4-2019

Weak σ-convergence: Theory and applications

Jianning KONG
Shandong University

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

Donggyu SUL
University of Texas at Dallas

Citation

KONG, Jianning; PHILLIPS, Peter C. B.; and SUL, Donggyu. Weak σ-convergence: Theory and applications. (2019). *Journal of Econometrics*. 209, (2), 185-207.
Available at: https://ink.library.smu.edu.sg/soe_research/2284

Published in Journal of Econometrics, Volume 209, Issue 2, April 2019, Pages 185-207.
Accepted November 2018
https://doi.org/10.1016/j.jeconom.2018.12.022

Weak σ- Convergence: Theory and Applications*

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

January 13, 2017

Abstract

The concept of relative convergence, which requires the ratio of two time series to converge to unity in the long run, explains convergent behavior when series share commonly divergent stochastic or deterministic trend components. Relative convergence of this type does not necessarily hold when series share common time decay patterns measured by evaporating rather than divergent trend behavior. To capture convergent behavior in panel data that do not involve stochastic or divergent deterministic trends, we introduce the notion of weak σ-convergence, whereby cross section variation in the panel decreases over time. The paper formalizes this concept and proposes a simple-to-implement linear trend regression test of the null of no σ-convergence. Asymptotic properties for the test are developed under general regularity conditions and various data generating processes. Simulations show that the test has good size control and discriminatory power. The method is applied to examine whether the idiosyncratic components of 90 disaggregate personal consumption expenditure (PCE) price index items σ-converge over time. We find strong evidence of weak σ-convergence in the period after 1992, which implies that cross sectional dependence has strenthened over the last two decades. In a second application, the method is used to test whether experimental data in ultimatum games converge over successive rounds, again finding evidence in favor of weak σ-convergence. A third application studies convergence and divergence in US States unemployment data over the period 2001-2016.

Keywords: Asymptotics under misspecified trend regression, Cross section dependence, Evaporating trend, Relative convergence, Trend regression, Weak $\sigma-$convergence.

JEL Classification: C33

*Phillips acknowledges NSF support under Grant No. SES-1285258.
[†]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

"The real test of a tendency to convergence would be in showing a consistent diminution of variance", Hotelling (1933) cited in Friedman (1992)

1 Introduction

The notion of convergence is a prominent element in many branches of economic analysis. In macroeconomics and financial economics, for instance, the influence of transitory (as distinct from persistent) shocks on an equilibrium system diminishes over time. The effects of such shocks is ultimately eliminated when the system is stable, absorbs their impact, and restores an equilibrium position. In microeconomics, particularly in experiments involving economic behavior, heterogeneous subject outcomes may be expected under certain conditions to converge to some point (or a set of points) or to diverge when those conditions fail. The object of much research in experimental economics is to determine by econometric analysis whether or not predictions from game theory, finance, or micro theory hold up in experimental data. While the general idea of convergence in economic behavior is well-understood in broad terms in economics, empirical analysis requires more specific formulation and embodiment of the concept of convergence over time to facilitate econometric testing.

The idea of cointegration as it developed in the 1980s for studying co-movement among nonstationary trending time series bears an important general relationship to convergence. Cointegrated series match one another in the sense that over the long run some linear relationship of them is a stationary rather than a nonstationary time series. But while the cointegration concept has proved extraordinarily useful in practical time series work, cointegration itself does not explain trends in the component variables. These are embodied implicitly in the system's unit roots and deterministic drifts.

The empirical task of determining convergence among time series has moved in a distinct direction from the theory and application of cointegration in the last two decades. Convergence studies flourished particularly in cross country economic growth analyses during the 1990s when economists became focused on long run behavioral comparisons of variables such as real per capita GDP across countries and the potential existence of growth convergence clubs where countries might be grouped according to the long run characteristics of their GDP or consumption behavior. This research led to several new concepts, including 'conditional convergence' and 'absolute convergence' as well as specific measures such as σ(sigma)-convergence for evaluating convergence characteristics in practical work – see Barro (1991), Barro and Sala-I-Martin (1991, 1992), Evans (1996, 1998), and the overview by Durlauf and Quah (1999), among many others in what is now a large literature.

The σ-convergence concept measures gaps among time series by examining whether cross sectional variation decreases over time, as would be anticipated if two series converge. Conditional convergence interestingly requires divergence among the growth rates to ensure catch up and convergence in levels. Thus, for poor countries to catch up with rich countries, poor countries need to grow faster than rich countries. Econometric detection of convergence therefore has to deal with this subtlety in the data. To address this difficulty Phillips and Sul (2007, hereafter PS) used the concept of 'relative convergence' and developed a simple econometric regression test to assess this mode of convergence. Two series converge relatively over time when the time series share the same stochastic or deterministic trend elements in the long run, so that the ratio of the two series eventually converge to unity.

The PS regression trend regression test for convergence has been popular in applications. But neither conditional nor relative convergence concepts are well suited to characterize convergence among time series that do not have (common) divergent deterministic or stochastic trend elements such as polynomial time trends or integrated time series. Instead, many economic time series, especially after differencing (such as growth rates), do not display evidence of deterministic growth or the random wandering behavior that is the primary characteristic of integrated data. In addition, much laboratory experimental data are non-integrated by virtue of their construction in terms of bounded responses, and much macro data during the so-called Great Moderation from the mid 1980s show less evidence of persistent trend behavior. Researchers interested in empirical convergence properties of such times series need an alternative approach that accommodates panels of asymptotically stationary or weakly dependent series, where the concept of convergence involves an explicit time decay function that may be common across series in the panel.

The present paper seeks to address that need by working directly with convergence issues in a panel of non-divergent trending time series and by developing an empirical test for convergence that is suited to such panels. Interestingly, the original concept of $\sigma-$convergence that is based on cross section sample variation is suitable for analyzing such panels for convergence properties in the data and our work builds on this concept by developing a simple regression test procedure. The main contributions of the paper are fourfold: (i) we introduce a concept of weak $\sigma-$convergence whereby cross section variation in the panel decreases over time; (ii) we propose a simple linear trend regression test to assess evidence for such convergence; (iii) we develop an asymptotic theory for inference with this test in practical work; and (iv) we provide empirical applications of the new procedure to personal consumption expenditure price index data, to US States unemployment data, and to experimental data involving ultimatum games.

There are two major differences between the approach used in PS, which is based on the so-called $\log t$ regression, and the trend decay regressions advocated in the present paper for asymptotically weakly dependent data. First, the $\log t$ regression approach uses sample cross sectional variation in the relative transition curves and a logarithmic trend regression for detection of convergence. By contrast, the method proposed here uses linear trend regression to detect trend decay in the sample cross section variation after the elimination of common components. This objective matches precisely the 'real test' of 'showing a consistent diminution of variance' suggested originally by Hotelling (1933) and cited in the header of this article. One of the advantages of linear trend regression in addition to its obvious simplicity in practice is that the sign of the fitted slope coefficient captures trend decay even though the regression is misspecified.

Second, the asymptotic properties of the two procedures are very different. Trend regression is used in the present paper as a detective device in an intentionally misspecified regression so that test outcomes signal convergence or divergence of cross section averages over time by virtue of the sign behavior of the trend slope coefficient and its associated t-test statistic. This behavior in turn reflects the nature of the dominant trend or trend decay that is present in the data. The asymptotic properties of these misspecified trend regression statistics are of some independent interest, but it is their effectiveness in detecting trend decay convergence that is the primary focus of the present paper.

The remainder of the paper is organized as follows. The next section provides a non-technical introduction to convergence testing and begins with an empirical example to motivate the introduction of a new concept of weak $\sigma-$convergence that accords with the notion suggested by Hotelling (1933) in the header. The section briefly reviews existing tests for convergence, explains the need for a new concept of the Hotelling type that is useful in economic, social and experimental applications, and provides the simple linear trend regression mechanism that is proposed in this paper for testing convergence. Section 3 provides a formal development of the concept of weak $\sigma-$convergence, discusses various matters of formulation and interpretation in the context of several useful prototypical decay function models of convergence, and introduces the linear trend regression approach and associated t-ratio test of convergence. Section 4 derives asymptotic theory for the proposed test under null and alternative hypotheses (of both convergence and divergence). Several new results on power function trend regression asymptotics are obtained in these derivations, which may well be of wider interest. Section 5 reports some numerical calculations to demonstrate the contrasting test behavior under these two alternatives. Section 6 reports the results of Monte Carlo simulations to assess the finite sample performance of the test procedure. Section 7 illustrates the use of the new test in two additional empirical applications. Section 8 concludes. Technical derivations and proofs are in the Appendix. Supplementary materials (intended for online reference) that include the proofs of supporting lemmas and further numerical calculations and simulations are given in Appendix S.

2 Empirical Motivation and Modeling Preliminaries

2.1 Divergence and Convergence in US State Unemployment rates

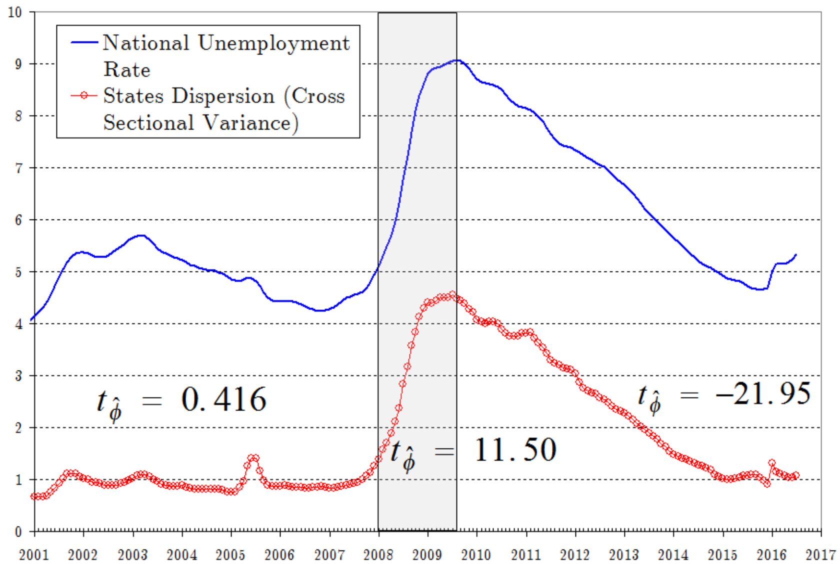
We begin with a motivating example. Figure 1 (upper panel) shows national unemployment rate data for the US over 2001:M1 to 2016:M7. The figure also plots the monthly sample cross section variance of unemployment rates in the 48 contiguous US States. The data are obtained from the Bureau of Labor Statistics.

The focus of economic interest concerns the behavior of State unemployment rates over the whole period and certain subperiods, particularly those preceding and following the subprime mortgage crisis. The periods prior to, during, and following the subprime mortgage crisis are of special interest because of the onset and impact of the great recession coupled with the distinct time series behavior in unemployment rates in these subperiods.

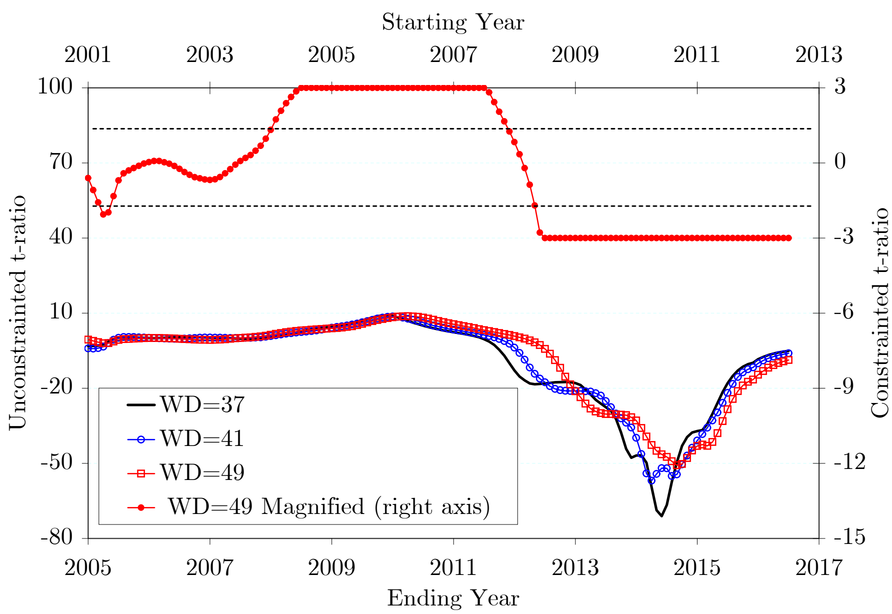
Evidently, the temporal patterns of the national unemployment rate and the cross section variation of State unemployment rates show some stability over 2001-2007. Both rise sharply during the crisis, and both fall steadily in the crisis aftermath. These patterns suggest a period of stationary fluctuations in unemployment rates, followed by divergence during the crisis, followed then by a steady decline in variation with convergence to pre-crisis levels. The tests we develop provide a quantitative analysis to buttress this descriptive commentary on the divergence and convergence of unemployment rates over this 15 year period.

The official period of the recession precipitated by the subprime mortgage crisis is December 2007 to June 2009 (the gray-shaded area in the figure). Over this period, cross section variation in State unemployment rates rose rapidly from a range of 4.6% (high: Michigan 7.3%; low: South Dakota 2.7%) in December 2007 to more than twice that figure reaching 10.7% in June 2009 (high: Michigan 14.9%; low: Nevada 4.2%). Almost immediately following the recession, cross section variation in unemployment rates started to decline and continued to do so until the national unemployment rate reached pre-crisis levels.

The methodology we propose to test convergence and divergence in the present paper involves the use of a simple linear trend regression with cross section variation as the dependent variable. The idea is related to Hotelling's (1933) original suggestion. Evans (1996) pursued a related idea in the context of economic growth using a linear trend regression to distinguish divergence from stationarity. Our methodology also uses a linear trend regression but explicitly allows for misspecification in that regression in making inferences and also permits a focus on convergence in cross section variances as well as divergence, thereby providing econometric methods to meet Hotelling's idea of a 'real test of a tendency to convergence'.



Panel A: Variance of US Unemployment Rates (with t-ratio convergence tests for the pre-, mid- and post- crisis periods)



Panel B: Rolling window recursive t-ratio statistics for various window widths (showing magnified values constrained to the interval [-3,3])

Figure 1: Impact of the Subprime Mortgage Crisis on Unemployment rates across 48 contiguous United States

Allowance for misspecification is important in applications. For instance, in the present example of unemployment rates, the mechanism generating the data is a complicated dynamic system that is further complicated by cross section averaging over the heterogeneous contiguous States of the USA. This mechanism is inexorably more complex than a simple linear trend regression, even allowing for segmentation of the regression over different subperiods of data such as the pre- and post- crisis periods. Such empirical formulations are therefore inevitably misspecified. But they are nonetheless useful and, indeed, commonly appear in empirical work allowing for both exogenously determined and endogenously determined break points. Our methodology is designed specifically to allow for such misspecification in the regression and we develop a limit theory for the usual t-ratio test statistic for the linear trend coefficient ($\hat{\phi}$, in later notation) in these misspecified regressions that distinguishes behavior in such subperiods and enables inference.

As we will show, the limit theory enables us to differentiate between periods of divergence, periods of convergence, and periods of stable fluctuations, based on signs and critical values from a standard normal distribution. The methodology is very easy to apply and involves a standard linear time trend regression of the form shown in equation (28) below and the computation of a heteroskedastic and autocorrelation robust t-ratio statistic $t_{\hat{\phi}}$.

For the empirical example of cross section variation in US States unemployment rates, the results of these tests are displayed in Figure 1. The top panel of the figure reports the t-ratio statistics $t_{\hat{\phi}} = 0.416, 11.50, -21.95$ for the pre-crisis, mid-crisis and post-crisis periods, exogenously determined according to the official period of the recession shown in the shaded region. As explained below, t statistics outside the standard normal critical values signal variation divergence in the right tail, variation convergence in the left tail, and stable variation within standard $\mathcal{N}(0,1)$ critical values centred on the origin. Even with the relatively short time series trajectories available in the three subperiods, the empirical results strongly confirm the heuristic visual evidence in the data trajectories of a rapid divergence from a stable period to 2007, followed by a steady decline in variation after mid 2009.

The lower Panel B of Figure 1 provides plots of recursive calculations of the same robust t ratio statistics computed from linear trend regressions with various rolling window (WD) widths. Three cases are shown in the figure, corresponding to $37, 41$, and 49 month rolling window widths. The starting date of the window (when $WD = 49$) is detailed in the upper horizontal axis and the end date is located on the lower horizontal axis.

As the rolling window width increases, the t-ratio recursion pattern becomes smoother and the absolute value of the t-ratio also tends to decrease. To magnify the scale of the recursive plot, the upper pane of Panel B shows the t-ratio recursion for $WD = 49$, constraining realized values to the interval $[-3, 3]$. The upper and lower 5% critical values of ± 1.65 appear as dotted lines in the figure on the right hand axis scale. These recursive tests enable the data to determine break dates where stability changes to divergence (February, 2008) and subsequently to convergence (May 2012) in terms of first crossing times of the critical values (c.f., Phillips, Wang, Yu, 2011; Phillips, Shi, Yu, 2015). Evidently, the recursive regression tests lead to broadly similar conclusions to those in which the break dates are given exogenously by the official dates of the recession, although the endogenously determined dates delay both the onset of the crisis impact on the divergence of unemployment rate variation and the onset of the decline in variation and convergence.

Two further empirical applications of our methods are given later in the paper in Section 7. That section also includes detailed procedures for implementation. Stata and Gauss codes are available at www.utdallas.edu/~d.sul/papers/weak_pro.zip.

2.2 Existing Tests and Weak $\sigma-$Convergence

As discussed in the Introduction, various notions of convergence were developed in the economic growth literature during the 1990s with multiple concepts suited to various empirical applications, including $\beta-$convergence and $\sigma-$convergence (for an overview, see Barro and Sala-i-Martin, 1991). These methods have been widely applied. Tests for $\beta-$convergence typically rely on cross section linear least squares regressions of the type

$$T^{-1}\sum\nolimits_{t=1}^{T}\left(y_{it}-y_{i0}\right)=\hat{a}+\hat{\beta}y_{i0}+z_i'\hat{\gamma}+\text{error}_i,\quad i=1,...,n, \tag{1}$$

of average historical growth rates $T^{-1}\sum_{t=1}^{T}\left(y_{it}-y_{i0}\right)$ on initial observations y_{i0} of log level real incomes and covariates z_i that control for idiosyncratic factors affecting growth in country i. These regressions test for a significantly negative slope coefficient $\hat{\beta}$ in the fitted regression. Significance in this coefficient suggests that countries with higher initial incomes have lower average growth rates facilitating catch-up by less developed economies with lower initial incomes. However, when $\hat{\gamma}$ differs significantly from zero, the limiting outcomes for countries i and j may differ. Evans (1996) explained why growth regressions like (1) provide valid guidance regarding convergence only under strict conditions.

To address these complications and provide a more general framework for analysis, Phillips and Sul (2007) formulated a nonlinear panel model of the form

$$y_{it}=b_{it}\theta_t,\quad \text{for } t=1,...,T; i=1,...,n, \tag{2}$$

where b_{it} is the ith individual slope coefficient at time t, which may be interpreted as a time varying loading coefficient attached to a common trend function θ_t, which may involve deterministic and stochastic trends. Individual countries share in the common trend driver θ_t to a greater or lesser extent over time depending on the loading coefficient b_{it}. This formulation accommodates many different generating mechanisms and allows for a convenient 'relative convergence' concept, which is defined as

$$\text{plim}_{t\to\infty}\frac{y_{it}}{y_{jt}}=1 \text{ for any } i\neq j. \tag{3}$$

The relative convergence condition may be tested using an empirical least squares regression of the following form involving a '$\ln t$' regressor

$$\ln\left(H_1/H_t\right)-2\ln\ln t=\hat{a}+\hat{\gamma}\ln t+\hat{u}_t, \tag{4}$$

where $H_t=n^{-1}\sum_{i=1}^{n}\left(h_{it}-1\right)^2$ and $h_{it}=y_{it}/(n^{-1}\sum_{i=1}^{n}y_{it})$ is the relative income of country i. If the estimate $\hat{\gamma}$ is significantly positive, then this 'log$t$ test' provides evidence supporting relative convergence. The test is primarily useful in contexts where the panel data involve stochastic and deterministic trends such as θ_t that may originate in common technological, educational, multinational, and trade-related drivers of growth. When the data do not involve such trends as θ_t, then the concept of relative convergence in (3) is far less useful.

In such cases, a more appropriate notion is $\sigma-$convergence, a concept that is defined in terms of declining cross sectional dispersion over time. This concept was originally suggested by Hotelling (1933), as indicated in the quotation that heads this article. It is naturally appealing in many contexts, such as the US States unemployment rate example just studied where there is a direct focus on cross section variation and its behavior over time. At present, however, there is no convenient and statistically rigorous test or asymptotic theory available for inference concerning

$\sigma-$convergence. Evans (1996) used cross sectional variance primarily to test divergence, and Evans and Karras (1996), and Hobijn and Franses (1999) tested $\sigma-$convergence by considering differences between dyadic pairs of y_{it} rather than cross section variance or standard deviation.

The $\sigma-$convergence concept is not applicable when panel observations involve stochastic or deterministic trends. Consider, for instance, the simple panel model

$$y_{it} = a_i + \left(b + \varepsilon_i t^{-1/2}\right) t + \epsilon_{it} t^{-\beta}$$

where $\epsilon_{it} \sim iid\left(0, \sigma_\epsilon^2\right)$ over (i, t), $\varepsilon_i \sim iid\left(0, \sigma_\varepsilon^2\right)$, and the components $(a_i, \varepsilon_i, \epsilon_{it})$ are all independent. It is easy to see that relative convergence holds but not $\sigma-$convergence. In particular, taking dyadic pairs (y_{it}, y_{jt}) we have

$$\text{plim}_{t\to\infty} \frac{y_{it}}{y_{jt}} = \text{plim}_{t\to\infty} \frac{b + \varepsilon_i t^{-1/2}}{b + \varepsilon_j t^{-1/2}} = 1,$$

but when considering cross section variances, such as $K_{nt}^y = n^{-1} \sum_{i=1}^n (y_{it} - n^{-1} \sum_{i=1}^n y_{it})^2$, we have

$$\text{plim}_{n\to\infty} K_{nt}^y = \sigma_a^2 + \sigma_\varepsilon^2 t + \sigma_\epsilon^2 t^{-2\beta},$$

where σ_a^2 is the variance of a_i. When $\sigma_\varepsilon^2 > 0$, the cross sectional dispersion of y_{it} increases over time. Thus, $\varepsilon_i \neq \varepsilon_j$ for some $i \neq j$ is sufficient to prevent $\sigma-$convergence.

A formal test of $\sigma-$convergence requires a well-defined concept and econometric machinery for inference. Quah (1996) defined $\sigma-$convergence in terms of the cross section variance K_{nt}^y by the condition

$$K_{nt}^y \leq K_{nt-1}^y \text{ for all } t. \tag{5}$$

Evidently, the definition (5) partly accords with Hotelling's suggestion but does not require 'consistent diminution in variance'. Moreover, the temporal monotonicity of (5) is restrictive in most applications because it does not allow for subperiod fluctuation or short-period temporal divergence. In place of (5), our approach introduces a weaker condition that focuses on the asymptotic behavior of the sample covariance

$$\widehat{Cov}\left(K_{nt}^y, t\right) = T^{-1} \sum\nolimits_{t=1}^T \tilde{K}_{nt}^y \tilde{t} < 0, \tag{6}$$

where $\tilde{K}_{nt}^y = K_{nt}^y - T^{-1} \sum_{t=1}^T K_{nt}^y$, and $\tilde{t} = t - T^{-1} \sum_{t=1}^T t$. Under suitable standardization, as we show below, the asymptotic behavior of the sample covariance $\widehat{Cov}\left(K_{nt}^y, t\right)$ may be used to define a concept of 'weak $\sigma-$convergence' which is less restrictive than monotonicity and which is amenable to inference. Section 3 of the paper provides a formal definition of this concept. Importantly, the condition (6) does not imply that cross sectional dispersion tends to zero eventually and it allows for local temporal divergences and subperiod fluctuation. However, the concept captures the notion that cross section dispersion shows a tendency to decrease over time and in doing so may be characterized as displaying *weak $\sigma-$convergence.*

2.3 Testing for Weak $\sigma-$Convergence

The simple idea involved in testing weak $\sigma-$convergence is to assess by a trend regression whether cross section dispersion declines over time. Since the mechanism of decline is not formulated in an explicit data generating process, the test is performed via a linear time trend regression of the form

$$K_{nt}^y = \hat{a}_{nT} + \hat{\phi}_{nT} t + \hat{u}_t, \tag{7}$$

and a simple robust t ratio test of whether the fitted coefficient $\hat{\phi}_{nT}$ is significantly less than zero. The technical difficulty in justifying such a test involves allowance for misspecification in the regression (7), which is formulated as an empirical regression not an underlying data generating process for the sample cross section variances $K_{nt}^y = n^{-1}\sum_{i=1}^n (y_{it} - n^{-1}\sum_{i=1}^n y_{it})^2$. The latter process is inevitably complex because it involves both the generation of the individual components y_{it} and the construction of the sample average K_{nt}^y. For the purposes of testing whether cross section dispersion declines over time, a suitably general time decay specification of the type

$$K_{nt}^y = a_n + b_n t^{-\lambda} + v_{n,t} \geq 0 \text{ for all } t, \tag{8}$$

may be employed, wherein the component coefficients a_n, b_n, and $v_{n,t}$ embody the import of the individual data generating processes of the y_{it} and the cross section averaging involved in K_{nt}^y. Under stability conditions as $n \to \infty$ that ensure $\text{plim}_{n\to\infty}(a_n, b_n) = (a, b) > 0$ and $v_{n,t}$ is asymptotically stationary, then cross section dispersion K_{nt}^y will decline over time when the 'convergence parameter' $\lambda > 0$. On the other hand, when $\lambda < 0$, K_{nt}^y increases over time indicating that some of the component time series y_{it} have nonstationary or heterogeneously divergent trend characteristics.[1] When $\lambda > 0$, then cross section dispersion decreases over time and ultimately fluctuates about the level $a > 0$, and weak $\sigma-$convergence holds. When $\lambda = 0$, there is no apparent tendency to diverge or converge and K_{nt}^y fluctuates in an asymptotically stationary manner about the level a.

Using the empirical regression (7), a robust t-ratio test on the fitted coefficient $\hat{\phi}_{nT}$ can be used to assess evidence for weak $\sigma-$convergence, divergence or asymptotic stability in K_{nt}^y. This statistic takes the standard form for the time trend regressor case, viz.,

$$t_{\hat{\phi}_{nT}} = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_u^2 / \sum_{t=1}^T \tilde{t}^2}}, \tag{9}$$

where $\hat{\Omega}_u^2$ is a typical long run variance estimate based on the residuals $\hat{u}_t = K_{nt}^y - \hat{a}_{nT} - \hat{\phi}_{nT} t$ from (7), such as the Bartlett-Newey-West (BNW) estimate

$$\hat{\Omega}_u^2 = \frac{1}{T}\sum_{t=1}^T \hat{u}_t^2 + 2\frac{1}{T}\sum_{\ell=1}^L \vartheta_{\ell L} \sum_{t=1}^{T-\ell} \hat{u}_t \hat{u}_{t+\ell}, \tag{10}$$

where $\vartheta_{\ell L}$ are the Bartlett lag kernel weights and the lag truncation parameter $L = \lfloor T^\kappa \rfloor$ for some small $\kappa > 0$.

The test based on the t-ratio (9) is straightforward to implement using the empirical regression (7). As shown in Theorem 2, one sided critical values from the standard normal distribution $\mathcal{N}(0,1)$ are used in testing to detect convergence ($t_{\hat{\phi}_{nT}}$ significantly negative) and divergence ($t_{\hat{\phi}_{nT}}$ significantly positive) from the null of fluctuating variation. Validity conditions for the test are discussed in detail in the following sections. Essentially, these require that the number of cross sectional units (n) be larger than the number of time series observations (T). In particular, if $\frac{n}{T} \to \infty$ as $n, T \to \infty$, the t-ratio diverges to negative infinity under weak $\sigma-$convergence, leading to a consistent test. Under divergence, the t-ratio diverges to positive infinity as $n, T \to \infty$ without any restriction on the n/T ratio.

An interesting case occurs when the rate of decline in the variation K_{nt}^y is extremely fast. In this case, the second term in (8) converges to zero rapidly because $\lambda > 0$ is large. In such situations, the

[1] If $\text{plim}_{n\to\infty} b_n = b < 0$, $a - b > 0$ and $\lambda > 0$, then K_{nt} increases over time but stabilizes to fluctuate about some constant level a as $n, T \to \infty$, which is a form of asymptotic stability over time. See Section 3.1 for more discussion.

cross sectional dispersion decreases fast initially and soon stabilizes to fluctuations around some long run level because of the rapid convergence. Such cases of convergence are typically much harder to detect because of the inevitable small sample effects of the rapid trend decay in the regressor and potential failure of the persistent excitation condition[2] in asymptotics. Nonetheless and somewhat remarkably, the t-ratio in (9) is shown to remain negative and significant at the one-sided 5% level[3] even as $\lambda \to \infty$.

2.4 Further Uses and Empirical Applications

There are advantages to the use of trend regression in convergence studies beyond the simplicity of the approach and the robustness to misspecification. Importantly, no specific data ordering of the cross sectional units is required in calculating the sample cross section variance K_{nt}. Moreover, the approach is robust to the number and the members of the cross sectional individuals, which may differ in each time period as long as the cross sectional sample sizes $\{n_t : t = 1, ..., T\}$ satisfy the uniform divergence condition $\min_{t \leq T}(n_t)/T \to \infty$ relative to the time series sample size T. This flexibility is particularly useful in analyzing experimental data where the subjects and number of subjects in each session may differ for each experimental round.

A second advantage is that the weak $\sigma-$convergence test does not require any model pre-tests or preliminary regression. The approach is equally able to detect divergence as convergence and no separate specification is required to allow for this. Thus, if the variable of interest is stochastically nonstationary or has a divergent trend, then the t-ratio test reveals this property by virtue of a right-sided test significance, as discussed earlier.

The following two empirical areas provide additional examples where the notion of weak $\sigma-$convergence is useful in applications.

The Law of One Price There is now a large empirical literature, especially in international finance, on the law of one price. Many of the empirical studies on this topic focus on the behavior of key rates and ratios over time. For example, relative purchasing power parity (PPP) states that real currency depreciation rates should converge over time; and the uncovered interest parity condition states that excess returns should converge. To fix ideas, define y_{it} as the deviation from a parity condition such as $y_{it} = \Delta s_{it} - \pi_{it} + \pi_t^*$ where Δs_{it} is the depreciation rate of the nominal exchange rate of the ith currency, π_{it} is the inflation rate of the ith country, and π_t^* is the inflation rate of the numeraire country. The law of one price requires that y_{it} converges over time. Since y_{it} can be positive or negative for any t and ratios y_{it}/y_{jt} may not converge to unity for $i \neq j$, the concept of relative convergence is not readily applicable to such data. But under certain conditions, weak $\sigma-$ convergence can be applied when the panel elements y_{it} share a common component F_t with a residual component x_{it} that satisfies (13), viz.,

$$y_{it} = F_t + x_{it}, \tag{11}$$

[2] The persistent excitation condition in simple linear regression requires that the sample variance of the regressor diverges as the sample size tends to infinity, so that the regressor is increasingly informative about the slope of the regression. In the present case of the trend decay model (8), note that $\sum_{t=1}^{T} t^{-2\lambda} \to \sum_{t=1}^{\infty} t^{-2\lambda} < \infty$ for all $\lambda > 0.5$. As shown in the proof of Lemma 4 and Theorem 2, this failure in persistent excitation, in contrast to that of the empirical regression (7) where $\sum_{t=1}^{T} t^2 \to \infty$, impacts the asymptotic behavior of both the regression coefficient $\hat{\phi}_{nT}$ and the t-ratio (9). Nonetheless, the t-ratio statistic remains negative and does not converge to zero as $T \to \infty$. So the t-ratio is still a useful in testing convergence even in this case.

[3] In fact, as shown later in Theorem 2, the asymptotic form of the t -ratio is $t_{\hat{\phi}_{nT}} \to_p -\sqrt{3}$ from below as $\lambda \to \infty$.

and then weak $\sigma-$convergence of x_{it} implies that y_{it} converges to follow F_t over time.

Alternatively, y_{it} may have the following static common factor structure.

$$y_{it} = \theta_i' F_t + x_{it}. \tag{12}$$

Weak $\sigma-$convergence of x_{it} in this case implies that the deviation of the law of one price can be explained by the common components embodied in the factors F_t because the variance of x_{it} shrinks over time under weak $\sigma-$convergence. Usually F_t is unobservable. So weak $\sigma-$convergence should be tested with the estimated idiosyncratic components.[4] Importantly, weak $\sigma-$convergence of x_{it} does not necessarily imply weak $\sigma-$convergence of y_{it} in (12). For when there is cross section heterogeneity and $\theta_i \neq \theta$ for some i, then y_{it} is not weak $\sigma-$convergent unless $\theta_i - \theta \to_p 0$ at a suitably fast rate relative to the factor F_t.

Convergence in Experimental Studies The concept of weak $\sigma-$convergence is useful for assessing outcomes in certain experimental studies. As summarized in the overviews by Ledyard (1995) and Chaudhuri (2011), repeated public good games typically display the property that the fractional contribution to the public account decreases over time, which makes weak $\sigma-$convergence towards a zero contribution of interest. In the ultimatum game, for instance, each subject in the experiment is asked what fraction of the endowment the subject wants to contribute to the partner. Many experimental studies (Gale et al., 1995, Cooper and Dutcher, 2011, Avrahami et al., 2010) reveal that the offers in such games converge toward a threshold acceptance point in the long run. Whether or not the offers converge over time can then be subjected to an empirical test of weak $\sigma-$convergence. Similarly, in a repeated Prisoner's dilemma game, it is of primary interest whether the estimated probability of the dominant strategy converges over time. In almost all such laboratory experimental studies, the data lie in a bounded interval (typically between zero and unity) and interest focuses on convergence behavior within such an interval, making weak $\sigma-$convergence a more relevant concept than relative convergence. Section 7 pursues an experimental data application showing how trend regression can be used to test convergence in economic experiments.

---

[4]Let $C_{it} = \theta_i' F_t$ and $\hat{x}_{it} = y_{it} - \hat{C}_{it}$ where $\hat{C}_{it} = \hat{\theta}_i' \hat{F}_t$. From Bai (2003), $\hat{C}_{it} - C_{it} = O_p\left(m_{nT}^{-1}\right)$ where $m_{nT} = \min\left[\sqrt{n}, \sqrt{T}\right]$, and so the estimation error $\hat{C}_{it} - C_{it} \to_p 0$ as $m_{nT} \to \infty$ can be treated as an asymptotically negligible component. Then weak $\sigma-$convergence of $\hat{x}_{it} = x_{it} - (\hat{C}_{it} - C_{it})$ implies weak $\sigma-$convergence of x_{it} from condition (ii) in (13) in the next section. Let $K_{nt}(\hat{x}) = n^{-1}\sum_{i=1}^n \hat{x}_{it}^2$, assume that x_{it} is weak $\sigma-$convergent, and set $\bar{K}_t^x = \text{plim}_{n\to\infty} K_{nt}^x$ and $a = \text{plim}_{t\to\infty} \bar{K}_t^x \in [0, \infty)$. Then

$$\begin{aligned} \frac{1}{n}\sum\nolimits_{i=1}^n \hat{x}_{it}^2 &= \frac{1}{n}\sum\nolimits_{i=1}^n x_{it}^2 + \frac{1}{n}\sum\nolimits_{i=1}^n \left(\hat{C}_{it} - C_{it}\right)^2 - 2\frac{1}{n}\sum\nolimits_{i=1}^n x_{it}\left(\hat{C}_{it} - C_{it}\right) \\ &= \frac{1}{n}\sum\nolimits_{i=1}^n x_{it}^2 + \frac{1}{n}\sum\nolimits_{i=1}^n \left(\hat{C}_{it} - C_{it}\right)^2 + o_p(1), \end{aligned}$$

and the three conditions of (13) are all satisfied regardless of the relative size of n and T. First, take the case where $n > T$. We have at most

$$\frac{1}{n}\sum\nolimits_{i=1}^n \left(\hat{C}_{it} - C_{it}\right)^2 = O_p\left(T^{-1}\right), \ \frac{1}{n}\sum\nolimits_{i=1}^n x_{it}\left(\hat{C}_{it} - C_{it}\right) = O_p\left(n^{-1/2}\right),$$

and then $\text{plim}_{n\to\infty} K_{nt}(\hat{x}) = \bar{K}_t(x) + O_p\left(T^{-1}\right) < \infty$. Hence, the first and second conditions of (13) are satisfied and the final condition holds by the weak $\sigma-$convergence of x_{it} since $\hat{C}_{it} - C_{it} = o_p(1)$. When $n \leq T$ we have $n^{-1}\sum_{i=1}^n \hat{x}_{it}^2 = n^{-1}\sum_{i=1}^n x_{it}^2 + O_p\left(n^{-1/2}\right)$, and all three conditions of (13) again hold.

3 Modeling and Testing Weak $\sigma-$ Convergence

This section begins with a formal definition of weak $\sigma-$convergence that embodies the ideas described above and that is designed to have sufficient generality to be useful in many different applications where convergence may not be monotone over time. To assist in interpreting the empirical regression used to test for convergence, we investigate a class of prototypical data generating processes each of which satisfies the weak $\sigma-$convergence conditions. These processes provide a foundation for developing asymptotics of the regression test under potential misspecification of the empirical regression relation.

To craft a suitably general concept of convergence, we are motivated by Hotelling's idea of decay in cross sectional variation over time. To avoid a strict requirement of monotonicity and to assist in developing a trend regression approach to testing convergence, it is useful to define a measure based on the behavior of the sample covariance of cross section dispersion with a simple linear trend. To fix ideas, let $\{x_{it} : i = 1, .., n; t = 1, ..., T\}$ be a panel of cross section and time series observations, let $\bar{x}_{\cdot t} := n^{-1} \sum_{i=1}^{n} x_{it}$, and define $\tilde{x}_{it} := x_{it} - \bar{x}_{\cdot t}$.

Definition (Weak $\sigma-$convergence): *Let $K_{nt}^{x} = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{it}^{2}$. The panel x_{it} is said to $\sigma-$converge weakly if the following conditions hold*

$$\begin{array}{l} \text{(i) } \text{plim}_{n\to\infty} K_{nt}^{x} = \bar{K}_{t}^{x} < \infty, \; a.s. \text{ for all } t \\ \text{(ii) } \text{plim}_{t\to\infty} \bar{K}_{t}^{x} = a \in [0, \infty), \\ \text{(iii) } \limsup_{T\to\infty} \gamma_{T}\left(\bar{K}_{t}^{x}, t; c_{T}\right) < 0 \; a.s., \end{array} \tag{13}$$

where $\gamma_{T}\left(\bar{K}_{t}^{x}, t; c_{T}\right) := \frac{1}{c_{T}} \sum_{t=1}^{T} \widetilde{\bar{K}_{t}^{x}} \tilde{t}$ is a time series sample covariance of $\bar{K}_{t}^{x}$ with a linear time trend t normalized by some suitable increasing sequence $c_{T} \to \infty$.

The first two conditions in (13) state that the probability limit of the cross section sample variance exists for each t and that its time limit exists and is non-negative. We allow for the case where the limit $\bar{K}_{t}^{x}$ is stochastic. For example, we may have $K_{nt}^{x} \to_{p} \bar{K}_{t}^{x} = \mathbb{E}\left\{x_{it}|\mathcal{C}\right\}$ where $\mathcal{C}$ is the invariant sigma field of shocks that are common across section, in which case $\bar{K}_{t}^{x}$ is a random sequence that is stochastically dependent on t (c.f. Phillips and Sul, 2003, 2007; Andrews, 2005). If in place of (ii) $\bar{K}_{t}^{x}$ diverges, then we say that x_{it} is $\sigma-$divergent. When x_{it} includes either deterministic trends like time polynomials or integrated random components like random walks, x_{it} is $\sigma-$divergent.

To illustrate $\sigma-$divergence, suppose the nonlinear panel model (2) is the generating mechanism with explicit deterministically trending panel form

$$x_{it} = a_{i} + \mu_{i} t^{\alpha} + \varepsilon_{it} = \left(\mu_{i} + \frac{a_{i} + \varepsilon_{it}}{t^{\alpha}}\right) t^{\alpha} = b_{it} t^{\alpha}, \tag{14}$$

where (a_{i}, μ_{i}) are *iid* idiosyncratic components distributed with finite support, ε_{it} is a zero mean covariance stationary time series for each i and where ε_{it} is *iid* and independent of $\left(a_{j}, \mu_{j}\right)$ for all i and j. In this case, it is straightforward to show conditions under which $\bar{K}_{t}^{x}$ is an increasing function of t. Let $\sigma_{a\mu}$ be the covariance between a_{i} and μ_{i}, σ_{a}^{2} be the variance of a_{i}, σ_{μ}^{2} be the variance of μ_{i}, and σ_{ε}^{2} be the variance of ε_{it}. By straightforward calculation we obtain $\bar{K}_{t}^{x} = \sigma_{a}^{2} + 2\sigma_{a\mu} t^{\alpha} + \sigma_{\mu}^{2} t^{2\alpha} + \sigma_{\varepsilon}^{2} = O\left(t^{2\alpha}\right)$, so that, as $t \to \infty$, $\bar{K}_{t}^{x} \sim \sigma_{\mu}^{2} t^{2\alpha}$ and $\bar{K}_{t}^{x}$ therefore increases over time eventually as long as $\alpha > 0$ and $\sigma_{\mu}^{2} > 0$. For integrated components, we may replace the power trend t^{α} in the generating mechanism (14) by a partial sum series $\theta_{t} = \sum_{s=1}^{t} \xi_{s}$ for some stationary time series ξ_s with zero mean, spectrum $f_\xi(\lambda)$, and long run variance $\omega_\xi^2 = 2\pi f_\xi(0) > 0$. Since $\theta_t = O_p(t^{1/2})$ and $T^{-1/2}\theta_{t=\lfloor Tr \rfloor} \Rightarrow B_\theta(r)$, a Brownian motion with variance ω_ξ^2, we find that $\bar{K}_t^x = \sigma_a^2 + 2\sigma_{a\mu}\theta_t^\alpha + \sigma_\mu^2\theta_t^{2\alpha} + \sigma_\varepsilon^2$. In this case as $t \to \infty$, $\bar{K}_t^x \sim \sigma_\mu^2\theta_t^2 \sim_a \sigma_\mu^2 t B_\theta(1)^2 = O_p(t)$. Thus, when there a stochastic trend in the panel generating mechanism, $\bar{K}_t$ is eventually increasing with probability one as $t \to \infty$ provided $\sigma_\mu^2 > 0$. It follows that condition (13) fails in the presence of either deterministic trends like time polynomials or stochastic trends such as random walks.

Condition (iii) in (13) requires that a suitably weighted sample covariance between the cross section variance limit $\bar{K}_t^x$ and a linear time trend t be negative eventually. The weight sequence $c_T \to \infty$ delivers a normalization that is suited to the time series properties of $\bar{K}_t^x$ in relation to a linear time trend, without requiring specificity in the definition. This generality accommodates a wide range of empirical possibilities. For example, consider the case where $\bar{K}_t^x = At^{-\alpha}$ for some constants $A, \alpha > 0$. Then, as Lemma 2 in the Appendix shows, the normalization function c_T may be defined as

$$c_T = \begin{cases} T^{-2+\alpha} & \text{if } \alpha < 1 \\ (T\ln T)^{-1} & \text{if } \alpha = 1 \\ T^{-1} & \text{if } \alpha > 1 \end{cases}.$$

Thus, when $\alpha = 1$, we have $\bar{K}_t^x = At^{-1}$, and then the standardized temporal covariation is

$$\begin{aligned} \gamma_T\left(\bar{K}_t^x, t; c_T\right) &= \frac{A}{T\ln T}\sum_{t=1}^T \widetilde{t^{-1}}\tilde{t} = \frac{A}{T\ln T}\sum_{t=1}^T\left[t^{-1} - \frac{1}{T}\sum_{t=1}^T t^{-1}\right]t \\ &= \frac{A}{T\ln T}\sum_{t=1}^T\left[t^{-1} - \frac{T\ln T + \gamma + O\left(T^{-1}\right)}{T}\right]t \to -\frac{A}{2} \text{ as } T \to \infty, \end{aligned} \tag{15}$$

where γ is the Euler-Mascheroni constant. This result continues to apply when $\bar{K}_t^x = A_t t^{-1}$ with $A_t = A + v_{At}$, where $A > 0$ and v_{At} is covariance stationary supported on $[-A, \infty)$ with zero mean and $1-$summable autocovariance function as in condition (46) of Lemma 3 of the Appendix.[5]

If the cross section variation has some non-monotonic behavior as in $\bar{K}_t^x = \frac{A}{t}\sin^2(t)$, then again setting $c_T = T\ln T$, we obtain the following standardized time covariation

$$\begin{aligned} \gamma_T\left(\bar{K}_t^x, t; c_T\right) &= \frac{A}{T\ln T}\sum_{t=1}^T \widetilde{\frac{\sin^2 t}{t}}\tilde{t} = \frac{A}{T\ln T}\sum_{t=1}^T\left[\frac{\sin^2 t}{t} - \frac{1}{T}\sum_{t=1}^T\frac{\sin^2 t}{t}\right]t \\ &= \frac{A}{T\ln T}\sum_{t=1}^T \sin^2 t - \frac{A(T+1)}{2T\ln T}\sum_{t=1}^T\frac{\sin^2 t}{t} \\ &= -\frac{A}{2\ln T}\left\{\frac{1}{2}\ln T - \frac{1}{2}\operatorname{Ci}(2T)\right\} + o(1) \to -\frac{1}{4}A, \end{aligned} \tag{16}$$

where $\operatorname{Ci}(x) = -\int_x^\infty \frac{\cos t}{t}dt$ is the cosine integral, which has asymptotic expansion $\operatorname{Ci}(x) \sim \frac{\sin x}{x} + O\left(x^{-2}\right)$ as $x \to \infty$. If $\bar{K}_t^x = \frac{A_t}{t}\sin^2(t)$ with a random coefficient A_t as indicated above, then the

[5] In particular, we find that

$$\begin{aligned} \gamma_T\left(\bar{K}_t^x, t; c_T\right) &= \frac{1}{T\ln T}\sum_{t=1}^T A_t\widetilde{t^{-1}}\tilde{t} = \frac{A}{T\ln T}\sum_{t=1}^T\left[t^{-1} - \frac{1}{T}\sum_{t=1}^T t^{-1}\right]t + \frac{1}{T\ln T}\sum_{t=1}^T v_{At}\left[t^{-1} - \frac{1}{T}\sum_{t=1}^T t^{-1}\right]t \\ &= \frac{A}{T\ln T}\sum_{t=1}^T\left[t^{-1} - \frac{1}{T}\sum_{t=1}^T t^{-1}\right]t + O_p\left(\frac{1}{\sqrt{T}}\right) \to -\frac{A}{2} \text{ as } T \to \infty. \end{aligned}$$

normalization factor $c_T = T \ln T$ again applies and (16) continues to hold. In more general cases where we have $\bar{K}_t^x = \frac{A_t}{t^\alpha} \sin^2(t)$ for some α, then the standardizing factor $c_T = c_T(\alpha)$ depends on the value of α, as implied by Lemma 3 in the Appendix.

In order to ensure an indicative negative value to $\limsup_{T\to\infty} \gamma_T\left(\bar{K}_t^x, t; c_T\right)$, condition (iii) uses a standardized sample covariance function rather than a (self normalized) correlation function. To show the limitation of using a correlation function, consider the case above where $\bar{K}_t^x = At^{-1}$. Straightforward calculations reveal that the corresponding correlation function is

$$\rho_T\left(\bar{K}_t^x, t\right) := \frac{\sum_{t=1}^T \widetilde{t^{-1}}\tilde{t}}{\left\{\sum_{t=1}^T \left(\widetilde{t^{-1}}\right)^2 \sum_{t=1}^T \tilde{t}^2\right\}^{1/2}} \sim -\left[\frac{(1/2)(T+1)\ln T}{\left\{\frac{\pi^2}{6}\frac{T^3}{12}\right\}^{1/2}}\right] = O\left(\frac{\ln T}{T^{1/2}}\right) \to 0,$$

so that in the limit the correlation tends to zero. That is, although $\rho_T\left(t^{-1}, t\right)$ is negative for large T, the correlation coefficient tends to zero as $T \to \infty$. In regression applications where we regress cross sectional variation on a linear trend, it is the sign of the regression coefficient that is important in determining convergence over time. The asymptotic behavior of the regression test statistic then depends on the magnitude of the regression coefficient in relation to its standard error, just as in spurious regression limit theory (Phillips, 1986). The fact that the sign of the regression coefficient is revealed in a simple trend regression, even when that regression is misspecified, forms the basic motivation for the test developed later. This relationship is important in our empirical work and will become evident in the analysis of the convergence test given later.

The requirement of condition (iii) that $\limsup_{T\to\infty} \gamma_T\left(\bar{K}_t^x, t; c_T\right) < 0$ *a.s.* can be modified to assure divergence in variation over time, rather than convergence. In particular, if conditions (i) and (ii) hold and $\liminf_{T\to\infty} \gamma_T\left(\bar{K}_t^x, t; c_T\right) > 0$ *a.s.* then the panel x_{it} may be said to $\sigma-$diverge weakly. In this case, the sign of the coefficient in the regression of the sample cross section variation on a linear time trend is positive and this ultimately reveals divergent behavior even though the regression itself may be misspecified.

3.1 Modeling Weak $\sigma-$Convergence with Decay Functions

To fix ideas and develop a framework for asymptotic analysis and testing we introduce an explicit modeling framework for the panel data x_{it}. Following PS, we use a power law time decay function, which is a convenient formulation to study weak $\sigma-$convergence.[6] Here we consider cases where additive heterogeneous and exogenous shocks enter the panel x_{it} and how these shocks are neutralized over time under convergence. We consider three prototypical cases to illustrate alternative models of convergence.

First, we model weak $\sigma-$convergence using a decay function for the mean of the panel in the following DGP, which is analogous to (14) but with temporal decay, and where temporal shocks influence only the mean level of the panel x_{it}

$$\textbf{Model M1}: x_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it}. \tag{17}$$

[6] Other decay functions are possible. For example, the exponential function $e^{c/t} \to 1$ as $t \to \infty$ is useful in capturing multiplicative decay, and the geometric function ρ^t with $|\rho| < 1$ is useful in capturing faster forms of decay than power laws.

Here a_i is the mean of x_{it}, μ_i is an initial (period 1) shock to the ith unit, and ϵ_{it} has zero mean and variance $\mathbb{E}\epsilon_{it}^2 = \sigma_{\epsilon,i}^2$. The power decay parameter $\alpha > 0$ (in contrast to (14)) and, as earlier, the idiosyncratic components (a_i, μ_i) are *iid* with finite support and are independent of the ϵ_{it}. The cross sectional variation of x_{it} in this case can be broken down into the following components

$$K_{nt}^x = \sigma_{a,n}^2 + \sigma_{\epsilon,nt}^2 + 2\sigma_{a\mu,n} t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} + e_{n,t}, \tag{18}$$

where $\sigma_{a,n}^2 = n^{-1}\sum_{i=1}^n \tilde{a}_i^2$, $\tilde{a}_i := a_i - \bar{a}_.$, $\sigma_{\epsilon,nt}^2 = n^{-1}\sum_{i=1}^n \tilde{\epsilon}_{it}^2$, $\tilde{\epsilon}_{it} := \epsilon_{it} - \bar{\epsilon}_{.t}$, $\sigma_{a\mu,n} = n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\mu}_i$, $\sigma_{\mu,n}^2 = n^{-1}\sum_{i=1}^n \tilde{\mu}_i^2$ and $e_{n,t} = 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha} \to_p 0$ as $n \to \infty$.

Examination of K_{nt}^x in (18) shows that weak $\sigma-$convergence holds if $\alpha > 0$ and $\sigma_{a\mu} \geq 0$. When there is only constant cross section variation in the panel, as occurs for instance when $x_{it} = a + \mu t^{-\alpha} + \epsilon_{it}$ and $\sigma_{\epsilon,nt}^2 = n^{-1}\sum_{i=1}^n \tilde{\epsilon}_{it}^2 \to_p \sigma_\epsilon^2 > 0$, then $\bar{K}_t^x = \sigma_\epsilon^2$ and there is no weak $\sigma-$convergence over time. In fact, the cross section mean and variation are constant for each t so that the sample covariation $\sum_{t=1}^T \widetilde{\bar{K}_t^x}\widetilde{t} = 0$ and the upper limit $\limsup_{T\to\infty} \gamma_T\left(\bar{K}_t^x, t; c_T\right) = 0$ *a.s.* In such cases there is panel mean weak convergence of the form $x_{it} \Rightarrow a + \epsilon_{i\infty}$ where the weak limit has constant variation σ_ϵ^2 over time. Thus, even though the variation does not shrink over time, we get individual element panel convergence in mean up to a homogeneously varying error. To eliminate such trivial cases, we henceforth assume that $\sigma_{a,n}^2 \to_p \sigma_a^2 > 0$ and $\sigma_{\mu,n}^2 \to_p \sigma_\mu^2 > 0$.

If $\alpha < 0$, then x_{it} is $\sigma-$divergent. In this case, the $t^{-2\alpha}$ term eventually dominates the $t^{-\alpha}$ term in (18) for large t.[7] This domination may also hold when $\alpha > 0$ if $\mathbb{E}\left(a_i\mu_i\right) = 0$, as then $\sigma_{a\mu} = \text{plim}_{n\to\infty}\sigma_{a\mu,n} = 0$ and $\sigma_{a\mu,n}t^{-\alpha} = O_p\left(n^{-1/2}t^{-\alpha}\right) = o_p\left(t^{-2\alpha}\right)$ uniformly in $t \leq T$ provided $T^{2\alpha}/n \to 0$. When $\sigma_{a\mu} \neq 0$, the sign of $\sigma_{a\mu}$ is also relevant in assessing convergence or divergence of variation. For instance, if $\alpha > 0$ and $\sigma_{a\mu} < 0$, the $t^{-\alpha}$ term dominates the $t^{-2\alpha}$ term in (18) as $t \to \infty$ and K_{nt}^x increases over time and eventually stabilizes to fluctuate around $\sigma_a^2 + \sigma_\epsilon^2$ as $n, T \to \infty$.

We also consider a variant of the DGP in (17) where K_{nt}^x may converge in an oscillatory way to a constant, as in the following example where

$$x_{it} = a_i + \mu_{it}t^{-\alpha} + \epsilon_{it}, \quad \epsilon_{it} \sim iid\left(0, \sigma_\epsilon^2\right),$$

with $a_i \sim iid\left(a, \sigma_a^2\right)$, $\mu_{it} = \mu_i \sin\left(\lambda\pi t\right)$, $\mu_i \sim iid\left(\mu, \sigma_\mu^2\right)$ and all these random coefficients are independent of ϵ_{it}. Then

$$K_{nt}^x = \sigma_{a,n}^2 + \sigma_{\epsilon,nt}^2 + 2\sigma_{a\mu,n}\sin\left(\lambda\pi t\right)t^{-\alpha} + \sigma_{\mu,n}^2\left[\sin\left(\lambda\pi t\right)\right]^2 t^{-2\alpha} + e_{n,t},$$

where $e_{n,t} = 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it}\sin\left(\lambda\pi t\right)t^{-\alpha} \to_p 0$ as $n \to \infty$. The probability limit of K_{nt}^x as $n \to \infty$ is

$$\bar{K}_t^x = \sigma_a^2 + \sigma_\epsilon^2 + \sigma_\mu^2\left[\sin\left(\lambda\pi t\right)\right]^2 t^{-2\alpha}, \tag{19}$$

under the assumptions above concerning the components and when a_i is uncorrelated with μ_i. Figure 2 below shows two examples of $\bar{K}_t^x$, setting $\sigma_a^2 + \sigma_\epsilon^2 = 1$, $\sigma_\mu^2 = 1$, $\lambda \in [1/5, 1/10]$, and $\alpha \in [0.5, 0.25]$. More complex versions of such processes, for example by setting $\mu_{it} = \mu_i \sin\left(\lambda_i\pi t\right)$ with random λ_i, would introduce differences in oscillatory behavior across section that would allow for cyclical variations across i in the data. Irregularities in the limiting form of the cyclical behavior over time

[7] When $\sigma_{a\mu} < 0$ and $\alpha < 0$, the variation K_t^x may follow a $U-$shaped time path if $|\sigma_{a\mu}| > \sigma_\mu^2$. In such cases, K_t^x may initially decrease before beginning to increase over time. When $|\sigma_{a\mu}| \leq \sigma_\mu^2$, then K_t^x increases monotonically over time.

as $n \to \infty$ can be introduced using functional versions such as $\sin(\pi\lambda_i(t/T))$ for $\lambda_i(t/T) \in \Lambda$, where Λ is the class of monotonically increasing functions over the unit interval $[0,1]$.

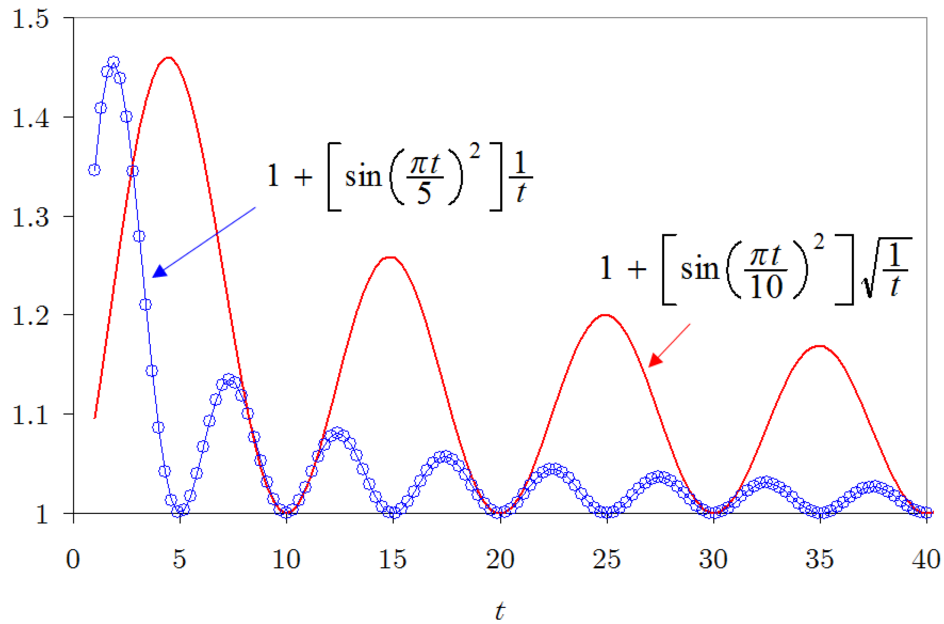


Figure 2: Two examples of oscillatory convergence illustrating (19)

As a further example, we may consider stochastic trend weak $\sigma-$convergence for panels of the form

$$x_{it} = a_i + b_{it}\frac{1}{\sum_{s=1}^{t}\xi_s} + \epsilon_{it}, \quad \epsilon_{it} \sim iid\left(0,\sigma^2\right),$$

and

$$(a_i, b_{it}) \sim iid\left(\begin{bmatrix} a \\ b_t \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_{bt}^2 \end{bmatrix}\right) \text{ over } i,$$

with the ϵ_{it} independent of (a_i, b_{it}), and b_t, $\sigma_{bt}^2 = O_p(1)$, reflecting the presence of common shocks over time on the mean and variance of b_{it}. Under general conditions on the time series ξ_s (e.g., Phillips and Solo, 1992), partial sums of ξ_s satisfy an invariance principle upon standardization, so that $T^{-1/2}\sum_{s=1}^{\lfloor T\cdot\rfloor}\xi_s \Rightarrow B_\xi(\cdot)$, a Brownian motion with variance $\omega^2 > 0$. Then, since

$$\begin{aligned} K_{nt}^x &\to_{a.s.} \sigma_a^2 + \frac{\sigma_{bt}^2 b_t^2 t^{-1}}{\left(t^{-1/2}\sum_{s=1}^{t}\xi_s\right)^2} \text{ as } n \to \infty \\ &\sim \sigma_a^2 + \frac{\sigma_{bt}^2 b_t^2 t^{-1}}{B_\xi(1)^2} \to_p \sigma_a^2, \text{ as } t \to \infty \end{aligned} \tag{20}$$

in which case we again get weak $\sigma-$convergence as $t \to \infty$ to σ_a^2. Observe that the ratio $1/B_\xi(1)^2$ in (20) is the reciprocal of a χ_1^2 variate and therefore has no finite integer moments, so that realizations of the second component of (20) have potentially heavy tails. Moreover, depending on the realization of the coefficient function b_t^2 there may be increasing or decreasing behavior en route to the limit value σ_a^2.

The second prototypical case arises when shocks influence only the cross sectional variation of x_{it} not the mean level of the panel x_{it}, as in the following DGP.[8]

$$\textbf{Model M2: } x_{it} = a_i + \epsilon_{it} t^{-\beta}. \tag{21}$$

The sample cross sectional variation K_{nt}^x is then

$$K_{nt}^x = \sigma_{a,n}^2 + \sigma_{\epsilon,nt}^2 t^{-2\beta} + e_{n,t}, \tag{22}$$

where $\sigma_{\epsilon,nt}^2 = n^{-1}\sum_{i=1}^n \tilde{\epsilon}_{it}^2$, and $e_{n,t} = 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} \to_p 0$ as $n \to \infty$. If $n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it} \to_p 0$ and $\beta > 0$ the power decay rate in (22) is 2β. When $\beta < 0$, there is divergence and x_{it} has increasing cross section variation over time. For instance, if $\beta = -1/2$, $x_{it} = O_p(\sqrt{t})$ and x_{it} has stochastic order similar to a unit root process.

The last prototypical model combines level and scale time changes and has the following form

$$\textbf{Model M3: } x_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it} t^{-\beta}, \tag{23}$$

where the idiosyncratic components satisfy the same assumptions as before. The sample cross sectional variation of x_{it} in this case is given by

$$K_{nt}^x = \sigma_{a,n}^2 + 2\sigma_{a\mu,n}t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} + \sigma_{\epsilon,nt}^2 t^{-2\beta} + e_{n,t}, \tag{24}$$

where $e_{n,t} = 2n^{-1}t^{-\beta}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it} + 2n^{-1}t^{-\alpha-\beta}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it} \to_p 0$ as $n \to \infty$, and $\sigma_{a,n}^2 = n^{-1}\sum_{i=1}^n \tilde{a}_i^2$, $\sigma_{a\mu,n} = n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\mu}_i$, and $\sigma_{\mu,n}^2 = n^{-1}\sum_{i=1}^n \tilde{\mu}_i^2$, as before in (18). Weak $\sigma-$convergence clearly holds if $\alpha > 0$ and $\beta > 0$.

The outcomes for the sample cross section variation in these models may be summarized as follows:

$$K_{nt}^x = a_n + \eta_{n,t} + \varepsilon_{n,t}, \tag{25}$$

where

$$a_n = \begin{cases} \sigma_{a,n}^2 + \sigma_{\epsilon,nT}^2 & \text{for M1,} \\ \sigma_{a,n}^2 & \text{for M2,} \\ \sigma_{a,n}^2 & \text{for M3,} \end{cases} \quad \eta_{n,t} = \begin{cases} 2\sigma_{a\mu,n}t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} & \text{for M1,} \\ \sigma_{\epsilon,nT}^2 t^{-2\beta} & \text{for M2,} \\ 2\sigma_{a\mu,n}t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} + \sigma_{\epsilon,nT}^2 t^{-2\beta} & \text{for M3,} \end{cases} \tag{26}$$

and

$$\varepsilon_{n,t} = \begin{cases} 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) & \text{for M1,} \\ 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right)t^{-2\beta} & \text{for M2,} \\ 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right)t^{-2\beta} & \text{for M3.} \end{cases} \tag{27}$$

Clearly, Model M3 in (23) nests the first two models. M3 becomes M1 when $\beta = 0$ and $\alpha \neq 0$, and M3 becomes M2 when $\alpha = 0$ and $\beta \neq 0$.

[8] The model may be modified to allow for a composite error of the form $\epsilon_{it}^* + \epsilon_{it}t^{-\beta}$ with an additional stationary component ϵ_{it}^* This error augmentation does not alter subsequent results and is therefore not included in (21).

3.2 Testing and Application of Weak $\sigma-$Convergence

3.2.1 Direct Nonlinear Regression

An obvious initial possibility for testing weak $\sigma-$convergence is to run a nonlinear regression based on the form of the implied decay function of K_{nt} given in (24) and carry out tests on the coefficients and the sign of the power trend parameters. The parameters of interest are σ_a^2, α, β, σ_μ^2 and σ_ϵ^2. If these parameters were identifiable and estimable using nonlinear least squares, testing weak $\sigma-$convergence might be possible by this type of direct model specification, fitting, and testing. However, the parameters are not all identifiable or asymptotically identifiable in view of the multifold identification problem that is present in models with multiple power trend parameters. Readers are referred to Baek, Cho and Phillips (2015) and Cho and Phillips (2015) for a recent study of this multifold identification problem, and more general issues of identification and testing analysis in time series models with power trends of the type that appear in (23) and (25).

Even if restrictions were imposed to ensure that all parameters were identified in a direct model specification of convergence, formulation of a suitable null hypothesis presents further difficulties. Our interest centres on the possible presence of weak $\sigma-$convergence, which holds in the model when $\beta > 0$ and $\alpha > 0$. Hence, the conditions for weak $\sigma-$convergence are themselves multifold, which further complicates testing. Further, it is well known that nonlinear estimation of the power trend parameters α and β is inconsistent when $\alpha, \beta > 0.25$ because of weakness in the signal that is transmitted from a decay trend regressor (see Malinvaud, 1970, Wu, 1981, Phillips, 2007, and Lemma 1 below). Finally, a parametric nonlinear regression approach relies on a given specification, whereas in practical work the nature of data and its generating mechanism across section and over time are generally so complex that any given model will be misspecified. In consequence, econometric tests based on the direct application of nonlinear regression to a given model will suffer from specification bias resulting in size distortion. It is therefore of considerable interest and importance in applications to be able to provide a convergence test without providing a complete model specification for the panel.

In view of these manifold difficulties involved in direct model specification and testing, we pursue a convenient alternative approach to test for weak $\sigma-$convergence. The idea is to employ a simple linear trend regression that is capable of distinguishing convergence from divergence, even though a linear trend regression is misspecified under the convergence hypothesis. In fact, a linear trend may be interpreted as a form of spurious trend under the convergence hypothesis. Yet this type of empirical regression provides asymptotically revealing information about convergence, as we now explain, just as spurious regressions typically reveal the presence of trend in the data through the use of another coordinate system (Phillips, 1998, 2005).

3.2.2 Linear Trend Regression

The idea is to run a least squares regression of cross section sample variation[9] K_{nt} on a linear trend giving

$$K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT} t + \hat{u}_t, \ t = 1, ..., T \tag{28}$$

where $\hat{u}_t$ is the fitted residual, and to perform a simple significance test on the fitted trend slope coefficient $\hat{\phi}_{nT}$. This regression enables us to test the key defining property of weak $\sigma-$convergence. In particular, according to the definition, if $\text{plim}_{n\to\infty} K_{nt}$ exists and K_{nt} is a decreasing function

[9] In what follows we remove the variable name affix and write K_{nt}^x simply as K_{nt}.

of t, then weak $\sigma-$convergence holds. In this event, in terms of the regression (28), we expect the slope coefficient $\hat{\phi}_{nT}$ to be significantly negative, whereas if $\hat{\phi}_{nT}$ is not significantly different from zero or is greater than zero, then the null of no $\sigma-$convergence cannot be rejected.

In order to construct a valid significance test, allowance must be made for the fact that the model (28) is generally misspecified when K_{nt} satisfies a model such as (24), so that a robust test of significance must allow for the presence of serially correlated and heteroskedastic residuals. Further, as we will show under certain regularity conditions, the corresponding robust t-ratio statistic $t_{\hat{\phi}_{nT}}$ diverges to negative infinity in the presence of weak $\sigma-$convergence, so that this simple regression t-test is consistent.

The misspecification implicit in the trend regression (28) complicates the asymptotic properties of the estimates and the t-ratio statistic, so that the limit behavior of both $\hat{\phi}_{nT}$ and $t_{\hat{\phi}_{nT}}$ depends on the values of α and β and the relative sample sizes n and T. This limit behavior is examined next.

4 Asymptotic Properties

This section provides asymptotic properties of the suggested test in the previous section. We start with asymptotics for the slope coefficient estimator $\hat{\phi}_{nT}$ and then develop the limit theory for the t-ratio statistic. To proceed in the analysis we impose the following conditions on the components of the system given by model M3 in (23), which is convenient to use in what follows because it subsumes models M1 and M2.

Assumption A:

(i) The model error term, ϵ_{it}, is independently distributed over i with finite fourth moment and is strictly stationary over t with autocovariance sequence $\gamma_i(h) = \mathbb{E}(\epsilon_{it}\epsilon_{it+h})$ satisfying the summability condition $\sum_{h=1}^{\infty} h|\gamma_i(h)| < \infty$ and with long run variance $\Omega_e^2 = \sum_{h=-\infty}^{\infty}\gamma_i(h) > 0$.

(ii) The slope coefficients, a_i and μ_i, are cross sectionally independent and have finite second moments.

(iii) $\mathbb{E}a_i\epsilon_{jt} = \mathbb{E}\mu_i\epsilon_{jt} = \mathbb{E}\epsilon_{it}\epsilon_{jt} = 0$ *for all* $i, j,$*and* t,with $i \neq j$.

The cross section independence over i and stationarity over t in (i) are restrictive but are also fairly common. It seems likely that both conditions may be considerably relaxed and cross sectional dependence in ϵ_{it} and some heterogeneity over t may permitted, for example under suitable uniform moment and mixing conditions that assure the validity of our methods. For simplicity we do not pursue these extensions in the present work.

In what follows it is useful to note that as $T \to \infty$ sums of reciprocal powers of the integers have the following asymptotic form (see Lemma 1 in the Appendix)

$$\tau_T(\lambda) = \sum_{t=1}^{T} t^{-\lambda} = \begin{cases} \frac{1}{1-\lambda}T^{1-\lambda} + O(1) & \text{if } \lambda < 1, \\ \ln T + O(1) & \text{if } \lambda = 1, \\ \zeta(\lambda) = O(1) & \text{if } \lambda > 1. \end{cases}$$

As is well known, $\tau_T(\lambda)$ is $O(1)$ for $\lambda > 1$, has a representation by Euler-Maclaurin summation in terms of Bernoulli numbers, and can be simply bounded. Lemma 1 provides more detail about the Riemann zeta function limit $\zeta(\lambda)$ and the various asymptotic representations of $\tau_T(\lambda)$, which turn out to be useful in our asymptotic development.

The least squares coefficient $\hat{\phi}_{nT}$ in the trend regression (28) can be decomposed into deterministic and random component parts as follows. We use the general framework for the sample cross section variation K_{nt}^x given by (25) - (27). We may write $\eta_{n,t}$ as

$$\eta_{n,t} = \eta_t + \xi_{n,t} = \eta_t + O_p\left(n^{-1/2}\right), \tag{29}$$

where η_t is the $n-$probability limit of $\eta_{n,t}$, specifically

$$\eta_t = \begin{cases} 2\sigma_{a\mu}t^{-\alpha} + \sigma_\mu^2 t^{-2\alpha} & \text{for M1,} \\ \sigma_\epsilon^2 t^{-2\beta} & \text{for M2,} \\ 2\sigma_{a\mu}t^{-\alpha} + \sigma_\mu^2 t^{-2\alpha} + \sigma_\epsilon^2 t^{-2\beta} & \text{for M3,} \end{cases} \tag{30}$$

where $\sigma_{a\mu} = \text{plim}_{n\to\infty}\sigma_{a\mu,n}$, $\sigma_\mu^2 = \text{plim}_{n\to\infty}\sigma_{\mu,n}^2$, and $\sigma_\epsilon^2 = \text{plim}_{n\to\infty}\sigma_{\epsilon,nT}^2$. We further define the quantities

$$\xi_{a\mu,n} : \ = \sigma_{a\mu,n} - \sigma_{a\mu} = n^{-1}\sum\nolimits_{i=1}^n \left(\tilde{a}_i\tilde{\mu}_i - \sigma_{a\mu}\right) = O_p\left(n^{-1/2}\right), \tag{31}$$

$$\xi_{\sigma,n} : \ = \sigma_{\mu,n}^2 - \sigma_\mu^2 = n^{-1}\sum\nolimits_{i=1}^n \left(\tilde{\mu}_i^2 - \sigma_\mu^2\right) = O_p\left(n^{-1/2}\right), \tag{32}$$

so that the residual in (29) can be written as $\xi_{n,t} := 2\xi_{a\mu,n}t^{-\alpha} + \xi_{\sigma,n}t^{-2\alpha} = O_p\left(n^{-1/2}\right)$ uniformly in t for all $\alpha > 0$ for M1.

Setting $a_{tT} = \tilde{t}/\left(\sum_{s=1}^T \tilde{s}^2\right)$ and using (29), the trend regression coefficient $\hat{\phi}_{nT}$ in (28) can be decomposed into three components as follows

$$\hat{\phi}_{nT} = \sum\nolimits_{t=1}^T a_{tT}\tilde{\eta}_t + \sum\nolimits_{t=1}^T a_{tT}\tilde{\xi}_{n,t} + \sum\nolimits_{t=1}^T a_{tT}\tilde{\varepsilon}_{n,t} =: I_A + I_B + I_C, \tag{33}$$

where $\tilde{\eta}_t = \widetilde{bt^{-\lambda}}$, $\tilde{\xi}_{n,t} = \xi_{n,t} - T^{-1}\sum_{t=1}^T \xi_{nt}$, $\tilde{\varepsilon}_{n,t} = \varepsilon_{n,t} - T^{-1}\sum_{t=1}^T \varepsilon_{nt}$ and λ represents the relevant decay parameter, and b is the corresponding coefficient in that term. To be specific, the values of λ and b in the three model cases M1-M3 are summarized in the Table M below.

Case	M1		M2		M3	
	b	λ	b	λ	b	λ
$\alpha, \beta > 0$,and $\sigma_{a\mu} \neq 0$	$2\sigma_{a\mu}$	α	σ_ϵ^2	2β	$2\sigma_{a\mu}$ for α, σ_ϵ^2 for 2β	$\min[\alpha, 2\beta]$
$\alpha, \beta > 0$,and $\sigma_{a\mu} = 0$	σ_μ^2	2α	σ_ϵ^2	2β	σ_μ^2 for 2α, σ_ϵ^2 for 2β	$\min[2\alpha, 2\beta]$
$\alpha < 0$ or $\beta < 0$	σ_μ^2	2α	σ_ϵ^2	2β	σ_μ^2 for 2α, σ_ϵ^2 for 2β	$\min[2\alpha, 2\beta]$

Table M: Parameter Specifications for Models M1 - M3

As is apparent in the table, for model M3 there are two possible sources of decay (or divergence) and the relevant value of the parameter λ is determined by the majorizing force. These possibilities are accounted for in the proofs of the results that follow.

It is convenient to define the conditional order-rate element

$$\mathcal{O}_{T\lambda} = -\begin{cases} L_\lambda T^{-1-\lambda} & \text{if } \lambda < 1, \\ 6T^{-2}\ln T & \text{if } \lambda = 1, \\ 6\zeta(\lambda)T^{-2} & \text{if } \lambda > 1. \end{cases} \tag{34}$$

where $L_\lambda = 6\lambda[(2-\lambda)(1-\lambda)]^{-1}$. The limit behavior of $\hat{\phi}_{nT}$ in the regression equation (28) is characterized more easily in terms of $\mathcal{O}_{T\lambda}$ in the following result. Since the linear trend regression (28) is typically misspecified, interest centers on the asymptotic behavior of $\hat{\phi}_{nT}$ under the various potential models of data generation, the possible values of the rate parameters (α, β) in the trend decay functions of M1, M2, and M3, and the sample size divergence rates $n, T \to \infty$.

Theorem 1 (Linear Trend Regression Limit Behavior)

Under assumption A and as $(n, T) \to \infty$ jointly, the limit behavior of the fitted coefficient $\hat{\phi}_{nT}$ in regression (28) is characterized in the following results.

(i) Under weak $\sigma-$convergence (with $\lambda > 0$ and $b > 0$), then $\hat{\phi}_{nT} = b \times \mathcal{O}_{T\lambda} < 0$ for $\frac{1}{n} + \frac{T}{n} \to 0$ and the respective values of λ given in Table M.

(ii) Under $\sigma-$divergence (with $\lambda > 0$ and $b < 0$), then $\hat{\phi}_{nT} = b \times \mathcal{O}_{T\lambda} > 0$ for $\frac{1}{n} + \frac{T}{n} \to 0$; or $\hat{\phi}_{nT} = b \times L_\lambda T^{-1-\lambda} > 0$ if $\lambda < 0$ with no restiction on the n/T ratio as $(n, T) \to \infty$.

(iii) Under the null hypothesis of neither convergence nor divergence ($\lambda = 0$), then $\hat{\phi}_{nT} = O_p\left(n^{-1/2} T^{-3/2}\right)$, irrespective of the n/T ratio.

The decay parameter λ which governs convergence behavior is unknown and is not estimated. Since the empirical trend regression equation (28) is generally misspecified when $\lambda \neq 0$, the key point of interest is whether the fitted coefficient $\hat{\phi}_{nT}$ and its associated t-ratio in regression (28) have asymptotically distinguishable behavior that reveal weak $\sigma-$convergence in the data. When the deterministic component ($I_A = \sum_{t=1}^{T} a_{tT} \eta_t$) of $\hat{\phi}_{nT}$ dominates (33) as it typically does (see the discussion below), it turns out that there is identifiable behavior in the sign of $\hat{\phi}_{nT}$ and this property is used as the basis of a convergence test.

In establishing the results of the theorem, the proof examines the components of (33) to assess the main contribution to the asymptotic behavior of $\hat{\phi}_{nT}$. The proof of the theorem provides detailed calculations and examines the various cases implied by the different parameter configurations. The outcomes for the three models that emerge from this theory are summarized below with discussion about the role of the decay parameter λ and its specific parametric values $(\alpha, 2\alpha, 2\beta)$ in determining the respective outcomes.

First, in model M1 the dominant term depends on the value of α, and the I_A term dominates provided $nT^{-2\alpha} \to \infty$ if $\alpha < 1/2$ and provided $n/T \to \infty$ if $\alpha \geq 1/2$. In model M2, the decay parameter is 2β. As long as $\beta > 0$, $\mathbb{E}\hat{\phi}_{nT}$ passes to zero as $T \to \infty$, and the other terms pass to zero faster than the first term I_A in (33) regardless of the n/T ratio. Thus, regardless of the value of $\beta > 0$ and the n/T ratio, the first term always dominates the other two terms. In model M3, the situation is more complex because there are two potential sources of decay or divergence; but, in general, the behavior under M3 is similar to that of M1 and the I_A term dominates provided $n/T \to \infty$.

As is apparent in the statement of the theorem, discriminating behavior in the fitted slope coefficient $\hat{\phi}_{nT}$ (and, as we will see, test consistency) typically require the rate condition that $n/T \to \infty$. This condition ensures that the sample cross section variation has stabilized sufficiently (for large enough n) to facilitate the identification of trend decay or divergence in the variation over time (for large T). It is of some interest whether this rate condition might be relaxed if a more flexible power trend regression of the form

$$K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT} t^{\psi} + \hat{u}_t, \ t = 1, ..., T, \text{ and some given } \psi > 0, \tag{35}$$

were used in place of the linear trend regression equation (28). In fact, as discussed in Appendix S, use of a power trend regressor t^{ψ} in the empirical regression instead of a simple linear trend does not lead to different rate requirements regarding (n, T). Simulations with various values of the exponent parameter ψ confirmed that there is also no reason based on finite sample performance to use a value of ψ different from unity in the empirical regression.

With the asymptotic behavior of $\hat{\phi}_{nT}$ in hand, limit theory can be developed for the corresponding t-ratio in the regression (28). We use the robust form of the test statistic given earlier in (9) which employs a standard long run variance estimate $\hat{\Omega}_u^2$ constructed by lag kernel methods as in (10) from the regression residuals $\hat{u}_t = K_{nt} - \hat{a}_{nT} - \hat{\phi}_{nT} t$. Since the trend regression equation is misspecified, $\hat{\Omega}_u^2$ does not consistently estimate the long run variance Ω_e^2 of the errors ϵ_{it} in models M1,M2, or M3 as $n, T \to \infty$ unless the parameters $\alpha = \beta = 0$ in those models and there is no decay function in the generating model. That special case is taken as the null hypothesis of no convergence or divergence, viz., $\mathbb{H}_0 : \alpha = \beta = 0$, under which consistency $\hat{\Omega}_u^2 \to_p \Omega_e^2$ follows by standard methods.

The primary focus of interest in testing is not the null $\mathbb{H}_0 : \alpha = \beta = 0$ but the alternative hypothesis $\mathbb{H}_A : \alpha \neq 0$ or $\beta \neq 0$ under which there is convergence or divergence in the cross section sample variation. Under $\mathbb{H}_A$, the linear trend regression specification is no longer maintained and the relevant asymptotic behavior is that of the long run variance estimate $\hat{\Omega}_u^2$ under misspecification of the trend regression. To capture the misspecification effect, it is convenient to decompose the regression residual into two primary components as

$$\hat{u}_t = \left(\tilde{\eta}_{n,t} - \hat{\phi}_{nT}\tilde{t}\right) + \tilde{\varepsilon}_{nT} =: \tilde{\mathcal{M}}_{nt} + \tilde{\varepsilon}_{nT}, \tag{36}$$

where $\tilde{\eta}_{n,t} = \eta_{n,t} - T^{-1}\sum_{t=1}^{T}\eta_{n,t}$ and $\tilde{\varepsilon}_{nt} = \varepsilon_{nt} - T^{-1}\sum_{t=1}^{T}\varepsilon_{nt}$. Using (29)-(32) we have $\eta_{n,t} = \eta_t + \xi_{n,t} = \eta_t + O_p\left(n^{-1/2}\right)$ uniformly in t for all $\alpha > 0$ for M1 and M3. Then,

$$\tilde{\eta}_{n,t} = \tilde{\eta}_t + \tilde{\xi}_{n,t} = b\widetilde{t^{-\lambda}} + \tilde{\xi}_{n,t},$$

using the simplified summary notation of Table M. More specifically, from Lemma 5 in the Appendix, we have

$$\tilde{\xi}_{n,t} = \begin{cases} 2\xi_{a\mu,n}\widetilde{t^{-\alpha}} + \xi_{\mu,n}\widetilde{t^{-2\alpha}} = o_p\left(\widetilde{t^{-2\alpha}}\right) & \text{for M1,} \\ \xi_{\sigma,n}\widetilde{t^{-2\beta}} = o_p\left(\widetilde{t^{-2\beta}}\right) & \text{for M2,} \\ 2\xi_{a\mu,n}\widetilde{t^{-\alpha}} + \xi_{\mu,n}\widetilde{t^{-2\alpha}} + \xi_{\sigma,n}\widetilde{t^{-2\beta}} = o_p\left(\min\left(\widetilde{t^{-2\alpha}}, \widetilde{t^{-2\beta}}\right)\right) & \text{for M3,} \end{cases} \tag{37}$$

which may be expressed in the simple form that $\tilde{\xi}_{n,t} = o_p\left(\tilde{\eta}_t\right)$ uniformly in t as $n/T \to \infty$. Since the trend regression coefficient $\hat{\phi}_{nT}$ satisfies the decomposition (33), we find that

$$\begin{aligned} \tilde{\mathcal{M}}_{nt} &= \tilde{\eta}_{n,t} - \hat{\phi}_{nT}\tilde{t} = \tilde{\eta}_t + \tilde{\xi}_{n,t} - \left(I_A + I_B + I_C\right)\tilde{t} \\ &= \tilde{\eta}_t - I_A\tilde{t} + \tilde{\xi}_{n,t} - \tilde{t}\left(I_B + I_C\right) \\ &= \tilde{m}_t + R_{nt}, \end{aligned}$$

with deterministic part $\tilde{m}_t = \tilde{\eta}_t - I_A\tilde{t}$ and random part $R_{nt} = \tilde{\xi}_{n,t} - \tilde{t}\left(I_B + I_C\right)$. As $n/T \to \infty$, we show in the Appendix in the proof of Theorem 1 that I_A dominates I_B and I_C for all three models; and, from above, $\tilde{\xi}_{n,t} = o_p\left(\tilde{\eta}_t\right)$ uniformly in t as $n/T \to \infty$. It follows that $R_{nt} = o_p\left(\tilde{m}_t\right)$ uniformly in t as $n/T \to \infty$.

Under model M2, the term $\tilde{\mathcal{M}}_{nt}$ in (36) always dominates the second term asymptotically in the behavior of $\hat{\Omega}_u^2$ as $(n,T) \to \infty$, irrespective of the n/T ratio. In models M1 and M3, $\tilde{\mathcal{M}}_{nt}$ continues to dominate the behavior of $\hat{\Omega}_u^2$ as $(n,T) \to \infty$ provided $n/T \to \infty$. Thus, $\tilde{\mathcal{M}}_{nt}$ can be rewritten

$$\tilde{\mathcal{M}}_{nt} = b\left[\widetilde{t^{-\lambda}} - \tilde{t}\left(\sum\nolimits_{t=1}^{T} \widetilde{\tilde{t}t^{-\lambda}}\right)\left(\sum\nolimits_{t=1}^{T} \tilde{t}^2\right)^{-1}\right] + R_{nt}, \tag{38}$$

where R_{nt} is a smaller order term. Thus, when $\tilde{\mathcal{M}}_{nt}$ dominates the behavior of $\hat{\Omega}_u^2$ as $(n,T) \to \infty$, the asymptotic behavior of the t-ratio is determined as follows

$$t_{\hat{\phi}_{nT}} = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_u^2 / \sum_{t=1}^{T} \tilde{t}^2}} \sim \frac{\hat{\phi}_{nT}}{\sqrt{\Omega_{\mathcal{M}}^2 / \sum_{t=1}^{T} \tilde{t}^2}} = \frac{\left(b\sum_{t=1}^{T} \widetilde{\tilde{t}t^{-\lambda}}\right)\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1/2}}{\sqrt{\Omega_{\mathcal{M}}^2}}, \tag{39}$$

making the t-ratio a function of only λ, κ, and T asymptotically when $n/T \to \infty$. In (39) the quantity $\Omega_{\mathcal{M}}^2$ is constructed in the usual manner as a long run variance estimate, viz.,

$$\Omega_{\mathcal{M}}^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T} \tilde{m}_t^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\tilde{m}_t\tilde{m}_{t+\ell}, \tag{40}$$

as in (10) with lag truncation parameter L; and, being a function of $\tilde{m}_t$, $\Omega_{\mathcal{M}}^2$ is a deterministic function of t. The asymptotic equivalence in (39) is established in the proof of the following result which gives the asymptotic behavior of $t_{\hat{\phi}_{nT}}$ under the null and alternative hypotheses.

Theorem 2 (Asymptotic Properties of the $t_{\hat{\phi}_{nT}}$ ratio)

Under Assumption A, the t-ratio statistic $t_{\hat{\phi}_{nT}}$ in the empirical regression (28) has the following asymptotic behavior as $n, T \to \infty$:

(i) Under weak $\sigma-$convergence ($\lambda > 0$ and $b > 0$) and when $n/T \to \infty$,

$$t_{\hat{\phi}_{nT}} \to -\tau_\lambda^* = \begin{cases} -\infty & \text{if } 0 < \lambda < 1, \\ -\sqrt{6/\kappa^2} & \text{if } \lambda = 1, \\ -\mathbb{Z}(\lambda)\sqrt{3} & \text{if } 1 < \lambda < \infty, \\ -\sqrt{3} & \text{if } \lambda \to \infty. \end{cases} \tag{41}$$

where $\kappa > 0$ is defined by the lag truncation parameter $L = \lfloor T^\kappa \rfloor$ in the long run variance estimator (10). The function $\mathbb{Z}(\lambda) := \zeta(\lambda)\left(\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda, t)\right)^{-1/2} > 1$ for all $\lambda > 1$, where $\zeta(\lambda) = \sum_{t=1}^{\infty} t^{-\lambda}$ and $\zeta(\lambda, t) = \sum_{s=1}^{\infty}(s+t)^{-\lambda}$ are the Riemann and Hurwitz zeta functions, respectively.

(ii) Under $\sigma-$divergence, as $n, T \to \infty$,

$$t_{\hat{\phi}_{nT}} \to \begin{cases} +\infty & \text{if } \lambda < 0 \textit{ regardless of the } n/T \textit{ ratio}, \\ \tau_\lambda^* & \text{if } \sigma_{a\mu} < 0 \textit{ with } \lambda > 0 \textit{ and } n/T \to \infty. \end{cases} \tag{42}$$

(iii) Under the null hypothesis $\mathbb{H}_0 : \lambda = 0$ (neither convergence nor divergence), as $n, T \to \infty$ irrespective of the n/T ratio,

$$t_{\hat{\phi}_{nT}} \to^d \mathcal{N}(0,1). \tag{43}$$

As indicated in (41) and (42), the precise limit behavior of the t-ratio statistic depends on the parameter λ, the lag truncation constant $\kappa > 0$ in $L = \lfloor T^{\kappa} \rfloor$, and certain other constants when $\lambda \geq 1$. When the Bartlett-Newey-West estimate is used in constructing $\hat{\Omega}_u^2$, the constant κ is commonly set to 1/3.

Theorem 2 (ii) defines t-ratio behavior under $\sigma-$divergence when $\lambda < 0$ and the limit theory is expected. For when $\lambda \in \{2\alpha, 2\beta\}$ and is negative, the dominant term is either $t^{-2\alpha}$ or $t^{-2\beta}$, so that cross section variation diverges permanently and the t-ratio is positive and increasing as $n, T \to \infty$. Theorem 2 (ii) also shows that when $\lambda > 0$ and $\sigma_{a\mu} < 0$, the behavior of the t ratio is a mirror image of part (i). Theorem 2 (iii) gives the standard result for a correctly specified model with weakly dependent errors. Thus, when $\alpha = \beta = 0$, the trend regression is well defined as a simple model with a slope coefficient of zero, and the t-ratio is asymptotically $\mathcal{N}(0, 1)$ by standard nonparametrically studentized limit theory.

Theorem (i) is the key result of most relevance in empirical studies of convergence. The explicit limit behavior shown in (41) derives from the fact that the t-ratio takes asymptotically the deterministic form (39), whose limit form can be well characterized. As long as the deterministic component in the estimator $\hat{\phi}_{nT}$ is dominant, the results given in (41) hold. Remarkably, the t-ratio is completely free of nuisance parameters in the limit because the scale parameter b appears in both numerator and denominator of the t-ratio and thereby cancels, making the limiting form of the t-ratio a function only of the value of λ and the bandwidth parameter κ used in the construction of the long run variance estimate. This property makes the test statistic especially convenient and auspicious for practical work.

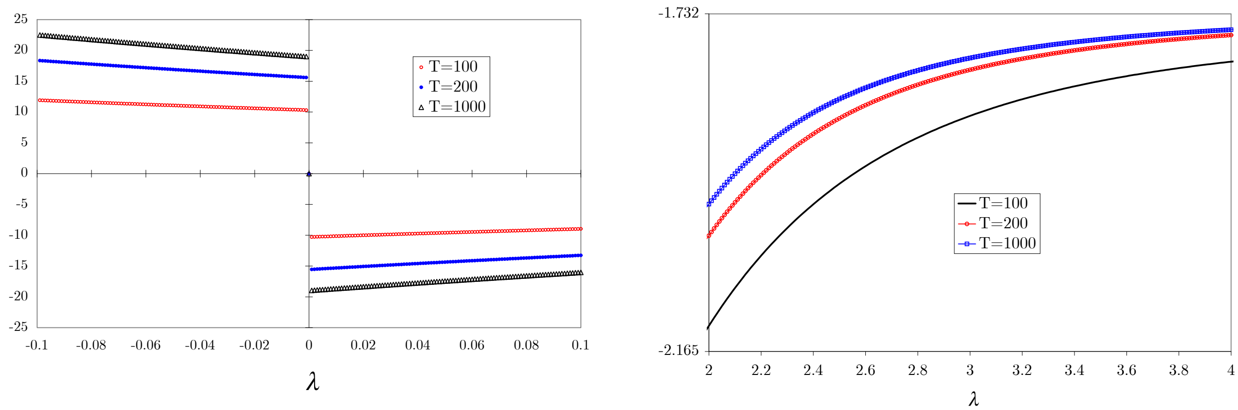
When a 5% one-sided test is used, the critical value of the test for convergence is -1.65. Then, even if $\lambda \to \infty$ and convergence is extremely fast (making convergence in the data extremely hard to detect because of the effective small sample property of the convergence behavior), the maximum value of the t-ratio $t_{\hat{\phi}_{nT}}$ is $-\sqrt{3} = -1.73$, which is significant at the 5% level. Hence, although the the t-test is not consistent in this case, it is still capable of detecting convergence with high probability asymptotically even under these difficult conditions. When $\lambda \in (0, 1)$, the test is consistent for convergence behavior and when $\lambda < 0$ the test is consistent for divergence as $(n, T) \to \infty$ irrrespective of the behavior of the ratio n/T.

5 Numerical Calculations

To demonstrate the contrasting test behavior under the alternatives of convergence and divergence, we use the following numerical calculation. When $n \to \infty$ the probability limit of K_{nt} under M1 - M3 is the following deterministic function of t

$$\text{plim}_{n\to\infty} K_{nt} = K_t = a + \eta_t = a + bt^{-\lambda}, \tag{44}$$

for some non-zero constants a and b. We calculate the t-ratio under this asymptotic $(n \to \infty)$ deterministic DGP (44) for various sample sizes T and refer to it as the t_T^{∞}-ratio. Figure 3 shows how t_T^{∞} behaves for various values of λ. In the vicinity of $\lambda \sim 0$, Panel A of Figure 3 shows that $t_T^{\infty} \to \pm\infty$ as $T \to \infty$, according as $\lambda \lesseqgtr 0$. The distinction between the two alternatives is strongly evident, even for $T = 100$. Panel B of Figure 3, shows the behavior of t_T^{∞} as λ increases for various values of T. The approach of t_T^{∞} to the asymptote $-\sqrt{3}$ as $\lambda \to \infty$ is clearly evident and becomes stronger as T increases.



Panel A: Behavior in the vicinity of $\lambda = 0$ Panel B: Behavior as $\lambda \to \infty$

Figure 3: Asymptotic behavior of the t_T^∞ ratio ($\kappa = 1/3$, $\alpha = \beta$, $\sigma_\mu^2 = \sigma_\epsilon^2 = 1, n \to \infty$)

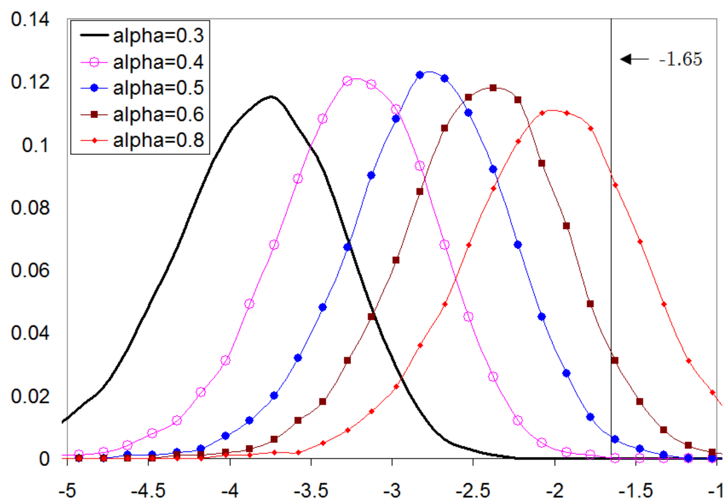


Figure 4: Empirical distribution of $t_{\hat{\phi}_{nT}}$ under M1
($n = 1000$, $T = 100$, $\sigma_{a\mu} = 0$, $\sigma_a^2 = \sigma_\mu^2 = 1$, $\epsilon_{it} \sim iid\mathcal{N}(0,1)$, $\kappa = 1/3$)

To explore behavior of the test in the vicinity $\lambda \sim 0$, Figure 4 plots the density of the t-ratio for various values of α in model M1 with $n = 1000$ and $T = 100$. We set $\sigma_{a\mu} = 0$, $\sigma_\mu^2 = 1$ and $\kappa = 1/3$ in (17) and use draws of $\epsilon_{it} \sim iid\mathcal{N}\left(0, \sigma_\epsilon^2\right)$, $\mu_i \sim iid\mathcal{N}\left(0, \sigma_\mu^2\right)$, and $a_i \sim iid\mathcal{N}\left(0, \sigma_a^2\right)$ with 50,000 replications. Evidently for $\alpha = 0.5$ the density lies almost completely to the left of the 5% critical value -1.65 even for the moderate time series sample size $T = 100$. For $\alpha = 0.3, 0.4$, the distribution shifts further to the left and the test is even more powerful, whereas for $\alpha \geq 0.5$, the distribution moves to the right and the rejection frequency starts to decline. Test power continues to decline as α departs further from 0.5. The same pattern applies as n or T increases.

Figure 5 shows the power function over a range of α values for different n, with $T = 50$, $\sigma_\mu^2 = 4$, $\sigma_\epsilon^2 = 1$, and $100,000$ replications. Rapid movements in the power function occur around $\alpha = 0$ as the model parameter changes from a divergent alternative through the null hypothesis ($\alpha = 0$) to a convergent alternative. Observe that for moderate values of α with $\alpha < 1$ (equivalently $\lambda < 2$)

the power function is close to unity. But when $\lambda \geq 2$, the convergence rate is fast and, as discussed above, the discriminatory power of the test is reduced because of an effective small sample problem. Indeed, for Model M1 with $\lambda = 2$ (i.e. $\alpha = 1$) the half life in mean levels is just one period and the half life in the variation is less than one period[10]. Nevertheless, even in this rather extreme case of rapid convergence, test power is well in excess of 50%.

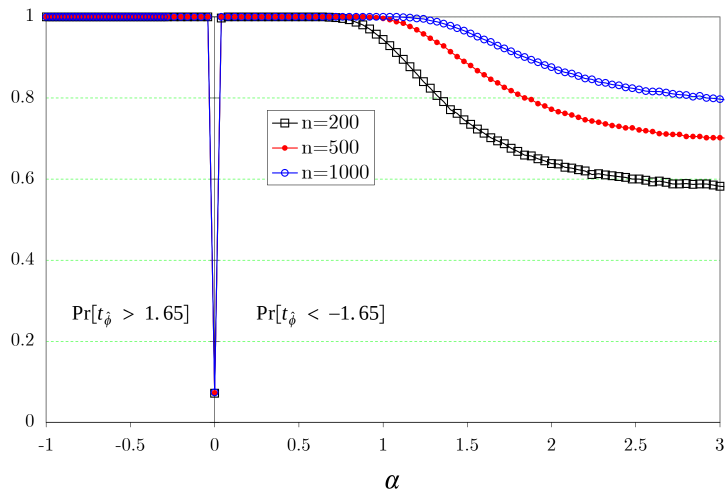


Figure 5: Test Rejection Frequencies over $\lambda = 2\alpha$ in model M1
($T = 50$, $\sigma_{a\mu} = 0$, $\sigma_\mu^2 = 4$, $\sigma_a^2 = \sigma_\epsilon^2 = 1$, $\kappa = 1/3$)

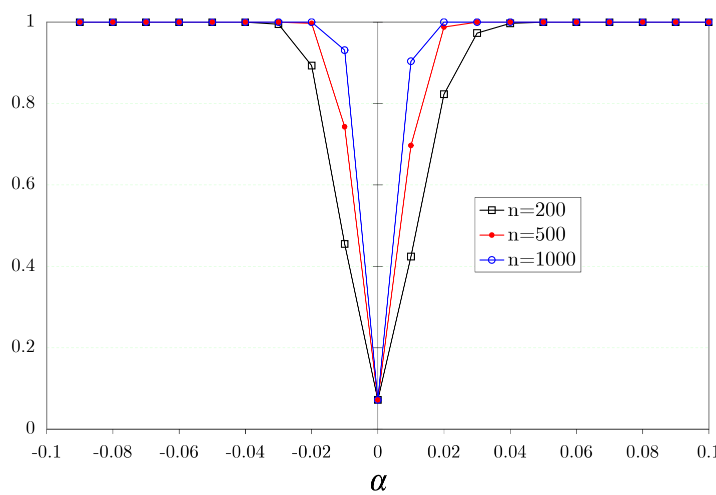


Figure 6: Test power curves near $\alpha = 0$ for various n in model M1

[10] The mean level in model M1 has the form $\mathbb{E}(x_{it}) = a + \mu t^{-\alpha} \underset{t\to\infty}{\rightarrow} a$, when $\alpha > 0$. Then $\mathbb{E}(x_{i1} - x_{i2}) = \mu\left(1 - 2^{-\alpha}\right)$ $= 2^{-\alpha}\left(2^{\alpha} - 1\right)\mathbb{E}(x_{i1} - x_{i\infty}) = \mathbb{E}(x_{i1} - x_{i\infty})/2$ for $\alpha = 1$ and the half life in mean level from $t = 1$ is just one period when $\alpha = 1$. The limiting variation when $n \to \infty$ has the form $K_{\infty,t} = b + \psi t^{-2\alpha}$, so that $K_{\infty,1} - K_{\infty,2}$ $= \psi\left(1 - 2^{-2\alpha}\right) = 2^{-2\alpha}\left(2^{2\alpha} - 1\right)(K_{\infty,1} - K_{\infty,\infty}) = (3/4)(K_{\infty,1} - K_{\infty,\infty})$, and the half life in the variation $K_{\infty,t}$ from $t = 1$ is less than one period.

As is apparent in Figure 5, the test rejection frequency changes rapidly from the nominal 5% at the null where $\alpha = 0$ to virtually 100% for even small departures from the null. This behavior in the power function is sensitive, at least in the immediate vicinity of $\alpha = 0$ to the extent of cross section averaging. To demonstrate, Figure 6 magnifies the region around $\alpha = 0$ from Figure 5 to reveal the extent of this sensitivity to the cross section sample size n. Evidently, with greater cross section information as n increases, the distinction between the null and the alternative becomes more sharply defined, increasing test power as expected.

Similar features to those discussed above apply for tests based on data generated by models M2 and M3. These findings are given in the Appendix S as supplementary material to this paper.

6 Monte Carlo Simulations

We investigate the finite sample performance of the trend regression test of convergence and divergence using the following data generating process

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

where

$$\begin{aligned} a_i &\sim iid\mathcal{N}\left(0, \sigma_a^2\right), \ \mu_i \sim iid\mathcal{N}(0,1), \\ \epsilon_{it} &= \rho_i \epsilon_{it-1} + v_{it}, \ v_{it} \sim iid\mathcal{N}(0,1), \ \rho_i \sim U[0, 0.5], \end{aligned}$$

and $\theta_i \sim iid\mathcal{N}(0,1)$ or $\theta_i = 1$ for all i. The fixed parameter settings are: $\sigma_a \in [1, 2, 5, 10]$, and $\alpha, \beta \in [-0.1, 0, 0.1, 0.5]$. The experimental design for each model and restrictions on the parameter values are as follows:

Model M1: ($\beta = 0$) We take the case where $\theta_i = 1$ for all i as the case $\theta_i \neq \theta_j$ for $i \neq j$ is considered in Model 2. This model is useful in studying panel data convergence when cross sectional dependence is homogeneous (here via the common factor F_t). We consider two cases depending on the value of $\sigma_{a\mu}$, one case with $\sigma_{a\mu} = 0.45$ and the other case with $\sigma_{a\mu} = 0$. Comparison of these cases highlights the impact of $\sigma_{a\mu}$ on test performance where asymptotics are known to be affected through the differing values of the rate parameter λ (see Table M).

Model M2: ($\alpha = 0$) Two cases are considered. In the first case $\theta_i = 1$ for all i, whereas in the second case θ_i is generated from $iid\mathcal{N}(0,1)$ and idiosyncratic components must be estimated to eliminate common factor F_t. More specifically, we use estimates of x_{it} defined by

$$\hat{x}_{it} = y_{it} - \hat{\theta}_i \hat{F}_t,$$

where $\hat{\theta}_i$ and $\hat{F}_t$ are obtained by principal component methods. In this experiment, the number of common factors is assumed to be known. Bai and Ng (2002) showed that the number of common factors can be sharply determined by suitable information criteria when sample sizes of n and T are moderate and this was confirmed in our simulations in the present case, so these results are not reported.

Model M3: $(\alpha = \beta)$ For brevity, we consider only the case $\alpha = \beta$. Simulation results for other cases are available online[11]. As in Model M1, we consider two cases depending on the value of $\sigma_{a\mu}$.

Table 1 reports size and power of the one-sided convergence test in model M1 with settings $\kappa = 1/3$ and $L = \text{int}(T^{\kappa})$ in the long run variance calculation. When $\alpha < 0$ or $\beta < 0$, the size of the one-sided test is expected to be zero and this is confirmed in Table 1 (with $\alpha = -0.1$) and in Table 2 (with $\beta = -0.1$) for model M2. Moreover, test size in M1 and M2 is very similar, again as expected because of the null hypothesis setting $\alpha = \beta = 0$. The Table 1 results show that test power is dependent on $\sigma_{a\mu}$. When $\sigma_{a\mu} \neq 0$, the test is consistent when $(n, T) \to \infty$ irrespective of the n/T ratio if $\alpha < 0.5$, as demonstrated in the Appendix. Otherwise, test power increases with n but may decrease as T increases with n fixed. For example, when $\alpha = 0.3, 0.5$ and $\sigma_{a\mu} = 0$, test power decreases as T increases for any fixed n. This is explained by the fact that when $\sigma_{a\mu} = 0$ the decay parameter $\lambda = 2\alpha > 0.5$ in these experiments, so that convergence is faster and discriminatory power is correspondingly reduced as T increases with n fixed. On the other hand, when $\sigma_{a\mu} \neq 0$, the power of the test increases as T increases.

Table 2 shows test size in model M2, which is comparable with that of Table 1 for model M1. When $\beta = -0.1$, the test size is virtually zero, which is expected for the one-sided test because the t-ratio tends to infinity in this case and large positive values of the statistic are expected. When $\beta = 0$, there is some mild size distortion for small T, which does not seem to rise or fall as n increases, but which diminishes quickly as T increases. Test size does not seem sensitive to σ_a^2 or when estimated idiosyncratic elements are estimated, which perhaps to be expected given the robust limit theory in Theorem 2.

Table 3 reports test power for model M2. Interestingly, power is smaller for $\beta = 0.1$ than when $\beta = 0.5$. The test statistic densities reveal (see Figure S1) that as β increases the variance of the t-ratio decreases but at the same time the mean of the t-ratio decreases in absolute value. This reduction in variance of the test statistic seems to affect finite sample power performance more than reduction in mean. Also, Table 3 shows that test power decreases as the variance of a_i increases, which is explained by the fact that as σ_a^2 increases there is greater fluctuation in the panel data level for all t, and this induced noise reduces discriminatory power in the test. When $\theta_i \sim iid\mathcal{N}(0, 1)$ and idiosyncratic components are estimated, test power is similar to the fixed $\theta_i = 1$ case. In general, the findings show that as long as $\beta < 1$ test power increases with T for fixed n and increases as n increases for fixed T.

Table 4 shows test power for model M3. Test size is not reported in this case becauses the results are very similar to those of models M1 and M2 and we report only the case where $\alpha = \beta$ as the results are similar for other cases. The main finding is that test power increases as n increases regardless of the value of $\sigma_{a\mu}$ and generally increases as T increases for fixed n. The exception occurs when $\sigma_{a\mu} = 0$ and $\alpha = \beta = 0.5$ where there is evidence of a minor attentuation in power as T increases, which is explained as earlier by the fact that when $\sigma_{a\mu} = 0$ the decay parameter $\lambda = 2\alpha > 0.5$ and test discriminatory power is reduced because of the faster convergence rate and the implied small sample effect as T increases with n fixed.

7 Empirical Examples

We demonstrate two practical applications of the proposed test. The first data set is a balanced panel consisting of 90 disaggregated personal consumption expenditure (PCE) items. The second

[11] www.utdallas.edu/~d.sul/papers/Monte_res_9_17.xls

application involves a balanced pseudo-panel data set. The proposed test remains valid in pseudo-panels as long as the sample cross sectional variation approximates well the true cross sectional variance in each time period.

7.1 Weak $\sigma-$Convergence with 90 PCE inflation Rates

Here we report a very interesting empirical fact about the weak $\sigma-$convergence with 90 disaggregate PCE inflation rates. The source of the data is the annual PCE (Table 2.5.5) and our full data set covers 90 disaggregated series over the period 1960 to 2014. During the latter sample period 1992 to 2014, PCE inflation rates experienced much smaller variation than in the 1970s and 1980s.

Following the common factor literature, we assume that the PCE inflation rates have a static factor structure of the form

$$\pi_{it} = a_i + \theta_i' F_t + \pi_{it}^o, \tag{45}$$

with common factors F_t, factor loadings θ_i, individual series fixed effects a_i, and idiosyncratic inflation rate π_{it}^o. Our main concern is whether or not the idiosyncratic components of the 90 disaggregated PCE inflation rates manifest weak $\sigma-$convergence over time. We start by estimating the number of the static common factors using Bai and Ng's (2002) IC_2 criterion (up to a potential maximum of 8 factors). Three factors are found over the entire sample period from 1960 to 2014. Next, we obtain estimates of the idiosyncratic components by using principal components.[12]

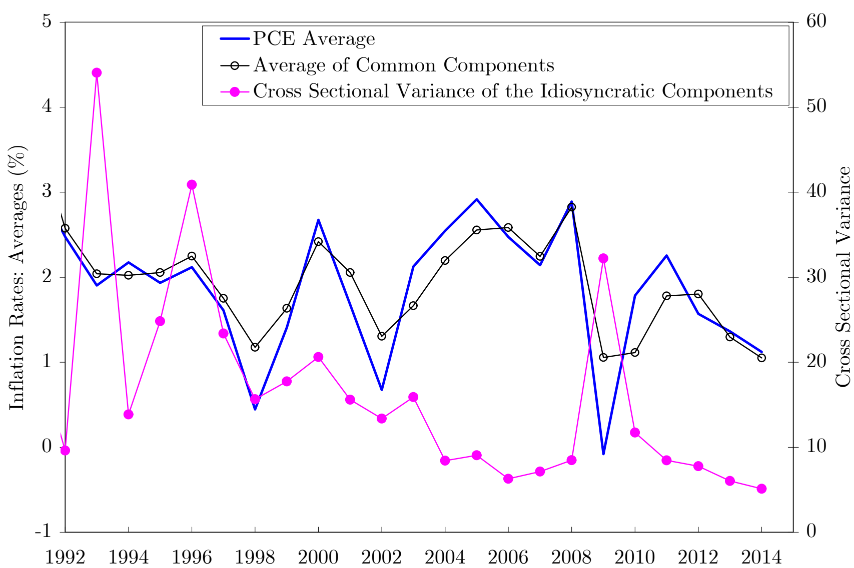


Figure 7: Cross Sectional Means and Variances of 90 PCE items

Figure 7 plots the PCE average inflation rates (heavy dark blue line) for the 90 disaggregated series, the average of the estimated common components (thin black line with empty circles), and

[12] In determining the number of the common factors, we standardize the sample observations for each i (dividing π_{it} by its standard deviation for each i) before calculating the IC_2 criterion and estimating the common factors. Let $\hat{F}_t$ be the principal component estimates obtained from the standardized sample. Once the common factors are estimated, the factor loadings are estimated by regression of the original sample data, π_{it}, on a constant and $\hat{F}_t$ (45) for each i. The final estimated idiosyncratic components are calculated by taking residuals $\hat{\pi}_{it}^r = \pi_{it} - \hat{\theta}_i' \hat{F}_t$, so that fixed effects are embodied in $\hat{\pi}_{it}^r$. That is, $\hat{\pi}_{it}^r = a_i + \pi_{it}^o + \theta_i' F_t - \hat{\theta}_i' \hat{F}_t$.

the sample variance of the estimated idiosyncratic components (thin pink line with solid circles) over the period 1992 - 2014. Evidently, the cross sectional variance is generally decreasing over this time period but with some fluctuations, most notably during 2009.

Table 5 reports the weak $\sigma-$convergence test results with the whole sample (from 1960 to 2014) and two subsamples (before and after 1992). For the sample after 1992, the null of no $\sigma-$convergence is rejected even at the 1% level. Two different lag truncation parameter settings ($L = 3, 6$) were used in the construction of the long run variance estimates used in the tests and, as is apparent in the table, the test outcomes and evidence for $\sigma-$convergence in the data are robust to lag choice. The selected common factor dimension (k) is also varied from 1 to 8, and again all cases support evidence for $\sigma-$convergence.

Test results for the sample prior to 1992 and for the entire sample are very different. It is well known that inflation rates reached a peak in the early1980s and displayed time series wandering characteristics over the 1970s and 1980s. Common factors to inflation rates estimated for the 1970s and 1980s therefore tend to behave rather like random walks and, using the entire sample of data, it is hard to reject the null of a unit root in the inflation rates. If the series are integrated, then the null of no $\sigma-$convergence should not be rejected, as discussed earlier in the paper. Application of the convergence test confirms this intuition. As is evident from Table 5, irrespective of the choice of k and L, the null of no $\sigma-$convergence is not rejected in any case either for the whole sample or the subsample from 1960 to 1992. For the latter subsample regressions, the t-ratios all exceed the right side critical value 1.65, leading to the conclusion that inflation rates before 1992 were diverging. For the entire sample, there is no evidence for divergence (and the t ratio is negative for most choices of k) and the null of no $\sigma-$convergence is not rejected in any of the parameter settings.

Table 6 shows the trend regression test results for various starting years and with various lag parameter settings of L and with $k = 3$. As the starting year rises the number of time series observations T declines. But even with much smaller values of T, the null of no $\sigma-$convergence is rejected in all cases. The table also shows that average (absolute) cross sectional correlation also increases as we move closer to the end of the sample by raising the starting date in the regression.

7.2 Convergence in Ultimatum Games

One of the most studied games in experimental economics is the ultimatum game. A standard ultimatum game consists of two players: a leader (proposer) and a follower (responder). The leader offers a portion (x) of a fixed pie (money) to the follower. If the offer is accepted, then the pie is divided as proposed. Otherwise, both players receive nothing. The game theory prediction on the optimal offer is near zero since all positive offers are expected to be accepted. Since the pioneering study by Güth, Schmittberger and Schwartz (1982), more than 2,000 experimental studies have shown that leaders usually offer around 40% of the pie, and offers lower than 30% of the pie are often rejected. See Güth (1995), Bearden (2001), Cooper and Kagel (2013) and Cooper and Dutcher (2011) for surveys of this literature.

A natural question is whether offers tend to converge over rounds in repeated games. We use the experimental data from Ho and Su (2009) to examine evidence for the convergence. Ho and Su ran 24 rounds of Ultimatum games with 4 sections. Each section had between 15 and 21 subjects, and each subject played the game 24 times. For each round subjects were randomly matched with others. So one subject could be a follower in one round, but become a leader in another round. For each round, there are three players in the Ho-Su experiment: one leader and two followers. From their data, we form a pseudo panel of 25 subjects over 24 rounds. Figure 8 shows the cross sectional average and variance over rounds. Interestingly, the offer fraction seems to follow a slow decaying function: initial offers were slightly higher than 40%, but with more rounds the offers seem to fall and stabilize slightly above 30%. Cross sectional variation clearly fluctuates but is evidently slowly decreasing over time.

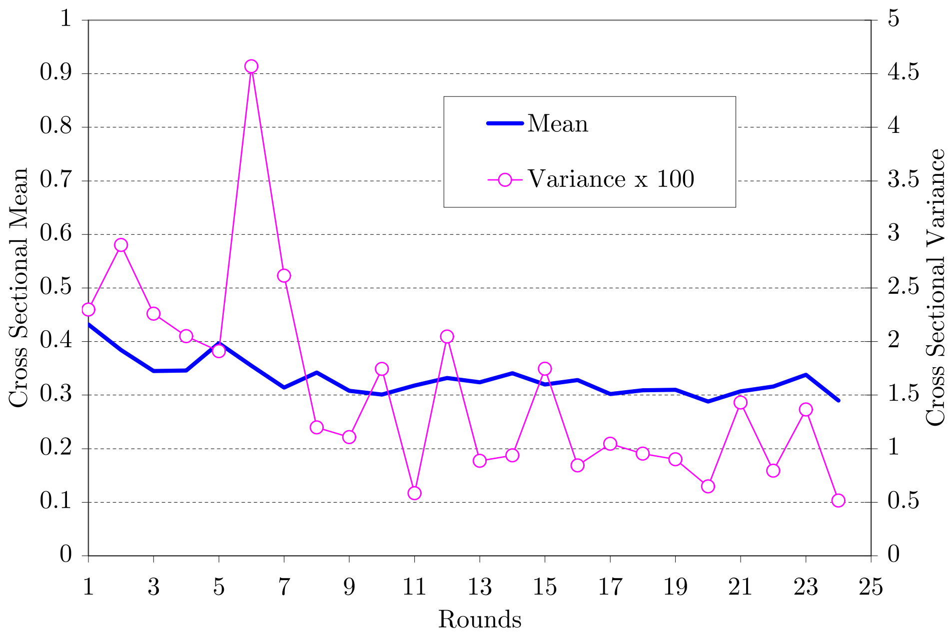


Figure 8: Cross Sectional Average and Variance
(From Ho & Su (2009))

We ran trend regressions with the cross sectional variance from these data. The results are reported in Table 7 and allow for various starting points in the regression. When the initialization is set at the round 1 game, the point estimate is $\hat{\phi}_{nT} = -0.087$ with t-ratios $t_{\hat{\phi}_{nT}}(L) \leq -4.299$ for all values $L \in \{1, 3, 5, 7\}$ of the lag truncation parameter. The null hypothesis of no $\sigma-$convergence is therefore rejected even at the 0.1% level. This finding confirms that as the ultimatum game is repeated, cross section variation in the offer rates declines. In further investigation, the trend regression was performed with initializations set at later rounds of the game. Due to the high peak in the variance at round 6, the point estimates $\hat{\phi}_{nT}$ remain close to the same level -0.09 until the 6th round sample observations are discarded. Commencing from later initializations, the regression point estimates drop to -0.05 and show evidence of some further decline thereafter. Nonetheless, the t-ratios all lead to rejections of the null of no $\sigma-$convergence at close to the 1% level.

8 Conclusion

Concepts of convergence have proved useful in studying economic phenomena at both micro and macro levels and have wider applications in the social, medical, and natural sciences. Of particular interest in empirical work is whether given data across a body of individual units show a tendency toward convergence in the sense of a persistent diminution in their variation over time, an idea that was clearly articulated by Hotelling (1933) in the header to this article. The concept of weak

$\sigma-$convergence introduced in the present paper gives analytic characterization to this concept and, more importantly for implementation, one that is amenable to convenient econometric testing. The approach relies on a simple linear trend regression which is correctly specified only when the data is subject to no change or evolution over time, but which leads to a statistical test of convergence that has discriminatory power when there is either diminution or dilation of variation over time.

When a system is disturbed and cross section variation is affected, the convergence test is an empirical mechanism for assessing whether the disturbances influence the system over time in a directional manner that diminishes or raises variance. In the event that there is no directional impact, the slope coefficient in the trend regression is zero and the test does not register any evolutionary change. But if the disturbances are neutralized and variation is reduced over time, the estimated slope coefficient is negative and the test registers diminution in variance even when the precise mechanism is unknown. When the directional impact is positive and variation rises over time, the estimated slope coefficient is positive and the test registers rising variation. Asymptotic theory in the paper justifies this simple approach to testing convergence and divergence in panel data when the underlying stochastic processes are unknown but fall within some general categories of models with evaporating or dilating trends in variation.

The methodology applies whether or not the observed data are cross sectionally dependent, under general regularity conditions for which a law of large numbers holds. Moreover, the data may be drawn from panels or pseudo-panels where observations may relate to different individuals or cross sectional units in each time period. The main technical requirement on the panel is that the respective sample sizes $(n, T) \to \infty$ and that $\frac{n}{T} \to \infty$, although the latter rate condition is not always required. Simulations show that the methods provide good discriminatory power in most cases of convergence and divergence, even when the time series sample and cross section sample sizes are of comparable size.

References

[1] Andrews, D. (2005). Cross section regression with common shocks. *Econometrica*, 73, 1551–1585.

[2] Avrahami, J., W. Güth, R. Hertwig, Y. Kareev, and H. Otsubo (2010). Learning (not) to Yield: an Experimental Study of Evolving Ultimatum Game Behavior. *Jena Economic Research Papers,* 2010-092.

[3] Baek, Y. I., J. S. Cho, and P. C. B. Phillips (2015). Testing Linearity using Power Transforms of Regressors. *Journal of Econometrics*, 187, 376-384.

[4] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71, 135–172.

[5] Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrics,* 70(1), 191-221.

[6] Barro, R.J. (1991). Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics,* 106, 407-43.

[7] Barro, R.J. and X. Sala-I-Martin (1991). Convergence Across States and Regions. *Brookings Papers on Economic Activity,* 1, 107–158.

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

[9] Bearden, N. (2001), Ultimatum Bargaining Experiments: The State of the Art. *Decision Sciences.*

[10] Chaudhuri, A. (2011). Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature. *Experimental Economics*, 14(1), 47–83.

[11] Cho, J. S., and P. C. B. Phillips (2015). Sequentially Testing Polynomial Regression Hypotheses using Power Transforms of Regressors. Working Paper, Yale University.

[12] Cooper, D. and J. Kagel (2013). Other Regarding Preferences: A Selective Survey of Experimental Results. *Handbook of Experimental Economics,* 2.

[13] Cooper, D. and G. Dutcher (2011). The dynamics of responder behavior in ultimatum games: a meta-study. *Experimental Economics.* 14, 519–546.

[14] Durlauf, S. N., and D. T. Quah (1999). The New Empirics of Economic Growth. *Handbook of Macroeconomics,* Vol. 1A, Chap. 4.

[15] Evans, P. (1996). Using Cross-country Variances to Evaluate Growth Theories. *Journal of Economic Dynamics and Control,* 20, 1027-1049.

[16] Evans, P. (1998). Using Panel Data to Evaluate Growth Theories. *International Economics Review,* 39, 295–306.

[17] Evans, P. and G. Karras (1996). Convergence Revisited. *Journal of Monetary Economics*, 37, 249-265.

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

[19] Gale, J., K. Binmore, and L. Samuelson (1995). Learning to be Imperfect: the Ultimatum Game. *Games and Economic Behavior.* 8, 56–90.

[20] Güth, W. (1995). On Ultimatum Bargaining Experiments – a Personal Review. *Journal of Economic Behavior and Organization,* 27, 329–344.

[21] Ho, T. and X. Su (2009). Peer-Induced Fairness in Games. *American Economic Review,* 2022-2049.

[22] Hobijn, B. and P.H. Franses (2000). Asymptotically perfect and relative convergence of productivity. Journal of Applied Econometrics 15, 59–81

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

[24] Kac, V. and P. Cheung (2002). Euler-Maclaurin Formula. *Quantum Calculus,* Springer New York, 92-98.

[25] Ledyard, J. (1995). Public Goods: A Survey of Experimental Research. *The Handbook of Experimental Economics.* Roth and Kagel, eds. Princeton, NJ: Princeton University Press.

[26] Malinvaud, E. (1970). The Consistency of Nonlinear Regressions. *The Annals of Mathematical Statistics,* 41(3), 956-969.

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

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

[29] Phillips, P. C. B. (2005). Challenges of Trending Time Series Econometrics. *Mathematics and Computers in Simulation*, 68, 401-416.

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

[31] Phillips, P.C.B., S. Shi, and Y. Yu (2015). Testing for Multiple Bubbles: Historical episodes of exuberance and collapse in the S&P 500, *International Economic Review*, 56, 1077-1131.

[32] Phillips P. C. B., Y. Wu and J. Yu (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?. *International Economic Review,* 52, pp. 201-226.

[33] Phillips, P. C. B. and V. Solo (1992). Asymptotics for Linear Processess, *The Annals of Statistics*, 20, 971–1001.

[34] Phillips, P. C. B. and D. Sul (2007). Transition modeling and econometric convergence tests. *Econometrica,* 75(6), 1771-1855.

[35] Quah, D. (1996). Empirics for Economic Growth and Convergence, *European Economic Review*, 40, 1353–1375.

[36] Wu, Chien-Fu (1981). Asymptotic Theory of Nonlinear Least Squares Estimation. *The Annals of Statistics,* 9(3), 501-513.

Appendix

The following lemmas are useful in establishing Theorems 1 and 2. Proofs are given in the Supplementary Appendix S.

Lemma 1

Finite series of sums of powers of integers have the following asymptotic forms as $T \to \infty$

$$\tau_T(\alpha) = \sum_{t=1}^{T} t^{-\alpha} = \begin{cases} \frac{1}{1-\alpha} T^{1-\alpha} + O(1) & \text{if } \alpha < 1, \\ \ln T + O(1) & \text{if } \alpha = 1, \\ \mathcal{Z}_T(\alpha) = O(1) & \text{if } \alpha > 1, \end{cases}$$

$$H_T(\alpha, \ell) = \sum_{t=0}^{T} (t+\ell)^{-\alpha} = \begin{cases} \frac{1}{1-\alpha} (T+\ell)^{1-\alpha} + O(1) & \text{if } \alpha < 1, \\ \ln(T+\ell) - \ln \ell + O(1) & \text{if } \alpha = 1, \\ \zeta_T(\alpha, \ell) = O(1) & \text{if } \alpha > 1, \end{cases}$$

where, for $\alpha > 1$, $\ell \geq 1$,

$$\begin{aligned} \mathcal{Z}_T(\alpha) &\to \zeta(\alpha) = \sum_{t=1}^{\infty} \frac{1}{t^\alpha} = \frac{1}{\alpha-1} + \frac{1}{2} + \Delta_\alpha, \\ \zeta_T(\alpha, \ell) &\to \zeta(\alpha, \ell) = \sum_{t=0}^{\infty} \frac{1}{(t+\ell)^\alpha} = \frac{1}{\ell^\alpha} + \frac{1}{(1+\ell)^\alpha} \left(\frac{1}{2} + \frac{1+\ell}{\alpha-1} \right) + \Delta_{\alpha,\ell}. \end{aligned}$$

with Δ_α and $\Delta_{\alpha,\ell}$ defined below in (63) and (64) in the supplementary appendix, and where $\zeta(\alpha, \ell) \leq \zeta(\alpha)$ for all integer $\ell \geq 1$.

Lemma 2

Define $\tilde{t} = t - T^{-1} \sum_{t=1}^{T} t$, $\widetilde{t^{-\alpha}} = t^{-\alpha} - T^{-1} \sum_{t=1}^{T} t^{-\alpha}$, $\mathcal{T}_T(1, \alpha) = \sum_{t=1}^{T} \tilde{t}\widetilde{t^{-\alpha}}$, $\mathcal{S}_T(\alpha) = \sum_{t=1}^{T} \widetilde{t^{-\alpha}}\widetilde{t^{-\alpha}}$, and $\mathcal{B}_T(\alpha) = \frac{1}{T} \sum_{t=1}^{T} \left[\widetilde{t^{-\alpha}} - \tilde{t} \left(\sum \tilde{t}^2 \right)^{-1} \sum \tilde{t}\widetilde{t^{-\alpha}} \right]^2$. Then, as $T \to \infty$, we have

$$\mathcal{T}_T(1, \alpha) = \begin{cases} -\frac{\alpha}{2(\alpha-2)(\alpha-1)} T^{2-\alpha} + O\left(T^{1-\alpha}\right) & \text{if } \alpha < 1, \\ -\frac{1}{2} T \ln T + O(T) & \text{if } \alpha = 1, \\ -\frac{1}{2} \zeta(\alpha) T + O(1) & \text{if } \alpha > 1, \end{cases}$$

$$\mathcal{S}_T(\alpha) = \begin{cases} \frac{\alpha^2}{(\alpha-1)^2 (1-2\alpha)} T^{1-2\alpha} + O(1) & \text{if } \alpha < 1/2, \\ \ln T + O(1) & \text{if } \alpha = 1/2, \\ \zeta(2\alpha) + O\left(T^{-1}\right) & \text{if } \alpha > 1/2, \end{cases}$$

and

$$\mathcal{B}_T(\alpha) = \begin{cases} \frac{\alpha^2}{(\alpha-1)^2 (1-2\alpha)} T^{-2\alpha} + O\left(T^{-1}\right) & \text{if } \alpha < 1/2, \\ T^{-1} \ln T + O\left(T^{-1}\right) & \text{if } \alpha = 1/2, \\ T^{-1} \zeta(2\alpha) + o\left(T^{-1}\right) & \text{if } \alpha > 1/2, \end{cases} = \begin{cases} O\left(T^{-2\alpha}\right) & \text{if } \alpha < 1/2, \\ O\left(T^{-1} \ln T\right) & \text{if } \alpha = 1/2, \\ O\left(T^{-1}\right) & \text{if } \alpha > 1/2. \end{cases}$$

Lemma 3:

Let v_{it} be cross section independent over i and covariance stationary over t with mean zero and autocovariogram $\gamma_{h,v,i} = \mathbb{E}(v_{it}v_{it+h})$ satisfying the summability condition

$$\sum_{h=1}^{\infty} h\left|\gamma_{h,v,i}\right| < \infty, \tag{46}$$

for all i. Suppose $b_i \sim iid\left(0, \sigma_b^2\right)$. Then

$$\begin{aligned}
\sum_{t=1}^{T} v_{it}t^{-\alpha} &= O_p\left(\left[\tau_T(2\alpha)\right]^{1/2}\right), \\
\sum_{t=1}^{T} v_{it}\tilde{t}t^{-\alpha} &= O_p\left(T\left[\tau_T(2\alpha)\right]^{1/2}\right), \\
\sum_{t=1}^{T} b_i\tilde{t}t^{-\alpha} &= O_p\left(\mathcal{T}_T(1,\alpha)\right).
\end{aligned}$$

Lemma 4:

Let $m_t = t^{-\lambda} - t\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}$ and $L = \lfloor T^{\kappa}\rfloor$ for some $\kappa \in (0,1)$. *Then for $\lambda > 0$*

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

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

where $\tilde{m}_t = m_t - \frac{1}{T-\ell}\sum_{s=1}^{T-\ell} m_s$, $\tilde{m}_{t+\ell} = m_{t+\ell} - \frac{1}{T-\ell}\sum_{s=1}^{T-\ell} m_{s+\ell}$, $\zeta(\lambda,\ell)$ is the Hurwitz zeta function and $\zeta(2\lambda)$ is the Riemann zeta function

Lemma 5

Suppose $b_i \sim iid\left(b, \sigma_b^2\right)$. Let $\xi_{b,n} = n^{-1}\sum_{i=1}^{n}(b_i - b)$. Then as $n, t \to \infty$ with $n/T \to \infty$, we have

$$\xi_{b,n}\widetilde{t^{-\alpha}} = o_p\left(\widetilde{t^{-2\alpha}}\right). \tag{47}$$

which may be expressed in the simple form that $\tilde{\xi}_{n,t} = o_p(\tilde{\eta}_t)$ uniformly in t as $n/T \to \infty$. .

Proof of Theorem 1: The Asymptotic Limit of $\hat{\phi}_{nT}$

To analyze the asymptotic behavior of the trend regression coefficient $\hat{\phi}_{nT}$ we use the convenient decomposition (33), viz.,

$$\hat{\phi}_{nT} = \sum_{t=1}^{T} a_{tT}\eta_t + \sum_{t=1}^{T} a_{tT}\xi_{n,t} + \sum_{t=1}^{T} a_{tT}\varepsilon_{n,t} =: I_A + I_B + I_C,$$

where

$$a_{tT} = \frac{\tilde{t}}{\sum_{s=1}^{T} \tilde{s}^2} = \frac{t - T^{-1}\sum_{s=1}^{T} s}{T^3 \times \frac{1}{T^3}\sum_{s=1}^{T} \tilde{s}^2} = \frac{12}{T^3}\left(t - \frac{T+1}{2}\right)\left\{1 + O\left(T^{-1}\right)\right\},$$

The dominant term in η_t,denoted by $\eta_{t,d}$, can be classified according to the three models as follows

Case	M1	M2	M3
$\alpha, \beta > 0$,and $\sigma_{a\mu} \neq 0$	$2\sigma_{a\mu}t^{-\alpha}$	$\sigma_\epsilon^2 t^{-2\beta}$	$2\sigma_{a\mu}t^{-\alpha} + \sigma_\epsilon^2 t^{-2\beta}$
$\alpha, \beta > 0$,and $\sigma_{a\mu} = 0$	$\sigma_\mu^2 t^{-2\alpha}$	$\sigma_\epsilon^2 t^{-2\beta}$	$\sigma_\mu^2 t^{-2\alpha} + \sigma_\epsilon^2 t^{-2\beta}$
$\alpha < 0$ or $\beta < 0$	$\sigma_\mu^2 t^{-2\alpha}$	$\sigma_\epsilon^2 t^{-2\beta}$	$\sigma_\mu^2 t^{-2\alpha} + \sigma_\epsilon^2 t^{-2\beta}$

Using the general form $\eta_{t,d} = bt^{-\lambda}$,where λ represents the decay parameter and b is the corresponding coefficient in that term, we first obtain the following expression for I_A. We specify b and λ later in the case of each individual model.

$$\begin{aligned} I_A &= \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \sum_{t=1}^{T} \tilde{t}\eta_t \sim b\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \sum_{t=1}^{T} \tilde{t}t^{-\lambda} \\ &= -6b \times \begin{cases} \frac{\lambda}{(\lambda-2)(\lambda-1)} T^{-1-\lambda} + O\left(T^{-2-\lambda}\right) & \text{if } \lambda < 1, \\ T^{-2}\ln T + O\left(T^{-2}\right) & \text{if } \lambda = 1, \\ T^{-2} Z_T(\lambda)\left\{1 + O\left(T^{-1}\right)\right\} & \text{if } \lambda > 1. \end{cases} \end{aligned} \tag{48}$$

In particular:

(i) Case $\lambda < 1$

$$\begin{aligned} \sum_{t=1}^{T} a_{tT}t^{-\lambda} &= \frac{\sum_{t=1}^{T} t^{1-\lambda} - \frac{T+1}{2}\sum_{t=1}^{T} t^{-\lambda}}{\sum_{s=1}^{T} \tilde{s}^2} = \frac{\left\{\frac{T^{-\lambda-1}}{2-\lambda} - \frac{T^{-1-\lambda}}{2(1-\lambda)}\right\}\left\{1 + O\left(T^{-1}\right)\right\}}{T^{-3}\sum_{s=1}^{T} \tilde{s}^2} \\ &= -\left(\frac{6\lambda T^{-1-\lambda}}{(2-\lambda)(1-\lambda)}\right)\left\{1 + O\left(T^{-1}\right)\right\} \end{aligned}$$

(ii) Case $\lambda = 1$

$$\begin{aligned} \sum_{t=1}^{T} a_{tT}t^{-\lambda} &= \frac{T - \frac{T+1}{2}\sum_{t=1}^{T} t^{-1}}{\sum_{s=1}^{T} \tilde{s}^2} = \frac{\left\{T^{-2} - \frac{T^{-2}}{2}\ln T\right\}\left\{1 + O\left(T^{-1}\right)\right\}}{T^{-3}\sum_{s=1}^{T} \tilde{s}^2} \\ &= -6T^{-2}\ln T\left\{1 + O\left(T^{-1}\right)\right\} \end{aligned}$$

(iii) Case $\lambda > 1$

$$\begin{aligned} \sum_{t=1}^{T} a_{tT}t^{-\lambda} &= \frac{\sum_{t=1}^{T} t^{1-\lambda} - \frac{T+1}{2}\sum_{t=1}^{T} t^{-\lambda}}{\sum_{s=1}^{T} \tilde{s}^2} = \frac{\frac{T^{2-\lambda}}{2-\lambda}\left\{1 + O\left(T^{-1}\right)\right\} - \frac{T+1}{2} Z_T(\lambda)}{T^3 \times \frac{1}{T^3}\sum_{s=1}^{T} \tilde{s}^2} \\ &= -\left(\frac{\frac{T^{-2}}{2} Z_T(\lambda)\left\{1 + O\left(T^{-1}\right)\right\}}{\frac{1}{T^3}\sum_{s=1}^{T} \tilde{s}^2}\right) = -6T^{-2} Z_T(\lambda)\left\{1 + O\left(T^{-1}\right)\right\} \end{aligned}$$

Next consider $I_B = \sum_{t=1}^{T} a_{tT}\xi_{n,t}$.When $\sigma_{a\mu} \neq 0$, we have

$$I_B = O_p\left(n^{-1/2}\right) \times \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \sum_{t=1}^{T} \tilde{t}t^{-\lambda}$$

Thus, $I_B = O_p\left(n^{-1/2}\right) \times I_A$ and I_A dominates I_B always. When $\sigma_{a\mu} = 0$ (this only influences M1 and M3), term $2\sigma_{a\mu}t^{-\alpha}$ disappears in I_A, but term $2\sigma_{a\mu,n}t^{-\alpha}$ is still present in I_B. Hence, when $\sigma_{a\mu} = 0$,

$$I_B = \begin{cases} O_p\left(n^{-1/2}T^{-3}\right)\mathcal{T}_T(1,\alpha) & \text{for M1,} \\ O_p\left(n^{-1/2}T^{-3}\right)\mathcal{T}_T(1,\lambda^*), \text{ with } \lambda^* = \min[\alpha, 2\beta] & \text{for M3,} \end{cases}$$

since $n^{-1}\sum_{i=1}^{n} \tilde{a}_i\tilde{\mu}_i = O_p\left(n^{-1/2}\right)$ by (31), and

$$\begin{aligned} &\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \sum_{t=1}^{T} \sigma_{a\mu,n}\tilde{t}\widetilde{t^{-\alpha}} \\ = &\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \left(n^{-1}\sum_{i=1}^{n} \tilde{a}_i\tilde{\mu}_i\right)\sum_{t=1}^{T} \tilde{t}\widetilde{t^{-\alpha}} = \begin{cases} O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha < 1, \\ O_p\left(n^{-1/2}T^{-2}\ln T\right) & \text{if } \alpha = 1, \\ O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \alpha > 1. \end{cases} \end{aligned}$$

For term I_C, first recall that

$$\varepsilon_{n,t} = \begin{cases} 2n^{-1}\sum_{i=1}^{n}(\tilde{a}_i\tilde{\epsilon}_{it} + \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha}) + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right) & \text{for M1,} \\ 2n^{-1}\sum_{i=1}^{n}\tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right)t^{-2\beta} & \text{for M2,} \\ 2n^{-1}\sum_{i=1}^{n}(\tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha-\beta}) + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right)t^{-2\beta} & \text{for M3.} \end{cases}$$

Let $\tilde{\varsigma}_{it} = 2\tilde{a}_i\tilde{\epsilon}_{it}$. Then irrespective of whether $\sigma_{a\mu} = 0$, or $\sigma_{a\mu} \neq 0$, if $\alpha > 0$ and $\beta > 0$ the dominant term in $\varepsilon_{n,t}$ is as follows:

M1	M2	M3
$n^{-1}\sum_{i=1}^{n}\tilde{\varsigma}_{it} + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right)$	$n^{-1}\sum_{i=1}^{n}\tilde{\varsigma}_{it}t^{-\beta}$	$n^{-1}\sum_{i=1}^{n}\tilde{\varsigma}_{it}t^{-\beta}$

Using lemma 3, for M2 and M3, we have

$$I_C = \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \sum_{t=1}^{T} \left(n^{-1}\sum_{i=1}^{n} \tilde{\varsigma}_{it}\right)\tilde{t}\widetilde{t^{-\beta}} = \begin{cases} O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \beta < 1/2, \\ O_p\left(n^{-1/2}T^{-2}[\ln T]^{1/2}\right) & \text{if } \beta = 1/2, \\ O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \beta > 1/2, \end{cases}$$

and then, for M1,

$$I_C = \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{\varsigma}_{it}\tilde{t} + \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \sum_{t=1}^{T} \tilde{t}\left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right) = O_p\left(n^{-1/2}T^{-3/2}\right).$$

If $\alpha < 0$ or $\beta < 0$,the order of term I_C will be discussed under each case.

The Asymptotic Limit of $\hat{\phi}_{nT}$ When $\alpha > 0, \beta > 0$

When $\alpha > 0, \beta > 0$,we separate the proof when $\sigma_{a\mu} = 0$ from that when $\sigma_{a\mu} \neq 0$.

(i) The Asymptotic Limit of $\hat{\phi}_{nT}$ when $\sigma_{a\mu} = 0$

Recall that $\eta_{t,d} = bt^{-\lambda}$. Then, when $\sigma_{a\mu} = 0$, we have: $\lambda = 2\alpha$ and $b = \sigma_\mu^2$ in M1; $\lambda = 2\beta$, $b = \sigma_\epsilon^2$ in M2; and $\lambda = \min[2\alpha, 2\beta]$, with $b = \sigma_\mu^2$ if $\lambda = 2\alpha$, and $b = \sigma_\epsilon^2$ if $\lambda = 2\beta$ in M3. We take each model in turn to obtain the final results.

Under M1: We have

$$\begin{aligned}
\hat{\phi}_{nT} &= I_A + I_B + I_C \\
&= \begin{cases} O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha < 1/2, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha = 1/2, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } 1/2 < \alpha < 1, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha = 1, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha > 1, \end{cases} \\
&= \begin{cases} O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha < 1/2, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha = 1/2, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha > 1/2. \end{cases} \qquad (49)
\end{aligned}$$

So I_A dominates I_B when $n/T \to \infty$.

Under M2 For M2, I_A always dominates I_B as discussed above. Then

$$\begin{aligned}
\hat{\phi}_{nT} &= (I_A + I_C)\{1 + o(1)\} \\
&= \begin{cases} O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \beta < 1/2, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}[\ln T]^{1/2}\right) & \text{if } \beta = 1/2, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \beta > 1/2, \end{cases} \qquad (50)
\end{aligned}$$

so that I_A dominates I_C.

Under M3 We have

$$I_A = O\left(T^{-3}\right) \times \mathcal{T}_T(1, \lambda) = \begin{cases} O\left(T^{-1-\lambda}\right) & \text{if } \lambda < 1, \\ O\left(T^{-2}\ln T\right) & \text{if } \lambda = 1, \\ O\left(T^{-2}\right) & \text{if } \lambda > 1, \end{cases} \text{ with } \lambda = \min[2\alpha, 2\beta],$$

$$\begin{aligned}
I_B &= O_p\left(n^{-1/2}T^{-3}\right)\mathcal{T}_T(1, \lambda^*), \text{ with } \lambda^* = \min[\alpha, 2\beta], \\
&= \begin{cases} O_p\left(n^{-1/2}T^{-1-\lambda^*}\right) & \text{if } \lambda^* < 1, \\ O_p\left(n^{-1/2}T^{-2}\ln T\right) & \text{if } \lambda^* = 1, \\ O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \lambda^* > 1, \end{cases}
\end{aligned}$$

$$I_C = \begin{cases} O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \beta < 1/2, \\ O_p\left(n^{-1/2}T^{-2}[\ln T]^{1/2}\right) & \text{if } \beta = 1/2, \\ O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \beta > 1/2. \end{cases}$$

We need to consider the following two subcases.

Case 1: $(\alpha \leq \beta)$ Combining all the three terms, we have the following.

$$\begin{aligned}
\hat{\phi}_{nT} &= I_A + I_B + I_C \\
&= \begin{cases}
O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \alpha \leq \beta < 1/2, \\
O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \alpha < \beta = 1/2, \\
O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \alpha = \beta = 1/2, \\
O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1/2 = \alpha < \beta, \\
O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } 1/2 = \alpha = \beta, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1 < \alpha \leq \beta, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1/2 < \alpha < 1 \leq \beta, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1 = \alpha \leq \beta, \\
O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \alpha < 1/2 < \beta,
\end{cases} \\
&= \begin{cases}
O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha < 1/2, \\
O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha = 1/2, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } 1/2 < \alpha < 1, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\ln T\right) & \text{if } \alpha = 1, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \alpha > 1,
\end{cases} \qquad (51)
\end{aligned}$$

so that I_A dominates I_B and I_C when $n/T \to \infty$.

Case 2: $(\alpha > \beta)$ When $\sigma_{a\mu} = 0$ and $\alpha > \beta$, I_A always dominates I_C. When $2\beta \leq \alpha$,

$$\begin{aligned}
\hat{\phi}_{nT} &= I_A + I_B \\
&= \begin{cases}
O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \beta < 1/2, \\
O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \beta = 1/2, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \beta > 1/2.
\end{cases}
\end{aligned}$$

If $2\beta > \alpha > \beta$, we have

$$\begin{aligned}
\hat{\phi}_{nT} &= I_A + I_B \\
&= \begin{cases}
O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha < 2\beta < 1, \\
O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha < 2\beta = 1, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-1-\alpha}\right) & \text{if } \alpha < 1 < 2\beta, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\ln T\right) & \text{if } \alpha = 1 < 2\beta, \\
O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1 < \alpha < 2\beta,
\end{cases} \qquad (52)
\end{aligned}$$

so that I_A dominates I_B when $n/T \to \infty$.

(ii) The Asymptotic Limit of $\hat{\phi}_{nT}$ when $\sigma_{a\mu} \neq 0$

When $\sigma_{a\mu} \neq 0$,and $\alpha, \beta > 0$, we have: $\lambda = \alpha$ and $b = \sigma_{a\mu}$ for M1; $\lambda = 2\beta$ and $b = \sigma_{\epsilon}^2$ for M2; $\lambda = \min\left[\alpha, 2\beta\right]$ for M3, with $b = \sigma_{a\mu}$ if $\lambda = \alpha$, and $b = \sigma_{\epsilon}^2$ if $\lambda = 2\beta$ for M3. When $b = \sigma_{a\mu}$, the sign of I_A is consonant with that of $-b$. So that when $\sigma_{a\mu} > 0$, I_A is negative, and when $\sigma_{a\mu} < 0$, I_A is positive.

Under M1 $\hat{\phi}_{nT}$ can be written as

$$\hat{\phi}_{nT} = (I_A + I_C)\{1 + o_p(1)\} = \begin{cases} O\left(T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha < 1, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha = 1, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-3/2}\right) & \text{if } \alpha > 1, \end{cases} \tag{53}$$

so that if $\alpha < 0.5$, then I_A dominates I_C for any n. Otherwise, I_A dominates I_C when $n/T \to \infty$.

Under M2 For M2, the behavior of $\hat{\phi}_{nT}$ when $\sigma_{a\mu} \neq 0$ is the same as when $\sigma_{a\mu} = 0$.

Under M3 When $\sigma_{a\mu} \neq 0$, we have

$$I_A = \begin{cases} O\left(T^{-3}\mathcal{T}_T(1,\alpha)\right) & \text{if } \alpha \leq 2\beta, \\ O\left(T^{-3}\mathcal{T}_T(1,2\beta)\right) & \text{if } \alpha > 2\beta, \end{cases}$$

where $\mathcal{T}_T(1,\alpha) = \sum_{t=1}^{T} \widetilde{\tilde{t}t^{-\alpha}} = \sum_{t=1}^{T} t^{-(\alpha-1)} - \left(\frac{T+1}{2}\right)\sum_{t=1}^{T} t^{-\alpha}$ is defined in the proof of Lemma 2.

From the analysis above, the term I_C has the following order

$$I_C = \begin{cases} O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \beta < 1/2, \\ O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \beta = 1/2, \\ O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \beta > 1/2. \end{cases}$$

Case 1: ($\alpha \leq 2\beta$) When $\sigma_{a\mu} \neq 0$ and $\alpha \leq 2\beta$,

$$\begin{aligned} \hat{\phi}_{nT} &= \{I_A + I_C\}\{1 + o(1)\} \\ &= \begin{cases} O\left(T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \alpha \leq 2\beta < 1, \\ O\left(T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \alpha < 2\beta = 1, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \alpha = 2\beta = 1, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1 = \alpha < 2\beta, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } 1 = \alpha = 2\beta, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } 1 < \alpha \leq 2\beta, \\ O\left(T^{-1-\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \alpha < 1 < 2\beta, \end{cases} \\ &= \begin{cases} O\left(T^{-1-\alpha}\right) & \text{if } \alpha < 1, \\ O\left(T^{-2}\ln T\right) & \text{if } \alpha = 1, \\ O\left(T^{-2}\right) & \text{if } \alpha > 1. \end{cases} \end{aligned} \tag{54}$$

Note that since $O\left(T^{-1-\alpha+3/2+\beta}\right) = O\left(T^{-\alpha+1/2+\beta}\right) > O\left(T^{-2\beta+1/2+\beta}\right)$ when $\alpha \leq 2\beta$, the first term dominates the second term.

Case 2: $(\alpha > 2\beta)$ When $\sigma_{a\mu} \neq 0$ and $\alpha > 2\beta$,

$$\begin{aligned}
\hat{\phi}_{nT} &= I_A + I_C \\
&= \begin{cases} O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-3/2-\beta}\right) & \text{if } \beta < 1/2, \\ O\left(T^{-2}\ln T\right) + O_p\left(n^{-1/2}T^{-2}\left[\ln T\right]^{1/2}\right) & \text{if } \beta = 1/2, \\ O\left(T^{-2}\right) + O_p\left(n^{-1/2}T^{-2}\right) & \text{if } \beta > 1/2. \end{cases} \qquad (55)
\end{aligned}$$

The first term always dominates the second term.

The Asymptotic Limit of $\hat{\phi}_{nT}$ When $\alpha < 0$ or $\beta < 0$

Recall the OLS estimate is decomposed in the form

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

Note that when $\alpha < 0$ or $\beta < 0$, we have: $\lambda = 2\alpha$ and $b = \sigma_\mu^2$ in M1; $\lambda = 2\beta$ and $b = \sigma_\epsilon^2$ in M2; $\lambda = \min(2\alpha, 2\beta)$ in M3. As shown above, I_A always dominates I_B since $I_B = O_p\left(n^{-1/2}\right) I_A$.Using Lemma 2, we have

$$I_A = -b\left(\frac{6\lambda T^{-1-\lambda}}{(2-\lambda)(1-\lambda)}\right)\left\{1 + O\left(T^{-1}\right)\right\}.$$

Note the sign of I_A is positive when $\alpha < 0$ or $\beta < 0$.

Under Model M1 and M2:

First, we consider the term I_C. If $\alpha < 0$ (for M1) or $\beta < 0$ (for M2), the dominating term in ε_{nt} is $2n^{-1}\sum_{i=1}^{n}\tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha}$ in M1 and $\left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right)t^{-2\beta}$ in M2. By using lemma 3, we have

$$\begin{aligned}
I_C &= \left(\sum\nolimits_{t=1}^{T}\tilde{t}^2\right)^{-1}\sum\nolimits_{t=1}^{T}\tilde{t}\varepsilon_{nt} \\
&= \begin{cases} O_p\left(n^{-1/2}T^{-2}\left[\tau_T(2\alpha)\right]^{1/2}\right) & \text{for M1}, \\ O_p\left(n^{-1/2}T^{-2}\left[\tau_T(4\beta)\right]^{1/2}\right) & \text{for M2}. \end{cases}
\end{aligned}$$

Combining the two parts we have

$$\begin{aligned}
\hat{\phi}_{nT} &= I_A + I_C \\
&= \begin{cases} O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-3/2-\alpha}\right) & \text{for M1}, \\ O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-3/2-2\beta}\right) & \text{for M2}. \end{cases}
\end{aligned}$$

Thus, I_A dominates I_C for M1 and M2.

Under Model M3

We proceed case by case as follows.

Case 1: ($\alpha < 0$ **and** $\beta > 0$) If $\alpha < 0$ but $\beta > 0$, the dominating term is $\sigma_\mu^2 t^{-2\alpha}$ in η_t and is $2n^{-1}\sum_{i=1}^n \tilde{\mu}_i \tilde{\epsilon}_{it} t^{-\alpha-\beta}$ in ε_{nt}. Then $I_A = \left(\sum_{t=1}^T \tilde{t}^2\right)^{-1} \sum_{t=1}^T \tilde{t}\tilde{\eta}_t = O\left(T^{-1-2\alpha}\right)$ and

$$I_C = \left(\sum_{t=1}^T \tilde{t}^2\right)^{-1} \sum_{t=1}^T \tilde{t}\tilde{\varepsilon}_{nt} = O_p\left(n^{-1/2}T^{-2}\left[\tau_T\left(2\alpha+2\beta\right)\right]^{1/2}\right).$$

Note from Lemma 1 that $\tau_T(2\alpha+2\beta)$ reaches the largest order of $O\left(T^{1-2\alpha-2\beta}\right)$ when $\alpha+\beta < 1/2$. Even at this case, $I_C = O_p\left(n^{-1/2}T^{-3/2-\alpha-\beta}\right)$,and is still dominated by I_A.Hence, we have

$$\hat{\phi}_{nT} = O\left(T^{-1-2\alpha}\right) + O_p\left(n^{-1/2}T^{-2}\left[\tau_T\left(2\alpha+2\beta\right)\right]^{1/2}\right) = O\left(T^{-1-2\alpha}\right).$$

Case 2: ($\alpha > 0$ **and** $\beta < 0$) If $\alpha > 0$ and $\beta < 0$, the dominating term is $\sigma_\epsilon^2 t^{-2\beta}$ in η_t, and is $\left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) t^{-2\beta}$ in ε_{nt}. Then $I_A = O\left(T^{-1-2\beta}\right)$ and $I_C = O_p\left(n^{-1/2}T^{-2}\left[\tau_T\left(4\beta\right)\right]^{1/2}\right)$. Hence we have

$$\begin{aligned}\hat{\phi}_{nT} &= O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-2}\left[\tau_T\left(4\beta\right)\right]^{1/2}\right)\\ &= O\left(T^{-1-2\beta}\right) + O_p\left(n^{-1/2}T^{-3/2-2\beta}\right) = O\left(T^{-1-2\beta}\right).\end{aligned}$$

Case 3: ($\alpha < 0$ **and** $\beta < 0$) If $\alpha < 0$ and $\beta < 0$,the dominating terms are $\sigma_\mu^2 t^{-2\alpha} + \sigma_\epsilon^2 t^{-2\beta}$ in η_t, and $2n^{-1}\sum_{i=1}^n \tilde{\mu}_i \tilde{\epsilon}_{it} t^{-\alpha-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) t^{-2\beta}$ in ε_{nt}. Then we have

$$\begin{aligned}\hat{\phi}_{nT} &= O\left(T^{-3}\right)\left[\sigma_\mu^2 \mathcal{T}_T\left(1,2\alpha\right) + \sigma_\epsilon^2 \mathcal{T}_T\left(1,2\beta\right)\right] + O_p\left(n^{-1/2}T^{-2}\left[\tau_T\left(2\alpha+2\beta\right)\right]^{1/2}\right)\\ &\quad + O_p\left(n^{-1/2}T^{-2}\left[\tau_T\left(4\beta\right)\right]^{1/2}\right)\\ &= O\left(T^{-1-2\delta}\right),\end{aligned}$$

where $\delta = \min(\alpha, \beta)$.

The Asymptotic Limit of $\hat{\phi}_{nT}$ When $\alpha = \beta = 0$

This case is covered by standard theory and the proof is omitted.
□

Proof of Theorem 2: (t-ratio of $\hat{\phi}_{nT}$)

The proof under $\alpha = \beta = 0$ is standard and is omitted.

(i) The Asymptotic Limit of $t_{\hat{\phi}_{nT}}$ when $\alpha > 0, \beta > 0$

The asymptotic behavior of the t-ratio is determined by the three component factors of

$$t_{\hat{\phi}_{nT}} = \hat{\phi}_{nT}\left(\sum_{t=1}^T \tilde{t}^2\right)^{1/2} / \sqrt{\Omega_{\mathcal{M}}^2}.$$

The behavior of $\hat{\phi}_{nT}$ is described above in (49) - (55), and $\sum_{t=1}^T \tilde{t}^2 = \frac{1}{12}T^3\left\{1 + O\left(T^{-1}\right)\right\}$. The behavior of the long run variance estimate $\Omega_{\mathcal{M}}^2$ is obtained as follows. As in the proof of Theorem 1, it is convenient to use the following specifications of λ for each model.

Model specifications of λ

- M1: $\lambda = \alpha\,(\text{or } 2\alpha)$ when $\sigma_{a\mu} \neq 0$ (respectively, $\sigma_{a\mu} = 0)$;
- M2: $\lambda = 2\beta$;
- M3: $\lambda = \min[\alpha, 2\beta]$ (or $\min[2\alpha, 2\beta]$) when $\sigma_{a\mu} \neq 0$ (respectively, $\sigma_{a\mu} = 0$).

We write $\widetilde{\mathcal{M}}_{nt} = \tilde{m}_t + R_{nt}$, where the deterministic part is $\tilde{m}_t = \widetilde{t^{-\lambda}} - \tilde{t}\left(\sum_{t=1}^{T}\tilde{t}\widetilde{t^{-\lambda}}\right)\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}$ and the random part is $R_{nt} = \tilde{\xi}_{n,t} - \tilde{t}\,(I_B + I_C)$. Note that as $n/T \to \infty$, $R_{nt} = o_p(\tilde{m}_t)$ uniformly in t. The regression residual is given by

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

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

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

where

$$\hat{\Omega}_{\mathcal{M}}^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{\mathcal{M}}_{nt}^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)\tilde{\mathcal{M}}_{nt}\tilde{\mathcal{M}}_{nt+\ell}, \tag{56}$$

$$\Omega_{\mathcal{M}}^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{m}_t^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)\tilde{m}_t\tilde{m}_{t+\ell}. \tag{57}$$

Note that

$$\begin{aligned}
\hat{\Omega}_{\mathcal{M}}^2 &= \frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{\mathcal{M}}_{nt}^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)\tilde{\mathcal{M}}_{nt}\tilde{\mathcal{M}}_{nt+\ell} \\
&= \frac{1}{T}\sum\nolimits_{t=1}^{T}(\tilde{m}_t + R_{nt})^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)(\tilde{m}_t + R_{nt})(\tilde{m}_{t+\ell} + R_{nt+\ell}) \\
&= \frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{m}_t^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)\tilde{m}_t\tilde{m}_{t+\ell} \\
&\quad + \frac{1}{T}\sum\nolimits_{t=1}^{T}R_{nt}^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)R_{nt}R_{nt+\ell} \\
&\quad + 2\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{m}_tR_{nt} + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{T^\kappa + 1}\right)(\tilde{m}_tR_{nt+\ell} + R_{nt}\tilde{m}_{t+\ell})
\end{aligned}$$

We have shown that $R_{nt} = o_p(\tilde{m}_t)$ uniformly in $t \le T$, from which it follows that the dominating term in $\hat{\Omega}^2_{\mathcal{M}}$ is $\Omega^2_{\mathcal{M}}$, which is represented by $\hat{\Omega}^2_{\mathcal{M}} \sim \Omega^2_{\mathcal{M}}$.

From Lemma 4, we know that

$$\Omega^2_{\mathcal{M}} \sim \begin{cases} \frac{\lambda^2(\lambda+1)^2}{(1-2\lambda)\left(\lambda^2-3\lambda+2\right)^2} b^2 T^{-2\lambda+\kappa} & \text{if } \lambda < 1/2, \\ b^2 T^{-1+\kappa} \ln T & \text{if } \lambda = 1/2, \\ b^2 T^{-1+\kappa} \zeta(2\lambda) & \text{if } 1/2 < \lambda < 1/(1+\kappa), \\ \dfrac{1}{(1-\lambda)^2 (2-\lambda)} b^2 T^{-\lambda+\kappa-\lambda\kappa} & \text{if } 1/(1+\kappa) \le \lambda < 1, \\ \frac{\kappa^2}{2} b^2 T^{-1} \ln^2 T + O\left(T^{-2}\ln T\right) & \text{if } \lambda = 1, \\ T^{-1} b^2 \left\{\sum_{t=1}^{\infty} t^{-\lambda} \zeta(\lambda, \ell) - \zeta(2\lambda)\right\} & \text{if } \lambda > 1. \end{cases}$$

Recall that

$$\varepsilon_{n,t} = \begin{cases} 2n^{-1}\sum_{i=1}^n \tilde{a}_i \tilde{\epsilon}_{it} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i \tilde{\epsilon}_{it} t^{-\alpha} + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right) & \text{for M1}, \\ 2n^{-1}\sum_{i=1}^n \tilde{a}_i \tilde{\epsilon}_{it} t^{-\beta} + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right) t^{-2\beta} & \text{for M2}, \\ 2n^{-1}\sum_{i=1}^n \tilde{a}_i \tilde{\epsilon}_{it} t^{-\beta} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i \tilde{\epsilon}_{it} t^{-\alpha-\beta} + \left(\sigma^2_{\epsilon,nt} - \sigma^2_{\epsilon,nT}\right) t^{-2\beta} & \text{for M3}. \end{cases}$$

For fixed t, $\varepsilon_{n,t} = O_p\left(n^{-1/2}\right)$. When $\alpha > 0$ and $\beta > 0$, the dominant term in $\varepsilon_{n,t}$ is $2n^{-1}\sum_{i=1}^n \tilde{a}_i \tilde{\epsilon}_{it}$ for M1 and $2n^{-1}\sum_{i=1}^n \tilde{a}_i \tilde{\epsilon}_{it} t^{-\beta}$ for M2 and M3. From Lemma 3, we have

$$\begin{aligned} \hat{\Omega}^2_{\varepsilon} &= \frac{1}{T}\sum_{t=1}^T \tilde{\varepsilon}^2_{nt} + 2\frac{1}{T}\sum_{\ell=1}^L \vartheta_{\ell L} \sum_{t=1}^{T-\ell} \tilde{\varepsilon}_{nt}\tilde{\varepsilon}_{nt+\ell} \\ &= \begin{cases} O_p\left(n^{-1}T^{-1}\mathcal{S}(\beta)\right) & \text{in M2 \& M3}, \\ O_p\left(n^{-1}\right) & \text{in M1}. \end{cases} \end{aligned}$$

Comparing $\hat{\Omega}^2_{\mathcal{M}}$ and $\hat{\Omega}^2_{\varepsilon}$, it is evident that the order of $\hat{\Omega}^2_{\mathcal{M}}$ exceeds that of $\hat{\Omega}^2_{\varepsilon}$ as long as $n/T \to \infty$. Next, consider $\hat{\Omega}_{m\varepsilon}$. By Cauchy-Schwarz, $\hat{\Omega}_{m\varepsilon}$ is bounded above as

$$\frac{1}{T}\sum_{t=1}^T \tilde{\mathcal{M}}_{nt}\tilde{\varepsilon}_{nt} \le \left(\frac{1}{T}\sum_{t=1}^T \tilde{\mathcal{M}}^2_{nt}\right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^T \tilde{\varepsilon}^2_{nt}\right)^{1/2} = o_p\left(\frac{1}{T}\sum_{t=1}^T \tilde{\mathcal{M}}^2_{nt}\right),$$

since $\frac{1}{T}\sum_{t=1}^T \tilde{\varepsilon}^2_{nt} = o_p\left(\frac{1}{T}\sum_{t=1}^T \tilde{\mathcal{M}}^2_{nt}\right)$.

Combining these results, we find that

$$\hat{\Omega}^2_u = \hat{\Omega}^2_{\mathcal{M}} + \hat{\Omega}^2_{\varepsilon} + 2\hat{\Omega}_{\mathcal{M}\varepsilon} \sim \hat{\Omega}^2_{\mathcal{M}} \sim \Omega^2_{\mathcal{M}}, \tag{58}$$

and therefore

$$\hat{\Omega}^2_u = \begin{cases} \frac{\lambda^2(\lambda+1)^2}{(1-2\lambda)\left(\lambda^2-3\lambda+2\right)^2} b^2 T^{-2\lambda+\kappa} & \text{if } \lambda < 1/2, \\ b^2 T^{-1+\kappa} \ln T & \text{if } \lambda = 1/2, \\ b^2 T^{-1+\kappa} \zeta(2\lambda) & \text{if } 1/2 < \lambda < 1/(1+\kappa), \\ \dfrac{1}{(1-\lambda)^2 (2-\lambda)} b^2 T^{-\lambda+\kappa-\lambda\kappa} & \text{if } 1/(1+\kappa) \le \lambda < 1, \\ \frac{\kappa^2}{2} b^2 T^{-1} \ln^2 T + O\left(T^{-2}\ln T\right) & \text{if } \lambda = 1, \\ T^{-1} b^2 \left\{\sum_{t=1}^{\infty} t^{-\lambda} \zeta(\lambda, \ell) - \zeta(2\lambda)\right\} & \text{if } \lambda > 1. \end{cases} \tag{59}$$

Using (58), the t ratio $t_{\hat{\phi}_{nT}}$ has the same asymptotic form as

$$t_{\hat{\phi}_{nT}} \sim \frac{\hat{\phi}_{nT}\left(\sum \tilde{t}^2\right)^{1/2}}{\sqrt{\hat{\Omega}^2_{\mathcal{M}} + \hat{\Omega}^2_{\varepsilon} + 2\hat{\Omega}_{\mathcal{M}\varepsilon}}} \sim \frac{\hat{\phi}_{nT}\left(\sum \tilde{t}^2\right)^{1/2}}{\sqrt{\Omega^2_{\mathcal{M}}}},$$

where $\Omega^2_{\mathcal{M}}$ is deterministic. Therefore, using (59) as $n/T \to \infty$ with $\kappa = 1/3$ and a finite λ, the asymptotic behavior of the t-ratio in all models can be represented as follows:

$$t_{\hat{\phi}_{nT}} = \begin{cases} \dfrac{-\sqrt{3}\,(1-2\lambda)^{1/2}}{(\lambda+1)} T^{1/2-\kappa/2} = (-1)\times O\left(T^{1/2-\kappa/2}\right) \to -\infty & \text{if } \lambda < 1/2, \\ -\frac{2}{3}\sqrt{3}T^{1/2-\kappa/2}\left(\ln T\right)^{-1/2} = (-1)\times O\left(T^{1/2-\kappa/2}\left(\ln T\right)^{-1/2}\right) \to -\infty & \text{if } \lambda = 1/2, \\ \dfrac{-\sqrt{3}\lambda}{(\lambda-2)(\lambda-1)}\left(\zeta(2\lambda)\right)^{-1/2} T^{1-\lambda-\kappa/2} = (-1)\times O\left(T^{1-\lambda-\kappa/2}\right) \to = -\infty & \text{if } 1/2 < \lambda < 1/(1+\kappa), \\ \dfrac{-\sqrt{3}\lambda}{(2-\lambda)^{1/2}} T^{(1-\lambda)(1-\kappa)/2} = (-1)\times O\left(T^{(1-\lambda)(1-\kappa)/2}\right) = -\infty & \text{if } 1/(1+\kappa) \le \lambda < 1, \\ \dfrac{-6bT^{-2}\ln T\left(\frac{1}{12}T^3\right)^{1/2}}{\left(\dfrac{\kappa^2}{2}b^2T^{-1}\ln^2 T\right)^{1/2}} = -3\sqrt{6} & \text{if } \lambda = 1, \\ \dfrac{-6bT^{-2}\zeta(\lambda)\left(\frac{1}{12}T^3\right)^{1/2}}{\left(T^{-1}b^2\left\{\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,t)\right\}\right)^{1/2}} = \dfrac{-\zeta(\lambda)\sqrt{3}}{\left(\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,t)\right)^{1/2}} > -\sqrt{3} & \text{if } \lambda > 1. \end{cases} \tag{60}$$

Note that with $\kappa = 1/3$, we have $1-\lambda-\kappa/2 > 0$ as long as $1/2 < \lambda < 1/(1+\kappa)$. The last inequality in (60) is obtained by noting that when $t > 1$, $\zeta(\lambda, t) < \zeta(\lambda)$ and hence

$$\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,t) < \sum_{t=1}^{\infty} t^{-\lambda}\sum_{s=1}^{\infty}\frac{1}{s^{\lambda}} = \zeta(\lambda)^2,$$

giving

$$\frac{\zeta(\lambda)}{\left(\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,t)\right)^{1/2}} > 1 \text{ for all } \lambda > 1.$$

Moreover, when $\lambda \to \infty$ we have

$$\begin{aligned} \lim_{\lambda\to\infty}\left\{\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,t) - \zeta(2\lambda)\right\} &= \lim_{\lambda\to\infty}\left\{\sum_{t=1}^{\infty}\sum_{s=0}^{\infty}\frac{1}{t^{\lambda}(t+s)^{\lambda}} - \sum_{t=1}^{\infty}\frac{1}{t^{2\lambda}}\right\} \\ &= \lim_{\lambda\to\infty}\left\{\sum_{t=1}^{\infty}\sum_{s=1}^{\infty}\frac{1}{t^{\lambda}(t+s)^{\lambda}}\right\} = 0. \end{aligned}$$

Then, as $\lambda \to \infty$ we have

$$\lim_{\lambda\to\infty} t_{\hat{\phi}_{nT}} = -\lim_{\lambda\to\infty}\frac{\zeta(\lambda)\sqrt{3}}{\left(\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,t)\right)^{1/2}} = -\lim_{\lambda\to\infty}\frac{\zeta(\lambda)\sqrt{3}}{\zeta(2\lambda)^{1/2}} = -\sqrt{3},$$

giving a sharp result for the t-ratio for large λ.

On the other hand, as $\lambda \to 0$, the t-ratio has the following limit

$$\lim_{\lambda\to+0} t_{\hat{\phi}_{nT}} = \lim_{\lambda\to+0}\frac{-\sqrt{3}\,(1-2\lambda)^{1/2}}{(\lambda+1)}T^{1/2-\kappa/2} = -\sqrt{3}T^{1/2-\kappa/2} \to -\infty \text{ as } T \to \infty.$$

(ii) The Asymptotic Limit of $t_{\hat{\phi}_{nT}}$ when $\alpha < 0$ or $\beta < 0$

Recall the representation

$$\tilde{\mathcal{M}}_{nt} = b\left[\widetilde{t^{-\lambda}} - \tilde{t}\left(\sum\nolimits_{t=1}^{T}\tilde{t}^2\right)^{-1}\sum\nolimits_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right] + R_{nt} = b\tilde{m}_t + R_{nt}, \tag{61}$$

where $\lambda = 2\alpha$ for M1, $\lambda = 2\beta$ for M2, and $\lambda = \min[2\alpha, 2\beta]$ for M3. The factor b in (61) is positive since $b = \sigma_\epsilon^2$ or σ_μ^2. Then we have

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{m}_t^2 = b^2\mathcal{B}(\lambda) = O_p\left(T^{-2\lambda}\right),$$

since from Lemma 2, $\mathcal{B}(\lambda) = O\left(T^{-2\lambda}\right)$ when $\lambda < 0$. Note also that when $\lambda < 0$,

$$\Psi_{\ell,11} = \sum\nolimits_{t=1}^{T-\ell}\left(t^2 + t\ell\right)^{-\lambda} = \sum\nolimits_{t=1}^{T-\ell} t^{-\lambda}(t+\ell)^{-\lambda} < \sum\nolimits_{t=1}^{T-\ell}(t+\ell)^{-2\lambda} \sim \frac{T^{1-2\lambda}}{1-2\lambda}$$

Then

$$\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\left(1 - \frac{\ell}{L+1}\right)\Psi_{\ell,11} < \frac{T^{-2\lambda}L}{1-2\lambda} = O\left(T^{-2\lambda+\kappa}\right).$$

This implies from lemma 4 that

$$\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\tilde{m}_t\tilde{m}_{t+\ell} = O\left(T^{-2\lambda+\kappa}\right).$$

The order of $\Omega_{\mathcal{M}}^2$ is at least $O_p\left(T^{-2\lambda}\right)$ and is less than $O\left(T^{-2\lambda+\kappa}\right)$,where $\lambda = 2\alpha$ for M1, $\lambda = 2\beta$ for M2, and $\lambda = \min[2\alpha, 2\beta]$ for M3. From Lemma 3, we have

$$\hat{\Omega}_\varepsilon^2 = \begin{cases} O_p\left(n^{-1}T^{-1}\mathcal{S}(\alpha)\right) = O_p\left(n^{-1}T^{-2\alpha}\right) & \text{in M1,} \\ O_p\left(n^{-1}T^{-1}\mathcal{S}(2\beta)\right) = O_p\left(n^{-1}T^{-4\beta}\right) & \text{in M2,} \\ \max\left[O_p\left(n^{-1}T^{-2(\alpha+\beta)}\right), O_p\left(n^{-1}T^{-4\beta}\right)\right] & \text{in M3.} \end{cases}$$

Hence, $\hat{\Omega}_{\mathcal{M}}^2$ dominates $\hat{\Omega}_\varepsilon^2$.And $\hat{\Omega}_{\mathcal{M}}^2$ also dominates $\hat{\Omega}_{m\varepsilon}$ since

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{\mathcal{M}}_{nt}\tilde{\varepsilon}_{nt} \le \left(\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{\mathcal{M}}_{nt}^2\right)^{1/2}\left(\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{\varepsilon}_{nt}^2\right)^{1/2} = o_p\left(\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{\mathcal{M}}_{nt}^2\right),$$

Hence the long run variance has the following order

$$\hat{\Omega}_u^2 = O\left(T^{-2\lambda+\kappa}\right).$$

Then

$$t_{\hat{\phi}_{nT}} = \frac{\hat{\phi}_{nT}\sqrt{\sum_{t=1}^{T}\tilde{t}^2}}{\sqrt{\hat{\Omega}_u^2}} = \frac{O\left(T^{-1-\lambda}\right)}{O\left(T^{\kappa/2-\lambda}\right)}O\left(T^{3/2}\right) = O\left(T^{1/2-\kappa/2}\right).$$

In this event, $t_{\hat{\phi}_{nT}} \to +\infty$ as $n, T \to \infty$, since the sign of $t_{\hat{\phi}_{nT}}$ is consistent with $\hat{\phi}_{nT}$ which is positive when $\lambda < 0$.

Table 1: Size and Power of the Test in M1

	$\sigma_{a\mu}$	α	σ_a^2	$T\backslash n$	25	50	100	200	500	1000
Size	0	0	1	25	0.105	0.111	0.109	0.113	0.104	0.117
				50	0.091	0.090	0.089	0.092	0.091	0.091
				100	0.070	0.074	0.075	0.076	0.072	0.071
				200	0.069	0.072	0.066	0.063	0.067	0.070
	0	-0.1	1	25	0.012	0.005	0.000	0.000	0.000	0.000
				50	0.002	0.000	0.000	0.000	0.000	0.000
				100	0.000	0.000	0.000	0.000	0.000	0.000
				200	0.000	0.000	0.000	0.000	0.000	0.000
Power	0	0.3	2	25	0.268	0.360	0.489	0.644	0.881	0.970
				50	0.272	0.345	0.479	0.625	0.884	0.979
				100	0.263	0.342	0.462	0.635	0.882	0.981
				200	0.259	0.340	0.465	0.637	0.876	0.982
	0.45	0.3	2	25	0.526	0.704	0.892	0.982	1.000	1.000
				50	0.580	0.781	0.941	0.993	1.000	1.000
				100	0.635	0.841	0.973	0.999	1.000	1.000
				200	0.705	0.898	0.989	1.000	1.000	1.000
	0	0.5	2	25	0.276	0.336	0.431	0.565	0.778	0.915
				50	0.221	0.296	0.362	0.495	0.712	0.863
				100	0.212	0.241	0.296	0.417	0.612	0.806
				200	0.180	0.211	0.279	0.348	0.522	0.712
	0.45	0.5	2	25	0.564	0.742	0.896	0.983	1.000	1.000
				50	0.555	0.760	0.917	0.988	1.000	1.000
				100	0.575	0.764	0.929	0.994	1.000	1.000
				200	0.555	0.773	0.938	0.994	1.000	1.000

Table 2: Size of the Test in M2

θ_i	β	σ_a^2	$T\backslash n$	25	50	100	200	500	1000
1	0	1	25	0.103	0.116	0.111	0.107	0.105	0.110
			50	0.082	0.098	0.085	0.092	0.091	0.091
			100	0.077	0.074	0.076	0.087	0.069	0.074
			200	0.072	0.063	0.067	0.064	0.076	0.067
1	-0.1	1	25	0.004	0.001	0.000	0.000	0.000	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
			200	0.000	0.000	0.000	0.000	0.000	0.000
1	0	5	25	0.119	0.113	0.119	0.114	0.115	0.114
			50	0.095	0.096	0.094	0.100	0.093	0.099
			100	0.082	0.085	0.082	0.080	0.087	0.082
			200	0.066	0.068	0.067	0.077	0.069	0.071
1	-0.1	5	25	0.018	0.006	0.002	0.000	0.000	0.000
			50	0.005	0.001	0.000	0.000	0.000	0.000
			100	0.001	0.000	0.000	0.000	0.000	0.000
			200	0.000	0.000	0.000	0.000	0.000	0.000
$iid\mathcal{N}(0.5, 1)$	0	1	25	0.106	0.102	0.100	0.100	0.106	0.115
			50	0.083	0.085	0.086	0.092	0.091	0.091
			100	0.077	0.078	0.076	0.074	0.074	0.075
			200	0.065	0.065	0.068	0.066	0.061	0.068

Table 3: Power of the Test in M2

θ_i	β	σ_a^2	$T\backslash n$	25	50	100	200	500	1000
1	0.1	1	25	0.452	0.623	0.815	0.958	0.999	1.000
			50	0.574	0.786	0.943	0.997	1.000	1.000
			100	0.752	0.937	0.996	1.000	1.000	1.000
			200	0.907	0.991	1.000	1.000	1.000	1.000
1	0.5	1	25	0.934	0.992	1.000	1.000	1.000	1.000
			50	0.967	0.998	1.000	1.000	1.000	1.000
			100	0.984	1.000	1.000	1.000	1.000	1.000
			200	0.992	1.000	1.000	1.000	1.000	1.000
1	0.1	5	25	0.181	0.207	0.247	0.307	0.461	0.618
			50	0.165	0.211	0.282	0.368	0.580	0.790
			100	0.193	0.240	0.334	0.479	0.744	0.933
			200	0.224	0.297	0.426	0.647	0.916	0.992
1	0.5	5	25	0.335	0.417	0.544	0.697	0.911	0.989
			50	0.352	0.452	0.589	0.782	0.950	0.996
			100	0.375	0.485	0.625	0.815	0.974	0.999
			200	0.393	0.518	0.681	0.853	0.986	0.999
$iid\mathcal{N}(0.5,1)$	0.1	1	25	0.419	0.567	0.753	0.914	0.993	1.000
			50	0.560	0.765	0.934	0.995	1.000	1.000
			100	0.738	0.931	0.995	1.000	1.000	1.000
			200	0.905	0.990	1.000	1.000	1.000	1.000
$iid\mathcal{N}(0.5,1)$	0.5	1	25	0.908	0.987	1.000	1.000	1.000	1.000
			50	0.950	0.997	1.000	1.000	1.000	1.000
			100	0.975	0.999	1.000	1.000	1.000	1.000
			200	0.991	0.999	1.000	1.000	1.000	1.000

Table 4: Power of the test in M3

$\sigma_{a\mu}$	$\alpha=\beta$	σ_a^2	$T\backslash n$	25	50	100	200	500	1000
0	0.3	2	25	0.799	0.932	0.993	1.000	1.000	1.000
			50	0.856	0.965	0.997	1.000	1.000	1.000
			100	0.880	0.972	0.999	1.000	1.000	1.000
			200	0.880	0.966	0.998	1.000	1.000	1.000
0.45	0.3	2	25	0.958	0.998	1.000	1.000	1.000	1.000
			50	0.991	1.000	1.000	1.000	1.000	1.000
			100	0.997	1.000	1.000	1.000	1.000	1.000
			200	0.999	1.000	1.000	1.000	1.000	1.000
0	0.5	2	25	0.834	0.949	0.994	1.000	1.000	1.000
			50	0.829	0.936	0.987	1.000	1.000	1.000
			100	0.806	0.906	0.976	0.999	1.000	1.000
			200	0.763	0.867	0.949	0.992	1.000	1.000
0.45	0.5	2	25	0.990	1.000	1.000	1.000	1.000	1.000
			50	0.996	1.000	1.000	1.000	1.000	1.000
			100	0.997	1.000	1.000	1.000	1.000	1.000
			200	0.998	1.000	1.000	1.000	1.000	1.000

Table 5: Evidence of weak $\sigma-$convergence
among personal consumption expenditure price index items

Factor number	Whole Sample			From 1960 to 1992			From 1992 to 2014		
k	$\hat{\phi}_{nT}$	$t_{\hat{\phi}_{nT}}(3)$	$t_{\hat{\phi}_{nT}}(6)$	$\hat{\phi}_{nT}$	$t_{\hat{\phi}_{nT}}(3)$	$t_{\hat{\phi}_{nT}}(6)$	$\hat{\phi}_{nT}$	$t_{\hat{\phi}_{nT}}(3)$	$t_{\hat{\phi}_{nT}}(6)$
1	0.212	1.680	1.471	0.792	3.789	3.487	-0.868	-3.309	-4.456
2	0.133	1.343	1.276	0.424	2.071	2.010	-0.696	-3.146	-4.039
3	0.000	0.001	0.001	0.399	3.012	2.878	-0.994	-4.113	-4.735
4	-0.004	-0.044	-0.038	0.376	2.862	2.722	-0.870	-3.420	-3.553
5	-0.022	-0.231	-0.200	0.351	3.064	3.152	-0.900	-3.786	-3.730
6	-0.031	-0.588	-0.541	0.252	2.679	2.979	-0.274	-2.267	-2.553
7	-0.033	-0.723	-0.652	0.203	2.512	2.731	-0.298	-3.898	-4.165
8	-0.026	-0.644	-0.579	0.198	2.906	3.444	-0.230	-3.447	-3.630

Notes: k stands for the number of the common factors; $t_{\hat{\phi}_{nT}}(3)$ and $t_{\hat{\phi}_{nT}}(6)$ are the t-ratios computed with $L=3,6$ truncation lags in the long run variance estimates.

Table 6: Trend Regressions with Various Starting Years (PCE data)

Starting Year	Average Correlation	$\hat{\phi}_{nT}$	$t_{\hat{\phi}_{nT}}$ (3)	$t_{\hat{\phi}_{nT}}$ (4)	$t_{\hat{\phi}_{nT}}$ (5)	$t_{\hat{\phi}_{nT}}$ (6)
1992	0.507	-0.994	-4.113	-4.226	-4.497	-4.735
1993	0.513	-1.223	-4.344	-4.241	-4.298	-4.39
1994	0.518	-0.873	-3.503	-3.668	-3.901	-4.139
1995	0.532	-1.027	-3.527	-3.594	-3.737	-3.868
1996	0.552	-1.024	-3.183	-3.243	-3.374	-3.493
1997	0.565	-0.686	-2.577	-2.785	-3.111	-3.384
1998	0.591	-0.585	-2.035	-2.251	-2.584	-2.901
1999	0.613	-0.618	-1.884	-2.070	-2.351	-2.619

Table 7: Trend Regression Results for Ultimatum Game data with Various Starting Rounds

Starting Rounds	$\hat{\phi}_{nT} \times 100$	$t_{\hat{\phi}_{nT}}$ (1)	$t_{\hat{\phi}_{nT}}$ (3)	$t_{\hat{\phi}_{nT}}$ (5)	$t_{\hat{\phi}_{nT}}$ (7)
1	-0.087	-4.299	-4.698	-5.034	-5.362
2	-0.090	-4.082	-4.472	-4.769	-4.975
3	-0.085	-3.543	-3.829	-4.246	-4.526
4	-0.086	-3.285	-3.539	-3.885	-4.124
5	-0.090	-3.133	-3.258	-3.518	-3.713
6	-0.096	-2.908	-2.917	-3.105	-3.269
7	-0.050	-3.153	-3.309	-3.914	-4.402
8	-0.028	-2.231	-2.296	-2.872	-3.210
9	-0.031	-2.256	-2.325	-2.891	-3.223
10	-0.037	-2.329	-2.426	-3.042	-3.283

Appendix S: Supplementary Material

(Included as an Online Supplement)

This document contains supplementary proofs, additional discussion on power trend regression, some further numerical calculations, and additional simulations to those reported in paper. We begin with the proofs of Lemmas 1-5.

Proofs of Lemmas

Proof of Lemma 1: By Euler-Maclaurin summation, the stated representations and large T approximations of $\tau_T(\alpha)$ and $H_T(\alpha,\ell)$ for $\alpha \leq 1$ are well known. When $\alpha > 1$, the exact expressions for the infinite sums are given by the Riemann and Hurwitz zeta functions

$$\begin{aligned} \zeta(\alpha) &= \textstyle\sum_{t=1}^{\infty} t^{-\alpha} = \frac{1}{\alpha-1} + \frac{1}{2} + \Delta_\alpha, \\ \zeta(\alpha,\ell) &= \frac{1}{\ell^\alpha} + \frac{1}{(1+\ell)^\alpha}\left(\frac{1}{2} + \frac{1+\ell}{\alpha-1}\right) + \Delta_{\alpha,\ell}, \end{aligned} \tag{62}$$

where

$$\Delta_\alpha = \sum_{j=1}^{J} \begin{pmatrix} \alpha+2j-2 \\ 2j-1 \end{pmatrix} \left(\frac{B_{2j}}{2j}\right) - (2J+1)! \begin{pmatrix} \alpha+2J \\ 2J+1 \end{pmatrix} \int_1^\infty \frac{P_{2J+1}(t)}{t^{\alpha+2J+1}} dt, \tag{63}$$

$$\begin{aligned} \Delta_{\alpha,\ell} = &\sum_{j=1}^{J} \begin{pmatrix} \alpha+2j-2 \\ 2j-1 \end{pmatrix} \left(\frac{B_{2j}}{2j}\right) \frac{1}{(1+\ell)^{\alpha+2j-1}} \\ &- (2J+1)! \begin{pmatrix} \alpha+2J \\ 2J+1 \end{pmatrix} \int_1^\infty \frac{P_{2J+1}(t)}{(t+\ell)^{\alpha+2J+1}} dt, \end{aligned} \tag{64}$$

the B_{2j} are Bernoulli numbers, and $P_{2J+1}(t) = (-1)^{J-1} \sum_{k=1}^{\infty} 2\sin(2k\pi t) / (2k\pi)^{2J+1}$. Thus, the expressions given in (62) provide upper bounds for $Z_T(\alpha)$ and $\zeta_T(\alpha,\ell)$. Simpler bounds are readily constructed (e.g., Kac and Cheung, 2002; KC). Indeed, since $t^{-\alpha}$ is positive, continuous, and tends to zero, the Euler-Maclaurin-Cauchy constant $\gamma_\alpha = \lim_{T\to\infty} \left\{ \sum_{t=1}^{T} t^{-\alpha} - \int_1^T t^{-\alpha} dt \right\}$ exists and is finite for all $\alpha > 1$. Note the explicit form

$$\textstyle\sum_{t=1}^{\infty} t^{-\alpha} = \frac{1}{\alpha-1} + \frac{1}{2} + \alpha \int_1^\infty \frac{\lfloor x \rfloor - x + \frac{1}{2}}{x^{\alpha+1}} dx, \tag{65}$$

where the floor function $\lfloor x \rfloor$ denotes the integer part of x. Since $-\frac{1}{2} < \lfloor x \rfloor - x + \frac{1}{2} \leq \frac{1}{2}$ we have the bound

$$\left| \int_1^\infty \frac{\lfloor x \rfloor - x + \frac{1}{2}}{x^{\alpha+1}} dx \right| \leq \int_1^\infty \frac{\frac{1}{2}}{x^{\alpha+1}} dx = \frac{1}{2\alpha},$$

from which we deduce that $|\Delta_\alpha| < \frac{1}{2}$ for all $\alpha > 1$. The first element of (65) is, of course, unbounded as $\alpha \to 1$. Finally, the inequality $\zeta(\alpha,\ell) \leq \zeta(\alpha)$ holds trivially for all $\ell \geq 1$ when $\alpha > 1$.
□

Proof of Lemma 2 The proof of lemma 2 follows in a straightforward way by direct calculation using lemma 1. In particular, we have as $T \to \infty$

$$\mathcal{T}_T(1,\alpha) = \sum_{t=1}^{T} \widetilde{\tilde{t}t^{-\alpha}} = \sum_{t=1}^{T} t^{-(\alpha-1)} - \left(\frac{T+1}{2}\right)\sum_{t=1}^{T} t^{-\alpha} \tag{66}$$

$$= \begin{cases} -\dfrac{\alpha}{2(\alpha-2)(\alpha-1)}T^{2-\alpha} + O\left(T^{1-\alpha}\right) & \text{if } \alpha < 1, \\ -\dfrac{1}{2}T\ln T + T + O(1) & \text{if } \alpha = 1, \\ -\dfrac{1}{2}\zeta(\alpha)T + O(1) & \text{if } \alpha > 1, \end{cases} \tag{67}$$

$$\begin{aligned} \mathcal{S}_T(\alpha) &= \sum_{t=1}^{T} \widetilde{t^{-\alpha}}\widetilde{t^{-\alpha}} = \sum_{t=1}^{T}\left(t^{-\alpha} - T^{-1}\sum_{t=1}^{T} t^{-\alpha}\right)t^{-\alpha} \\ &= \begin{cases} \dfrac{\alpha^2}{(\alpha-1)^2(1-2\alpha)}T^{1-2\alpha} + O(1) & \text{if } \alpha < 1/2, \\ \ln T + O(1) & \text{if } \alpha = 1/2, \\ \zeta(2\alpha) + o(1) & \text{if } \alpha > 1/2. \end{cases} \end{aligned} \tag{68}$$

For $\mathcal{B}_T(\alpha)$,we first simplify as follows

$$\begin{aligned} \mathcal{B}_T(\alpha) &= \frac{1}{T}\sum_{t=1}^{T}\left[\widetilde{t^{-\alpha}} - \tilde{t}\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\alpha}}\right]^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\left[\widetilde{t^{-\alpha}} - \tilde{t}\frac{\mathcal{T}_T(1,\alpha)}{\sum_{t=1}^{T}\tilde{t}^2}\right]^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{t^{-\alpha}}\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\left[\frac{\mathcal{T}_T(1,\alpha)}{\sum_{t=1}^{T}\tilde{t}^2}\right]^2 - 2\frac{1}{T}\sum_{t=1}^{T}\tilde{t}^2\left[\frac{\mathcal{T}_T(1,\alpha)}{\sum_{t=1}^{T}\tilde{t}^2}\right]^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{t^{-\alpha}}\right)^2 - \frac{1}{T}\frac{\left[\mathcal{T}_T(1,\alpha)\right]^2}{\sum_{t=1}^{T}\tilde{t}^2} \\ &= T^{-1}\mathcal{S}_T(\alpha) - 12T^{-4}\left[\mathcal{T}_T(1,\alpha)\right]^2. \end{aligned} \tag{69}$$

Then, using the results just established we have

$$\begin{aligned} \mathcal{B}_T(\alpha) &= T^{-1}\mathcal{S}_T(\alpha) - 12T^{-4}\left[\mathcal{T}_T(1,\alpha)\right]^2 = T^{-1}\mathcal{S}_T(\alpha)\left\{1 + O\left(T^{-1}\right)\right\} \\ &= \begin{cases} \dfrac{\alpha^2}{(\alpha-1)^2(1-2\alpha)}T^{-2\alpha} + O\left(T^{-1}\right) & \text{if } \alpha < 1/2, \\ T^{-1}\ln T + O\left(T^{-1}\right) & \text{if } \alpha = 1/2, \\ T^{-1}\zeta(2\alpha) + o\left(T^{-1}\right) & \text{if } \alpha > 1/2, \end{cases} \\ &= \begin{cases} O\left(T^{-2\alpha}\right) & \text{if } \alpha < 1/2, \\ O\left(T^{-1}\ln T\right) & \text{if } \alpha = 1/2, \\ O\left(T^{-1}\right) & \text{if } \alpha > 1/2, \end{cases} \end{aligned} \tag{70}$$

as required. □

Proof of Lemma 3 To derive the required asymptotic orders, it is sufficient to compute the order of the variances since all quantities have zero mean. By direct calculation

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} v_{it}t^{-\alpha}\right]^2 = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{T} v_{it}v_{is}t^{-\alpha}s^{-\alpha}\right] = \sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_{s-t,v,i}t^{-\alpha}s^{-\alpha} \\
= \sum_{h=-T+1}^{T-1}\gamma_{h,v,i}\sum_{t=1}^{T}\sum_{s=1}^{T} t^{-\alpha}s^{-\alpha}\mathbf{1}\{s-t=h\} \\
= \sum_{h=-T+1}^{T-1}\gamma_{h,v,i}\sum_{\substack{t=1 \\ 1\le t+h\le T}}^{T} t^{-\alpha}(t+h)^{-\alpha},
\end{aligned}
$$

and we deduce by summability and Cauchy-Schwarz that

$$
\mathbb{E}\left[\sum_{t=1}^{T} v_{it}t^{-\alpha}\right]^2 \le \sum_{h=-\infty}^{\infty}\left|\gamma_{h,v,i}\right|\left\{\left(\sum_{t=1}^{T} t^{-2\alpha}\right)^2\right\}^{1/2} = O\left(\tau_T(2\alpha)\right).
$$

Hence, $\sum_{t=1}^{T} v_{it}t^{-\alpha} = O_p\left([\tau_T(2\alpha)]^{1/2}\right)$ since $\mathbb{E}\left(\sum_{t=1}^{T} v_{it}t^{-\alpha}\right) = 0$. Next,

$$
\begin{aligned}
\sum_{t=1}^{T} v_{it}\widetilde{t^{-\alpha}} &= \sum_{t=1}^{T} v_{it}\left(t^{-\alpha} - \frac{1}{T}\sum_{t=1}^{T} t^{-\alpha}\right) \\
&= \sum_{t=1}^{T} v_{it}t^{-\alpha} - \frac{1}{T}\sum_{t=1}^{T} t^{-\alpha}\sum_{t=1}^{T} v_{it} \\
&= O_p\left([\tau_T(2\alpha)]^{1/2}\right) + O_p\left(T^{-1/2}\tau_T(\alpha)\right) = O_p\left([\tau_T(2\alpha)]^{1/2}\right).
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{t=1}^{T} \tilde{t}t^{-\alpha}v_{it} &= \sum_{t=1}^{T} v_{it}t^{1-\alpha} - \frac{T+1}{2}\sum_{t=1}^{T} v_{it}t^{-\alpha} \\
&= O_p\left(\left(\sum_{t=1}^{T} t^{2-2\alpha}\right)^{1/2}\right) + O_p\left(T\left(\sum_{t=1}^{T} t^{-2\alpha}\right)^{1/2}\right) \\
&= O_p\left(T[\tau_T(2\alpha)]^{1/2}\right).
\end{aligned}
$$

For the final result, note that

$$
\mathbb{E}\left[\sum_{t=1}^{T} b_i\tilde{t}t^{-\alpha}\right]^2 = \mathbb{E}\left[b_i\mathcal{T}_T(1,\alpha)\right]^2 = \sigma_b^2\mathcal{T}_T^2(1,\alpha),
$$

from which we deduce that $\sum_{t=1}^{T} b_i\tilde{t}t^{-\alpha} = O_p\left(\mathcal{T}_T(1,\alpha)\right)$. □

Proof of Lemma 4 Denote $t_\ell = t + \ell$ for any given integer $\ell \geq 1$ and observe that

$$G(T,\lambda) := \frac{1}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\left(1-\frac{\ell}{L+1}\right)\tilde{m}_t\tilde{m}_{t+\ell} \tag{71}$$

$$= \frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum_{t=1}^{T-\ell}\tilde{m}_t\left(m_{t+\ell}-\frac{1}{T-\ell}\sum_{s=1}^{T-\ell}m_{s+\ell}\right)$$

$$= \frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum_{t=1}^{T-\ell}\tilde{m}_t m_{t+\ell}$$

$$= \frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum_{t=1}^{T-\ell}\left[\widetilde{t^{-\lambda}}-\tilde{t}\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right]\times$$

$$\left[t_\ell^{-\lambda}-t_\ell\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right]$$

$$=: \frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\left(\Psi_{\ell,1}+\Psi_{\ell,2}\right),$$

where

$$\Psi_{\ell,1} := \sum_{t=1}^{T-\ell}\left[\widetilde{t^{-\lambda}}t_\ell^{-\lambda}-\widetilde{t^{-\lambda}}t_\ell\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)\right],$$

$$\Psi_{\ell,2} := \sum_{t=1}^{T-\ell}\left[-\tilde{t}t_\ell^{-\lambda}\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)\right.$$
$$\left.+\tilde{t}t_\ell\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-2}\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)^2\right].$$

We decompose the term $\Psi_{\ell,1}$ in (71) first, writing

$$\Psi_{\ell,1} = \sum_{t=1}^{T-\ell}\left[\widetilde{t^{-\lambda}}t_\ell^{-\lambda}-\widetilde{t^{-\lambda}}t_\ell\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)\right] \tag{72}$$

$$= \sum_{t=1}^{T-\ell}\left(t^{-\lambda}-\frac{1}{T-\ell}\sum_{t=1}^{T-\ell}t^{-\lambda}\right)t_\ell^{-\lambda}-\sum_{t=1}^{T-\ell}\left(\widetilde{t^{-\lambda}}t_\ell\right)\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)$$

$$= \sum_{t=1}^{T-\ell}\left(t^2+t\ell\right)^{-\lambda}-\frac{1}{T-\ell}\sum_{t=1}^{T-\ell}t_\ell^{-\lambda}\sum_{t=1}^{T-\ell}t^{-\lambda}$$
$$-\sum_{t=1}^{T-\ell}\left(\widetilde{t^{-\lambda}}t_\ell\right)\left(\sum_{t=1}^{T}\tilde{t}^2\right)^{-1}\left(\sum_{t=1}^{T}\widetilde{\tilde{t}t^{-\lambda}}\right)$$

$$=: \Psi_{\ell,11}+\Psi_{\ell,12}+\Psi_{\ell,13}$$

When $\lambda < 1$, as $T \to \infty$ with a finite $\ell \geq 1$, we have

$$\sum_{t=1}^{T-\ell}(t+\ell)^{-2\lambda} < \sum_{t=1}^{T-\ell}\left(t^2+t\ell\right)^{-\lambda} < \min\left(\sum_{t=1}^{T-\ell}t^{-2\lambda},\sum_{t=1}^{T-\ell}(t\ell)^{-\lambda}\right).$$

We calculate the upper bound first. Note that

$$\sum_{t=1}^{T-\ell}t^{-2\lambda} = \begin{cases}\dfrac{1}{1-2\lambda}(T-\ell)^{1-2\lambda}+O(1) & \text{if } \lambda < 1/2,\\ \ln(T-\ell)+O(1) & \text{if } \lambda = 1/2,\\ \zeta(2\lambda)=O(1) & \text{if } \lambda > 1/2,\end{cases}$$

so that

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

Meanwhile for $\lambda<1$,

$$\begin{aligned}
&\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}\left(t\ell\right)^{-\lambda}<\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\ell^{-\lambda}\sum\nolimits_{t=1}^{T}t^{-\lambda} &(73)\\
=\ &\frac{1}{1-\lambda}\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\left[T^{1-\lambda}\ell^{-\lambda}+O\left(1\right)\right]\\
\sim\ &\frac{1}{1-\lambda}T^{-\lambda}\sum\nolimits_{\ell=1}^{T^{\kappa}}\ell^{-\lambda}-\frac{1}{1-\lambda}T^{-\lambda-\kappa}\sum\nolimits_{\ell=1}^{T^{\kappa}}\ell^{1-\lambda}\\
\sim\ &\frac{1}{\left(1-\lambda\right)^{2}}T^{-\lambda}T^{\kappa(1-\lambda)}-\frac{1}{\left(1-\lambda\right)\left(2-\lambda\right)}T^{-\lambda-\kappa}T^{\kappa(2-\lambda)}\\
=\ &\frac{1}{\left(1-\lambda\right)^{2}\left(2-\lambda\right)}T^{-\lambda+\kappa-\lambda\kappa}=O\left(T^{-\lambda+\kappa-\lambda\kappa}\right) &(74)
\end{aligned}$$

uniformly in $\ell\leq L=\lfloor T^{\kappa}\rfloor$ with $\kappa<1$. Hence

$$\begin{aligned}
&\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}\left(t^{2}+t\ell\right)^{-\lambda} &(75)\\
<\ &\begin{cases}\min\left[O\left(T^{-2\lambda+\kappa}\right),O\left(T^{-\lambda+\kappa-\lambda\kappa}\right)\right]=O\left(T^{-2\lambda+\kappa}\right) & \text{if } \lambda<1/2,\\ \min\left[O\left(T^{\kappa-1}\ln T\right),O\left(T^{\kappa-1/2-\kappa/2}\right)\right]=O\left(T^{\kappa-1}\ln T\right) & \text{if } \lambda=1/2,\\ \min\left[O\left(T^{\kappa-1}\right),O\left(T^{-\lambda+\kappa-\lambda\kappa}\right)\right]=O\left(T^{\kappa-1}\right) & \text{if } 1/2<\lambda<1/\left(1+\kappa\right),\\ \min\left[O\left(T^{\kappa-1}\right),O\left(T^{-\lambda+\kappa-\lambda\kappa}\right)\right]=O\left(T^{-\lambda+\kappa-\lambda\kappa}\right) & \text{if } 1/\left(1+\kappa\right)\leq\lambda<1.\end{cases}
\end{aligned}$$

Next, we consider the lower bound. We have

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

Then

$$\begin{aligned}
&\frac{1}{T}\sum\nolimits_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum\nolimits_{t=1}^{T-\ell}\left(t^{2}+t\ell\right)^{-\lambda} &(76)\\
>\ &\begin{cases}\dfrac{T^{-2\lambda}L}{1-2\lambda} & \text{if } \lambda<1/2\\ \frac{L}{T}\ln T & \text{if } \lambda=1/2\\ T^{-1+\kappa}\zeta\left(2\lambda\right) & \text{if } \lambda>1/2\end{cases}=\begin{cases}O\left(T^{-2\lambda+\kappa}\right) & \text{if } \lambda<1/2,\\ O\left(T^{-1+\kappa}\ln T\right) & \text{if } \lambda=1/2,\\ O\left(T^{-1+\kappa}\right) & \text{if } 1>\lambda>1/2.\end{cases}
\end{aligned}$$

Combining (76) with (75) yields

$$\frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\sum_{t=1}^{T-\ell}\left(t^2+t\ell\right)^{-\lambda} \tag{77}$$
$$\sim \begin{cases} \dfrac{T^{-2\lambda+\kappa}}{1-2\lambda} & \text{if } \lambda<1/2, \\ T^{-1+\kappa}\ln T & \text{if } \lambda=1/2, \\ T^{-1+\kappa}\zeta\left(2\lambda\right) & \text{if } 1/2<\lambda<1/\left(1+\kappa\right), \\ \dfrac{1}{\left(1-\lambda\right)^2\left(2-\lambda\right)}T^{-\lambda+\kappa-\lambda\kappa} & \text{if } 1/\left(1+\kappa\right)\leq\lambda<1, \end{cases}$$
$$= \begin{cases} O\left(T^{-2\lambda+\kappa}\right) & \text{if } \lambda<1/2, \\ O\left(T^{\kappa-1}\ln T\right) & \text{if } \lambda=1/2, \\ O\left(T^{\kappa-1}\right) & \text{if } 1/2<\lambda<1/\left(1+\kappa\right), \\ O\left(T^{-\lambda+\kappa-\lambda\kappa}\right) & \text{if } 1/\left(1+\kappa\right)\leq\lambda<1. \end{cases}$$

In the first three cases, the asymptotic order of the sum is the same as the asymptotic order of the upper and lower bounds, as confirmed in the above derivation. In the last case when $1/\left(1+\kappa\right)\leq\lambda<1$, we use the upper bound. Note that when $\lambda<\kappa/2$, this term increases as T increases, but when $\lambda\geq\kappa/2$, this term decreases as T increases.[13]

When $\lambda=1$ and for $\ell\to\infty$ we have

$$\begin{aligned} \Psi_{\ell,11} &= \sum_{t=1}^{T-\ell}\frac{1}{t^2+t\ell}=\frac{1}{\ell}\sum_{t=1}^{T-\ell}\left(\frac{1}{t}-\frac{1}{t+\ell}\right)=\frac{1}{\ell}\left\{\left(\sum_{t=1}^{T-\ell}\frac{1}{t}\right)-\left(\sum_{t=1}^{T}\frac{1}{t}-\sum_{s=1}^{\ell}\frac{1}{s}\right)\right\} \\ &= \frac{1}{\ell}\left\{\left(\ln\left(T-\ell\right)+\gamma_e\right)-\left(\ln T+\gamma_e-\left[\ln\ell+\gamma_e\right]\right)+O\left(\frac{1}{\ell}+\frac{\ell}{T}\right)\right\} \\ &= \frac{1}{\ell}\left(\ln\ell+\gamma_e\right)+O\left(\frac{1}{\ell^2}+\frac{1}{T}\right), \end{aligned} \tag{78}$$

and when ℓ is fixed we have

$$\begin{aligned} \Psi_{\ell,11} &= \frac{1}{\ell}\sum_{s=1}^{\ell}\frac{1}{s}-\frac{1}{\ell}\sum_{t=T-\ell+1}^{T}\frac{1}{t}=\frac{1}{\ell}\sum_{s=1}^{\ell}\frac{1}{s}+\frac{1}{\ell}\left\{\ln\frac{T-\ell}{T}+O\left(T^{-1}\right)\right\} \\ &= \frac{1}{\ell}\sum_{s=1}^{\ell}\frac{1}{s}+O\left(\frac{1}{T}\right). \end{aligned} \tag{79}$$

Then, using (78), (79), and with $L=T^{\kappa}\to\infty$ as $T\to\infty$, we obtain

$$\begin{aligned} &\frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\Psi_{\ell,1}\sim\frac{1}{T}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\left(\frac{1}{\ell}\sum_{s=1}^{\ell}\frac{1}{s}\right) \\ =&\ \frac{1}{T}\sum_{\ell=1}^{L}\ell^{-1}\sum_{s=1}^{\ell}s^{-1}-\frac{1}{T\left(L+1\right)}\sum_{\ell=1}^{L}\sum_{s=1}^{\ell}s^{-1} \\ \sim&\ \frac{1}{T}\int_{\ell=1}^{L}\ell^{-1}\int_{s=1}^{\ell}s^{-1}ds\,d\ell-\frac{1}{T\left(L+1\right)}\int_{\ell=1}^{L}\int_{s=1}^{\ell}s^{-1}ds\,d\ell \\ =&\ \frac{1}{T}\int_{\ell=1}^{L}\ell^{-1}\left(\ln\ell\right)d\ell-\frac{1}{T\left(L+1\right)}\int_{\ell=1}^{L}\ln\ell\, d\ell \\ =&\ \frac{1}{2T}\ln^2 L+O\left(T^{-1}\ln T\right)=\frac{\kappa^2}{2T}\ln^2 T+O\left(T^{-1}\ln T\right). \end{aligned}$$

[13] A graphical demonstration of the relevance of the decay rate (73) in determining the behavior of $G\left(T,\lambda\right)$ as $T\to\infty$ is given in Fig. A1 at the end of this Appendix.

When $\lambda > 1$, we find from Lemma 1 that

$$\begin{aligned}
\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\Psi_{\ell,11} &= \lim_{T\to\infty}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\left(1-\frac{\ell}{L+1}\right)\left(t^2+t\ell\right)^{-\lambda} \\
&= \lim_{L\to\infty}\sum_{\ell=1}^{L}\sum_{t=1}^{L}\left(1-\frac{\ell}{L+1}\right)\frac{1}{t^{\lambda}\left(t+\ell\right)^{\lambda}} \\
&= \lim_{L\to\infty}\sum_{t=1}^{L}\frac{1}{t^{\lambda}}\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\frac{1}{\left(t+\ell\right)^{\lambda}} \\
&= \lim_{L\to\infty}\sum_{t=1}^{L}\frac{1}{t^{\lambda}}\sum_{\ell=1}^{L}\frac{1}{\left(t+\ell\right)^{\lambda}}+O\left(L^{1-\lambda}\right) \\
&= \sum_{t=1}^{\infty}\frac{1}{t^{\lambda}}\left\{\zeta\left(\lambda,t\right)-\frac{1}{t^{\lambda}}\right\}+O\left(L^{1-\lambda}\right) \\
&= \sum_{t=1}^{\infty}\frac{\zeta\left(\lambda,t\right)}{t^{\lambda}}-\zeta\left(2\lambda\right)
\end{aligned}$$

where $\zeta\left(\lambda,\ell\right)$ is the Hurwitz zeta function which is well defined for all $\lambda > 1$ and $t > 0$. Note, in particular, that

$$\zeta\left(\lambda,t\right)=\sum_{\ell=0}^{\infty}\frac{1}{\left(t+\ell\right)^{\lambda}}=\sum_{\ell=1}^{\infty}\frac{1}{\left(t+\ell\right)^{\lambda}}+\frac{1}{t^{\lambda}}<\zeta\left(\lambda\right)+\frac{1}{t^{\lambda}},$$

with strict inequality showing that[14]

$$\sum_{t=1}^{\infty}\frac{\zeta\left(\lambda,t\right)}{t^{\lambda}}<\sum_{t=1}^{\infty}\frac{\zeta\left(\lambda\right)}{t^{\lambda}}+\sum_{t=1}^{\infty}\frac{1}{t^{2\lambda}}$$

$$\sum_{t=1}^{\infty}\frac{\zeta\left(\lambda,t\right)}{t^{\lambda}}-\zeta\left(2\lambda\right)<\zeta\left(\lambda\right)^2=O\left(1\right).$$

Our next step is to calculate $\sum_{\ell=1}^{L}\left(1-\frac{\ell}{L+1}\right)\left(\Psi_{\ell,12}+\Psi_{\ell,13}+\Psi_{\ell,2}\right)$ but before doing so we provide summation results for $\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\tilde{t}t_{\ell}^{-\lambda}$ and $\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\widetilde{t^{-\lambda}}t_{\ell}$. These results and their orders

[14] It is not needed in the results for Lemma 4 but is interesting to note (and will be used later) that as $\lambda\to\infty$, we get $\lim_{\lambda\to\infty}\left\{\sum_{t=1}^{\infty}t^{-\lambda}\zeta\left(\lambda,t\right)-\zeta\left(2\lambda\right)\right\}=1-1=0$, which corresponds to

$$\lim_{\lambda\to\infty}\lim_{M\to\infty}\sum_{\ell=1}^{M}\sum_{t=1}^{M}\frac{1}{t^{\lambda}\left(t+\ell\right)^{\lambda}}=\lim_{M\to\infty}\sum_{\ell=1}^{M}\sum_{t=1}^{M}\lim_{\lambda\to\infty}\frac{1}{t^{\lambda}\left(t+\ell\right)^{\lambda}}=0,$$

whereas $\lim_{\lambda\to\infty}\zeta\left(\lambda\right)^2=1$. Therefore, as $\lambda\to\infty$ we find that the serial correlation terms

$$\frac{1}{T}\sum_{\ell=1}^{M}\sum_{t=1}^{T-\ell}\left(1-\frac{\ell}{M+1}\right)\tilde{m}_t\tilde{m}_{t+\ell}\to 0.$$

of magnitude are given as follows:

$$\begin{aligned}
\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\tilde{t}t_{\ell}^{-\lambda} &= \sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(t+\ell-\frac{1}{T-\ell}\sum\nolimits_{t=1}^{T-\ell}(t+\ell)\right)t_{\ell}^{-\lambda} \\
&= \sum\nolimits_{\ell=1}^{L}\sum\nolimits_{s=\ell+1}^{T}\tilde{s}s^{-\lambda} = \sum\nolimits_{\ell=1}^{L}\left(\sum\nolimits_{s=1}^{T}\tilde{s}s^{-\lambda}-\sum\nolimits_{s=1}^{\ell}\tilde{s}s^{-\lambda}\right) \\
&= \sum\nolimits_{\ell=1}^{L}\left(\mathcal{T}_{T}(1,\lambda)-\mathcal{T}_{\ell}(1,\lambda)\right) \\
&\sim \begin{cases} -\dfrac{\lambda}{2(\lambda-2)(\lambda-1)}T^{2-\lambda+\kappa} & \text{if } \lambda<1, \\ -\dfrac{1}{2}T^{1+\kappa}\ln T & \text{if } \lambda=1, \\ -\dfrac{1}{2}\zeta(\lambda)T^{1+\kappa} & \text{if } \lambda>1, \end{cases}
\end{aligned} \tag{80}$$

and

$$\begin{aligned}
& \sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\widetilde{t^{-\lambda}}t_{\ell} \\
=\ & \sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(\left(t^{-\lambda}-\frac{1}{T-\ell}\sum\nolimits_{t=1}^{T-\ell}t^{-\lambda}\right)(t+\ell)\right) \\
=\ & \sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(t^{-\lambda}-\frac{1}{T-\ell}\sum\nolimits_{t=1}^{T-\ell}t^{-\lambda}\right)t \\
=\ & \sum\nolimits_{\ell=1}^{L}\mathcal{T}_{T-\ell}(1,\lambda) \\
\sim\ & \begin{cases} -\dfrac{\lambda}{2(\lambda-2)(\lambda-1)}T^{2-\lambda+\kappa} & \text{if } \lambda<1, \\ -\dfrac{1}{2}T^{1+\kappa}\ln T & \text{if } \lambda=1, \\ -\dfrac{1}{2}\zeta(\lambda)T^{1+\kappa} & \text{if } \lambda>1, \end{cases} \\
=\ & \begin{cases} O\left(T^{\kappa+2-\lambda}\right) & \text{if } \lambda<1, \\ O\left(T^{\kappa+1}\ln T\right) & \text{if } \lambda=1, \\ O\left(T^{\kappa+1}\right) & \text{if } \lambda>1. \end{cases}
\end{aligned} \tag{81}$$

Then, using results (80) - (81) and Lemmas 1 and 2, we have

$$\begin{aligned}
\sum\nolimits_{\ell=1}^{L}\Psi_{\ell,12} &= -\sum\nolimits_{\ell=1}^{L}\frac{1}{T-\ell}\left(\sum\nolimits_{t=1}^{T-\ell}t^{-\lambda}\right)\sum\nolimits_{t=1}^{T-\ell}t_{\ell}^{-\lambda} \\
&= -\sum\nolimits_{\ell=1}^{L}\frac{1}{T-\ell}\left(\sum\nolimits_{t=1}^{T-\ell}t^{-\lambda}\right)\sum\nolimits_{\ell+1}^{T}t^{-\lambda} \\
&= -\sum\nolimits_{\ell=1}^{L}\frac{1}{T-\ell}\left(\sum\nolimits_{t=1}^{T-\ell}t^{-\lambda}\right)\left(\sum\nolimits_{t=1}^{T}t^{-\lambda}-\sum\nolimits_{t=1}^{\ell}t^{-\lambda}\right) \\
&= \begin{cases} O\left(T^{\kappa+1-2\lambda}\right) & \text{if } \lambda<1, \\ O\left(T^{\kappa-1}\ln^{2}T\right) & \text{if } \lambda=1, \\ O\left(T^{\kappa-1}\right) & \text{if } \lambda>1, \end{cases}
\end{aligned}$$

Note, when $\lambda < 1, \sum_{\ell=1}^{L} \Psi_{\ell,12} \sim -\frac{1}{(1-\lambda)^2} T^{\kappa+1-2\lambda}$.

$$\sum_{\ell=1}^{L} \Psi_{\ell,13} = -\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \left(\sum_{t=1}^{T} \widetilde{\tilde{t}t^{-\lambda}}\right) \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left(\widetilde{t^{-\lambda}} t_\ell\right)$$
$$= \begin{cases} O\left(T^{-3}\right) O\left(T^{2-\lambda}\right) O\left(T^{\kappa+2-\lambda}\right) = O\left(T^{\kappa+1-2\lambda}\right) & \text{if } \lambda < 1, \\ O\left(T^{-3}\right) O\left(T \ln T\right) O\left(T^{\kappa+1} \ln T\right) = O\left(T^{\kappa-1} \ln^2 T\right) & \text{if } \lambda = 1, \\ O\left(T^{-3}\right) O\left(T\right) O\left(T^{\kappa+1}\right) = O\left(T^{\kappa-1}\right) & \text{if } \lambda > 1. \end{cases}$$

When $\lambda < 1, \sum_{\ell=1}^{L} \Psi_{\ell,13} \sim -\dfrac{3\lambda^2}{(\lambda-2)^2 (\lambda-1)^2} T^{1-2\lambda+\kappa}$.

$$\sum_{\ell=1}^{L} \Psi_{\ell,2} = -\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} \left(\sum_{t=1}^{T} \widetilde{\tilde{t}t^{-\lambda}}\right) \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \tilde{t} t_\ell^{-\lambda}$$
$$+ \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-2} \left(\sum_{t=1}^{T} \widetilde{\tilde{t}t^{-\lambda}}\right)^2 \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \tilde{t} t_\ell$$
$$= \begin{cases} O\left(T^{-3}\right) O\left(T^{2-\lambda}\right) O\left(T^{\kappa+2-\lambda}\right) + O\left(T^{-6}\right) O\left(T^{4-2\lambda}\right) O\left(T^{\kappa+3}\right) & \text{if } \lambda < 1, \\ O\left(T^{-3}\right) O\left(T \ln T\right) O\left(T^{\kappa+1} \ln T\right) + O\left(T^{-6}\right) O\left(T^2 \ln^2 T\right) O\left(T^{\kappa+3}\right) & \text{if } \lambda = 1, \\ O\left(T^{-3}\right) O\left(T\right) O\left(T^{\kappa+1}\right) + O\left(T^{-6}\right) O\left(T^2\right) O\left(T^{\kappa+3}\right) & \text{if } \lambda > 1, \end{cases}$$
$$= \begin{cases} O\left(T^{\kappa+1-2\lambda}\right) & \text{if } \lambda < 1, \\ O\left(T^{\kappa-1} \ln^2 T\right) & \text{if } \lambda = 1, \\ O\left(T^{\kappa-1}\right) & \text{if } \lambda > 1. \end{cases}$$

When $\lambda < 1, \sum_{\ell=1}^{L} \Psi_{\ell,2} \sim -3\left(\dfrac{\lambda}{(\lambda-2)(\lambda-1)}\right)^2 T^{1-2\lambda+\kappa} + 3T^{-6}\left(\dfrac{\lambda}{(\lambda-2)(\lambda-1)} T^{2-\lambda}\right)^2 T^{3+\kappa} = 0$.

Combining these three terms we find that

$$\sum_{\ell=1}^{L} \left(1 - \frac{\ell}{L+1}\right) \left(\Psi_{\ell,12} + \Psi_{\ell,13} + \Psi_{\ell,2}\right) = \begin{cases} O\left(T^{\kappa+1-2\lambda}\right) & \text{if } \lambda < 1, \\ O\left(T^{\kappa-1} \ln^2 T\right) & \text{if } \lambda = 1, \\ O\left(T^{\kappa-1}\right) & \text{if } \lambda > 1. \end{cases} \quad (82)$$

When $\lambda < 1, \sum_{\ell=1}^{L} \left(1 - \frac{\ell}{L+1}\right) \left(\Psi_{\ell,12} + \Psi_{\ell,13} + \Psi_{\ell,2}\right) \sim -4 \frac{\lambda^2 - \lambda + 1}{\left(\lambda^2 - 3\lambda + 2\right)^2} T^{1-2\lambda+\kappa}$.

Hence, we have

$$\frac{1}{T} \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left(1 - \frac{\ell}{L+1}\right) \tilde{m}_t \tilde{m}_{t+\ell}$$
$$= \sum_{\ell=1}^{L} \left(1 - \frac{\ell}{L+1}\right) \left(\Psi_{\ell,11} + \Psi_{\ell,12} + \Psi_{\ell,13} + \Psi_{\ell,2}\right)$$
$$= \begin{cases} \frac{\lambda^2 (\lambda+1)^2}{(1-2\lambda)\left(\lambda^2 - 3\lambda + 2\right)^2} T^{-2\lambda+\kappa} & \text{if } \lambda < 1/2, \\ T^{-1+\kappa} \ln T & \text{if } \lambda = 1/2, \\ T^{-1+\kappa} \zeta\left(2\lambda\right) & \text{if } 1/2 < \lambda < 1/\left(1+\kappa\right), \\ \dfrac{1}{(1-\lambda)^2 (2-\lambda)} T^{-\lambda+\kappa-\lambda\kappa} & \text{if } 1/\left(1+\kappa\right) \le \lambda < 1, \\ \frac{\kappa^2}{2} T^{-1} \ln^2 T + O\left(T^{-2} \ln T\right) & \text{if } \lambda = 1, \\ T^{-1} \left\{\sum_{t=1}^{\infty} t^{-\lambda} \zeta\left(\lambda, \ell\right) - \zeta\left(2\lambda\right)\right\} & \text{if } \lambda > 1. \end{cases}$$

□

Proof of Lemma 5 Since $\xi_{b,n} = O_p\left(n^{-1/2}\right)$ and $n/T \to \infty$, we have

$$\xi_{b,n}\widetilde{t^{-\alpha}} = O_p\left(n^{-1/2}\widetilde{t^{-\alpha}}\right) = o_p\left(T^{-1/2}\widetilde{t^{-\alpha}}\right).$$

Note that

$$\widetilde{t^{-\alpha}}T^{-1/2} = \begin{cases} t^{-\alpha}T^{-1/2} - \dfrac{1}{1-\alpha}T^{-1/2}T^{-\alpha} & \text{if } \alpha < 1, \\ t^{-\alpha}T^{-1/2} - T^{-1}T^{-1/2}\ln T & \text{if } \alpha = 1, \\ t^{-\alpha}T^{-1/2} - \mathcal{Z}_T(\alpha)\,T^{-1}T^{-1/2} & \text{if } \alpha > 1, \end{cases}$$

and

$$\widetilde{t^{-2\alpha}} = \begin{cases} t^{-2\alpha} - \dfrac{1}{1-2\alpha}T^{-2\alpha} & \text{if } \alpha < 1/2, \\ t^{-2\alpha} - T^{-1}\ln T & \text{if } \alpha = 1/2, \\ t^{-2\alpha} - \mathcal{Z}_T(2\alpha)\,T^{-1} & \text{if } \alpha > 1/2. \end{cases}$$

Hence

$$\begin{aligned} &\widetilde{t^{-2\alpha}} - \widetilde{t^{-\alpha}}T^{-1/2} \\ = &\begin{cases} t^{-2\alpha} - \dfrac{1}{1-2\alpha}T^{-2\alpha} - t^{-\alpha}T^{-1/2} + \dfrac{1}{1-\alpha}T^{-1/2}T^{-\alpha} & \text{if } \alpha < 1/2, \\ t^{-2\alpha} - T^{-1}\ln T - t^{-\alpha}T^{-1/2} + \dfrac{1}{1-\alpha}T^{-1/2}T^{-\alpha} & \text{if } \alpha = 1/2, \\ t^{-2\alpha} - \mathcal{Z}_T(2\alpha)\,T^{-1} - t^{-\alpha}T^{-1/2} + \dfrac{1}{1-\alpha}T^{-1/2}T^{-\alpha} & \text{if } 1/2 < \alpha < 1, \\ t^{-2\alpha} - \mathcal{Z}_T(2\alpha)\,T^{-1} - t^{-\alpha}T^{-1/2} + T^{-1}T^{-1/2}\ln T & \text{if } \alpha = 1, \\ t^{-2\alpha} - \mathcal{Z}_T(2\alpha)\,T^{-1} - t^{-\alpha}T^{-1/2} + \mathcal{Z}_T(\alpha)\,T^{-1}T^{-1/2} & \text{if } \alpha > 1, \end{cases} \\ = &\ \widetilde{t^{-2\alpha}} + o\left(\widetilde{t^{-2\alpha}}\right). \end{aligned}$$

For example, when $\alpha = 1/2$, as $T \to \infty$ we have

$$\begin{aligned} \lim_{T\to\infty} \frac{-\widetilde{t^{-\alpha}}T^{-1/2}}{\widetilde{t^{-2\alpha}}} &= \lim_{T\to\infty} \frac{-t^{-\alpha}T^{-1/2} + \dfrac{1}{1-\alpha}T^{-1/2}T^{-\alpha}}{\widetilde{t^{-2\alpha}}} = \lim_{T\to\infty} \frac{-t^{-1/2}T^{-1/2} + 2T^{-1}}{t^{-1} - T^{-1}\ln T} \\ &= \lim_{T\to\infty} \frac{-\left(t/T\right)^{1/2} + 2\left(t/T\right)}{1 - \left(t/T\right)\ln T} \to 0, \end{aligned}$$

since $t/T \in (0,1]$ as $T \to \infty$ for $1 \le t \le T$. Hence, for all $t \le T$, $-\widetilde{t^{-\alpha}}T^{-1/2} = o\left(\widetilde{t^{-2\alpha}}\right)$. The required result (47) now follows.

□

Power Trend Regression

We explore the impact on Theorem 1 of using a power trend regression of the form (35) in place of a linear trend regression. In (35) the empirical regression involves the power trend regressor t^{ψ} for some given power parameter $\psi > 0$. Direct calculations extending the results in Theorem 1 show that the asymptotic behavior of the regression coefficient $\hat{\phi}_{nT}$ in this case is as follows:

$$\hat{\phi}_{nT} = \begin{cases} O_p\left(T^{-\psi-\lambda}\right) & \text{if } 0 < \lambda < 1, \\ O_p\left(T^{-1-\psi}\ln T\right) & \text{if } \lambda = 1, \\ O_p\left(T^{-1-\psi}\right) & \text{if } \lambda > 1. \end{cases} \tag{83}$$

rather than $\hat{\phi}_{nT} = O_p(\mathcal{O}_{T\lambda})$, where λ in (83) is as given in Theorem 1. Upon calculation, we find that

$$\left(\sum_{t=1}^{T}\left[t^{\psi} - \frac{1}{T}\sum_{t=1}^{T} t^{\psi}\right]^2\right)^{-1} \sum_{t=1}^{T}\left[t^{\psi} - \frac{1}{T}\sum_{t=1}^{T} t^{\psi}\right]\varepsilon_{n,t} = O_p\left(n^{-1/2}T^{-1/2-\psi}\right),$$

and then

$$n^{1/2}T^{1/2+\psi}\hat{\phi}_{nT} = \begin{cases} O\left(n^{1/2}T^{1/2-\lambda}\right) & \text{if } 0 < \lambda < 1, \\ O\left(n^{1/2}T^{-1/2}\ln T\right) & \text{if } \lambda = 1, \\ O\left(n^{1/2}T^{-1/2}\right) & \text{if } \lambda > 1. \end{cases}$$

Hence divergence of the scaled statistic $n^{1/2}T^{1/2+\psi}\hat{\phi}_{nT}$ requires $n/T \to \infty$ regardless of the value of ψ. Thus using a power trend regression with regressor t^{ψ} instead of a simple linear trend does not lead to different requirements regarding (n, T).

Additional Numerical Calculations

We extend the numerical calculations given in Section 5 of the paper for model M1 by conducting related computations for models M2 and M3. As either n or β increases, the variance of $e_{n,t}$ shrinks to zero for given T, so that the t-ratio diverges to a negative infinity under the alternative as $n \to \infty$ with a fixed β, or approaches the limit value of $-\sqrt{3}$ given in Theorem 2 as $\beta \to \infty$. We first investigate the finite sample behavior of the t-ratio for given n and T. Figure S1 plots the empirical density functions of the t-ratio with various β values in M2. We set $n = 100$, $T = 200$, $\sigma_a^2 = 1$, and $\epsilon_{it} \sim iid\mathcal{N}(0,1)$. As β decreases, the variance of the t ratio increases and the mean of the distribution moves to the left. Even for moderately large n and T, the entire empirical distribution of the t-ratio still lies in the left side of the critical value, -1.65, with for $\beta = 2$. As β passes to infinity, the empirical distribution collapses to a mass point at $-\sqrt{3} = -1.73$. For $\beta = 0.5$, x_{it} is convergent, the trend regression test is consistent, and its strong discriminatory power is evident in the density shown in Figure S1.

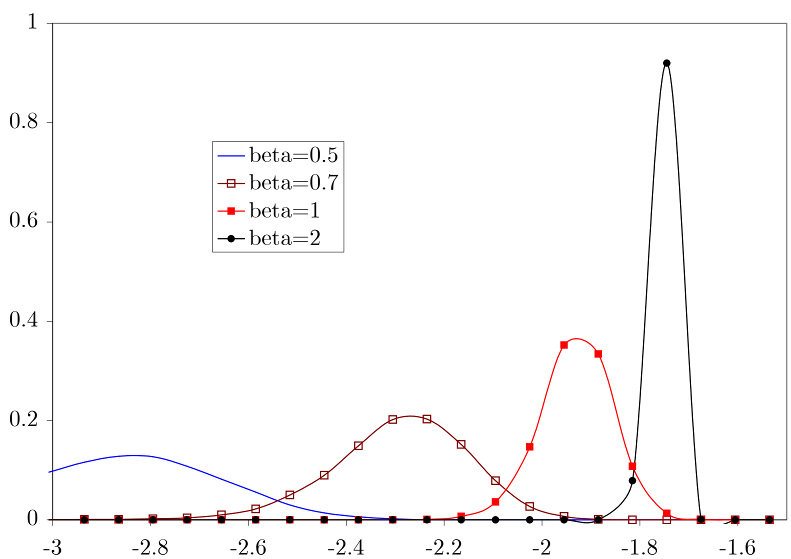


Figure S1: Empirical distribution of $t_{\hat{\phi}_{nT}}$ in M2
$T = 200$, $n = 100$, $\sigma_a^2 = 1$, $\epsilon_{it} \sim iid\mathcal{N}(0,1)$, $\kappa = 1/3$

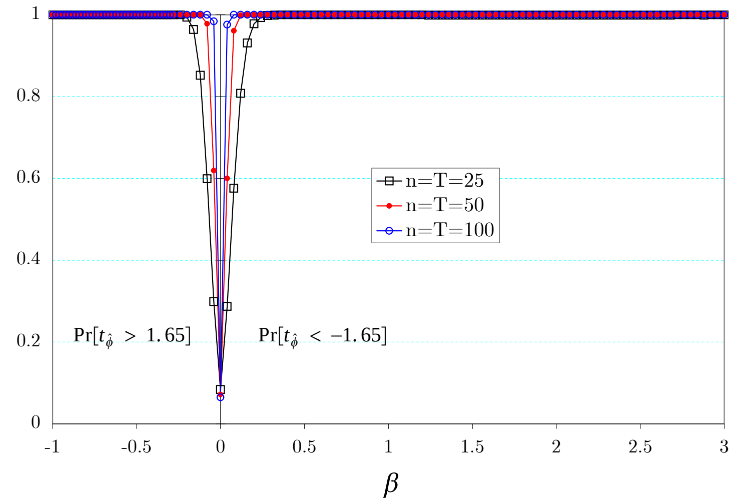


Figure S2: Rejection Frequencies plotted against β in Model M2 for a 5% level test with $\sigma_a^2 = 1$, $\sigma_\epsilon^2 = 4$.

Figure S2 displays the rapid changes in the power function of a 5% level test near $\beta = 0$ as n and T increase with $T = n$. As Theorem 2 indicates, no asymptotic n/T ratio condition is required for test consistency in this case. Evidently, as sample size increases, the rapid movement in the power function near $\beta = 0$ becomes more accentuated. The power function is unity outside a small neighborhood of $\beta = 0$ even for $n = T = 25$ because the empirical distribution of the t-ratio is well separated from the test critical value of -1.65.

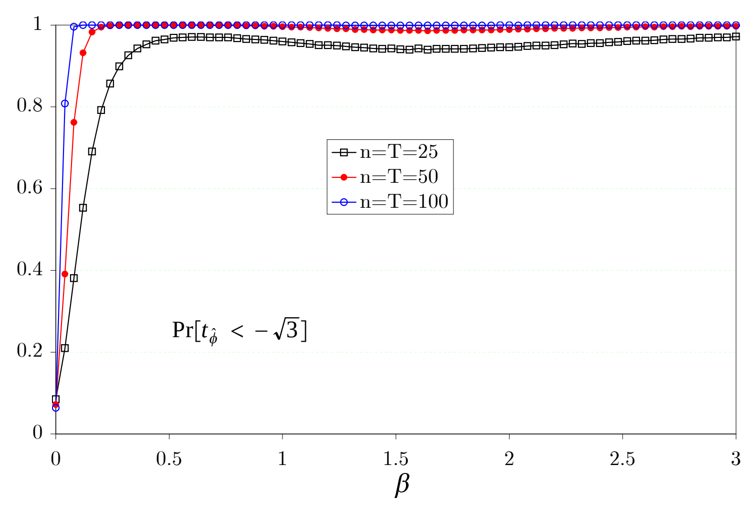


Figure S3: The Rejection Frequencies with the critical value of $-\sqrt{3}$ over $\lambda = 2\beta$ in Model 2 ($\sigma_a^2 = 1$, $\sigma_\epsilon^2 = 2$)

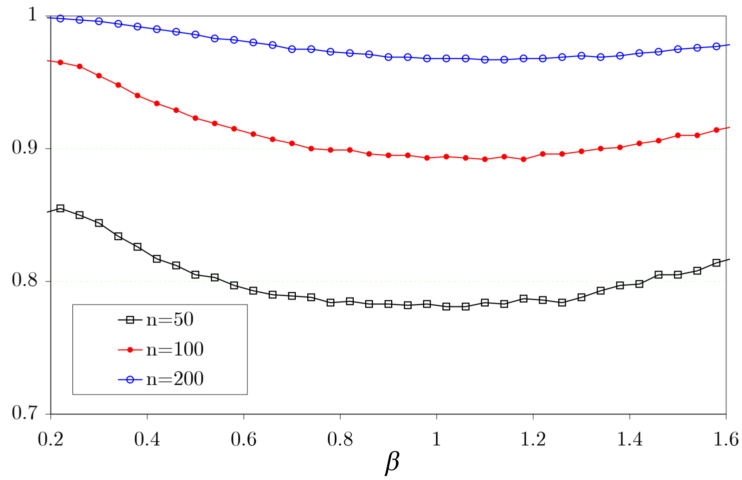


Figure S4: Test Rejection Frequencies in model M3
$\left(\alpha = 2\beta - 0.1,\ \sigma_a^2 = 8,\ \sigma_{a\mu}^2 = 0,\ \sigma_\epsilon^2 = \sigma_\mu^2 = 1\right)$

Since Figure S2 reports the rejection probability for the 5% level test with critical value of -1.65, which exceeds $-\sqrt{3}$, this figure does not reveal the relationship of the density function of the t-ratio to the limit value $-\sqrt{3}$. To explore this issue, we set the critical value of the test to $-\sqrt{3}$ and plot the associated rejection frequencies in Figure S3. When n, T is small, some portion of the t-ratio is in fact larger than $-\sqrt{3}$ so that in that case the rejection frequency is less than unity. However as n increases, the power function reaches unity rapidly, thereby indicating that the asympotic theory holds well in finite samples.

Finally, we consider model M3. As shown in Theorems 1 and 2, when either $2\beta > \alpha > \beta$ or $\alpha \leq \beta < 1$, the n/T ratio matters in the limit theory and we explore finite sample performance in this case, setting $\alpha = 2\beta - 0.1$, $\sigma_a^2 = 8$, and fixing $T = 50$ for all values of n. Figure S4 displays the power of the test for various values of β. The power functions are seen in the figure to have a mild $U-$shape and minimum power is found around $\beta = 1$. When $\beta > 1$, power increases as β increases. It is also apparent from Figure S4 that as n increases with T fixed, the n/T ratio increases and test power approaches unity around $\beta = 1$.

Approximation accuracy of (74) for $G(T, \lambda)$

We can assess the adequacy of the approximation (74) in a graphical demonstration by using the deterministic DGP in (44) to characterize the large T behavior of $G(T, \lambda) = \frac{1}{T} \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left(1 - \frac{\ell}{L+1}\right) \tilde{m}_t \tilde{m}_{t+\ell}$ when $\lambda < 1$. We let $\kappa = 1/3$ and compute the ratio $\frac{G(10^k, \lambda)}{G(10^6, \lambda)}$ for various values of λ and $k = 3, 4, 5$. The plots are shown in Figure A1. Evidently, the ratios exceed unity for large λ but rapidly decrease as λ decreases. The threshold value of λ for $G(T, \lambda)$ to decay as $T \to \infty$ is $\lambda > \kappa/2 = 1/6 \simeq 0.167$ for $\kappa = 1/3$, which is evidently well-matched in the figure, corroborating the limit behavior

$G(T, \lambda) = O\left(T^{-\lambda+\kappa-\lambda\kappa}\right) \searrow 0$ given in (74).

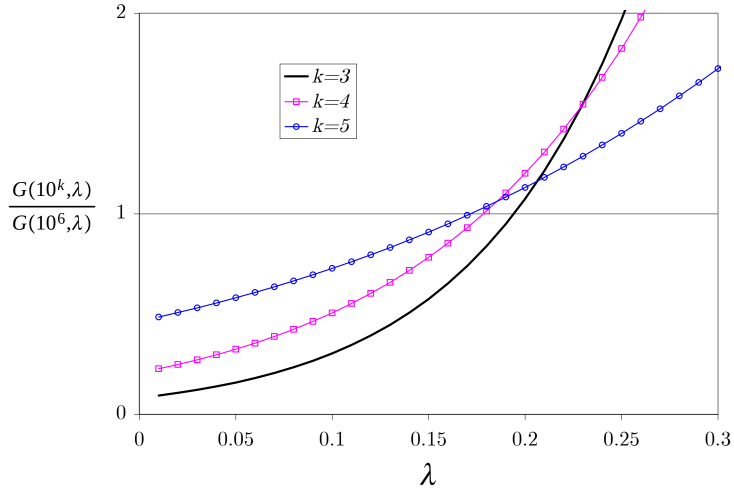


Figure A.1: Approximation Accuracy of (74) for $G(T, \lambda)$