{\Omega}_L^2 = \frac{1}{T}\sum_{t=1}^{T} \tilde{\varkappa}_t^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\tilde{\varkappa}_t \tilde{\varkappa}_{t+\ell}, \tag{18}$$

$$\hat{\Omega}_2^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 + \frac{2}{T}\sum_{\ell=1}^{M}\sum_{t=1}^{T-\ell}\left(1 - \frac{\ell}{M+1}\right)\hat{u}_t \hat{u}_{t+\ell}. \tag{19}$$

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 $M = \lfloor bT \rfloor$) in place of a HAC estimator. The two statistics t_{HAC} and t_{HAR} use sandwich formulae for the construction of the variance, with the HAR estimate $\hat{\Omega}_M^2$ in t_{HAR} and the HAC estimate $\hat{\Omega}_L^2$ in t_{HAC}. Under the null hypothesis, the asymptotic behavior of the t_{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_{HAR} and t_2 statistics differ from that t_{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 $\{t_1, t_2, t_{\mathrm{HAC}}, t_{\mathrm{HAR}}\}$. It is convenient for testing to consider asymptotic behavior under the null hypothesis of no convergence or divergence, i.e. when $\lambda = 0$. This null is useful because it enables directional test sof convergence ($\lambda > 0$) and divergence ($\lambda < 0$).

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 ϵ_{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* $\Omega_\phi^2 = 4\sigma_a^2\Omega_\epsilon^2 + \Omega_{\epsilon^2}^2$, Ω_ϵ^2 *is the long run variance of* ϵ_t, $\Omega_{\epsilon^2}^2$ *is the long run variance of* ϵ_t^2, *and* $\sigma_a^2 = \mathbb{E}\left(a_i - a\right)^2$.
>
> (ii) $t_1 \rightsquigarrow \mathcal{N}(0,1)$.
>
> (iii) $$t_2 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_2^2\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} \rightsquigarrow \frac{Z}{\left\{\int_0^1\int_0^1\left(1 - \frac{|r-s|}{b}\right)dW^\tau(r)\,dW^\tau(s)\right\}^{1/2}}.$$
>
> (iv) $$t_{\mathrm{HAR}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}T\hat{\Omega}_M^2\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} \rightsquigarrow \frac{Z}{\left\{\int_0^1\int_0^1\left(1 - \frac{|r-s|}{b}\right)\tilde{r}\tilde{s}dW^\tau(r)\,dW^\tau(s)\right\}^{1/2}}.$$

[4] When $\lambda = 0$, $\widetilde{t^{-0}} = t^0 - T^{-1}\sum_{t=1}^{T} t^0 = 0$ so that $\tilde{\eta}_t = \sigma_\epsilon^2\widetilde{t^{-0}} = 0$. Also note that $\xi_{n,t} = \eta_{n,t} - \eta_t = \left(\sigma_{\epsilon,nT}^2 - \sigma_\epsilon^2\right)t^{-\lambda} = \sigma_{\epsilon,nT}^2 - \sigma_\epsilon^2$, and $\tilde{\xi}_{n,t} = \xi_{n,t} - T^{-1}\sum_{t=1}^{T}\xi_{nt} = 0$.

(v) $t_{\mathrm{HAC}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1} T\hat{\Omega}_L^2 \left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} \rightsquigarrow \mathcal{N}(0,1)$.

In (iv) $\tilde{r} = r - \int_0^1 sds$. In (iii) and (iv) the stochastic process $W^{\tau}(\cdot)$ is the generalized standard Brownian Bridge $W^{\tau}(r) = W(r) - \alpha_B - \beta_B r$, which is the linearly $L_2[0,1]$ detrended version of the standard Brownian motion $W(r)$. The coefficients (α_W, β_W) are the solution of the $L_2[0,1]$ optimization problem (Phillips, 1988; Park and Phillips, 1988, 1989;)

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

As discussed in the proof, two long run variance components $\left(\Omega_{\epsilon}^2, \Omega_{\epsilon^2}^2\right)$ appear in the asymptotic variance Ω_{ϕ}^2 of $\sqrt{n}T^{3/2}\hat{\phi}_{nT}$. This complication arises because the residual term $\varepsilon_{n,t}$ in the expression for the dependent variable K_{nt} involves the two second order moment quantities $2n^{-1}\sum_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it}$ and $n^{-1}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_{\epsilon}^2\right)$ that contribute to long run variation through the variable $\zeta_{it} = 2\tilde{a}_i\tilde{\epsilon}_{it} + \left(\tilde{\epsilon}_{it}^2 - \sigma_{\epsilon}^2\right)$. The quantity $\Omega_{\phi}^2 = 4\sigma_a^2\Omega_{\epsilon}^2 + \Omega_{\epsilon^2}^2$ is the long run variance of ζ_{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_{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_{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_{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 $a_{nT} \sim_a c_{nT}$ if $a_{nT}/c_{nT} \to_p 1$ as $n, T \to \infty$. Then

$$\hat{\phi}_{nT} \sim_a I_A = \sum_{t=1}^{T} a_{tT}\tilde{\eta}_t = \{1 + o_p(1)\} \begin{cases} -\sigma_\epsilon^2 L_\lambda T^{-1-\lambda} & \text{if } \lambda < 1, \\ -6\sigma_\epsilon^2 T^{-2} \ln T & \text{if } \lambda = 1, \\ -6\sigma_\epsilon^2 \mathcal{Z}(\lambda) T^{-2} & \text{if } \lambda > 1. \end{cases} \tag{20}$$

where $L_\lambda = 6\lambda[(2-\lambda)(1-\lambda)]^{-1}$ and $\mathcal{Z}(\lambda) = \sum_{t=1}^{\infty} t^{-\lambda}$ is the Riemann zeta function. As $n, T \to \infty$, $\hat{\phi}_{nT}$ 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 σ_ϵ^2 and the smoothing parameters κ 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 $\hat{u}_t$ and κ. As shown in KPS (2019), when $n, T \to \infty$, the dominant term of $\hat{u}_t$ becomes

$$\hat{u}_t = \tilde{\eta}_{n,t} - \hat{\phi}_{nT}\tilde{t} + \tilde{\varepsilon}_{nT} \sim_a \tilde{\eta}_{n,t} - \hat{\phi}_{nT}\tilde{t},$$

but

$$\tilde{\eta}_{n,t} - \hat{\phi}_{nT}\tilde{t} = \tilde{\eta}_t - I_A\tilde{t} + \tilde{\xi}_{n,t} - \tilde{t}(I_B + I_C) \sim_a \tilde{\eta}_t - I_A\tilde{t}.$$

Hence the residual can be approximated as

$$\hat{u}_t \sim_a \tilde{\eta}_t - \left[\sum_{t=1}^{T} a_{tT}\tilde{\eta}_t\right]\tilde{t}, \tag{21}$$

which is a linearly detrended form of $\tilde{\eta}_t$. From the definitions following (10), $\tilde{\eta}_t = \sigma_\epsilon^2 \widetilde{t^{-\lambda}}$ and is a linear function of σ_ϵ^2. So, in conjunction with (20), it is apparent that the t_1 test statistic is free from the scale nuisance parameter of σ_ϵ^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 σ-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 $n, T \to \infty$:

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

where the primary term of $O(1)$ with $\lambda \geq 1$ is given by

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

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 $\mathbb{Z}\left(\lambda\right) := \mathcal{Z}\left(\lambda\right)\left(\sum_{t=1}^{\infty} t^{-\lambda}\mathcal{Z}\left(\lambda, t\right)\right)^{-1/2} > 1$ for all $\lambda > 1$, where $\mathcal{Z}\left(\lambda, t\right) = \sum_{s=1}^{\infty}\left(s+t\right)^{-\lambda}$ is the Hurwitz zeta function.

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

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

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

As $(n, T) \to \infty$ the statistics t_1, t_2, t_{HAR} and t_{HAC} are all negative under convergence since their signs are determined by the trend regression coefficient $\hat{\phi}_{nT}$ 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 κ 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_{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_{HAR} statistics depends on the value of λ and the smoothing parameter b. This behavior is explored in the numerical simulations that follow.

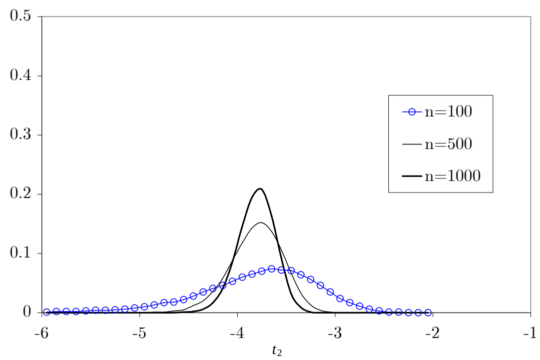

Figure 1a: Densities of the t_2 statistic
$T = 100$, $\beta = 0.25$ and $b = 0.3$

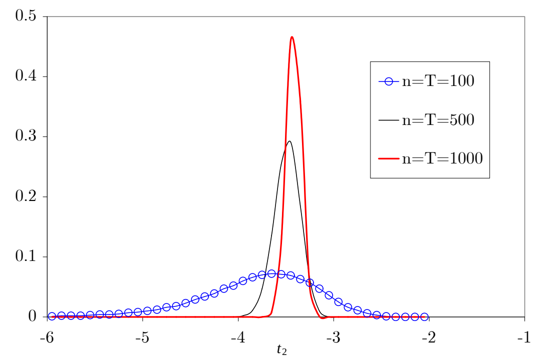

Figure 1b: Densities of the t_2 statistic
$\beta = 0.25$ and $b = 0.3$

The orders of magnitude of the t_{HAC} and the t_{HAR} statistics given in (25) bear a similar relation to those of the t_a^{HAC} and the t_a^{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 $L = \lfloor T^{\kappa} \rfloor$, the order of magnitude of the t_{HAC} statistic is $O_p\left(\sqrt{T/L}\right) = O_p\left(T^{(1-\kappa)/2}\right)$ just as in (3); and $t_{\mathrm{HAR}} = O_p(1)$ as in the case of t_a^{HAR}. Note that the asymptotic properties of both t_a^{HAC} and t_a^{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_{HAC} and the t_{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 $\eta_t = \sigma_\epsilon^2 t^{-\lambda}$ 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_{HAR} and t_{HAC} given in (25) and to asymptotic behavior analogous to the t_a^{HAC} and t_a^{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 λ increases, the fitted linear trend regression equation (9) reverts to the null specification as the effective regressor $\eta_t = \sigma_\epsilon^2 t^{-\lambda}$ in the true model has negligible deterministic trend properties when λ increases and the trend decay component is ultimately zero as $\lambda \to \infty$. Thus, as λ 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 $K_{nt} = n^{-1} \sum_{i=1}^n \left(x_{it} - \frac{1}{n} \sum_{i=1}^n x_{it} \right)^2$ become nonstochastic as $n \to \infty$. In particular, from (6) we have $K_{nt} = \sigma_{a,n}^2 + \eta_{n,t} + \varepsilon_{n,t}$ and it is easy to see that $\sigma_{a,n}^2 \to_p \sigma_a^2$, $\eta_{n,t} \to_p \sigma_\epsilon^2 t^{-\lambda}$, and $\varepsilon_{n,t} \to_p 0$ as $n \to \infty$. Hence, for large n we have the deterministic representation $K_{nt} \sim_a \sigma_a^2 + \sigma_\epsilon^2 t^{-\lambda}$. This large n setting for K_{nt} provides a convenient mechanism for studying the behavior of the trend coefficient estimate $\hat{\phi}_{nT}$ 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 λ 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 λ 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 λ. 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: $|t_1| \geq |t_{\mathrm{HAC}}| \geq |t_2| \geq |t_{\mathrm{HAR}}|$ for all values of λ with $b = 1/3$. Only for very large λ 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_{HAR} statistics with big changes in T and as λ changes. We maintain the setting $\kappa = b = 1/3$ for comparability. Figure 3 shows the t_2 and t_{HAR} statistic values as λ 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 $K_{nt} \sim_a \sigma_a^2 + \sigma_\epsilon^2 t^{-\lambda}$ under the alternative, as λ increases, we can expect the discriminatory power for detecting convergence to dissipate rapidly for large t because the impact on K_{nt} of $\sigma_\epsilon^2 t^{-\lambda}$ is small when t is large and λ 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 dispays the t_1 and t_{HAC} statistics for the same two values of T. As Theorem 1 predicts, both t_1 and t_{HAC} are time series sample size dependent as $T \to \infty$. As T increases, both test statistics become noticeably larger in absolute value when λ is moderately small. With a large λ, the effects of rising T are attenuated for the reason explained above and both statistics converge to $-\sqrt{3}$ as λ increases indefinitely.

Figure 5 explains how the smoothing parameter b influences the t_{HAR} statistic across various λ values. At a given λ, the absolute value of the t_{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_{HAR} on b is highly nonlinear. The behavior is somewhat expected since it is known that as b approaches to zero, the t_{HAR} statistic approaches the t_{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.

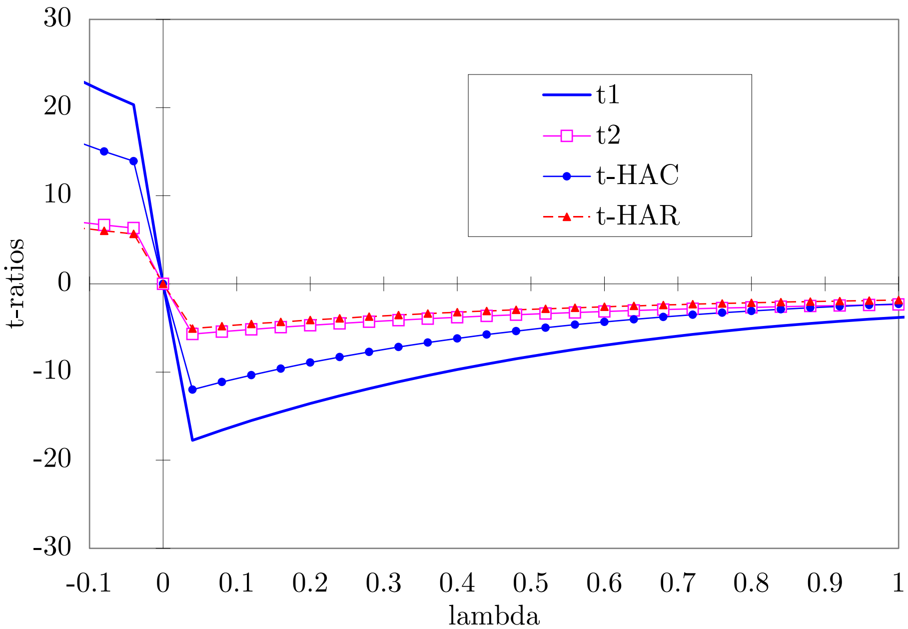

Figure 2: Test statistic numerical calculations ($b = \kappa = 1/3$, $T = 1,000$)

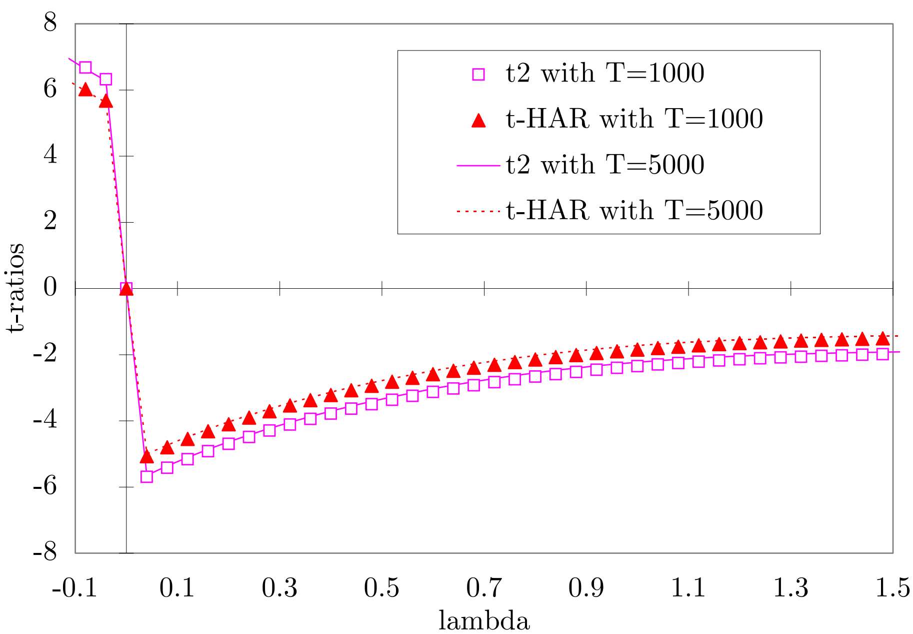

Figure 3: Impact of large T on tests t_2 and t_{HAR} ($b = \kappa = 1/3$)

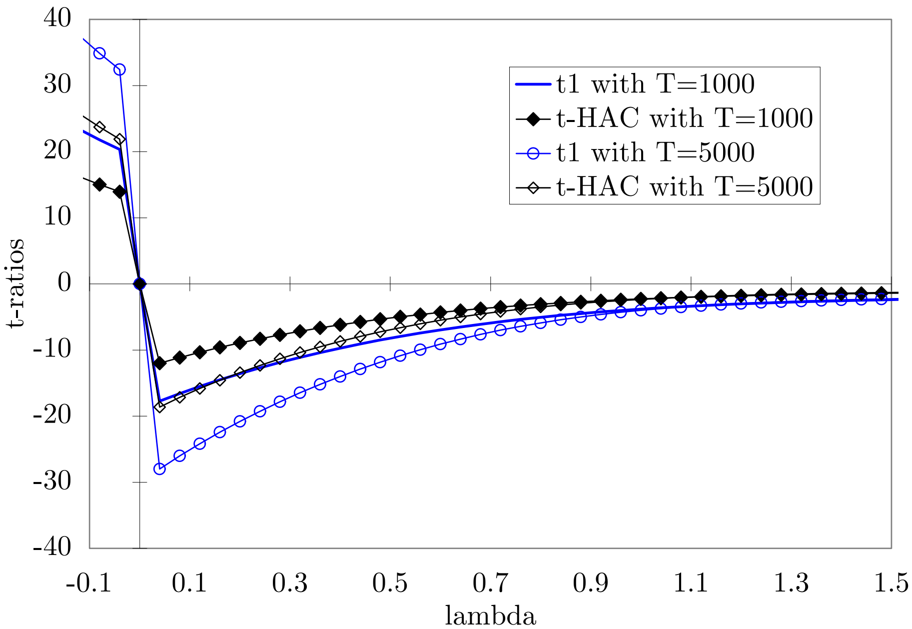

Figure 4: Time varying behavior of t_1 and t_{HAC} $(b = \kappa = 1/3)$

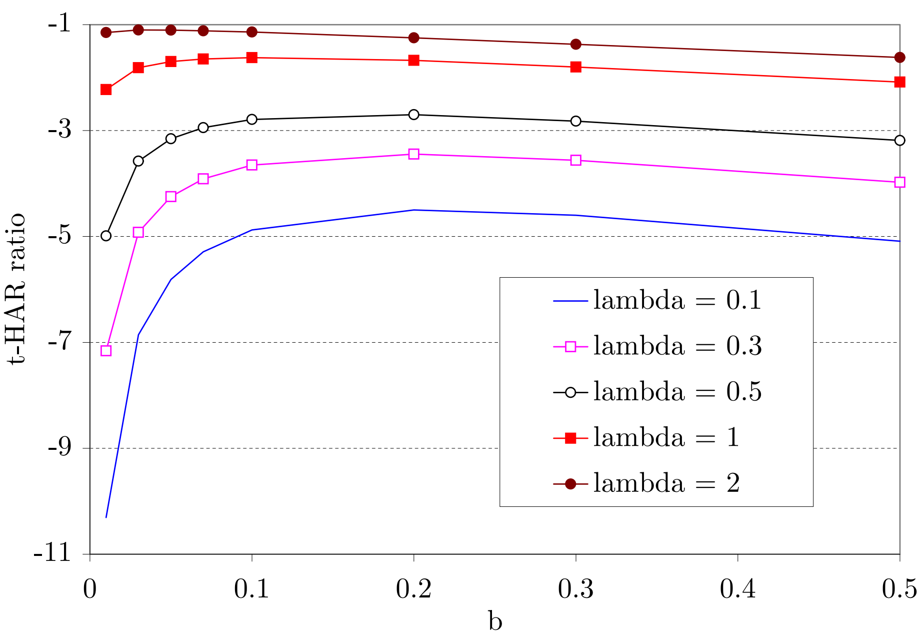

Figure 5: Limits of t_{HAR} for various b and λ $(T = 1,000)$.

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.,

$$y_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it} t^{-\beta}, \tag{27}$$

where $\epsilon_{it} = \rho\epsilon_{it-1} + v_{it}$, $a_i \sim iid\mathcal{N}(0,1)$, $\mu_i \sim iid\mathcal{N}(0,1)$, $v_{it} \sim iid\mathcal{N}(0,1)$, 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}

t_2			t_{HAR}		
$b = 0.1$	$b = 0.2$	$b = 0.3$	$b = 0.1$	$b = 0.2$	$b = 0.3$
-2.155	-2.499	-2.938	-2.341	-2.746	-3.118

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_{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_{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_{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 $t_{\hat{\phi}}$ 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

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

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

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

This crude estimate of λ 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 λ is less than $1/2$, the large $n-$asymptotic limits of both the t_{HAR} and t_2 tests approach negative constants.

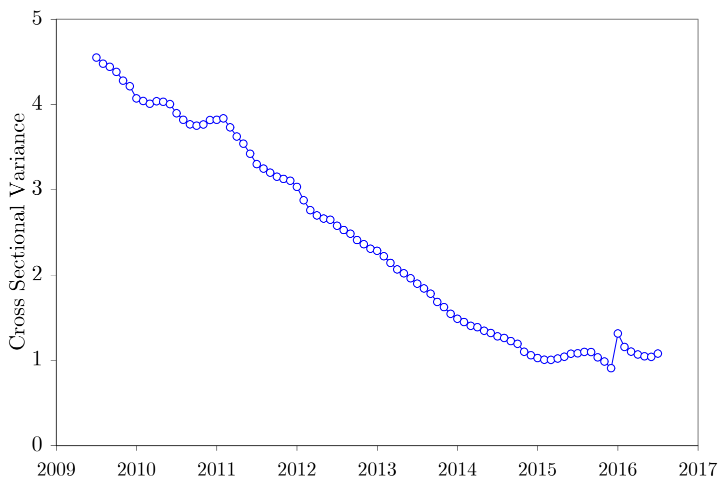

Figure 6: Cross section variance of unemployment rates over the 48 US contiguous States

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_{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_{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_{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

		k	$t-$ratio			k	$t-$ratio
t_1		4	-21.948	t_{HAC}		4	-19.636
t_2	$b=0.1$	9	-16.860	t_{HAR}	$b=0.1$	9	-15.931
	$b=0.2$	17	-14.446		$b=0.2$	17	-14.767
	$b=0.3$	30	-14.254		$b=0.3$	30	-15.732

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.

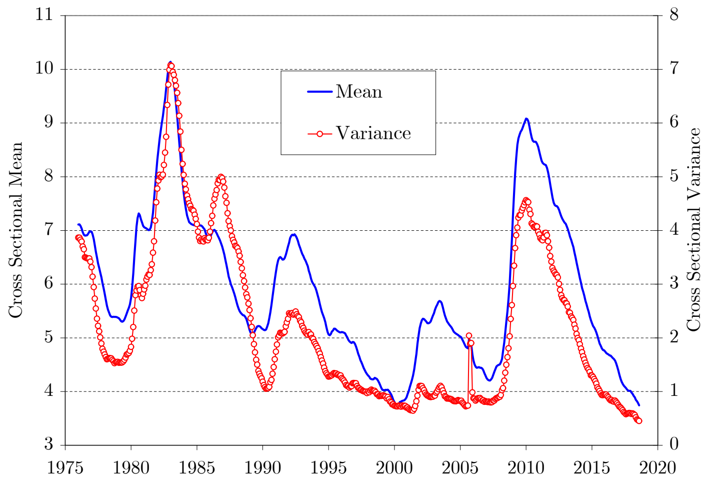

Figure 7: Time Varying Patterns of Cross Section Mean and Variance in US State Unemployment Rates over 42 years

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 σ-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 power.

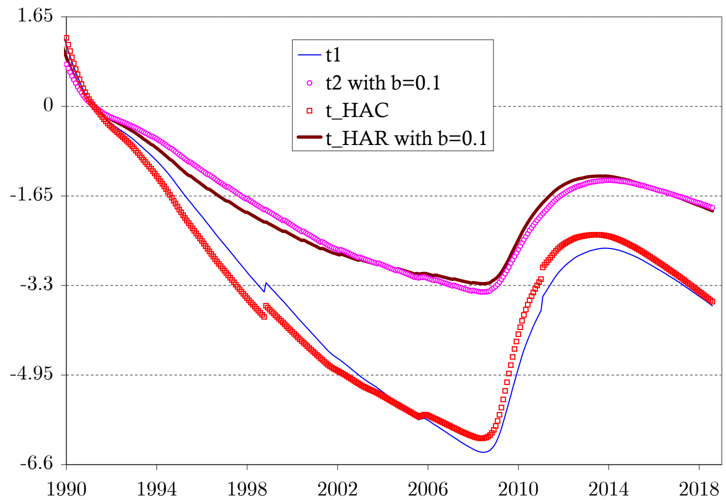

Figure 8: Recursive t-ratios

6 Concluding Remarks

We investigate the use of various HAC, HAR and sandwich-type long run variance estimators in testing weak σ-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 σ-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%)

T	n	t_1	t_{HAC}	t_2 b =0.1	t_2 b =0.2	t_2 b =0.3	t_{HAR} b =0.1	t_{HAR} b =0.2	t_{HAR} b =0.3
25	25	0.131	0.154	0.089	0.076	0.075	0.094	0.076	0.073
25	50	0.131	0.149	0.079	0.070	0.070	0.084	0.070	0.065
25	100	0.146	0.165	0.085	0.073	0.073	0.083	0.071	0.069
25	200	0.135	0.155	0.088	0.074	0.074	0.094	0.078	0.071
25	500	0.132	0.152	0.081	0.073	0.074	0.088	0.071	0.069
50	25	0.114	0.128	0.061	0.060	0.059	0.063	0.059	0.056
50	50	0.100	0.117	0.057	0.055	0.054	0.060	0.053	0.051
50	100	0.104	0.119	0.056	0.053	0.052	0.059	0.053	0.051
50	200	0.110	0.128	0.059	0.056	0.053	0.062	0.058	0.054
50	500	0.105	0.121	0.062	0.057	0.055	0.063	0.059	0.055
100	25	0.089	0.101	0.048	0.048	0.047	0.047	0.043	0.044
100	50	0.096	0.107	0.048	0.050	0.049	0.048	0.048	0.045
100	100	0.092	0.102	0.048	0.047	0.048	0.046	0.046	0.046
100	200	0.083	0.097	0.048	0.050	0.045	0.049	0.049	0.046
100	500	0.092	0.101	0.046	0.046	0.042	0.047	0.045	0.044
200	25	0.076	0.083	0.035	0.038	0.037	0.038	0.037	0.036
200	50	0.076	0.081	0.040	0.043	0.042	0.040	0.040	0.042
200	100	0.086	0.093	0.041	0.043	0.044	0.036	0.040	0.040
200	200	0.080	0.086	0.041	0.041	0.041	0.040	0.041	0.040
200	500	0.078	0.084	0.043	0.046	0.043	0.044	0.047	0.046

Table 3: Power Comparison under M2 (Size Adjusted): $\lambda = 2\beta = 0.1$

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

Table 4: Power Comparison under M1 (Size Adjusted): $\lambda = \alpha = 0.05$

				t_2			t_{HAR}		
t	n	t_1	t_{HAC}	$b = 0.1$	$b = 0.2$	$b = 0.3$	$b = 0.1$	$b = 0.2$	$b = 0.3$
25	25	0.081	0.086	0.081	0.080	0.082	0.086	0.084	0.083
25	50	0.111	0.111	0.111	0.108	0.102	0.111	0.109	0.108
25	100	0.135	0.128	0.135	0.128	0.124	0.128	0.127	0.124
25	200	0.182	0.183	0.182	0.176	0.176	0.183	0.177	0.176
25	500	0.306	0.303	0.306	0.293	0.284	0.303	0.285	0.290
50	25	0.102	0.098	0.100	0.103	0.100	0.099	0.097	0.094
50	50	0.135	0.126	0.131	0.123	0.120	0.125	0.120	0.118
50	100	0.187	0.185	0.185	0.168	0.159	0.183	0.170	0.171
50	200	0.262	0.253	0.256	0.243	0.235	0.250	0.239	0.234
50	500	0.472	0.446	0.462	0.417	0.390	0.440	0.405	0.391
100	25	0.129	0.127	0.123	0.119	0.112	0.122	0.113	0.112
100	50	0.185	0.189	0.179	0.172	0.158	0.179	0.170	0.165
100	100	0.262	0.266	0.252	0.238	0.236	0.255	0.237	0.234
100	200	0.407	0.395	0.390	0.353	0.325	0.365	0.348	0.337
100	500	0.732	0.688	0.694	0.627	0.592	0.658	0.592	0.563
200	25	0.167	0.173	0.165	0.158	0.153	0.163	0.158	0.154
200	50	0.252	0.247	0.240	0.227	0.217	0.236	0.219	0.217
200	100	0.399	0.390	0.366	0.337	0.319	0.343	0.314	0.310
200	200	0.624	0.596	0.586	0.528	0.490	0.526	0.495	0.470
200	500	0.918	0.884	0.871	0.792	0.729	0.806	0.716	0.672

References

[1] Barro, R.J. and X. Sala-I-Martin (1992). Convergence. *Journal of Political Economy*, 100, 223–251.

[2] Bonhomme, S. and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184.

[3] Bunzel, H. and Vogelsang, T. J.: (2005). Powerful trend function tests that are robust to strong serial correlation with an application to the prebisch-singer hypothesis, *Journal of Business and Economic Statistics,* 23, 381–394.

[4] Campbell, J. Y., Lettau, M., Malkiel, B. G. and Xu, Y.: (2001), Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *The Journal of Finance,* 56, 1–43.

[5] Chen, Y. and Y. Tu (2019). "Is Stock Price Correlated with Oil Price? Spurious Regressions with Moderately Explosive Processes," *Oxford Bulletin of Economics and Statistics* (forthcoming).

[6] Durlauf, S. N. and P. C. P. Phillips (1988). "Trends versus random walks in time series analysis," *Econometrica* 56, 1333–1354.

[7] Friedman, M. (1992). Do old fallacies ever die, *Journal of Economic Literature*, 30, 2129-2132.

[8] Hotelling, H. (1933). Review of *The triumph of mediocrity in business* by Horace Secrist. *Journal of the American Statistical Association, 28, 463-465.*

[9] Hsiao, C. (2014). *Analysis of Panel Data* (Third Ed.). New York: Cambridge University Press.

[10] Hwang, J. and Y. Sun (2018). Simple, Robust, and Accurate F and t Tests in Cointegrated Systems. *Econometric Theory,* 34, 949-984.

[11] Jansson, M. (2004). On the error of rejection probability in simple autocorrelation robust tests. *Econometrica*, 72, 937-946.

[12] Kiefer, N., T.J. Vogelsang, and H. Bunzel (2000). Simple Robust Testing of Regression Hypotheses, *Econometrica*, 69, 695-714

[13] Kiefer, N. and T.J. Vogelsang (2002), Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation, *Econometrica*, 70, 2093-2095.

[14] Kiefer, N. and T.J. Vogelsang (2002). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory,* 18:1350-1366.

[15] Kiefer, N. and T. J. Vogelsang (2005). A New Asymptotic Theory for Heteroskedasticity autocorrelation Robust Tests. Econometric Theory, 21, 1130–1164

[16] Kong, J., P.C.B. Phillips, and D. Sul (2019). Weak σ convergence: Theory and Applications, *Journal of Econometrics,* 209, 185-207.

[17] Lazarus, E., D. Lewis, J. Stock, and M.W. Watson (2018). HAR Inference: Recommendations for Practice, forthcoming in *Journal of Business and Economic Statistics,* (forthcoming).

[18] Müller, U.K. (2014), HAC Corrections for Strongly Autocorrelated Time Series, *Journal of Business and Economic Statistics*, 32, 311-322.

[19] Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics* 33, 311–340.

[20] Phillips, P. C. B. (1988). "Multiple regression with integrated processes." In N. U. Prabhu, (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics* 80, 79–106.

[21] Phillips, P. C. B. (1998). New Tools for Understanding Spurious Regressions. *Econometrica*, 66, 1299-1326.

[22] Phillips, P. C. B.(2005b). Automated Discovery in Econometrics. Econometric Theory, Vol. 21, No. 1 February, 2005, pp. 3-20

[23] Phillips, P.C.B. (2005a). HAC estimation by automated regression. *Econometric Theory,* 21:116–142.

[24] Phillips, P.C.B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, 23, 557–614.

[25] Phillips, P.C.B. and J.Y. Park (1988). Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors, Journal of the American Statistical Association, Vol. 83, No. 401: 111–115

[26] Phillips, P. C. B. and H. R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data , Econometrica, Vol. 67, No. 5: 1057–1111.

[27] Park, J.Y., and P.C.B. Phillips. (1988) Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4, No. 3: 468-497.

[28] Park, J.Y., and P.C.B. Phillips. (1989) Statistical inference in regressions with integrated processes: Part 2. Econometric Theory, Vol.5, No.1: 95–132.

[29] Phillips, P. C. B. and D. Sul (2007a). Transition Modeling and Econometric Convergence Tests, *Econometrica*, 75, 1771-1855.

[30] Phillips, P. C. B. and D. Sul (2007b). Some Empirics on Economic Growth under Heterogeneous Technology, *Journal of Macroeconomics*, 29, 455-469.

[31] Phillips, P. C. B. and D. Sul (2009). Economic Transition and Growth. *Journal of Applied Econometrics*, 24, 1153-1185.

[32] Phillips, P.C.B., Sun, Y., and Jin, S. (2007). Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Journal of Statistical Planning and Inference,* 137, 985-1023

[33] Phillips, P.C.B., Y. Zhang and X. Wang (2012). Limit theory of three $t-$statistics in spurious regressions. Mimeo, Singapore Management University.

[34] Robinson, P.M. (1995). Log-periodogram regression of time series with long range dependence. Annals of Statistics, 23, 1048–1072.

[35] Secrist, H. (1933). The triumph of mediocrity in business. Bureau of Business Research, Northwestern University.

[36] Su, L., Z. Shi and P. C. B. Phillips (2016). "Identifying Latent Structures in Panel Data", *Econometrica*, 84, 2215-2264.

[37] Sul, D., 2019. Panel Data Econometrics: Common Factor Analysis for Empirical Researchers. Routledge.

[38] Sun, Y. (2004). A convergent t-statistic in spurious regression. *Econometric Theory,* 20:943-962.

[39] Sun, Y. (2014). Let's Fix It: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. *Journal of Econometrics* 178, 659-677.

[40] Sun, Y. (2018). Comment on "HAR Inference: Recommendations for Practice", *Journal of Business and Economic Statistics,* (forthcoming).

[41] Sun, Y., P.C.B. Phillips and S. Jin (2008). "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing", *Econometrica*, 76, 175-194.

[42] Wang, Y., P. C. B. Phillips and Y. Tu (2019). Spurious Functional-coefficient Regression Models and a Robust Solution, Working Paper, University of Auckland.

[43] Wang, W., P. C. B. Phillips and L. Su (2019). The Heterogeneous Effects of the Minimum Wage on Employment Across States, *Economics Letters*, 2019 (forthcoming).

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

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

Assumption B:

(i) $k(x) : \mathbb{R} \to [0, 1]$ *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

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+,\ g_{q_0} = \lim_{x\to 0}\frac{1-k(x)}{|x|^{q_0}} < \infty\} \tag{29}$$

is greater than or equal to 1.

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

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

$$K_{nt} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(x_{it} - \frac{1}{n}\sum\nolimits_{i=1}^{n} x_{it}\right)^2 = \sigma_{a,n}^2 + \eta_{n,t} + \varepsilon_{n,t}, \tag{30}$$

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

$$\varepsilon_{n,t} = 2n^{-1}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right)t^{-2\beta}, \tag{31}$$

with $\sigma_{\epsilon,nt}^2 = n^{-1}\sum_{i=1}^{n}\tilde{\epsilon}_{it}^2$, $\sigma_{\epsilon,nT}^2 = T^{-1}\sum_{t=1}^{T}\sigma_{\epsilon,nt}^2$, and where the notation $\tilde{\epsilon}_{it} = \epsilon_{it} - n^{-1}\sum_{j=1}^{n}\epsilon_{jt}$ 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 $\eta_t = \sigma_\epsilon^2 t^{-\lambda}$, $\lambda = 2\beta$, $\xi_{n,t} = \eta_{n,t} - \eta_t = \left(\sigma_{\epsilon,nT}^2 - \sigma_\epsilon^2\right)t^{-\lambda}$ and σ_ϵ^2 is the variance of ϵ_{it}.

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

$$\begin{aligned}
\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2 &= n^{-1}\sum_{i=1}^{n}\tilde{\epsilon}_{it}^2 - T^{-1}\sum_{t=1}^{T}\sigma_{\epsilon,nt}^2 \\
&= n^{-1}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) - \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) \\
&= n^{-1}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) + O_p\left(\frac{1}{\sqrt{nT}}\right),
\end{aligned}$$

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

$$\begin{aligned}
K_{nt} &= \sigma_{a,n}^2 + \sigma_\epsilon^2 + 2n^{-1}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) + \left(\sigma_{\epsilon,nT}^2 - \sigma_\epsilon^2\right) \\
&= \sigma_n^2 + 2n^{-1}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + n^{-1}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) + O_p\left(\frac{1}{\sqrt{nT}}\right),
\end{aligned} \tag{32}$$

with $\sigma_n^2 = \sigma_{a,n}^2 + \sigma_\epsilon^2 + \left(\sigma_{\epsilon,nT}^2 - \sigma_\epsilon^2\right)$. We can write (32) as

$$K_{nt} = \sigma_n^2 + u_{nt}, \text{ with } u_{nt} = \frac{1}{\sqrt{n}}\zeta_{nt} + O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{33}$$

and

$$\zeta_{nt} = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} \tilde{a}_i\tilde{\epsilon}_{it} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_{it}, \tag{34}$$

$$\zeta_{it} = 2\tilde{a}_i\tilde{\epsilon}_{it} + \left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) \tag{35}$$

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

$$\widetilde{K_{nt}} = K_{nt} - T^{-1}\sum_{t=1}^{T} K_{nt} = u_{nt} - T^{-1}\sum_{t=1}^{T} u_{nt} = \widetilde{u_{nt}} = u_{nt} + O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{36}$$

since

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

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

$$K_{nt} = \sigma_n^2 + u_{nt} = \sigma_n^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_{it} + O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{37}$$

which is a simple location model in which the effective error u_{nt} involves ζ_{it} and hence both first and second centred sample moments of ϵ_{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

$$K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT}t + \hat{u}_t. \tag{38}$$

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

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

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

$$\begin{aligned}
\hat{\phi}_{nT} &= \sum\nolimits_{t=1}^{T} c_{tT}\widetilde{K_{nt}} = \sum\nolimits_{t=1}^{T} c_{tT}u_{nt} \\
&= \sum\nolimits_{t=1}^{T} c_{tT}\left[2n^{-1}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + n^{-1}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)\right] \\
&= \frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{T} c_{tT}\left[\frac{2}{\sqrt{n}}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\right].
\end{aligned}$$

Then,

$$\begin{aligned}
\sqrt{n}T^{3/2}\hat{\phi}_{nT} &= \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\frac{\frac{\tilde{t}}{T}}{\frac{\sum_{s=1}^{T}\tilde{s}^2}{T^3}}\left[\frac{2}{\sqrt{n}}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\right] \\
&= \frac{12}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\frac{\tilde{t}}{T}\left\{\frac{2}{\sqrt{n}}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)\right\} + o_p(1) \\
&\rightsquigarrow_{(n,T\to\infty)} \mathcal{N}\left(0, 12\Omega_\phi^2\right). \qquad (39)
\end{aligned}$$

We proceed to prove (39). In this expression for the limit theory the variance Ω_ϕ^2 depends on the two components $\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it}$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)$. 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 $\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it}$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)$ are asymptotically normal and independent in view of the independence of a_i and ϵ_{it} as

$$\mathbb{E}\left\{(a_i - a)\epsilon_{it}\left(\epsilon_{is}^2 - \sigma_\epsilon^2\right)\right\} = \mathbb{E}(a_i - a)\,\mathbb{E}\left\{\epsilon_{it}\left(\epsilon_{is}^2 - \sigma_\epsilon^2\right)\right\} = 0, \text{ for all } t, s$$

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

$$\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{t}}{T}\tilde{a}_i\tilde{\epsilon}_{it} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{t}}{T}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) \end{bmatrix} \rightsquigarrow_{(T\to\infty)} \mathcal{N}\left(0, \frac{1}{12}\begin{bmatrix} \sigma_a^2\Omega_\epsilon^2 & 0 \\ 0 & \Omega_{\epsilon^2}^2 \end{bmatrix}\right). \qquad (40)$$

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

$$\begin{bmatrix} \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\tilde{t}}{T}\tilde{a}_i\tilde{\epsilon}_{it} \\ \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\tilde{t}}{T}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right) \end{bmatrix} \rightsquigarrow_{(n,T\to\infty)} \mathcal{N}\left(0, \frac{1}{12}\begin{bmatrix} \sigma_a^2\Omega_\epsilon^2 & 0 \\ 0 & \Omega_{\epsilon^2}^2 \end{bmatrix}\right), \qquad (41)$$

where $\sigma_a^2 = \mathbb{E}(a_i - a)^2$. The long run variance of $\zeta_{it} = \left\{2\tilde{a}_i\tilde{\epsilon}_{it} + \left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)\right\}$ is $\Omega_\phi^2 = 4\sigma_a^2\Omega_\epsilon^2 + \Omega_{\epsilon^2}^2$ and it follows from (41) that

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum\nolimits_{t=1}^{T}\zeta_{it} = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum\nolimits_{t=1}^{T}\left\{2\tilde{a}_i\tilde{\epsilon}_{it} + \left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)\right\} \rightsquigarrow_{(n,T\to\infty)} \mathcal{N}\left(0, \Omega_\phi^2\right). \tag{42}$$

We deduce that

$$\begin{aligned} \sqrt{n}T^{3/2}\hat{\phi}_{nT} &= \frac{12}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\frac{\tilde{t}}{T}\left[\frac{2}{\sqrt{n}}\sum\nolimits_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)\right] + o_p(1) \\ &\rightsquigarrow_{(n,T\to\infty)} Z_\phi \equiv \mathcal{N}\left(0, 12\Omega_\phi^2\right), \quad \text{with } \Omega_\phi^2 = 4\sigma_a^2\Omega_\epsilon^2 + \Omega_{\epsilon^2}^2. \end{aligned} \tag{43}$$

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

In (43) the dependence of the asymptotic variance Ω_ϕ^2 of $\sqrt{n}T^{3/2}\hat{\phi}_{nT}$ on the two long run variance components $\left(\Omega_\epsilon^2, \Omega_{\epsilon^2}^2\right)$ 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 $2n^{-1}\sum_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it}$ and $n^{-1}\sum_{i=1}^{n}\left(\tilde{\epsilon}_{it}^2 - \sigma_\epsilon^2\right)$, which in turn lead to the more complex limiting variance matrix Ω_ϕ^2, which depends on fourth moments of ϵ_{it}.

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

$$t_1 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_1^2\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} = \frac{\sqrt{n}T^{3/2}\hat{\phi}_{nT}}{\sqrt{n\hat{\Omega}_1^2\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}},$$

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

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

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

$$\begin{aligned}\hat{a}_{nT} &= \frac{1}{T}\sum_{t=1}^{T}\left(K_{nt}-\hat{\phi}_{nT}t\right)=\sigma_n^2+T^{-1}\sum_{t=1}^{T}u_{nt}-\frac{1}{2}\sqrt{n}T^{3/2}\hat{\phi}_{nT}\frac{(T+1)}{\sqrt{n}T^{3/2}} \\ &= \sigma_n^2+O_p\left(\frac{1}{\sqrt{nT}}\right),\end{aligned}$$

using (36) and (39). Hence, the residual in (38), $\hat{u}_t$, becomes

$$\begin{aligned}\hat{u}_t &= K_{nt}-\hat{a}_{nT}-\hat{\phi}_{nT}t=\widetilde{u_{nt}}-\hat{\phi}_{nT}\tilde{t} \\ &= \widetilde{u_{nt}}-\left(\sum\nolimits_{t=1}^{T}c_{tT}\tilde{u}_{nt}\right)\tilde{t},\end{aligned}\qquad (44)$$

where $\hat{u}_t$ is simply linearly detrended u_{nt}. Since $u_{nt}=\frac{1}{\sqrt{n}}\zeta_{nt}$, it follows that $\hat{u}_t=n^{-1/2}\zeta_{nt}^{\tau}$ where $\zeta_{nt}^{\tau}=\widetilde{\zeta_{nt}}-\left(\sum_{t=1}^{T}c_{tT}\widetilde{\zeta_{nt}}\right)\tilde{t}$ is linearly detrended $\widetilde{\zeta_{nt}}$ where $\widetilde{\zeta_{nt}}=\zeta_{nt}-T^{-1}\sum_{t=1}^{T}\zeta_{nt}$. In view of the independence over i and weak dependence over t, we have, following the analysis in Theorem 1 (i) and (42), the limit theory

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\zeta_{nt}=\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\zeta_{it}+o_p(1)\rightsquigarrow_{(n,T\to\infty)}\mathcal{N}\left(0,\Omega_\zeta^2\right),\qquad (45)$$

where $\Omega_\zeta^2=\Omega_\phi^2=4\sigma_a^2\Omega_\epsilon^2+\Omega_{\epsilon^2}^2$ is the long run variance of $\zeta_{it}=2\tilde{a}_i\epsilon_{it}+\left(\epsilon_{it}^2-\sigma_\epsilon^2\right)$, Ω_ϵ^2 is the long run variance of ϵ_{it}, and $\Omega_{\epsilon^2}^2$ is the long run variance of ϵ_{it}^2. Similarly, we have the functional laws

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\zeta_{nt}\rightsquigarrow_{(n,T\to\infty)}B_{\Omega_\zeta^2}(\cdot)\quad\text{and}\quad\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\zeta_{nt}^{\tau}\rightsquigarrow_{(n,T\to\infty)}B_{\Omega_\zeta^2}^{\tau}(\cdot),\qquad (46)$$

where $B_{\Omega_\zeta^2}(\cdot)$ is Brownian motion with variance $\Omega_\zeta^2=\Omega_\phi^2$, and $B_{\Omega_\zeta^2}^{\tau}(\cdot)$ is the generalized Brownian Bridge process $B_{\Omega_\phi^2}^{\tau}(r)=B_{\Omega_\zeta^2}(r)-\alpha_B-\beta_B r$, which is the linearly $L_2[0,1]$ detrended version of $B_{\Omega_\zeta^2}(r)$ in which the coefficients (α_B,β_B) are the solution of the $L_2[0,1]$ optimization problem (Phillips, 1988; Park and Phillips, 1988, 1989.)

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

It follows that second moments of the residuals $\hat{u}_{nt}$ have asymptotic behavior determined by $\zeta_{nt}=n^{-1/2}\sum_{i=1}^{n}\zeta_{it}+o_p(1)\rightsquigarrow_{n\to\infty}\zeta_t^0$ where ζ_t^0 has the same time series behavior as the stationary process ζ_{it}. We are effectively demeaning and detrending the errors u_{nt} by the trend regression which, from (44), gives the residuals

$$\hat{u}_t=\widetilde{u_{nt}}-\left(\sum\nolimits_{t=1}^{T}c_{tT}\widetilde{u_{nt}}\right)\tilde{t}=\widetilde{u_{nt}}-\frac{\tilde{t}}{T}\times O_p\left(\frac{1}{\sqrt{nT}}\right)$$

so that, since $u_{nt} = n^{-1/2}\zeta_{nt}$, we have

$$\sqrt{n}\hat{u}_t = \zeta_{nt} - \frac{\tilde{t}}{T} \times O_p\left(\frac{1}{\sqrt{T}}\right) = \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n} \zeta_{it} + o_p(1) \rightsquigarrow_{n\to\infty} \zeta_t^0.$$

Thus, the sample second moments of the residuals $\hat{u}_t$ are asymptotically, as $n \to \infty$, equivalent to those of ζ_t^0. It follows that the sample second moments and autocovariances of $\hat{u}_t$ have the following limit behavior after scaling by n

$$\begin{aligned} n\hat{\gamma}_{\hat{u}}(0) &= \frac{n}{T}\sum_{t=1}^{T}\hat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n}\zeta_{it} + o_p(1)\right\}^2 \to_p \mathbb{E}\left(\zeta_{it}^2\right) \\ &= 4\mathbb{E}\left\{(a_i - a)^2\epsilon_{it}^2\right\} + \mathbb{E}\left(\epsilon_{it}^2 - \sigma_\epsilon^2\right)^2 = 4\sigma_a^2\sigma_\epsilon^2 + \sigma_{\epsilon^2}^2. \end{aligned} \tag{47}$$

Similarly, setting $\hat{\gamma}_{\hat{u}}(j) := \frac{n}{T}\sum_{1\le t,t-j\le T}\hat{u}_t\hat{u}_{t-j}$, we have

$$\begin{aligned} n\hat{\gamma}_{\hat{u}}(j) &= \frac{n}{T}\sum_{1\le t,t-j\le T}\hat{u}_t\hat{u}_{t-j} = \frac{1}{T}\sum_{1\le t,t-j\le T}\zeta_{nt}^\tau\zeta_{nt-j}^\tau \\ &\to_p \mathbb{E}\left(\zeta_{it}\zeta_{it-j}\right) = 4\mathbb{E}(a_i - a)^2\,\mathbb{E}\left(\epsilon_{it}\epsilon_{it-j}\right) + \mathbb{E}\left(\epsilon_{it}^2 - \sigma_\epsilon^2\right)\left(\epsilon_{it-j}^2 - \sigma_\epsilon^2\right) \\ &= 4\sigma_a^2\Gamma_\epsilon(j) + \Gamma_{\epsilon^2}(j), \end{aligned} \tag{48}$$

where we use the notation $\Gamma_v(j) = \mathbb{E}\left\{(v_t - \mathbb{E}v_t)(v_{t-j} - \mathbb{E}v_{t-j})\right\}$. We deduce that the HAC estimator has the following asymptotic behavior as $T \to \infty$ with $L = \lfloor T^{1/3}\rfloor$ as in the Bartlett-Newey-West estimator

$$\begin{aligned} n\hat{\Omega}_1^2 &= \frac{1}{T}\sum\nolimits_{t=1}^{T}\hat{u}_t^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\hat{u}_t\hat{u}_{t+\ell} \\ &= \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\zeta_t^0\right)^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\zeta_t^0\zeta_{t+\ell}^0 + o_p(1) \\ &\to_p \sum_{j=-\infty}^{\infty}\left\{4\sigma_a^2\Gamma_\epsilon(j) + \Gamma_{\epsilon^2}(j)\right\} = \Omega_\zeta^2 = \Omega_\phi^2. \end{aligned} \tag{49}$$

The final result for the limit distribution of the t_1 statistic

$$t_1 = \frac{\sqrt{n}T^{3/2}\hat{\phi}_{nT}}{\sqrt{n\hat{\Omega}_1^2\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} \rightsquigarrow \frac{\mathcal{N}\left(0, 12\Omega_\phi^2\right)}{\sqrt{12\Omega_\phi^2}} \equiv \mathcal{N}(0,1),$$

now follows.

Proof of (iii) We next consider the analogous t statistic

$$t_2 = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_2^2\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} = \frac{\sqrt{n}T^{3/2}\hat{\phi}_{nT}}{\sqrt{n\hat{\Omega}_2^2\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}},$$

in which the fixed-b HAR long run variance estimate

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

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

$$\begin{aligned} n\hat{\Omega}_2^2 &= n\sum\nolimits_{j=1-T}^{T-1} k\left(\frac{j}{T}\right)\hat{\gamma}_{\hat{u}}(j) = \frac{n}{T}\sum\nolimits_{j=1-T}^{T-1} k\left(\frac{j}{T}\right)\sum_{1\le t,t-j\le T}\hat{u}_t\hat{u}_{t-j} \\ &= \frac{n}{T}\sum\nolimits_{t=1}^{T}\sum\nolimits_{p=1}^{T} k\left(\frac{t-p}{T}\right)\hat{u}_t\hat{u}_p = \frac{1}{T}\sum\nolimits_{t=1}^{T}\sum\nolimits_{p=1}^{T} k\left(\frac{t-p}{T}\right)\zeta_{nt}^{\tau}\zeta_{np}^{\tau}. \end{aligned}$$

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

$$n\hat{\Omega}_2^2 = \sum\nolimits_{t=1}^{T}\sum\nolimits_{p=1}^{T} k\left(\frac{t-p}{T}\right)\frac{\zeta_{nt}^{\tau}}{\sqrt{T}}\frac{\zeta_{nt}^{\tau}}{\sqrt{T}} \rightsquigarrow_{(n,T\to\infty)} \int_0^1\int_0^1 k(r-s)\,dB_{\Omega_\zeta^2}^{\tau}(r)\,dB_{\Omega_\zeta^2}^{\tau}(s). \tag{50}$$

Using $\Omega_\zeta^2=\Omega_\phi^2$, the HAR t statistic for testing the significance of the linear trend in the empirical regression (38) therefore has the following limit theory as $(n,T\to\infty)$

$$\begin{aligned} t_2 &= \frac{\sqrt{n}T^{3/2}\hat{\phi}_{nT}}{\sqrt{n\hat{\Omega}_2^2\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} \\ &\rightsquigarrow_{(n,T\to\infty)} \frac{Z_\phi}{\left\{12\int_0^1\int_0^1 k(r-s)\,dB_{\Omega_\phi^2}^{\tau}(r)\,dB_{\Omega_\phi^2}^{\tau}(s)\right\}^{1/2}} &(51) \\ &= \frac{\sqrt{12}\Omega_\phi Z}{\left\{12\Omega_\phi^2\int_0^1\int_0^1 k(r-s)\,dW^{\tau}(r)\,dW^{\tau}(s)\right\}^{1/2}} \equiv \frac{Z}{\left\{\int_0^1\int_0^1 k(r-s)\,dW^{\tau}(r)\,dW^{\tau}(s)\right\}^{1/2}} &(52) \end{aligned}$$

where $Z_\phi \equiv \mathcal{N}\left(0,12\Omega_\phi^2\right)$, $Z\equiv\mathcal{N}(0,1)$, and $B_{\Omega_\xi^2}^{\tau}(r)=\Omega_\xi W^{\tau}(r)$ where $W^{\tau}(r)$ is detrended standard brownian motion $W(r)\equiv BM(1)$. 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

$$t_{\mathrm{HAR}} \rightsquigarrow_{(n,T\to\infty)} \frac{Z}{\left\{\int_0^1\int_0^1 k\left(\frac{r-s}{b}\right)dW^{\tau}(r)\,dW^{\tau}(s)\right\}^{1/2}},$$

just as in Kiefer and Vogelsang (2002, 2005), Phillips et al. (2007), Sun et al. (2008) and Sun (2014).

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

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

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

In $\hat{\Omega}_M^2$, the variate $\tilde{\varkappa}_t = \hat{u}_t\tilde{t} = \frac{1}{\sqrt{n}}\zeta_{nt}^{\tau}\tilde{t}$. Setting $\hat{\gamma}_{\tilde{\varkappa}}(j) := \dfrac{n}{T}\sum_{1\le t,t-j\le T}\tilde{\varkappa}_t\tilde{\varkappa}_{t-j}$, and proceeding as in (47) - (48), we have

$$\frac{n}{T^2}\hat{\gamma}_{\tilde{\varkappa}}(0) = \frac{n}{T}\sum_{1\le t\le T}\hat{u}_t^2\left(\frac{\tilde{t}}{T}\right)^2 = \frac{1}{T}\sum_{1\le t\le T}\left(\zeta_{nt}^{\tau}\right)^2\left(\frac{\tilde{t}}{T}\right)^2 \tag{53}$$

$$\to_p 12\mathbb{E}\left\{\left(\zeta_t^0\right)^2\right\} = 12\left[4\sigma_a^2\Gamma_\epsilon(0) + \Gamma_{\epsilon^2}(0)\right], \tag{54}$$

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

$$\begin{aligned}\frac{n}{T^2}\hat{\gamma}_{\tilde{\varkappa}}(j) &= \frac{n}{T}\sum_{1\le t,t-j\le T}\hat{u}_t\hat{u}_{t-j}\frac{\tilde{t}}{T}\left(\frac{\widetilde{t-j}}{T}\right) = \frac{1}{T}\sum_{1\le t,t-j\le T}\zeta_{nt}^{\tau}\zeta_{nt-j}^{\tau}\frac{\tilde{t}}{T}\left(\frac{\widetilde{t-j}}{T}\right)\\ &\to_p 12\mathbb{E}\left(\zeta_t^0\zeta_{t-j}^0\right) = 12\left[4\mathbb{E}a_i^2\mathbb{E}\left(\epsilon_{it}\epsilon_{it-j}\right) + \mathbb{E}\left(\epsilon_{it}^2-\sigma_\epsilon^2\right)\left(\epsilon_{it-j}^2-\sigma_\epsilon^2\right)\right] `\\ &= 12\left[4\sigma_a^2\Gamma_\epsilon(j) + \Gamma_{\epsilon^2}(j)\right].\end{aligned} \tag{55}$$

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

$$\begin{aligned}\frac{n}{T^2}\hat{\Omega}_M^{\mathrm{b}=1} &= \frac{n}{T^2}\sum_{j=1-T}^{T-1} k\left(\frac{j}{T}\right)\hat{\gamma}_{\tilde{\varkappa}}(j) = \sum_{t=1}^{T}\sum_{p=1}^{T} k\left(\frac{t-p}{T}\right)\frac{\zeta_{nt}^{\tau}}{\sqrt{T}}\frac{\zeta_{np}^{\tau}}{\sqrt{T}}\frac{\tilde{t}}{T}\left(\frac{\tilde{p}}{T}\right)\\ &\rightsquigarrow_{(n,T\to\infty)} \int_0^1\int_0^1 k(r-s)\,\tilde{r}\tilde{s}\,d\widetilde{B_{\Omega_\zeta^2}}(r)\,d\widetilde{B_{\Omega_\zeta^2}}(s),\end{aligned}$$

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 $b\in(0,1)$

$$\frac{n}{T^2}\hat{\Omega}_M^{\mathrm{b}} := \frac{n}{T^2}\sum\nolimits_{j=1-T}^{T-1} k_b\left(\frac{j}{T}\right)\hat{\gamma}_{\tilde{\varkappa}}(j) = \frac{n}{T^2}\sum\nolimits_{j=1-T}^{T-1} k\left(\frac{j}{bT}\right)\frac{1}{T}\sum_{1\le t,t-j\le T}\tilde{\varkappa}_t\tilde{\varkappa}_{t-j}$$

$$= \sum\nolimits_{t=1}^{T}\sum\nolimits_{p=1}^{T} k\left(\frac{t-p}{bT}\right)\frac{\zeta_{nt}^{\tau}}{\sqrt{T}}\frac{\zeta_{np}^{\tau}}{\sqrt{T}}\frac{\tilde{t}}{T}\left(\frac{\tilde{p}}{T}\right)$$

$$\rightsquigarrow_{(n,T\to\infty)} \int_0^1\int_0^1 k\left(\frac{r-s}{b}\right)\tilde{r}\tilde{s}dB_{\Omega_\xi^2}^{\tau}(r)\,dB_{\Omega_\xi^2}^{\tau}(s) = 12\left\{\int_0^1\int_0^1 k\left(\frac{r-s}{b}\right)\tilde{r}\tilde{s}dW^{\tau}(r)\,dW^{\tau}(s)\right\}$$

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

$$t_{\mathrm{HAR}_b} \rightsquigarrow_{(n,T\to\infty)} \frac{Z_\phi}{\left\{12\Omega_\xi^2\int_0^1\int_0^1 k\left(\frac{r-s}{b}\right)\tilde{r}\tilde{s}dW^{\tau}(r)\,dW^{\tau}(s)\right\}^{1/2}}$$

$$= \frac{Z}{\left\{\int_0^1\int_0^1 k\left(\frac{r-s}{b}\right)\tilde{r}\tilde{s}dW^{\tau}(r)\,dW^{\tau}(s)\right\}^{1/2}},$$

as stated.

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

$$t_{\mathrm{HAC}} = \frac{\hat{\phi}_{nT}}{\sqrt{\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}T\hat{\Omega}_L^2\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} = \frac{\sqrt{nT^3}\hat{\phi}_{nT}}{\sqrt{\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}nT^{-2}\hat{\Omega}_L^2\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}}$$

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 trunction $L = \lfloor T^{1/3}\rfloor$ is deduced as follows ,

$$\begin{aligned}
\frac{n}{T^2}\hat{\Omega}_L^2 &= \frac{n}{T}\sum\nolimits_{t=1}^{T}\tilde{\varkappa}_t^2 + \frac{2n}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{L+1}\right)\tilde{\varkappa}_t\tilde{\varkappa}_{t+\ell} \\
&= \frac{1}{T}\sum\nolimits_{t=1}^{T}(\zeta_{nt}^{\tau})^2\left(\frac{\tilde{t}}{T}\right)^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{L+1}\right)\zeta_{nt}^{\tau}\,\zeta_{nt+\ell}^{\tau}\frac{\tilde{t}}{T}\left(\frac{\widetilde{t+\ell}}{T}\right) \\
&= \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\frac{\tilde{t}}{T}\right)^2\mathbb{E}\left(\zeta_t^0\right)^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{L+1}\right)\frac{\tilde{t}}{T}\left(\frac{\widetilde{t+\ell}}{T}\right)\mathbb{E}\left(\zeta_t^0\zeta_{t+\ell}^0\right) + o_p(1) \\
&= \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\frac{\tilde{t}}{T}\right)^2\left\{\mathbb{E}\left(\zeta_t^0\right)^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{L+1}\right)\mathbb{E}\left(\zeta_t^0\zeta_{t+\ell}^0\right)\right\} + o_p(1) \\
&\to_p \frac{1}{12}\sum_{j=-\infty}^{\infty}\mathbb{E}\zeta_t^0\zeta_{t+j}^0 = \frac{1}{12}\sum_{j=-\infty}^{\infty}\left\{4\sigma_a^2\Gamma_\epsilon(j) + \Gamma_{\epsilon^2}(j)\right\} = \frac{1}{12}\Omega_\zeta^2 = \frac{1}{12}\Omega_\phi^2.
\end{aligned}$$

Similar results on consistency of the HAC estimator apply with other kernels satisfying Assumption B. It follows that

$$t_{\mathrm{HAC}} = \frac{\sqrt{nT^3}\hat{\phi}_{nT}}{\sqrt{\left(T^{-3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1} nT^{-2}\hat{\Omega}_L^2\left(\frac{1}{T^3}\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}}} \rightsquigarrow_{(n,T\to\infty)} \frac{Z_\phi}{\sqrt{12\Omega_\phi^2}} \equiv Z,$$

giving the stated result.

Proof of Theorem 2

Proof of (25)

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

$$\hat{u}_t = \widetilde{K_{nt}} - \hat{\phi}_{nT}\tilde{t} = \left(\tilde{\eta}_{n,t} - \hat{\phi}_{nT}\tilde{t}\right) + \tilde{\varepsilon}_{nt} =: \tilde{\mathcal{M}}_{nt} + \tilde{\varepsilon}_{nt}.$$

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

$$\begin{aligned}
\hat{\Omega}_L^2 &= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\hat{u}_t^2 + 2\frac{1}{T}\sum_{\ell=1}^{L}\vartheta_{\ell L}\sum_{t=1}^{T-\ell}\tilde{t}\widetilde{(t+\ell)}\hat{u}_t\hat{u}_{t+\ell} \\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\left(\tilde{\mathcal{M}}_{nt}+\tilde{\varepsilon}_{nt}\right)^2 + 2\frac{1}{T}\sum_{\ell=1}^{L}\vartheta_{\ell L}\sum_{t=1}^{T-\ell}\tilde{t}\widetilde{(t+\ell)}\left(\tilde{\mathcal{M}}_{nt}+\tilde{\varepsilon}_{nt}\right)\left(\tilde{\mathcal{M}}_{nt+\ell}+\tilde{\varepsilon}_{nt+\ell}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{\mathcal{M}}_{nt}^2 + 2\frac{1}{T}\sum_{\ell=1}^{L}\vartheta_{\ell L}\sum_{t=1}^{T-\ell}\tilde{t}\widetilde{(t+\ell)}\tilde{\mathcal{M}}_{nt}\tilde{\mathcal{M}}_{nt+\ell} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{\varepsilon}_{nt}^2 + 2\frac{1}{T}\sum_{\ell=1}^{L}\vartheta_{\ell L}\sum_{t=1}^{T-\ell}\tilde{t}\widetilde{(t+\ell)}\tilde{\varepsilon}_{nt}\tilde{\varepsilon}_{nt+\ell} \\
&\quad + 2\frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{\mathcal{M}}_{nt}\tilde{\varepsilon}_{nt} + 2\frac{1}{T}\sum_{\ell=1}^{L}\vartheta_{\ell L}\sum_{t=1}^{T-\ell}\tilde{t}\widetilde{(t+\ell)}\left(\tilde{\mathcal{M}}_{nt}\tilde{\varepsilon}_{nt+\ell}+\tilde{\varepsilon}_{nt}\tilde{\mathcal{M}}_{nt+\ell}\right) \\
&:= \hat{\Omega}_l^2 + \hat{\Omega}_\varepsilon^2 + 2\hat{\Omega}_{l\varepsilon},
\end{aligned}$$

where

$$\hat{\Omega}_l^2 = \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{\mathcal{M}}_{nt}^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}\tilde{\mathcal{M}}_{nt}\tilde{\mathcal{M}}_{nt+\ell}.$$

It has been shown in KPS (2019) that the dominating term in $\hat{\Omega}_L^2$ is $\hat{\Omega}_l^2$. Note that

$$\begin{aligned}
\hat{\Omega}_l^2 &= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{\mathcal{M}}_{nt}^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}\tilde{\mathcal{M}}_{nt}\tilde{\mathcal{M}}_{nt+\ell} \\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\left(\tilde{m}_t + R_{nt}\right)^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}\left(\tilde{m}_t + R_{nt}\right)\left(\tilde{m}_{t+\ell} + R_{nt+\ell}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{m}_t^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}\tilde{m}_t\tilde{m}_{t+\ell} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2 R_{nt}^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}R_{nt}R_{nt+\ell} \\
&\quad + 2\frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{m}_t R_{nt} + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}\left(\tilde{m}_t R_{nt+\ell} + R_{nt}\tilde{m}_{t+\ell}\right).
\end{aligned}$$

Let

$$\Omega_l^2 = \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{m}_t^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\vartheta_{\ell L}\tilde{t}\widetilde{(t+\ell)}\tilde{m}_t\tilde{m}_{t+\ell}.$$

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

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

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\tilde{p}_t^2 &= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\tilde{m}_t^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\left[\widetilde{t^{-\lambda}} - \tilde{t}\left(\sum_{t=1}^{T}\tilde{t}\widetilde{t^{-\lambda}}\right)\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\right]^2 \\
&= \begin{cases}
O\left(T^{2-2\lambda}\right) & \text{if } \lambda < 1 \\
O\left(1\right) & \text{if } \lambda = 1 \\
O\left(T^{2-2\lambda}\right) & \text{if } 1 < \lambda < 3/2 \\
O\left(T^{-1}\ln T\right) & \text{if } \lambda = 3/2 \\
O\left(T^{-1}\right) & \text{if } \lambda > 3/2
\end{cases}
\end{aligned}$$

Next, let

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

and

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

We expand P_M as

$$\begin{aligned}
& \frac{1}{T}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\tilde{p}_t\tilde{p}_{t+\ell} \\
= & \frac{1}{T}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\widetilde{t^{1-\lambda}}\widetilde{(t+\ell)^{1-\lambda}} \\
& -\mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\widetilde{t^{1-\lambda}}\widetilde{(t+\ell)^{2}} \\
& -\mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\widetilde{t^{2}}\widetilde{(t+\ell)^{1-\lambda}} \\
& +(\mathcal{T}_T(1,\lambda))^2\,T^{-7}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\widetilde{(t+\ell)^{2}}\widetilde{t^{2}} \\
:= & \Psi_1^M-\Psi_2^M-\Psi_3^M+\Psi_4^M,
\end{aligned}$$

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

$$\begin{aligned}
\Psi_1^M = & \frac{1}{T}\sum\nolimits_{\ell=1}^{M}\left(1-\frac{\ell}{M+1}\right)\sum\nolimits_{t=1}^{T-\ell}\widetilde{t^{1-\lambda}}\widetilde{(t+\ell)^{1-\lambda}} \\
= & \frac{1}{T}\sum\nolimits_{\ell=1}^{M}\left(1-\frac{\ell}{M+1}\right)\sum\nolimits_{t=1}^{T-\ell}\left(t^2+t\ell\right)^{1-\lambda} \\
& -\frac{1}{T}\sum\nolimits_{\ell=1}^{M}\left(1-\frac{\ell}{M+1}\right)\left(\frac{1}{T-\ell}\sum\nolimits_{t=1}^{T-\ell}t^{1-\lambda}\right)\sum\nolimits_{t=1}^{T-\ell}(t+\ell)^{1-\lambda} \\
:= & \Psi_{11}^M-\Psi_{12}^M,
\end{aligned}$$

$$\begin{aligned}
\Psi_2^M = & \mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\widetilde{t^{1-\lambda}}\,(t+\ell)^2 \\
= & \mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\left(1-\frac{\ell}{M+1}\right)\sum\nolimits_{t=1}^{T-\ell}t^{1-\lambda}\,(t+\ell)^2 \\
& -\mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\left(1-\frac{\ell}{M+1}\right)\frac{1}{T-\ell}\left(\sum\nolimits_{t=1}^{T-\ell}t^{1-\lambda}\right)\sum\nolimits_{t=1}^{T-\ell}(t+\ell)^2 \\
:= & \Psi_{21}^M-\Psi_{22}^M,
\end{aligned}$$

$$
\begin{aligned}
\Psi_3^M &= \mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\widetilde{t^2}\widetilde{(t+\ell)^{1-\lambda}} \\
&= \mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\sum\nolimits_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)t^2\,(t+\ell)^{1-\lambda} \\
&\quad -\mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{M}\left(1-\frac{\ell}{M+1}\right)\frac{1}{T-\ell}\left(\sum\nolimits_{t=1}^{T-\ell}t^2\right)\sum\nolimits_{t=1}^{T-\ell}(t+\ell)^{1-\lambda} \\
&:= \Psi_{31}^M-\Psi_{32}^M.
\end{aligned}
$$

Direct calculation gives the order of each term. Rather than record all the derivations we show here how to get the exact order of Ψ_{21}^L and Ψ_{21}^M. Note that

$$
\begin{aligned}
\Psi_{21}^L &= \mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}t^{1-\lambda}\,(t+\ell)^2 \\
&= \mathcal{T}_T(1,\lambda)\,T^{-4}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}\left(t^{3-\lambda}+t^{1-\lambda}\ell^2+2t^{2-\lambda}\ell\right).
\end{aligned}
$$

Consider each term.

$$
\begin{aligned}
\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}t^{3-\lambda} &= \sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\begin{cases}\frac{1}{4-\lambda}(T-\ell)^{4-\lambda} & \text{if } \lambda<4\\ \ln(T-\ell) & \text{if } \lambda=4\\ \zeta(\lambda-3) & \text{if } \lambda>4\end{cases} \\
&= \begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<4\\ O(L\ln T) & \text{if } \lambda=4\\ O(L) & \text{if } \lambda>4\end{cases},
\end{aligned}
$$

$$
\begin{aligned}
\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}t^{1-\lambda}\ell^2 &= \sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\ell^2\sum\nolimits_{t=1}^{T-\ell}t^{1-\lambda} \\
&= \sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\ell^2\begin{cases}\frac{1}{2-\lambda}(T-\ell)^{2-\lambda} & \text{if } \lambda<2\\ \ln(T-\ell) & \text{if } \lambda=2\\ \zeta(\lambda-1) & \text{if } \lambda>2\end{cases} \\
&= \begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<2\\ O(T^2L\ln T) & \text{if } \lambda=2\\ O(T^2L) & \text{if } \lambda>2\end{cases},
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum_{t=1}^{T-\ell}2t^{2-\lambda}\ell &= 2\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\ell\sum_{t=1}^{T-\ell}t^{2-\lambda} \\
&= 2\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\ell\begin{cases}\frac{1}{3-\lambda}\left(T-\ell\right)^{3-\lambda} & \text{if } \lambda<3\\ \ln\left(T-\ell\right) & \text{if } \lambda=3\\ \zeta\left(\lambda-2\right) & \text{if } \lambda>3\end{cases} \\
&= \begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<3\\ O\left(TL\ln T\right) & \text{if } \lambda=3\\ O\left(TL\right) & \text{if } \lambda>3\end{cases}.
\end{aligned}
$$

Combining all terms yeilds

$$
\begin{aligned}
\Psi_{21}^{L} &= \mathcal{T}_T\left(1,\lambda\right)T^{-4}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum_{t=1}^{T-\ell}t^{1-\lambda}\left(t+\ell\right)^2 \\
&= \mathcal{T}_T\left(1,\lambda\right)T^{-4}\begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<4\\ O\left(L\ln T\right) & \text{if } \lambda=4\\ O\left(L\right) & \text{if } \lambda>4\end{cases} + \mathcal{T}_T\left(1,\lambda\right)T^{-4}\begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<2\\ O\left(T^2L\ln T\right) & \text{if } \lambda=2\\ O\left(T^2L\right) & \text{if } \lambda>2\end{cases} \\
&\quad + \mathcal{T}_T\left(1,\lambda\right)T^{-4}\begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<3\\ O\left(TL\ln T\right) & \text{if } \lambda=3\\ O\left(TL\right) & \text{if } \lambda>3\end{cases} \\
&= \mathcal{T}_T\left(1,\lambda\right)T^{-4}\begin{cases}O\left(T^{4-\lambda}L\right) & \text{if } \lambda<2\\ O\left(T^2L\ln T\right) & \text{if } \lambda=2\\ O\left(T^2L\right) & \text{if } \lambda>2\end{cases}.
\end{aligned}
$$

Next, replacing L by T^{κ} leads to

$$
\begin{aligned}
\Psi_{21}^{L} &= \mathcal{T}_T\left(1,\lambda\right)\begin{cases}O\left(T^{-\lambda+\kappa}\right) & \text{if } \lambda<2\\ O\left(T^{-2+\kappa}\ln T\right) & \text{if } \lambda=2\\ O\left(T^{-2+\kappa}\right) & \text{if } \lambda>2\end{cases} \\
&= \begin{cases}T^{2-\lambda} & \text{if } \lambda<1\\ T\ln T & \text{if } \lambda=1\\ \zeta\left(\lambda\right)T & \text{if } \lambda>1\end{cases} \times \begin{cases}O\left(T^{-\lambda+\kappa}\right) & \text{if } \lambda<2\\ O\left(T^{-2+\kappa}\ln T\right) & \text{if } \lambda=2\\ O\left(T^{-2+\kappa}\right) & \text{if } \lambda>2\end{cases}
\end{aligned}
$$

$$= \begin{cases} O\left(T^{2-2\lambda+\kappa}\right) & \text{if } \lambda < 1 \\ O\left(T^{\kappa}\ln T\right) & \text{if } \lambda = 1 \\ O\left(T^{1-\lambda+\kappa}\right) & \text{if } 1 < \lambda < 2 \\ O\left(T^{-1+\kappa}\ln T\right) & \text{if } \lambda = 2 \\ O\left(T^{-1+\kappa}\right) & \text{if } \lambda > 2 \end{cases}.$$

To calculate the order of Ψ_{21}^M, we replace M by bT. That is,

$$\begin{aligned}
\Psi_{21}^M &= \mathcal{T}_T(1,\lambda)\,T^{-4}\begin{cases} O\left(T^{4-\lambda}M\right) & \text{if } \lambda < 2 \\ O\left(T^{2}M\ln T\right) & \text{if } \lambda = 2 \\ O\left(T^{2}M\right) & \text{if } \lambda > 2 \end{cases} \\
&= \mathcal{T}_T(1,\lambda)\begin{cases} O\left(T^{5-\lambda}\right) & \text{if } \lambda < 2 \\ O\left(T^{3}\ln T\right) & \text{if } \lambda = 2 \\ O\left(T^{3}\right) & \text{if } \lambda > 2 \end{cases} \\
&= \begin{cases} T^{2-\lambda} & \text{if } \lambda < 1 \\ T\ln T & \text{if } \lambda = 1 \\ \zeta(\lambda)\,T & \text{if } \lambda > 1 \end{cases} \times \begin{cases} O\left(T^{5-\lambda}\right) & \text{if } \lambda < 2 \\ O\left(T^{3}\ln T\right) & \text{if } \lambda = 2 \\ O\left(T^{3}\right) & \text{if } \lambda > 2 \end{cases} \\
&= \begin{cases} O\left(T^{3-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{2-\lambda}\ln T\right) & \text{if } \lambda = 1 \\ O\left(T^{2-\lambda}\right) & \text{if } 1 < \lambda < 2 \\ O\left(\ln T\right) & \text{if } \lambda = 2 \\ O(1) & \text{if } \lambda > 2 \end{cases}.
\end{aligned}$$

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

$$\Psi_{11}^M = \begin{cases} O\left(T^{3-2\lambda}\right) & \text{if } \lambda < 3/2 \\ O\left(\ln T\right) & \text{if } \lambda = 3/2 \\ O(1) & \text{if } \lambda > 3/2 \end{cases} \qquad \Psi_{11}^L = \begin{cases} O\left(T^{2-2\lambda+\kappa}\right) & \text{if } \lambda < 3/2 \\ O\left(T^{-1+\kappa}\ln T\right) & \text{if } \lambda = 3/2 \\ O\left(T^{-1+\kappa}\right) & \text{if } \lambda > 3/2 \end{cases}$$

$$\Psi_{12}^M = \begin{cases} O\left(T^{3-2\lambda}\right) & \text{if } \lambda < 2 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 2 \\ O\left(T^{-1}\right) & \text{if } \lambda > 2 \end{cases} \qquad \Psi_{12}^L = \begin{cases} O\left(T^{2-2\lambda+\kappa}\right) & \text{if } \lambda < 2 \\ O\left(T^{-2+\kappa}\ln^2 T\right) & \text{if } \lambda = 2 \\ O\left(T^{-2+\kappa}\right) & \text{if } \lambda > 2 \end{cases}$$

Hence

$$\Psi_1^M = \begin{cases} O\left(T^{3-2\lambda}\right) & \text{if } \lambda < 3/2 \\ O\left(\ln T\right) & \text{if } \lambda = 3/2 \\ O(1) & \text{if } \lambda > 3/2 \end{cases}, \text{ and } \Psi_1^L = \begin{cases} O\left(T^{2-2\lambda+\kappa}\right) & \text{if } \lambda < 3/2 \\ O\left(T^{-1+\kappa}\ln T\right) & \text{if } \lambda = 3/2 \\ O\left(T^{-1+\kappa}\right) & \text{if } \lambda > 3/2 \end{cases}$$

Next

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

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

Combining these two leads to

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

The third term becomes

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

$$\Psi_{32}^M = \begin{cases} O\left(T^{3-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{2-\lambda}\ln T\right) & \text{if } \lambda = 1 \\ O\left(T^{2-\lambda}\right) & \text{if } 1 < \lambda < 2 \\ O\left(\ln T\right) & \text{if } \lambda = 2 \\ O\left(1\right) & \text{if } \lambda > 2 \end{cases} \quad \Psi_{32}^L = \begin{cases} O\left(T^{2-2\lambda+\kappa}\right) & \text{if } \lambda < 1 \\ O\left(T^{1-\lambda+\kappa}\ln T\right) & \text{if } \lambda = 1 \\ O\left(T^{1-\lambda+\kappa}\right) & \text{if } 1 < \lambda < 2 \\ O\left(T^{-1+\kappa}\ln T\right) & \text{if } \lambda = 2 \\ O\left(T^{-1+\kappa}\right) & \text{if } \lambda > 2 \end{cases}$$

Combining these two terms yeilds

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

Last, the fourth term becomes

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

After combining all terms, we have

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

Finally,

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

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

Therefore we have

$$t_{\text{HAR}} \sim_a \begin{cases} O_p\left(\frac{T^{-1-\lambda}}{T^{1/2}T^{3/2-\lambda}}T^3\right) = O_p\left(1\right) & \text{if } \lambda < 1 \\ O_p\left(\frac{T^{-2}\ln T}{T^{1/2}T^{1/2}\ln T}T^3\right) = O_p\left(1\right) & \text{if } \lambda = 1 \\ O_p\left(\frac{T^{-2}}{T^{1/2}T^{1/2}}T^3\right) = O_p\left(1\right) & \text{if } \lambda > 1 \end{cases},$$

$$t_{\text{HAC}} \sim_a \begin{cases} O_p\left(\frac{T^{-1-\lambda}}{T^{1/2}T^{1-\lambda+\kappa/2}}T^3\right) = O_p\left(T^{(1-\kappa)/2}\right) & \text{if } \lambda < 1 \\ O_p\left(\frac{T^{-2}\ln T}{T^{1/2}\left(T^{\kappa}\ln^2 T\right)^{1/2}}T^3\right) = O_p\left(T^{(1-\kappa)/2}\right) & \text{if } \lambda = 1 \\ O_p\left(\frac{T^{-2}}{T^{1/2}T^{\kappa/2}}T^3\right) = O_p\left(T^{(1-\kappa)/2}\right) & \text{if } \lambda > 1 \end{cases}$$

using representations in terms of the dominant deterministic order.

Proof of (24)

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

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

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

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

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

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

This difference resultd in the difference in the order of Ψ_{11}^{M} and Ψ_{11}^{L}.

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

$$\Psi_{11}^M = \begin{cases} O\left(T^{-2\lambda+1}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases} \qquad \Psi_{12}^M = \begin{cases} O\left(T^{1-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases}$$

$$\Psi_{21}^M = \begin{cases} O\left(T^{1-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases} \qquad \Psi_{22}^M = \begin{cases} O\left(T^{1-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases}$$

$$\Psi_{31}^M = \begin{cases} O\left(T^{1-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases} \qquad \Psi_{32}^M = \begin{cases} O\left(T^{1-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases}$$

$$\Psi_{4}^M = \begin{cases} O\left(T^{1-2\lambda}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases}$$

Hence,

$$\begin{aligned} P_M(T,\lambda) &= \Psi_1 - \Psi_2 - \Psi_3 + \Psi_4 \\ &\sim_a \begin{cases} O\left(T^{-2\lambda+1}\right) & \text{if } \lambda < 1 \\ O\left(T^{-1}\ln^2 T\right) & \text{if } \lambda = 1 \\ O\left(T^{-1}\right) & \text{if } \lambda > 1 \end{cases} \end{aligned}$$

Then

$$\begin{aligned} t_2 &\sim_a \frac{\hat{\phi}_{nT}\left(\sum \tilde{t}^2\right)^{1/2}}{\sqrt{\Omega_m^2}} \\ &= \begin{cases} O_p\left(\frac{T^{-1-\lambda}}{T^{1/2-\lambda}}T^{3/2}\right) = O_p(1) & \text{if } \lambda < 1 \\ O_p\left(\frac{T^{-2}\ln T}{T^{-1/2}\ln T}T^{3/2}\right) = O_p(1) & \text{if } \lambda = 1 \\ O_p\left(\frac{T^{-2}}{T^{-1/2}}T^{3/2}\right) = O_p(1) & \text{if } \lambda > 1 \end{cases} \end{aligned}$$

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