Weak σ- Convergence: Theory and Applications*

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

January 4, 2019

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 46 disaggregate personal consumption expenditure (PCE) price inflation items σ-converge over time, finding strong evidence of weak σ-convergence in these data. 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

---

*The authors thank the Co-Editor, Associate Editor, and two referees for most helpful comments. Phillips acknowledges partial 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 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. 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. Also this section provides a formal development of the concept of weak $\sigma-$convergence, discusses various matters of formulation and interpretation in the context of several prototypical decay function models of convergence, and introduces the linear trend regression approach and an associated t-ratio test of convergence designed for practical implementation. Section 3 derives asymptotic theory for the proposed test under null and alternative hypotheses (of both convergence and divergence). Several new technical results on power function trend regression asymptotics are obtained in these derivations, which are of wider relevance than the concerns of the present paper. Section 4 reports some numerical calculations to demonstrate the contrasting test behavior under these two alternatives. Section 5 reports the results of Monte Carlo simulations to assess the finite sample performance of the test procedure. Section 6 illustrates the use of the new test in three empirical applications. Section 7 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. Stata and Gauss codes for the methods introduced in this paper are available at the author websites.

2 Empirical Motivation and Modeling Preliminaries

2.1 Testing and Definition of Weak $\sigma-$Convergence

As indicated in the quotation by Hotelling (1933) that heads this article, the notion of σ-convergence has been conceptually well understood since the early twentieth century. The concept 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 (2000) tested $\sigma-$convergence by considering differences between dyadic pairs of y_{it} rather than cross section variance or standard deviation.

To craft a suitably general concept of convergence, we may consider that the panel data of interest, y_{it}, can be decomposed into common and idiosyncratic components.

$$y_{it} = \theta_i' F_t + a_i + y_{it}^o = \theta_i' F_t + x_{it}. \tag{1}$$

We seek to examine $\sigma-$convergence in the panel idiosyncratic components x_{it} following the extraction of any common factors F_t using standard methods.[1] In an investigation of trends in volatility of individual stocks, for example, Campbell et al. (2001) used a panel model with a common trend factor of the form (1), tested for the presence of a strict linear trend using Vogelsang's (1998) robust t test, and examined convergence characteristics among the residual components x_{it}. Our approach formalizes the concept of decay in cross sectional variation over time without requiring a specific linear trend decay mechanism.

Let $\bar{x}_{\cdot t} := n^{-1} \sum_{i=1}^n x_{it}$, and define $\tilde{x}_{it} := x_{it} - \bar{x}_{\cdot t}$. We start with the following high level definition of weak $\sigma-$convergence that captures the key notion of a 'consistent diminution of cross section variation' over time. Primitive conditions that justify this formulation and provide a foundation for asymptotic theory are provided in Section 3.

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*

$$\begin{aligned}
&\text{(i) } \mathrm{plim}_{n\to\infty} K_{nt}^x = \bar{K}_t^x < \infty, \ a.s. \text{ for all } t \\
&\text{(ii) } \mathrm{plim}_{t\to\infty} \bar{K}_t^x = a \in [0,\infty), \\
&\text{(iii) } \limsup_{T\to\infty} \gamma_T\left(\bar{K}_t^x, t; c_T\right) < 0 \ a.s.,
\end{aligned} \tag{2}$$

[1] Importantly, the probability limit of the cross sectional variance of y_{it} as $n \to \infty$ is itself random. Indeed, we have $\mathrm{plim}_{n\to\infty} K_{nt}^y = \sigma_a^2 + \sigma_\theta^2 F_t^2 + \mathrm{plim}_{n\to\infty} K_{nt}^x$, which embodies the time series random common factor component F_t and the cross section common factor process $\bar{K}_t^x = \mathrm{plim}_{n\to\infty} K_{nt}^x$. Thus, the limiting cross section average dispersion $\mathrm{plim}_{n\to\infty} K_{nt}^y$of y_{it} may fluctuate over time according to the trajectories of $\left(F_t, \bar{K}_t^x\right)$.

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 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 following fitted form

$$K_{nt}^x = \hat{a}_{nT} + \hat{\phi}_{nT}t + \hat{u}_t. \tag{3}$$

In this regression a simple robust $t-$ratio test is conducted to assess whether the fitted slope coefficient $\hat{\phi}_{nT}$ is significantly less than zero, using a Newey-West type HAC estimator for the variance of $\hat{\phi}_{nT}$. Detailed discussion and asymptotic justification for this procedure are provided in Sections 3 and 5.

2.2 Existing Tests and Weak $\sigma-$Convergence

A typical formulation of $\beta-$convergence in terms of cross sectional growth regression can be written as

$$T^{-1}\left(x_{iT} - x_{i0}\right) = \hat{a} + \hat{\beta}x_{i0} + z_i'\hat{\varphi} + \text{residual}_i, \quad i = 1, ..., n, \tag{4}$$

where $T^{-1}\left(x_{iT} - x_{i0}\right)$ is the long run average, x_{i0} is initial log level real incomes, and z_i is a vector of auxiliary covariates. The regression permits tests for a significantly negative slope coefficient $\hat{\beta}$ in the fitted equation. 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 $\hat{\varphi}$ differs significantly from zero, the limiting outcomes for countries i and j may differ. Evans (1996) explained why growth regressions like (4) provide valid guidance regarding convergence only under strict conditions. Furthermore, as implied by Hotelling (1933) and Friedman (1992), the study of $\beta-$convergence does not provide a definitive test of a tendency to convergence in terms of a sustained 'diminution of variance'**.**

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

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

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), the notion of weak $\sigma-$convergence given above introduces a weaker condition that focuses on the asymptotic behavior of the sample covariance

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

where $\tilde{K}_{nt}^{x}=K_{nt}^{x}-T^{-1}\sum_{t=1}^{T}K_{nt}^{x}$, and $\tilde{t}=t-T^{-1}\sum_{t=1}^{T}t$.

A further existing test involves the idea of relative convergence. Phillips and Sul (2007) formulated a nonlinear panel model of the form

$$x_{it}=b_{it}\theta_{t},\quad \text{for } t=1,...,T;i=1,...,n, \tag{7}$$

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

$$\text{plim}_{t\to\infty}\frac{x_{it}}{x_{jt}}=1 \text{ for any } i\neq j. \tag{8}$$

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

$$\ln\left(H_1/H_t\right)-2\ln\ln t=\hat{a}+\hat{\gamma}\ln t+\hat{u}_t, \tag{9}$$

where $H_t=n^{-1}\sum_{i=1}^{n}\left(h_{it}-1\right)^2$ and $h_{it}=x_{it}/(n^{-1}\sum_{i=1}^{n}x_{it})$ 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 panel observations involve stochastic or deterministic trends, the relative convergence does not imply the weak $\sigma-$convergence. Consider, for instance, the simple panel model

$$x_{it}=a_i+b_{it}t+\epsilon_{it}t^{-\beta} \text{ with } b_{it}=b+\varepsilon_i t^{-1/2} \tag{10}$$

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 weak $\sigma-$convergence. Only when b_{it} converges b faster than t, (or $b_{it}=b+\varepsilon_i t^{-a}$ with $a>1$), the weak $\sigma-$convergence holds. Hence under the presence of distinct trending behavior, the weak $\sigma-$convergence is more restrictive than the relative convergence. Meanwhile when the data do not involve such trends as θ_t, then the concept of relative convergence in (8) is far less useful. For instance, relative convergence as indicated by (8) may not even exist in the case of panel data whose elements converge to zero.

2.3 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 x_{it}. Following PS, we use a power law time decay function, which is a convenient formulation to study weak $\sigma-$convergence.[2] 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. There are two convenient ways to accommodate such weak $\sigma-$convergence behavior: temporal shocks may influence only the mean level; and shocks may directly affect the cross sectional variance of the panel x_{it}. Combining these two mechanisms leads to the following model.[3]

$$x_{it} = a_i + \mu_i t^{-\alpha} + \epsilon_{it} t^{-\beta}, \tag{11}$$

were 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 $\alpha > 0$ and, as earlier, the idiosyncratic components (a_i, μ_i) are *iid* with finite support and are independent of the ϵ_{it}. Define $\tilde{a}_i = a_i - n^{-1}\sum_{i=1}^n a_i$ and similarly let $\tilde{\mu}_i$ and $\tilde{\epsilon}_{it}$ be deviations from their cross sectional means. Then the cross sectional variation of x_{it} in this case can be broken down into the following components.

$$K_{nt}^x = \sigma_{a,n}^2 + 2\sigma_{a\mu,n} t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} + \sigma_{\epsilon,nt}^2 t^{-2\beta} + e_{n,t}, \tag{12}$$

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

The statistical properties of the cross sectional dispersion of x_{it} hinge on the specific values of μ_i and β. In the following analysis, we consider the following three cases based on potential restrictions placed on μ_i and β.

Model	M1	M2	M3
Restriction	$\beta = 0$	$\mu_i = 0$	n/a

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

$$K_{nt}^x = a_n + \eta_{n,t} + \varepsilon_{n,t}, \tag{13}$$

[2] Other decay functions are possible. For example, for $c \in \mathbb{R}$ 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.

[3] This formulation does not include a remainder term of smaller order. For instance, if $\beta = 0$ and $\alpha = 1/3$, then x_{it} may take the more general form $x_{it} = a_i + \mu_i t^{-1/3} + \sum_{j=2}^p \mu_{ji} t^{-j/3} + \epsilon_{it}$ that involves higher order (smaller decay terms). In this event, the dominating decay term of x_{it} is $\mu_i t^{-1/3}$, and other terms can be written in residual form so that $x_{it} = a_i + \mu_i t^{-1/3} + \epsilon_{it} + o_p\left(t^{-1/3}\right)$ and smaller terms may be ignored in the development and asymptotics. Similarly, if $\mu_i = 0$ and $\beta = 1/3$, then x_{it} may take the more general form $x_{it} = a_i + \epsilon_{it} t^{-1/3} + \sum_{j=2}^q \epsilon_{it} t^{-j/3}$, or simply $x_{it} = a_i + \epsilon_{it} t^{-1/3} + o_p\left(t^{-1/3}\right)$, where the smaller order terms may again be ignored in the subsequent development and asymptotic theory.

where

$$a_n = \begin{cases} \sigma_{a,n}^2 + \sigma_{\epsilon,nT}^2 & \text{for M1,} \\ \sigma_{a,n}^2 & \text{for M2,} \\ \sigma_{a,n}^2 & \text{for M3,} \end{cases} \quad \eta_{n,t} = \begin{cases} 2\sigma_{a\mu,n}t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} & \text{for M1,} \\ \sigma_{\epsilon,nT}^2 t^{-2\beta} & \text{for M2,} \\ 2\sigma_{a\mu,n}t^{-\alpha} + \sigma_{\mu,n}^2 t^{-2\alpha} + \sigma_{\epsilon,nT}^2 t^{-2\beta} & \text{for M3,} \end{cases} \tag{14}$$

and

$$\varepsilon_{n,t} = \begin{cases} 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) & \text{for M1,} \\ 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) t^{-2\beta} & \text{for M2,} \\ 2n^{-1}\sum_{i=1}^n \tilde{a}_i\tilde{\epsilon}_{it}t^{-\beta} + 2n^{-1}\sum_{i=1}^n \tilde{\mu}_i\tilde{\epsilon}_{it}t^{-\alpha-\beta} + \left(\sigma_{\epsilon,nt}^2 - \sigma_{\epsilon,nT}^2\right) t^{-2\beta} & \text{for M3.} \end{cases} \tag{15}$$

We now discuss the differences in the temporal evolution of these models. From (13), the temporal decay character of the sample cross section variation K_{nt}^x is embodied in the component $\eta_{n,t}$. Evidently from (14), the temporal evolution of $\eta_{n,t}$ depends eventually on the dominant element as $t \to \infty$ among the terms that are present in $\eta_{n,t}$ for each model. This behavior is determined by the signs of the power parameters (α, β), their relative strengths, and the various coefficient values in (14) and their asymptotic behavior.

Model M2 is the simplest as $\mu_i = 0$ for all i and there is only a single term in $\eta_{n,t}$. The slope coefficient on $t^{-2\beta}$ in η_{nt} is $\sigma_{\epsilon,nT}^2$ as shown in (14), which depends on both n and T. Variation therefore reduces as $t \to \infty$ whenever $\beta > 0$ and $\sigma_{\epsilon,nT}^2 > 0$. This behavior does not depend on the n/T ratio because $\sigma_{\epsilon,nT}^2 \to_p \sigma_\epsilon^2$ when $(n, T) \to \infty$ irrespective of the relative divergence rates of (n, T).

The other models have multiple terms whose behavior can be more complex. In M1 $\beta = 0$, which implies that temporal effects on the system manifest through the component $a_i + \mu_i t^{-\alpha}$, which evolves according to $\mu_i t^{-\alpha}$ as $t \to \infty$.[4] The two terms $\left(2\sigma_{a\mu,n}t^{-\alpha}, \sigma_{\mu,n}^2 t^{-2\alpha}\right)$ that appear in $\eta_{n,t}$ for M1 have coefficients that depend on n and the asymptotic behavior of the dominant term is impacted by whether $\sigma_{a\mu,n} \to 0$. By further analysis of these terms, it is shown later (in Theorem 1 and in the ensuing discussion) that the dominating behavior is also influenced by the magnitude of the decay rate $\alpha > 0$ and the asymptotic behavior of the n/T ratio. The explanation is that the error term $\varepsilon_{n,t}$ in (15) involves weighted cross section sample averages of the errors ϵ_{it} and the scaled errors $\epsilon_{it}t^{-\alpha}$. The magnitude of these terms depends on n, T, and α. Thus, the convergence behavior of K_{nt}^x in this case evidently hinges on the sign of α and the relative importance of each of these terms, which in turn depends on the n/T ratio. Similar considerations influence the asymptotic behavior in model M3.

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

[4]The decay function $\mu_i t^{-\alpha}$ may be regarded as an evaporating trend factor component with idiosyncratic loadings μ_i.

$\sigma-$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 for large t.[5] 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 $\sigma_{\alpha\mu}$ 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 as $t \to \infty$ and K_{nt}^x increases over time and eventually stabilizes to fluctuate around $\sigma_a^2 + \sigma_\epsilon^2$ as $n, T \to \infty$.

Model M3 nests M1 and M2, and is particularly convenient for our theoretical development. In practice, simpler models like M1 or M2 may often provide useful characterizations. For instance, when common components are eliminated as in the US personal consumption expenditure item inflation rate and US State unemployment examples given in Section 6, M2 may characterize dynamic behavior that leads to weak σ-convergence or divergence. When no common element is eliminated, as in the Ultimatum game example of Section 6.2, M1 may be helpful in describing mean level convergence in the panel.

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

2.4.1 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 (12) 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 (13).

[5] 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.

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, 2005a).

2.4.2 Linear Trend Regression

The idea is to run a least squares regression of cross section sample variation[6] K_{nt} on a linear trend giving, as indicated earlier in (3), the fitted regression

$$K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT} t + \hat{u}_t, \ t = 1, ..., T \tag{16}$$

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 of t, then weak $\sigma-$convergence holds. In this event, in terms of the regression (16), 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.

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

In order to construct a valid significance test, allowance must be made for the fact that the model (16) is generally misspecified. Indeed, when K_{nt} satisfies a trend decay model such as (12), the regression may be considered spurious although, as is shown below, the asymptotic behavior of the fitted regression differs from that of a conventional spurious regression (Phillips, 1986). Nonetheless, 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 (16) 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.

3 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 (11), 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 uniform fourth moments, $\sup_i \mathbb{E}\left(\epsilon_{it}^4\right) < \infty$, and is strictly stationary over t with autocovariance sequence $\gamma_i(h) = \mathbb{E}(\epsilon_{it}\epsilon_{it+h})$ satisfying the summability condition $\sum_{h=1}^{\infty} h\left|\gamma_i(h)\right| < \infty$ 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 uniformly bounded 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 integrability moment and mixing conditions that assure the validity of our methods. For simplicity we do not pursue these extension details 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 (16) 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 (13) - (15). We may write $\eta_{n,t}$ as

$$\eta_{n,t} = \eta_t + \xi_{n,t} = \eta_t + O_p\left(n^{-1/2}\right), \tag{17}$$

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

$$\eta_t = \begin{cases} 2\sigma_{a\mu} t^{-\alpha} + \sigma_\mu^2 t^{-2\alpha} & \text{for M1}, \\ \sigma_\epsilon^2 t^{-2\beta} & \text{for M2}, \\ 2\sigma_{a\mu} t^{-\alpha} + \sigma_\mu^2 t^{-2\alpha} + \sigma_\epsilon^2 t^{-2\beta} & \text{for M3}, \end{cases} \tag{18}$$

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

$$\xi_{a\mu,n} : = \sigma_{a\mu,n} - \sigma_{a\mu} = n^{-1}\sum_{i=1}^{n}\left(\tilde{a}_i\tilde{\mu}_i - \sigma_{a\mu}\right) = O_p\left(n^{-1/2}\right), \tag{19}$$

$$\xi_{\sigma,n} : = \sigma_{\mu,n}^2 - \sigma_\mu^2 = n^{-1}\sum_{i=1}^{n}\left(\tilde{\mu}_i^2 - \sigma_\mu^2\right) = O_p\left(n^{-1/2}\right), \tag{20}$$

so that the residual in (17) 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 (17), the trend regression coefficient $\hat{\phi}_{nT}$ in (16) can be decomposed into three components as follows

$$\hat{\phi}_{nT} = \sum_{t=1}^{T} a_{tT}\tilde{\eta}_t + \sum_{t=1}^{T} a_{tT}\tilde{\xi}_{n,t} + \sum_{t=1}^{T} a_{tT}\tilde{\varepsilon}_{n,t} =: I_A + I_B + I_C, \tag{21}$$

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. The first term I_A is a purely deterministic term and depends only on the parameter λ. The second and third terms are random terms with zero means. If either of the second or third terms becomes dominant, then the sign of

$\hat{\phi}_{nT}$ is ambiguous, prevents a clear test conclusion. The glossary given in the Table C array (22) below summarizes the required conditions for first term dominance in (21).

Models	Cases $\sigma_{a\mu} \neq 0$	Cases $\sigma_{a\mu} = 0$
M1	$\frac{T}{n} \to 0$ with $\alpha \geq \frac{1}{2}$;or $\alpha < \frac{1}{2}$	$\frac{T}{n} \to 0$ with $\alpha \geq \frac{1}{2}$; or $\frac{T^{2\alpha}}{n} \to 0$ with $\alpha < \frac{1}{2}$
M2	n.a.	no restriction
M3	no restriction	$T/n \to 0$

(22)

Table C: Restrictions on the T/n Ratio in Various Cases

In M2, the first term in (21) dominates other terms, so that no restriction on the T/n ratio is required. In M1 and M3, when $\sigma_{a\mu} = 0$, the term $2\sigma_{a\mu}t^{-\alpha}$ is absent from I_A, but the term $2\sigma_{a\mu,n}t^{-\alpha}$ is present in I_B, which may dominate I_A if $T/n \nrightarrow 0$. When $\sigma_{a\mu} \neq 0$, the T/n ratio condition depends on the value of α in M1. When $\sigma_{a\mu} \neq 0$, no rate condition on the T/n ratio is required in model M3.

The values that λ and b take 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

$$\mathcal{O}_{T\lambda} = -\begin{cases} L_\lambda T^{-1-\lambda} & \text{if } \lambda < 1, \\ 6T^{-2}\ln T & \text{if } \lambda = 1, \\ 6\zeta(\lambda)T^{-2} & \text{if } \lambda > 1. \end{cases} \tag{23}$$

where $L_\lambda = 6\lambda[(2-\lambda)(1-\lambda)]^{-1}$. The limit behavior of $\hat{\phi}_{nT}$ in the regression equation (16) is characterized more easily in terms of $\mathcal{O}_{T\lambda}$ in the following result. Since the linear trend regression (16) 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 $n, T \to \infty$.

Since the empirical trend regression equation (16) 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 (16) 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 (21) as it typically does, 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. More formally, we can state the regression limit theory as follows.

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 (16) 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 restriction 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.

In establishing the results of the theorem, the proof examines the components of (21) to assess the main contribution to the asymptotic behavior of $\hat{\phi}_{nT}$. The proof of the theorem in Appendix provides detailed calculations and examines the various cases implied by the different parameter configurations.

With the asymptotic behavior of $\hat{\phi}_{nT}$ in hand, limit theory can be developed for the corresponding t-ratio in the regression (16), which takes the following standard form for the time trend regressor case, viz.,

$$t_{\hat{\phi}_{nT}} = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_u^2 / \sum_{t=1}^{T} \tilde{t}^2}}, \tag{24}$$

where $\hat{\Omega}_u^2$ is a typical long run variance estimate based on the residuals $\hat{u}_t = K_{nt}^x - \hat{a}_{nT} - \hat{\phi}_{nT} t$ from (16), such as the Bartlett-Newey-West (BNW) estimate

$$\hat{\Omega}_u^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T} \hat{u}_t^2 + 2\frac{1}{T}\sum\nolimits_{\ell=1}^{L} \vartheta_{\ell L} \sum\nolimits_{t=1}^{T-\ell} \hat{u}_t \hat{u}_{t+\ell}, \tag{25}$$

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

We use the robust form of the test statistic given in (24) which employs a standard long run variance estimate $\hat{\Omega}_u^2$ constructed by lag kernel methods as in (25) 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

$$\hat{u}_t = \left(\tilde{\eta}_{n,t} - \hat{\phi}_{nT} \tilde{t} \right) + \tilde{\varepsilon}_{nT} =: \tilde{\mathcal{M}}_{nt} + \tilde{\varepsilon}_{nT}, \tag{26}$$

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 (17)-(20) 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

$$\tilde{\xi}_{n,t} = \begin{cases} 2\xi_{a\mu,n}\widetilde{t^{-\alpha}} + \xi_{\mu,n}\widetilde{t^{-2\alpha}} = o_p\left(\widetilde{t^{-2\alpha}}\right) & \text{for M1,} \\ \xi_{\sigma,n}\widetilde{t^{-2\beta}} = o_p\left(\widetilde{t^{-2\beta}}\right) & \text{for M2,} \\ 2\xi_{a\mu,n}\widetilde{t^{-\alpha}} + \xi_{\mu,n}\widetilde{t^{-2\alpha}} + \xi_{\sigma,n}\widetilde{t^{-2\beta}} = o_p\left(\min\left(\widetilde{t^{-2\alpha}}, \widetilde{t^{-2\beta}}\right)\right) & \text{for M3,} \end{cases} \tag{27}$$

which may be expressed in the simple form that $\tilde{\xi}_{n,t} = o_p(\tilde{\eta}_t)$ uniformly in t as $n/T \to \infty$. Since the trend regression coefficient $\hat{\phi}_{nT}$ satisfies the decomposition (21), we find that

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

with deterministic part $\tilde{m}_t = \tilde{\eta}_t - I_A\tilde{t}$ and random part $R_{nt} = \tilde{\xi}_{n,t} - \tilde{t}\,(I_B + I_C)$. 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, $\tilde{\xi}_{n,t} = o_p(\tilde{\eta}_t)$ uniformly in t as $n/T \to \infty$. It follows that $R_{nt} = o_p(\tilde{m}_t)$ uniformly in $t \leq T$ as $n/T \to \infty$.

Under model M2, the term $\tilde{\mathcal{M}}_{nt}$ in (26) 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

$$\tilde{\mathcal{M}}_{nt} = b\left[\widetilde{t^{-\lambda}} - \tilde{t}\left(\sum\nolimits_{t=1}^{T} \tilde{t}\widetilde{t^{-\lambda}}\right)\left(\sum\nolimits_{t=1}^{T} \tilde{t}^2\right)^{-1}\right] + R_{nt}, \tag{28}$$

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

$$t_{\hat{\phi}_{nT}} = \frac{\hat{\phi}_{nT}}{\sqrt{\hat{\Omega}_u^2 / \sum_{t=1}^{T} \tilde{t}^2}} \sim \frac{\hat{\phi}_{nT}}{\sqrt{\Omega_{\mathcal{M}}^2 / \sum_{t=1}^{T} \tilde{t}^2}} = \frac{\left(b\sum_{t=1}^{T} \tilde{t}\widetilde{t^{-\lambda}}\right)\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1/2}}{\sqrt{\Omega_{\mathcal{M}}^2}}, \tag{29}$$

making the t-ratio a function of only λ, κ, and T asymptotically when $n/T \to \infty$. In (29) the quantity $\Omega_{\mathcal{M}}^2$ is constructed in the usual manner as a long run variance estimate, viz.,

$$\Omega_{\mathcal{M}}^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T} \tilde{m}_t^2 + \frac{2}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\tilde{m}_t\tilde{m}_{t+\ell}, \tag{30}$$

as in (30) with lag truncation parameter L; and, being a function of $\tilde{m}_t$, $\Omega_{\mathcal{M}}^2$ is a deterministic function of t.

Different lag truncation rules may be employed in (30) and other forms of t-ratio may be used in which different robust standard errors are used in (29), including heteroskedastic and autocorrelation robust (HAR) forms such as fixed-b and trend IV approaches (e.g., Kiefer and Vogelsang, 2002; Sun, 2004, 2018; Bunzel and Vogelsang, 2005; Phillips 2005b.)[7] The asymptotic equivalence in (29) 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.

[7] For example, instead of the simple t-ratio (29), one may consider alternative formulae such as

$$t_{\mathrm{HAC}} = \hat{\phi}_{nT}\left[\left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1} T\hat{\Omega}_M^2 \left(\sum_{t=1}^{T} \tilde{t}^2\right)^{-1}\right]^{-1/2}$$

which employ the HAC estimate

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

formed from the components $p_t = \hat{u}_t t$ and $\tilde{p}_t = p_t - T^{-1}\sum_{t=1}^{T} p_t$, or HAR estimates such as fixed-b methods with $M = bT$ for some fixed $b \in (0,1)$ in $\hat{\Omega}_M^2$. We do not report results here with alternate versions such as t_{HAC} and t_{HAR} since our findings indicate that overall the standard formula given in (29) provides better finite sample performance. Detailed analytic and simulation results for these cases are provided in Kong, Phillips and Sul (2017).

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 (16) has the following asymptotic behavior as $n, T \to \infty$:

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

$$\lim_{n,T\to\infty} t_{\hat{\phi}_{nT}} = -\tau^*_\lambda = \begin{cases} -\infty & \text{if } 0 < \lambda < 1, \\ -\sqrt{6/\kappa^2} & \text{if } \lambda = 1, \\ -\mathbb{Z}(\lambda)\sqrt{3} & \text{if } 1 < \lambda < \infty, \\ -\sqrt{3} & \text{if } \lambda \to \infty. \end{cases} \tag{31}$$

where $\kappa > 0$ is defined by the lag truncation parameter $L = \lfloor T^\kappa \rfloor$ in the long run variance estimator (25). 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 $n, T \to \infty$,

$$\lim_{n,T\to\infty} t_{\hat{\phi}_{nT}} = \begin{cases} +\infty & \text{if } \lambda < 0 \textit{ regardless of the } n/T \textit{ ratio}, \\ \tau^*_\lambda & \text{if } \sigma_{a\mu} < 0 \textit{ with } \lambda > 0 \textit{ and } n/T \to \infty. \end{cases} \tag{32}$$

(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,

$$t_{\hat{\phi}_{nT}} \to^d \mathcal{N}(0,1). \tag{33}$$

As indicated in (31) and (32), 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 2 (i) is the key result of most relevance in empirical studies of convergence. The explicit limit behavior shown in (31) derives from the fact that the t-ratio takes asymptotically the deterministic form (29), 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 (31) 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.

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. 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 $(n,T) \to \infty$ irrespective of the behavior of the ratio n/T.

Before we discuss the size and power properties of the t-ratio test, we make the following remarks about the implications of the above results for practical work on convergence testing.

Remark 1: (The Effects of Violation of the T/n Rate Requirement) Since Theorem 2 requires the rate condition $T/n \to 0$, it is naturally of interest to explore the consequences for the test when this condition is violated. First, from (22) and Theorem 1 it is evident that M2 does not require any T/n ratio requirement. So there are no adverse consequences for panel applications where this model is relevant. As shown in Section 6, with the exception of the experimental data application, the weak $\sigma-$convergence test is often performed after eliminating common components, which makes M2 the most relevant model in such cases.

Models M1 or M3 are typically more relevant for raw panel analyses in which common components are not present or cannot be eliminated and there is no strong trending behavior in the data, as distinct from possible trend decay in cross section variation. If $\lambda < 0$ in M1 or M3, as Theorem 2 indicates, the condition $T/n \to 0$ is not binding in this case either. Only when the conditions in (22) hold – for example when the idiosyncratic elements (α_i, μ_i) are uncorrelated ($\sigma_{a\mu} = 0$) – does the T/n ratio become relevant in influencing the asymptotic theory. As discussed later in the numerical simulation findings, when $\sigma_{a\mu} = 0$ the decay rate follows $t^{-2\alpha}$ (rather than $t^{-\alpha}$) in (14) and discriminatory power in testing for convergence may be attenuated by the faster convergence rate because this deterministic component may be dominated by the random component variation when n is not large relative to T.

Remark 2: (A Different Decay Function) Instead of the power decay function $t^{-\lambda}$ other formulations are possible, such as the geometric decay function ρ^t mentioned in footnote 2. To be specific, suppose

$$x_{it} = a_i + \mu_i \rho^t + \epsilon_{it}.$$

Since $\lim_{t\to\infty} \rho^t/t^{-\lambda} = 0$ for any $\lambda > 0$, geometric decay is faster. In earlier work, Kong and Sul (2013) showed that the main finding in Theorem 2 changes little in this event. In particular, the boundary limit of the t-ratio, $\pm\sqrt{3}$, is the same with the geometric decay function ρ^t. Importantly, the rate condition $T/n \to 0$ is likely to be more important for good test behavior in this case, for the reasons given above.

Remark 3: (Sub-convergent Clubs) As Phillips and Sul (2007) showed, convergence fails when there are multiple sub-convergent clubs and special methods are needed to identify club convergence. As in that work, classifying club membership becomes a feature of considerable empirical interest in practical work. When panel data include distinct stochastic trends the nonstationarity in the data assists in identifying club membership. However, if there are no distinct trends, it is much more challenging to sort individuals into multiple clubs, even when the decay function is known. Using model M3, for instance, a simple case of two sub-convergent clubs (G and G^c, say) involves a modified formulation of the type

$$x_{it} = \begin{cases} 0 + \mu_{Gi} t^{-\alpha} + \epsilon_{it} t^{-\beta} & \text{for } i \in G \\ 1 + \mu_{G^c i} t^{-\alpha} + \epsilon_{it} t^{-\beta} & \text{for } i \in G^c \end{cases}.$$

If $i \in G$, then $\mathbb{E}x_{it}$ converges zero, while if $i \notin G$, $\mathbb{E}x_{it}$ converges to unity. In this case x_{it} is $\sigma-$divergent, even though individual groups of the data do converge. To identify club membership, one may consider running a linear regression of x_{it} on $t^{-\alpha}$ for each i and use some classification method to subsequently group the estimates. However if $\alpha > 1/2$, least squares regression is inconsistent since $\sum_{t=1}^{T} t^{-2\alpha}$ is convergent as $T \to \infty$ and time series signal strength is too weak for consistent estimation. Hence, a different approach is needed to identify club membership. One possibility is to combine the methods of this paper with newly developed panel classification procedures, such as those in Su, Shi and Phillips (2016) that involve penalized regression to shrink the coefficient estimates towards empirically supported groupings. This is an interesting topic with substantial empirical relevance that we leave for future study.

Remark 4: (Power Trend Regression) 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 requires the rate condition $n/T \to \infty$. 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

$$K_{nt} = \hat{a}_{nT} + \hat{\phi}_{nT} t^{\psi} + \hat{u}_t, \ t = 1, ..., T, \text{ and some given } \psi > 0, \tag{34}$$

were used in place of the linear trend regression equation (16). 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.

4 Numerical Calculations

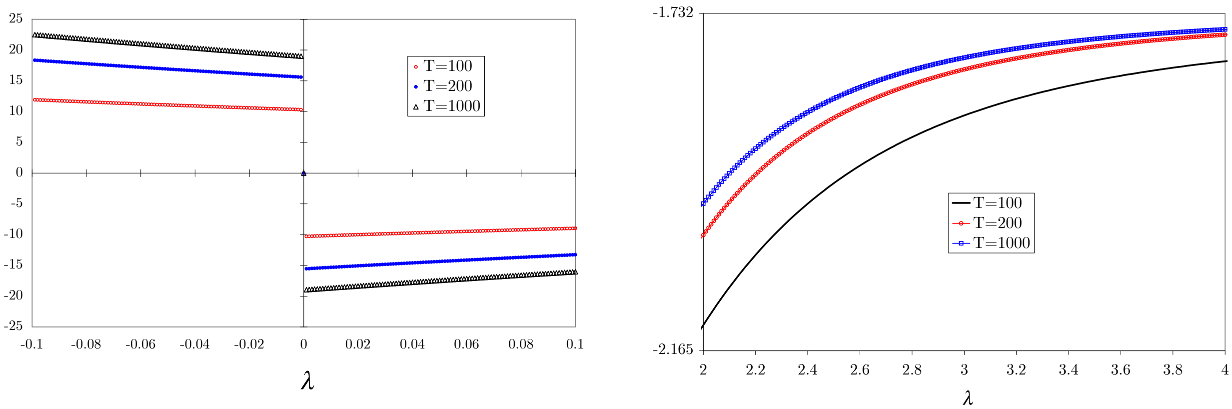
To demonstrate the contrasting test behavior under the alternatives of convergence and divergence, we report the following numerical calculations. These and the Monte Carlo simulations of the next section are designed to enable assessment of the size and power properties of the convergence test in relation to the magnitude of the decay parameter and sample size (n, T) configurations.

When $n \to \infty$ the probability limit of K_{nt} under M1 - M3 is the following deterministic function of t

$$\text{plim}_{n\to\infty} K_{nt} = K_t = a + \eta_t = a + bt^{-\lambda}, \tag{35}$$

for some non-zero constants a and b. We calculate the t-ratio under this asymptotic $(n \to \infty)$ deterministic DGP (35) for various sample sizes T and refer to it as the t_T^{∞}-ratio. Figure 1 shows how t_T^{∞} behaves for various values of λ. In the vicinity of $\lambda \sim 0$, Panel A of Figure 1 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 1, 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.

To explore behavior of the test in the vicinity $\lambda \sim 0$, Figure 2 plots the density of the t-ratio for various values of α in model M1 with $n = 1000$ and $T = 100$. We set $\sigma_{a\mu} = 0$, $\sigma_{\mu}^2 = 1$, $\kappa = 1/3$ 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.



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

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

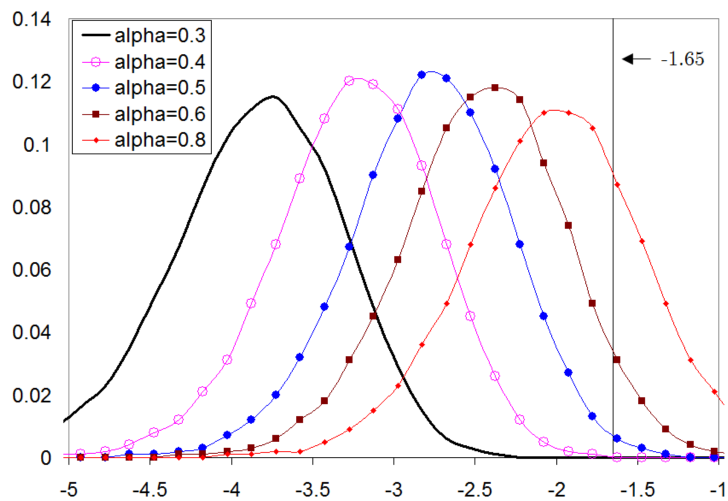


Figure 2: Empirical distribution of $t_{\hat{\phi}_{nT}}$ under M1
($n = 1000$, $T = 100$, $\sigma_{a\mu} = 0$, $\sigma_a^2 = \sigma_\mu^2 = 1$, $\epsilon_{it} \sim iid\mathcal{N}(0, 1)$, $\kappa = 1/3$)

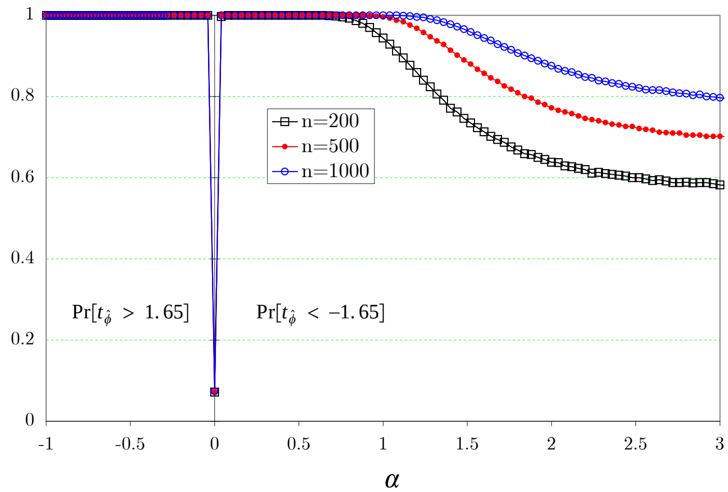


Figure 3: Test Rejection Frequencies over $\lambda = 2\alpha$ in model M1
($T = 50$, $\sigma_{a\mu} = 0$, $\sigma_\mu^2 = 4$, $\sigma_a^2 = \sigma_\epsilon^2 = 1$, $\kappa = 1/3$)

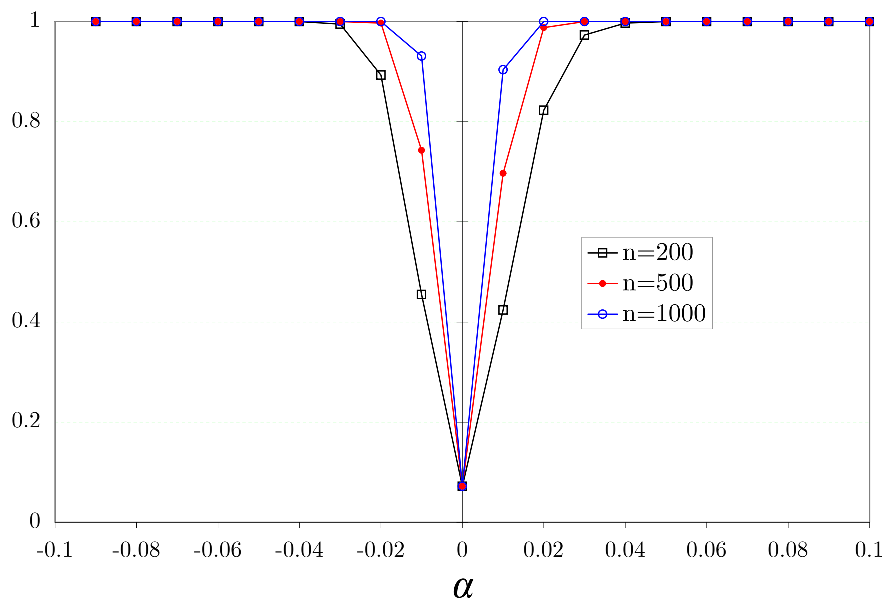


Figure 4: Test power curves near $\alpha = 0$ for various n in model M1

Figure 3 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[8]

As is apparent in Figure 3, 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 4 magnifies the region around $\alpha = 0$ from Figure 3 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.

5 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

[8] The mean level in model M1 has the form $\mathbb{E}(x_{it}) = a + \mu t^{-\alpha} \underset{t\to\infty}{\to} a$, when $\alpha > 0$. Then $\mathbb{E}(x_{i1} - x_{i2}) = \mu\left(1 - 2^{-\alpha}\right)$ $= 2^{-\alpha}(2^\alpha - 1)\mathbb{E}(x_{i1} - x_{i\infty}) = \mathbb{E}(x_{i1} - x_{i\infty})/2$ 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

$$\begin{aligned} K_{\infty,1} - K_{\infty,2} &= \psi\left(1 - 2^{-2\alpha}\right) = 2^{-2\alpha}\left(2^{2\alpha} - 1\right)\left(K_{\infty,1} - K_{\infty,\infty}\right) \\ &= (3/4)\left(K_{\infty,1} - K_{\infty,\infty}\right), \end{aligned}$$

and the half life in the variation $K_{\infty,t}$ from $t = 1$ is less than one period.

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.[9] 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[10]. As in Model M1, we consider two cases depending on the value of $\sigma_{a\mu}$.

[9] 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 (2). 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 (2) 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 (2) 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 (2) again hold.

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

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 because 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 attenuation 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.

6 Empirical Examples

We provide three empirical applications of the proposed test. The first data set is a balanced panel consisting of 46 disaggregated personal consumption expenditure (PCE) items. The second 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. The third example shows how the cross sectional dispersion of state level unemployment rates changed over a period that includes the subprime mortgage crisis.

6.1 Weak $\sigma-$Convergence with 46 PCE inflation Rates

Here we report an interesting empirical finding about weak $\sigma-$convergence with 46 disaggregate PCE inflation rates. The source of the data is the annual PCE (Table 2.4.4) obtained from the Bureau of Economic Analysis and our full data set covers 46 disaggregated series over the period 1978 to 2016.

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

$$\pi_{it} = a_i + \theta_i' F_t + \pi_{it}^o, \tag{36}$$

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 46 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). One factor is found over the entire sample period from 1979 to 2016 (loosing one sample observation in the conversion to inflation rates) after prewhitening and standardization. Next, we obtain estimates of the idiosyncratic components by using principal components.[11]

Figure 5 plots the PCE average inflation rates (shown by the heavy dark blue line) for the 46 disaggregated series and the sample variance of the estimated idiosyncratic components (thin

[11] 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 $\hat{F}_t$ (36) 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, $\hat{\pi}_{it}^r = a_i + \pi_{it}^o + \theta_i' F_t - \hat{\theta}_i' \hat{F}_t$.

pink line with solid circles) over the period 1979 - 2016. Evidently, the cross sectional variance is generally decreasing over this time period but with some fluctuations.

Table 5 reports the weak $\sigma-$convergence test results with the whole sample (from 1979 to 2016) and two subsamples (before and after 1992). For the sample after 1992, the null of no $\sigma-$convergence is rejected even at the 2.5% 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 3, and again all cases support evidence for $\sigma-$convergence.

Test results for the sample prior to 1992 and for the entire sample are different. It is well known that inflation rates reached a peak in the early1980s and displayed time series wandering characteristics over the 1980s. Common factors to inflation rates estimated for the 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 for the subsample from 1979 to 1992. On the other hand, for the full sample regressions over 1979-2016, the t-ratios are less than the right side critical value -1.65 for $k = 1, 2$, supporting the conclusion that inflation rates in the PCE are converging overall over the entire period.

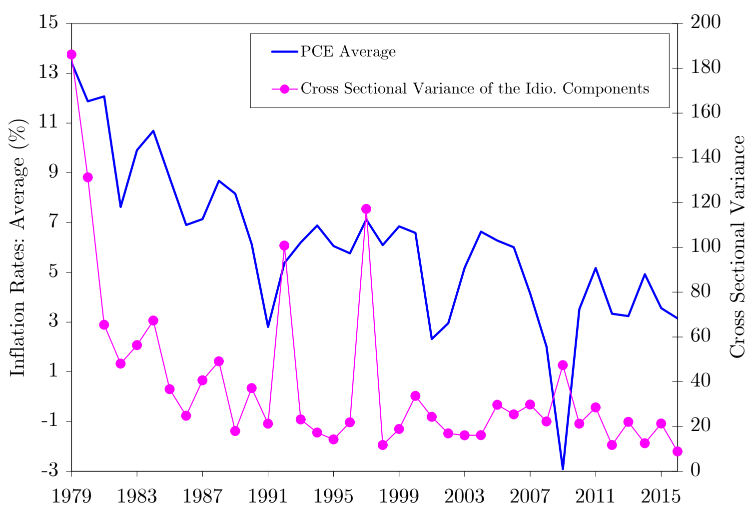


Figure 5: Cross Sectional Means and Variances of 46 PCE items

Table 6 shows the trend regression test results with samples dating from various starting years (each sample taken through to 2016), 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 the sample size reductions that this recursion involves, the null of no $\sigma-$convergence is rejected in all cases. These results support the overall conclusion of convergence in PCE inflation rates over this period.

6.2 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), Chaudhuri (2011), Cooper and Kagel (2013), 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 6 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.

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.

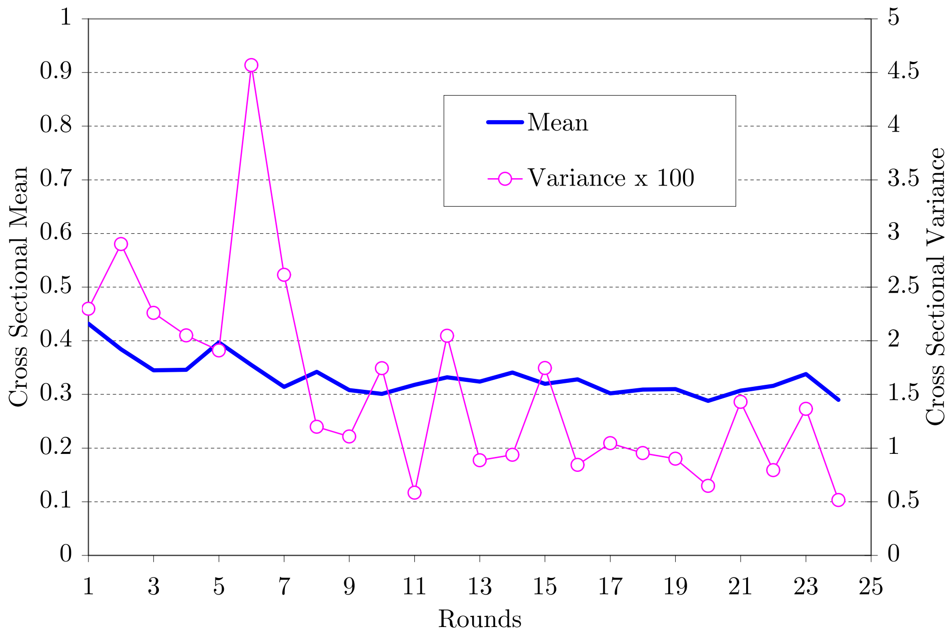


Figure 6: Cross Sectional Average and Variance
(Data from Ho & Su (2009))

6.3 Divergence and Convergence in US State Unemployment rates

Figure 7 (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.

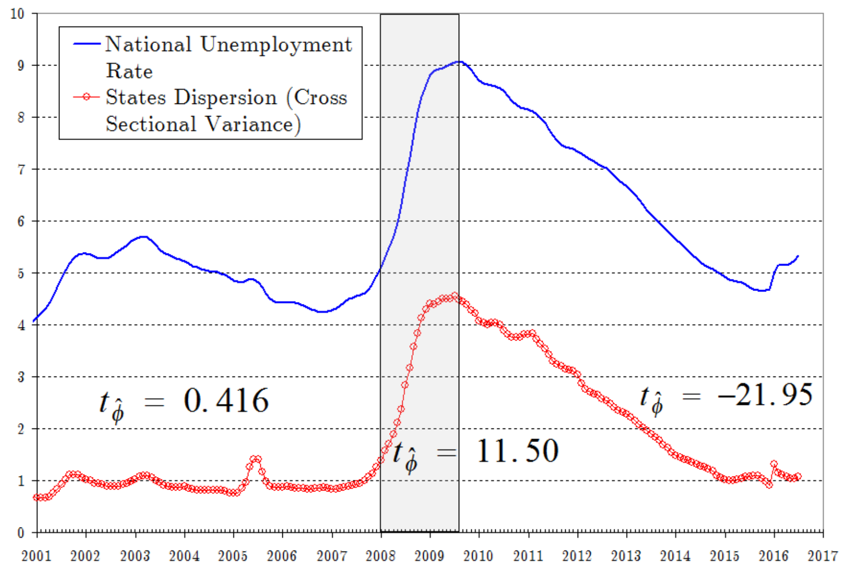
Evidently, the temporal patterns of the national unemployment rate and the cross section variation of State unemployment rates show some stability over 2001-2007. Both rise sharply during the crisis, and both fall steadily in the crisis aftermath. These patterns suggest a period of stationary fluctuations 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.

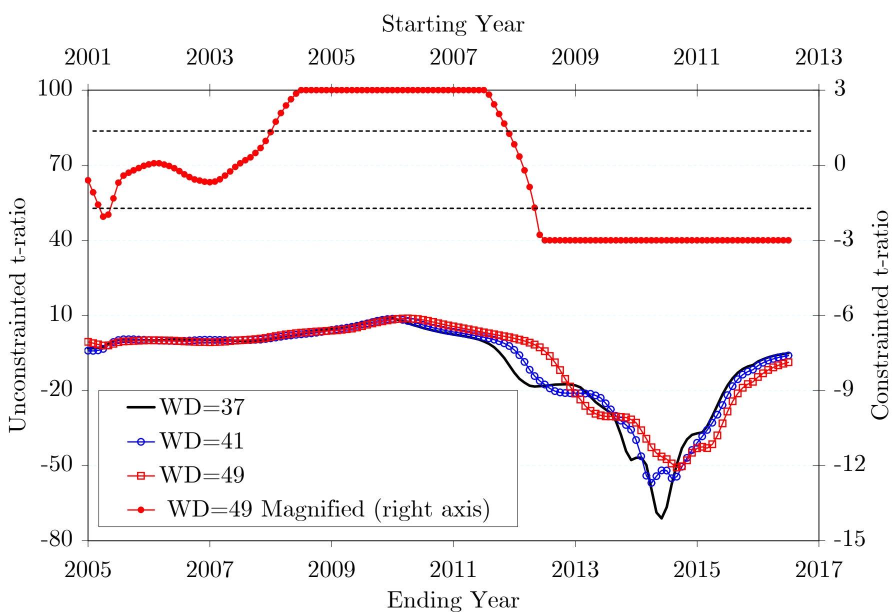
The top panel of the figure reports the t-ratio statistics $t_{\hat{\phi}} = 0.416, 11.50, -21.95$ for the pre-crisis, mid-crisis and post-crisis periods, which are exogenously determined according to the official period of the recession shown in the shaded region. As explained below, t statistics outside the standard normal critical values signal variation divergence in the right tail, variation convergence in the left tail, and stable variation within standard $(0, 1)$ 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 7 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, Wu, 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.



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

7 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] 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.

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

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

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

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

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

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

[8] Bunzel, H., and T. J. Vogelsang. (2005). Powerful trend function tests that are robust to strong serial correlation, with an application to the Prebisch–Singer hypothesis. *Journal of Business & Economic Statistics*, 23(4), 381–394.

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

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

[20] Güth, W., R. Schmittberger and B. Schwarze (1982). An Experiment Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization*, 3, 367-388.

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

[26] Kong, J., and D. Sul (2013). Estimation of treatment effects under multiple equilibria in repeated public good experiments. Mimeo, University of Texas at Dallas.

[27] Kong, J., P.C.B. Phillips and D. Sul (2018). Online Supplement to 'Weak $\sigma-$Convergence: Theory and Applications. Mimeo, University of Texas at Dallas.

[28] Kong, J., P.C.B. Phillips and D. Sul (2017). Testing weak $\sigma-$convergence based on HAR covariance matrix estimators. Mimeo, University of Texas at Dallas.

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

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

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

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

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

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

[35] 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.

[36] 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.

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

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

[39] Su, L., Shi, Z. and P. C. B. Phillips (2016). Identifying latent structures in panel data. Econometrica, 84(6), 2215–2264

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

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

[42] Sun, Y. (2018). Simple, Robust, and Accurate F and t Tests in Cointegrated Systems, *Econometric Theory* (to appear).

[43] Vogelsang, T. (1998). Trend function hypothesis testing in the presence of serial correla-tion, *Econometrica,* 66, 123–148.

[44] 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. All proofs including proofs of Theorem 1 and 2 are given in the online supplement by Kong, Phillips and Sul (2018), and rely on certain properties of infinite series and standard limit theory methods (e.g., Phillips and Solo, 1992).

Lemma 1

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

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

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

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* $\Delta_{\alpha,\ell}$ *are smaller order terms, which are defined 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*

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

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

and

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

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

$$\sum_{h=1}^{\infty} h\left|\gamma_{h,v,i}\right| < \infty, \tag{37}$$

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

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

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 $\lambda > 0$*

$$\begin{aligned} G(T,\lambda) &:= \frac{1}{T}\sum\nolimits_{\ell=1}^{L}\sum\nolimits_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\tilde{m}_t\tilde{m}_{t+\ell} \\ &= \begin{cases} O\left(T^{-2\lambda+\kappa}\right) & \text{if } \lambda < 1/2, \\ O\left(T^{\kappa-1}\ln T\right) & \text{if } \lambda = 1/2, \\ O\left(T^{\kappa-1}\right) & \text{if } 1/2 < \lambda < 1/(1+\kappa), \\ O\left(T^{-\lambda+\kappa-\lambda\kappa}\right) & \text{if } 1/(1+\kappa) \leq \lambda < 1, \\ \frac{\kappa^2}{2}T^{-1}\ln^2 T + O\left(T^{-2}\ln T\right) & \text{if } \lambda = 1, \\ T^{-1}\left\{\sum_{t=1}^{\infty} t^{-\lambda}\zeta(\lambda,\ell) - \zeta(2\lambda)\right\} & \text{if } \lambda > 1, \end{cases} \end{aligned}$$

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

$$\xi_{b,n}\widetilde{t^{-\alpha}} = o_p\left(\widetilde{t^{-2\alpha}}\right). \tag{38}$$

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

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 inflation items

Factor number	Whole Sample			From 1979 to 1992			From 1992 to 2016		
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	-1.243	-3.724	-3.646	-3.321	-1.644	-1.546	-1.049	-2.172	-2.498
2	-0.627	-2.950	-3.214	-0.055	-0.039	-0.037	-1.140	-2.556	-2.914
3	-0.352	-1.514	-1.481	1.585	1.351	1.267	-1.263	-2.868	-3.324

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	$\hat{\phi}_{nT}$	$t_{\hat{\phi}_{nT}}$ (3)	$t_{\hat{\phi}_{nT}}$ (4)	$t_{\hat{\phi}_{nT}}$ (5)	$t_{\hat{\phi}_{nT}}$ (6)
1979	-1.243	-3.724	-3.789	-3.631	-3.646
1981	-0.732	-3.107	-3.594	-3.481	-3.716
1983	-0.763	-2.861	-3.28	-3.179	-3.385
1985	-0.548	-1.919	-2.221	-2.110	-2.202
1987	-0.669	-2.128	-2.528	-2.425	-2.505
1989	-0.662	-1.820	-2.182	-2.090	-2.155
1991	-0.908	-2.173	-2.531	-2.423	-2.535
1992	-1.049	-2.172	-2.507	-2.450	-2.498

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