Identifying Common Trend Determinants in Panel Data*

Yoonseok Lee†, Peter C. B. Phillips‡, Suyong Song§ and Donggyu Sul¶

June 2025

Abstract

This paper develops a novel method for evaluating whether some observable economic variables are the determinants of the common latent trend in nonstationary panel data, which is typically removed or controlled for in two-way fixed effects regressions. By examining cross-sectional dispersion processes, we assess whether the majority of panel data series exhibit distributional convergence towards specific observed time series, identifying these as long-run determinants of the underlying latent common trend. This approach avoids the need for long panels and factor estimation. It also offers a new perspective on cointegration between time series and panel data, focusing on the relative variation of the panel data with respect to the cointegration error. Applying this method to U.S. state-level crime rates reveals that the percentage of young adults is a key determinant of violent crime trends, while the incarceration rate drives property crime trends. These findings, which differ from those obtained through standard two-way fixed effects analysis, provide a compelling explanation for the sharp decline in U.S. crime rates since the early 1990s.

Keywords: Latent Trend, Nonstationary Panel, Crime Rate, Two-Way Fixed Effects, Common Factors, Panel Cointegration

JEL Classification: C33, C38, K14.

*Phillips acknowledges research support from the NSF (Grant No. SES 18-50860) at Yale University and a Kelly Fellowship at the University of Auckland.

†Syracuse University; *E-mail*: ylee41@syr.edu

‡Yale University, University of Auckland, and Singapore Management University; *E-mail*: peter.phillips@yale.edu

§University of Iowa; *E-mail*: suyong-song@uiowa.edu

¶University of Texas at Dallas; *E-mail*: d.sul@utdallas.edu

1 Introduction

Panel data series typically share a common trend. However, research interest is often in estimating the marginal effects of control variables or predictors, and the traditional panel data regression analysis relies on two-way fixed effect models, where such a common trend is regarded as a nuisance parameter and controlled by the time fixed effect. For this reason, when the research question centers on identifying the drivers of the evolution (i.e., the leading trend itself) of the dependent variable, this practice cannot provide the desired answer. The time fixed effect will control for the leading trends of both the dependent and independent variables, and statistically significant association between them does not necessarily confirm that the independent variables are the trend determinants of the dependent variable.

Focusing on unveiling latent dynamics in panel data, this paper starts with the question: "What are the limitations of two-way fixed effects regression?" We specifically investigate the unobserved common trends among panel data series, aiming to unveil the underlying drivers of leading dynamics that are often neglected in standard two-way fixed effects analysis. In particular, we propose a novel method to verify if observed variables determine the underlying common trend of a large number of panel series, especially for nonstationary panel data series whose underlying (stochastic) trend is often hard to identify. Our approach avoids estimating the latent trend or latent nonstationary factors, making it computationally easy to implement. In these regards, this method is distinct from Bai and Ng (2006) and Parker and Sul (2016) that assume stationary panel data and require common factor estimation.

The key insight behind our method is the observation that many panel data series or their subgroup, even those with stochastic trends, tend to show converging patterns towards their common trend in the long run. This results in negative association between the cross-sectional dispersion and time, which is in a similar vein of the idea of weak $\sigma-$convergence by Phillips and Sul (2007) and Kong, Phillips, and Sul (2019). We leverage this concept of distributional convergence to identify observed time series that co-move with the latent trend when the majority of panel series converge towards those observed time series in the long run. These observed variables are then designated as the common trend determinants. Though we consider panel data whose time periods (T) and the cross sectional size (N) are large for asymptotic analysis, we only require N/T tends to infinity. Hence, the proposed approach can be implemented even for relatively short panel data, since it benefits from the information from the large cross section.

This study also contributes to empirical studies examining co-movement in panel data and the common factors driving such co-movement. For instance, studies on crime rates often investigate the decline in the U.S. national crime rate since early 1990s and the key economic variables associated with this trend change (e.g., Levitt (2004); Moody and Marvell (2010)). To demonstrate the effectiveness of our method, we analyze state-level crime rates in the U.S. Unlike the existing studies relaying on the two-way fixed effect analysis, we identify the demographic factor of the young adult population as a key determinant of the latent common trend of the violent crime rates, highlighting a previously overlooked factor compared to traditional focuses like police size or income disparity. For the property crime, on the other hand, we find the incarceration rate is a key determinant of the latent common trend.

The rest of the paper is organized as follows. Section 2 motivates our approach by addressing a limitation of the standard two-way fixed effects regression and illustrating the distributional convergence of panel data. Section 3 formally defines long-run trend determinants based on the concept of distributional convergence and outlines the procedure for identification. Section 4 derives limiting distributions of the main test statistic, providing theoretical justification of the proposed method. It also connects the key idea with panel cointegration, introducing a novel perspective on cointegration between time series and panel data. Section 5 addresses cases where only a subgroup of series exhibits distributional convergence and proposes a pre-screening method to identify the subgroup with the main common trend. Section 6 revisits the crime rate example and provides a detailed analysis, yielding results that differ from those obtained through standard two-way fixed effects estimation. Section 7 concludes with remarks. The Appendix provides a summary of the main procedure, proof of the main theorems, and the critical value tables. The online supplement provides proofs of technical lemmas, more results with linear trends, simulation results, and additional tables of the empirical analysis in Section 6.

2 Distributional Convergence of Panel Data

In the literature of determinant of crime rates, one research interest is to explain the decline of the U.S. national crime rates since the early 1990s. For instance, Figure 1 exhibits the time series of log national crime rates from 1985 to 2022 and shows that such sharp decline happened across all types of crimes.

Many studies seek the key economic variables that are associated with the crime rate trend to find the main drivers of such decline. See Levitt (2004), Moody and Marvell (2010), and the references therein for more discussions. To find such key determinants, existing studies often consider the following two-way fixed effects (TWFE) regression model:

$$y_{it} = \eta_i + \varrho_t + \beta' z_{it} + v_{it} \tag{1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where y_{it} is the log of crime rate of state i and year t, η_i is a state fixed effect, ϱ_t is a year fixed effect, and z_{it} is a vector of explanatory variables possibly including a lagged dependent variable. Instrumental variables regression is often employed to control for potential simultaneity. However, when y_{it} or z_{it} are unit root processes, this regression result should be interpreted with caution, particularly when y_{it} and z_{it} are not cointegrated. For such cases, first-differenced variables are often used, but the interpretation of β becomes different from the level regression.

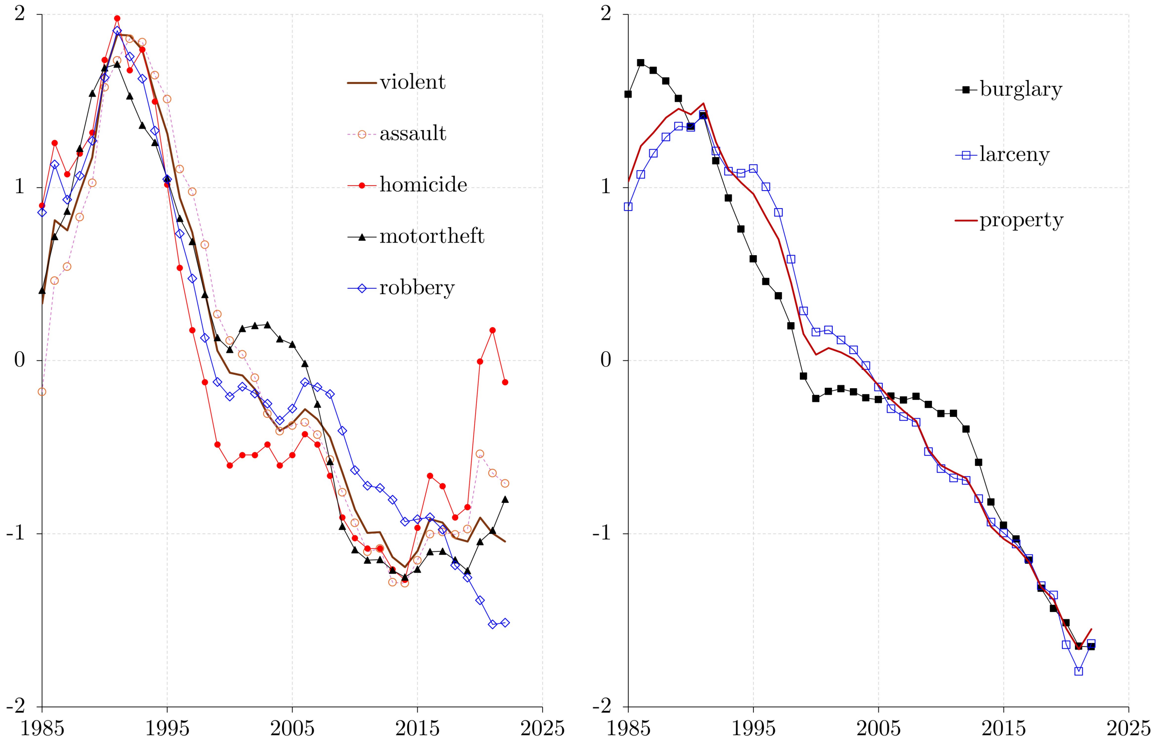
More importantly, even when both variables are stationary, controlling for the time effect ϱ_t removes all the (co-)trends between y_{it} and z_{it}. Hence, significant β in this regression model or in its aggregated form, $\overline{y}_t = \overline{\eta} + \varrho_t + \beta' \overline{z}_t + \overline{v}_t$ with $\overline{r}_t = n^{-1} \sum_{i=1}^{n} r_{it}$ for any panel series r_{it}, does not necessarily provide an evidence whether z_{it} (or $\overline{z}_t$) are the determinants of the declining trend of $\overline{y}_t$ or the national crime rate.

To see this, we suppose the data generating process of the panel series y_{it} is given by

$$y_{it} = \alpha_i + \tau_t + x_{it}^*, \tag{2}$$

where the idiosyncratic component x_{it}^* is assumed to be uncorrelated with α_i and τ_t. x_{it}^* satisfies the mean-reversion property though it can be heteroskedastic over i and t. The panel process y_{it} can be nonstationary and have stochastic trends, which thus allows that τ_t

Figure 1: National Crime Rates in the U.S.



(a) Violent Crimes and Motor Vehicle Theft (b) Property Crimes

Note: The figure shows the log of the number of offenses per 100,000 inhabitants from 1985 to 2022. All series are standardized to place them in one plot. Motor vehicle theft is not categorized as a violent crime, but its trajectory is very similar to those with violent crime rates. We exclude the rape whose definition by the FBI changed in 2013. Data source: The *Uniform Crime Report.*

be a unit-root process. In this representation, τ_t describes the latent common trends among the individual panel series y_{it}, which includes deterministic or stochastic trends. If we let

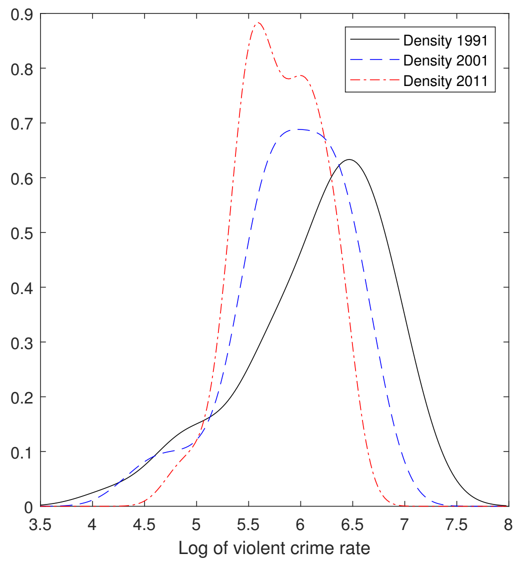
$$x_{it}^* = \beta' z_{it}^* + v_{it} \quad \text{and} \quad z_{it} = \alpha_{z,i} + \tau_{z,t} + z_{it}^* \tag{3}$$

with v_{it} being uncorrelated with $(\alpha_{z,i}, \tau_{z,t}, z_{it}^*)$, we can rewrite (2) as

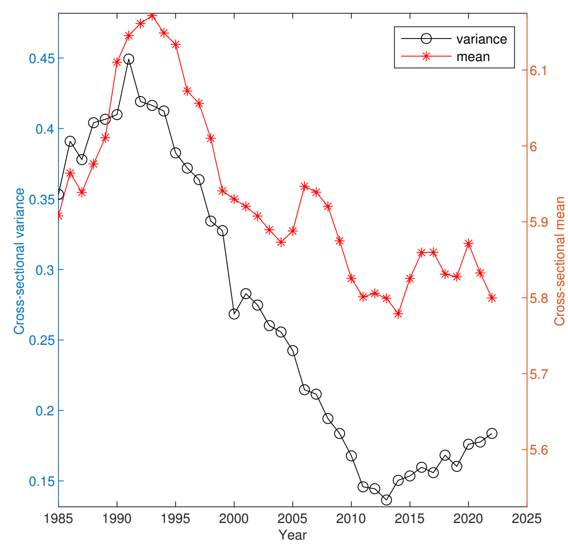
$$\begin{aligned} y_{it} &= \alpha_i + \tau_t + \beta' (z_{it} - \alpha_{z,i} - \tau_{z,t}) + v_{it} \\ &= (\alpha_i - \beta' a_{z,i}) + (\tau_t - \beta' \tau_{z,t}) + \beta' z_{it} + v_{it}, \end{aligned}$$

yielding the TWFE regression in (1), where the fixed effects, $\eta_i = (\alpha_i - \beta' \alpha_{z,i})$ and $\varrho_t =$

Figure 2: Distributional Convergence of Violent Crime Rates in the U.S.



(a) Changes of Densities



(b) Cross-sectional Mean and Variance

Note: Violent crime is composed of homicide, forcible rape, robbery and aggravated assault. The plots are based on the logarithm of the number of offenses per 100,000 population of each state. Figure 2(a) shows densities of the log violent crime rates in 1991, 2001 and 2011. The mode moves to the left, which implies decreasing of the average violent crime rates over time. At the same time, the cross-sectional dispersion shrinks as well. Figure 2(b) shows the sample cross-sectional mean (red asterisk) and variance (black circle) across 50 states. Both the mean and the variance decline since early 1990s.

$(\tau_t - \beta'\tau_{z,t})$, are arbitrarily correlated with the explanatory variables z_{it}. β in this TWFE regression tells the marginal effect of z_{it} to the idiosyncratic term x_{it}^* (i.e., the idiosyncratic deviation from the common trend) but it cannot explain the common trend τ_t of y_{it}.

This simple setup shows that finding significant elements of β in the standard TWFE regression model (1) cannot identify the key variables or factors explaining the crime rate trend. Instead, we suggest to examine the distributional dynamics of y_{it} to identify observed factors of the latent trend τ_t or the trend determinants.

To motivate this idea, we examine the state-level violent crime rates as an illustrating example. Figure 2 exhibits cross-sectional densities of the log violent crime rates, and their cross-sectional mean and variance trajectories across 50 states in the U.S. from 1985 to 2022. Interestingly, in addition to the sharp decline in the national or average crime rates since early 1990s, we can find the decline of cross-sectional dispersion (i.e., heterogeneity) of the crime rates. When τ_t shows a decreasing trajectory (if $\mathbb{E}\alpha_i$ is bounded) and the cross-sectional variance of x_{it}^* decreases over t, then the data generating process (2) well describes such distributional dynamics of the crime rates.

The latter feature reminds us the $\sigma-$convergence in the economic growth theory, which assumes monotonically shrinking cross-sectional distribution over t. Instead of considering such a strong concept of distributional convergence, we adopt the idea of weak $\sigma-$convergence by Kong, Phillips, and Sul (2019) and define the distributional convergence if the cross-sectional dispersion measure is negatively associated with time t. We formally define the following.[1]

Definition 1 (Weak $\sigma-$Convergence towards Common Trend) *A panel series y_{it} is weakly $\sigma-$convergent towards τ_t if the following conditions hold:*
(a) $\text{plim}_{n\to\infty} n^{-1}\sum_{i=1}^{n}(y_{it}-\tau_t)^2 = Q_t < \infty$ *a.s. for all t;*
(b) $\text{plim}_{t\to\infty} Q_t \in [0,\infty)$*;*
(c) $\limsup_{T\to\infty} c_T^{-1}\sum_{t=1}^{T}\widetilde{Q}_t\widetilde{t} = \gamma(Q_t,t) < 0$ *a.s. for some increasing sequence $c_T \to \infty$ as $T\to\infty$, where $\widetilde{z}_t = z_t - T^{-1}\sum_{t=1}^{T} z_t$ denotes the demeaned series of z_t.*

In this definition, the cross-section variation of the idiosyncratic component

$$x_{it} = \alpha_i + x_{it}^* = y_{it} - \tau_t \tag{4}$$

is negatively associated with t. Since the panel data y_{it} share a homogeneous trend τ_t, it hence implies that the cross-sectional distribution of y_{it} weakly converges towards τ_t as t gets large.

Apparently, the common trend τ_t is unobserved and unknown in most cases. An interesting question is then, as in the aforementioned crime rate study, how to tell observed time series variables θ_t to be related with τ_t, which can be identified as the key determinants of the dynamics of the panel series y_{it}. If y_{it} is stationary, one could regress y_{it} on some candidate variables θ_t and gauge their explanatory power to check if they are such trend determinants similarly as Chen, Roll, and Ross (1986). However, as noted by Bai and Ng

[1] When τ_t corresponds to the cross-sectional sample average of y_{it}, this definition is identical to the weak $\sigma-$convergence of y_{it} by Kong, Phillips, and Sul (2019).

(2006), this approach is valid only when the variance of regression error is small enough so that the correlation between y_{it} and θ_t is dominant. For this reason, Bai and Ng (2006) and Parker and Sul (2016) propose to study the correlation between an estimated trend $\widehat{\tau}_t$ and θ_t directly.[2]

We take an approach closer to Chen, Roll, and Ross (1986), which does not require estimating any latent common trend or factors. This approach is more suitable for nonstationary panel processes, which has not been considered in the aforementioned studies. To this end, we suppose that the cross-sectional variance of x_{it} (or equivalently that of x_{it}^*) can vary over t but the variance is negatively associated with t (i.e., x_{it} or x_{it}^* satisfies weak $\sigma-$convergence of Kong, Phillips, and Sul (2019)). This is equivalent to assume that the cross-sectional variation of y_{it} around the nonstationary time series τ_t is negatively associated with t as defined in Definition 1.[3]

If this negative association still holds even when we replace τ_t with some linear combination of the observed time series variables θ_t, then we conclude that θ_t are closely related with τ_t. In this case, we can identify such observed time series variables θ_t as long-run trend determinants of the panel data y_{it}. Importantly, this can be done without estimating the latent common trend τ_t or common nonstationary factors. We will formalize this idea in the next section.

Remark 1 *Our approach presumes y_{it} satisfies Definition 1 or x_{it} is weakly $\sigma-$convergent. Because $y_{it} - \overline{y}_t = x_{it} - \overline{x}_t$ from (4), this can be checked by examining whether y_{it} itself is weakly $\sigma-$convergent. To this end, recall that Kong, Phillips, and Sul (2019) suggest to*

[2]More precisely, they consider the common factor representation $y_{it} = \lambda_i' f_t + x_{it}^*$ and study the correlation between the estimated factor $\widehat{f}_t$ and θ_t, particularly when all the variables are stationary. In our nonstationary context, we can similarly consider the common factor structure as $y_{it} = \alpha_i + \lambda_i' f_t + x_{it}^*$ instead of (2), where f_t includes nonstationary factors. This common factor structure allows for heterogeneous influence of the common trend factors f_t and hence each individual can have its own latent trend $\tau_{it} = \lambda_i' f_t$. Our setup in (2) can be seen as a restricted common factor structure with $\lambda_i = \lambda$ for all i, but it is empirically more relevant to discuss common trend or co-movement in the framework of (2).

[3]Interestingly, such a distributional convergence behavior is often found in panel data series, which is mainly because panel data typically share a common attribute and interact with each other over evolution. Even when all the panel series do not converge to a common trend series, one can find subgroups that reveal convergence within each subgroup or club. See Section 5 for club convergence.

consider the t-test of ψ_K (say $\mathcal{T}^0_{\psi_K}$) from the auxiliary trend regression:

$$R_{n,t} = \psi_{K0} + \psi_K t + u_{K,t},\ \textit{where}\ R_{n,t} = \frac{1}{n}\sum_{i=1}^{n} (y_{it} - \overline{y}_t)^2 . \tag{5}$$

3 Evaluating Long-Run Trend Determinant

We now define the long-run trend determinant of panel data as follow. We let θ_t be an $m \times 1$ vector of observed time series and δ be an $m \times 1$ vector of parameters.

Definition 2 (Long-Run Trend Determinant) *θ_t is a vector of long-run trend determinants of panel series y_{it} if there exists non-zero δ such that y_{it} is weakly $\sigma-$convergent towards $\delta'\theta_t$ as defined in Definition 1.*

To develop a procedure to tell whether nonstationary time series θ_t is a vector of long-run trend determinant of nonstationary panel series y_{it}, where the latent common trend τ_t is unit root, we first suppose that τ_t and θ_t impose a long-run association. A standard way is to specify a cointegrating relation between them. More precisely, we suppose there exists a mean-zero stationary process ξ_t satisfying

$$\tau_t - \delta'\theta_t = \xi_t, \tag{6}$$

where there is no cointegration among θ_t. When τ_t is observed and T is large enough, a natural way to find such θ_t is the cointegration tests. Since τ_t is unobserved, however, we instead combine (2), (4), and (6) to have

$$y_{it} = \delta'\theta_t + \xi_t + x_{it}. \tag{7}$$

Definition 2 implies that θ_t become long-run trend determinants of y_{it} when the cross-sectional variation of the regression error, $y_{it} - \delta'\theta_t = \xi_t + x_{it}$, is negatively associated with t, provided δ is not zero.

Note that the cross-sectional sample variation of y_{it} from $\delta'\theta_t$ is decomposed as

$$S^{\delta}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} (y_{it} - \delta'\theta_t)^2 = \frac{1}{n}\sum_{i=1}^{n} (\xi_t + x_{it})^2 = \xi_t^2 + \frac{1}{n}\sum_{i=1}^{n} (x_{it} - \overline{x}_t)^2 + o_p(1), \tag{8}$$

where ξ_t and x_{it} are assumed to be mean zero and mutually uncorrelated. This expression implies that y_{it} is weakly $\sigma-$convergent towards $\delta'\theta_t$ when the sum $\xi_t^2 + n^{-1}\sum_{i=1}^n (x_{it} - \overline{x}_t)^2$ is negatively associated with t. Since we presume x_{it} is weakly $\sigma-$convergent, it requires the association between ξ_t^2 and t to be negative or to remain small enough so that the negative association between the sum $\xi_t^2 + n^{-1}\sum_{i=1}^n (x_{it} - \overline{x}_t)^2$ and t is intact. For this reason, we can identify long-run determinants θ_t even when they are not strictly cointegrated with τ_t as defined in (6), and hence it can be applied for more general cases. We only need the sum $\xi_t^2 + n^{-1}\sum_{i=1}^n (x_{it} - \overline{x}_t)^2$ to be negatively associated with t as discussed above, under which ξ_t is not necessarily mean-zero stationary for all t. For instance, τ_t and θ_t may not initially share common stochastic trends but become cointegrated after a certain time (cf. segmented cointegration, Kim (2003)). In such a case, the standard cointegration tests are likely to fail to detect the long-run relation (even when τ_t is observed), whereas our approach based on Definition 2 could successfully identify it (whether τ_t is observed or not).

In practice, as δ is unknown, we use

$$S_{n,t} = \frac{1}{n}\sum_{i=1}^{n}(y_{it} - \widehat{\delta}'\theta_t)^2 \tag{9}$$

for some consistent estimator $\widehat{\delta}$ instead of $S_{n,t}^{\delta}$ in (8). We can obtain $\widehat{\delta}$ from time series regression in the aggregated equation of (7):

$$\overline{y}_t = \alpha_0 + \delta'\theta_t + e_t, \tag{10}$$

where $\overline{y}_t = n^{-1}\sum_{i=1}^n y_{it}$ and e_t corresponds to the cross-sectional average of $\xi_t + x_{it}$.[4] When e_t is stationary, least squares estimation of (10) yields a consistent estimator of δ. Furthermore, if $\Delta\theta_t$ and e_s are uncorrelated (i.e., $\Delta\theta_t$ and ξ_s are uncorrelated in this setup) for all t and s,[5] the significance of each element in $\widehat{\delta} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_m)'$ can be checked using the standard

[4] This could be seen as replacing τ_t by $\overline{y}_t$ and run cointegrating regression. It is reminiscent of a common factor estimator by a cross-sectional average (e.g., Pesaran (2006)) or Mundlak (1978)'s approach for time fixed effects. In practice, we can use other (weighted) mean series of y_{it} instead of the cross-sectional sample mean series $\overline{y}_t$. For instance, in the crime rate example in Section 6, we can use the national crime rate instead of the cross-state average crime rate.

[5] When they are correlated, we can use the Fully Modified OLS by Phillips and Hansen (1990) or Canonical Cointegration Regression by Park (1992).

t-statistic (e.g., Phillips and Park (1988)). We do not consider $\theta_{j,t}$ for $j = 1, \ldots, m$ as a potential long-run trend determinant if its coefficient estimator is not significant. In such a case, we can reconstruct θ_t such that it only includes significant elements from this t-test and define $S_{n,t}$ in (9) using this reconstructed θ_t.[6]

The association between $S_{n,t}$ and t can be readily examined using the t-test of ϕ in an auxiliary trend regression:

$$S_{n,t} = \phi_0 + \phi t + u_t. \tag{11}$$

It is important to note that the trend regression (11) is most likely to be misspecified. However, the sign of ϕ can still tell the direction of association between $S_{n,t}$ and t, or between $n^{-1} \sum_{i=1}^{n} (y_{it} - \delta' \theta_t)^2$ and t. To this end, we consider the following robust t-statistic for ϕ:

$$\mathcal{T}_\phi(b) = \frac{\widehat{\phi}}{\left\{ \left(\sum_{t=1}^{T} (\widetilde{t})^2 \right)^{-1} T \widehat{\Omega}(b) \left(\sum_{t=1}^{T} (\widetilde{t})^2 \right)^{-1} \right\}^{1/2}}, \tag{12}$$

where $\widetilde{t} = t - T^{-1} \sum_{r=1}^{T} r$ and $\widehat{\Omega}(b)$ is the HAR (heteroskedasticity autocorrelation robust) long-run variance estimator for some fixed-b coefficient $b \in (0, 1]$ (e.g., Kiefer and Vogelsang (2005)). $\widehat{\Omega}(b)$ is defined as

$$\widehat{\Omega}(b) = \sum_{\ell=-(T-1)}^{T-1} K\left(\frac{\ell}{Tb} \right) \widehat{\Gamma}_\ell, \tag{13}$$

where

$$\widehat{\Gamma}_\ell = \frac{1}{T} \sum_{t=1}^{T-\ell} \varkappa_t \varkappa_{t+\ell} 1\{\ell \geq 0\} + \frac{1}{T} \sum_{t=-\ell+1}^{T} \varkappa_t \varkappa_{t+\ell} 1\{\ell < 0\}$$

with $\varkappa_t = \widehat{u}_t \widetilde{t}$, $\widehat{u}_t = S_{n,t} - \widehat{\phi}_0 - \widehat{\phi} t$, and a symmetric kernel function $K : \mathbb{R} \mapsto [0, 1]$ satisfying

[6] If e_t is unit-root and hence (10) becomes spurious regression, $\widehat{\delta}$ will appear to be significant with large T and all θ_t will be included in defining $S_{n,t}$. However, the auxiliary trend regression (11) will tell $S_{n,t}$ is not negatively associated with t, and hence such θ_t will be eventually deselected as long-run determinants.

$K(0) = 1$ and $\int K(\nu)\, d\nu = 1$.[7] If u_t is homoskedastic, $\mathcal{T}_\phi(b)$ is simplified as

$$\mathcal{T}_\phi^0(b) = \frac{\widehat{\phi}}{\left(\widehat{\Omega}_u(b) / \sum_{t=1}^{T} (\widetilde{t})^2\right)^{1/2}}, \tag{14}$$

where $\widehat{\Omega}_u(b)$ is obtained as $\widehat{\Omega}(b)$ in (13) by replacing $\varkappa_t$ with $\widehat{u}_t$.

When the one-side t-test in the auxiliary trend regression (11) rejects $\phi \geq 0$ against $\phi < 0$, we conclude that $S_{n,t}$ is negatively associated with t and hence y_{it} is weakly $\sigma-$convergent towards $\delta'\theta_t$. From the discussion above, this result yields that θ_t are long-run trend determinants of y_{it}. On the other hand, when the t-test cannot reject $\phi \geq 0$, $S_{n,t}$ is either positively associated with t or unassociated with t. This corresponds to the cases where the cross-sectional variation of $y_{it} - \delta'\theta_t$ increases over t, or neither increases nor decreases over t. Section 4 derives the limiting distribution of $\mathcal{T}_\phi(b)$ in this auxiliary trend regression and provides the simulated critical values.

Remark 2 *The interpretation in (8) offers an interesting perspective on cointegration between the panel series y_{it} and the time series θ_t. Even when the common trend τ_t of y_{it} is cointegrated with θ_t as in (6), y_{it} may not show weak $\sigma-$convergence towards $\delta'\theta_t$ (hence y_{it} does not appear to share a common stochastic trend with θ_t). This happens when the association between the squared cointegration error ξ_t^2 and t outweighs the negative association between the cross-sectional variance of y_{it} and t (see Remark 1). Therefore, Definition 2 accounts for the following two aspects jointly: the long-run relation between the common trend τ_t of y_{it} and the time series θ_t (i.e., the first-moment relation); and the non-dominating variation of the disequilibrium error ξ_t relative to that of the panel series y_{it} over t (i.e., the second-moment relation). In comparison, the standard cointegration only accounts for the first aspect. In this regard, our approach can be seen or used as a (more general) cointegration test between time series and panel series.*

[7] The required conditions of kernel functions are given in Assumption A. For the Bartlett kernel, the HAR estimator is given as

$$\widehat{\Omega}(b) = \frac{1}{T}\sum_{t=1}^{T} \varkappa_t^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\varkappa_t \varkappa_{t+\ell}$$

with $L = [bT]$, where $[c]$ is the largest integer smaller than or equal to c.

4 Limiting Distribution of the t-Test

In this section, we derive the limiting distributions of the t-statistic $\mathcal{T}_\phi(b)$ given in (12). It is important to note that, though this t-test is on the sign of ϕ in the trend regression (11), the restrictions on ϕ cannot be directly used to derive the limiting distribution of $\mathcal{T}_\phi(b)$. This is because the auxiliary trend regression (11) is a misspecified one and hence the true ϕ does not exist. For this reason, unlike the standard analysis, we drive the limiting distribution of $\mathcal{T}_\phi(b)$ by assuming some data generating processes that well characterizes the sign of the long-run association between $S_{n,t}$ and t.

More precisely, based on the decomposition given in (8), we can characterize the sign of the association between $S_{n,t}$ and t using the relative behaviors between $n^{-1}\sum_{i=1}^n (x_{it}-\overline{x}_t)^2$ and ξ_t^2. Note that $n^{-1}\sum_{i=1}^n (x_{it}-\overline{x}_t)^2$ is the same as $R_{n,t}=n^{-1}\sum_{i=1}^n (y_{it}-\overline{y}_t)^2$ as discussed in Remark 1. We let

$$R_t = \operatorname*{plim}_{n\to\infty} R_{n,t} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (y_{it}-\overline{y}_t)^2 .$$

Recall that we define the limit of the linear association between a time series z_t and t as

$$\gamma(z_t,t) = \lim\sup_{T\to\infty} \frac{1}{c_T}\sum_{t=1}^{T} \widetilde{z}_t\widetilde{t}$$

for some $c_T\to\infty$ in Definition 1-(c). We presume x_{it} is weakly $\sigma-$convergent, and thus $\gamma(R_t,t)<0$. We consider the following three cases:

$$\text{(i)}\;\; \xi_t \text{ is } I(1); \tag{15}$$

$$\text{(ii)}\;\; \xi_t \text{ is } I(0) \text{ and } \gamma(\xi_t^2,t) \text{ dominates } \gamma(R_t,t); \tag{16}$$

$$\text{(iii)}\;\; \xi_t \text{ is } I(0) \text{ and } \gamma(\xi_t^2,t) \text{ is dominated by } \gamma(R_t,t). \tag{17}$$

Under the case (i), $\gamma(\xi_t^2,t)$ is positive and dominates $\gamma(R_t,t)$, and hence the common trend of y_{it} is likely to reveal diverging trajectory from $\delta'\theta_t$. Under the case (ii), y_{it} shows neither weak $\sigma-$convergence toward $\delta'\theta_t$ nor divergence from $\delta'\theta_t$ because ξ_t is stationary. Under the case (iii), since $\gamma(R_t,t)<0$ is presumed, y_{it} reveals weak $\sigma-$convergence toward $\delta'\theta_t$. Therefore, the case (iii) is when $\mathcal{T}_\phi(b)$ rejects $\phi\geq 0$ in favor of $\phi<0$, whereas the cases (i) and (ii) are when $\mathcal{T}_\phi(b)$ cannot reject $\phi\geq 0$.

To formally specify these different cases, we suppose that x_{it} is a trend-stationary process for each i with shrinking variance over t, which is generated as

$$x_{it} = \alpha_i + \mu_i t^{-\kappa_1} + \epsilon_{it} + \varepsilon_{it} t^{-\kappa_2} \tag{18}$$

for some constants $\kappa_1, \kappa_2 \in (0, 1/2)$.[8] This specification is for weakly $\sigma-$ convergent x_{it}, similar to the models in Kong, Phillips, and Sul (2019) and Kong, Phillips, and Sul (2020), though it is more general than theirs. This setup maps the sign of ϕ in the trend regression (11) to the values of κ_1 and κ_2 in the data generating process (18). The values of κ_1 and κ_2 determine the decay rate of the cross-sectional variance of x_{it}, enabling flexible control over the relative behavior of $\gamma(\xi_t^2, t)$ and $\gamma(R_t, t)$, especially for the cases (ii) and (iii) above. In particular, we can verify that the case (ii) is specified with $\kappa_1, \kappa_2 > 1/4$, which ensures that the cross-sectional variance of x_{it} vanishes fast so that $\gamma(R_t, t)$ is small enough to be dominated by $\gamma(\xi_t^2, t)$. On the other hand, the case (iii) can be specified with $\kappa_1, \kappa_2 \leq 1/4$.

The first theorem below summarizes the limiting (null) distribution of $\mathcal{T}_\phi(b)$ when $S_{n,t}$ is not negatively associated with t, which is the case that θ_t are not long-run determinants of $y_{i,t}$. '$\Rightarrow$' denotes weak convergence of the associated probability measures and '$\equiv$' stands for the distributional equivalence. b is the fixed-b parameter and $K(\cdot)$ is the kernel function used in HAR estimator in (13). Let $\min\{\kappa_1, \kappa_2\} = \kappa_1 1\{\kappa_1 \leq \kappa_2\} + \kappa_2 1\{\kappa_1 > \kappa_2\}$, where $1\{\cdot\}$ is the binary indicator.

Theorem 1 *Suppose Assumptions 1-3 in the Appendix hold and $T/n \to 0$ as $n, T \to \infty$. Let x_{it} satisfy (18) with $\kappa_1, \kappa_2 \in (0, 1/2)$. For given $b \in (0, 1]$, when $S_{n,t}$ is not negatively associated with t (i.e., under the cases (15) or (16)), $\mathcal{T}_\phi(b)$ in (12) satisfies*

$$\begin{cases} \mathcal{T}_\phi(b) \Rightarrow F_1(b) & \text{if } \xi_t \sim I(1) \\ \mathcal{T}_\phi(b) \Rightarrow F_0(b) & \text{if } \xi_t \sim I(0) \text{ and } \min\{\kappa_1, \kappa_2\} > 1/4 \end{cases}$$

where

$$F_0(b) \equiv \frac{\mathcal{Z}}{\left\{12 \int_0^1 \int_0^1 K\left(\frac{t-s}{b}\right)\left(r - \frac{1}{2}\right)\left(s - \frac{1}{2}\right) dW^\tau(r)\, dW^\tau(s)\right\}^{1/2}} \tag{19}$$

[8] We exclude the case with $\kappa_1, \kappa_2 = 0$ because we presume x_{it} satisfies weak $\sigma-$convergence.

Table 1: One-Sided Asymptotic Critical Values (Bartlett kernel)

	$F_0(b)$ in (19)				$F_0^0(b)$ in (20)			
$b =$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
1%	-3.037	-3.758	-4.350	-4.861	-2.914	-3.598	-4.268	-4.988
2.5%	-2.488	-3.045	-3.500	-3.895	-2.385	-2.890	-3.407	-3.974
5%	-2.040	-2.467	-2.826	-3.135	-1.961	-2.340	-2.735	-3.181
10%	-1.554	-1.861	-2.117	-2.340	-1.501	-1.759	-2.035	-2.354
20%	-0.999	-1.181	-1.336	-1.472	-0.968	-1.117	-1.278	-1.469

Note: The values are the simulated percentiles of the limiting distribution of $\mathcal{T}_\phi(b)$ and $\mathcal{T}_\phi^0(b)$ given in (19) and (20), respectively, from 2 million replications. The Brownian motion is approximated by normalized sums of standard normal random variables using 10,000 steps and the Bartlett kernel is used for HAR estimation. Recall $F_0(b)$ allows for heteroskedasticity as (12) whereas $F_0^0(b)$ is under the homoskedasticity restriction as (14). Critical values for other $b \in (0,1]$ are available in the Appendix.

and $F_1(b)$ is given in (B.4) in the Appendix. $\mathcal{Z}$ is the standard normal random variable and $W^\tau(r)$ is the second-level Brownian bridge.[9]

Theorem **1** obtains the limiting distribution of $\mathcal{T}_\phi(b)$ under the two scenarios in (15) and (16). The first scenario (i.e., $F_1(b)$) is when $S_{n,t}$ is positively associated with t, in which τ_t and θ_t are not cointegrated. The second scenario (i.e., $F_0(b)$) is when $S_{n,t}$ is not associated with t though τ_t and θ_t are cointegrated. When u_t is homoskedastic, the t-statistic is simplified as $\mathcal{T}_\phi^0(b)$ in (14), whose limiting null distribution is given as

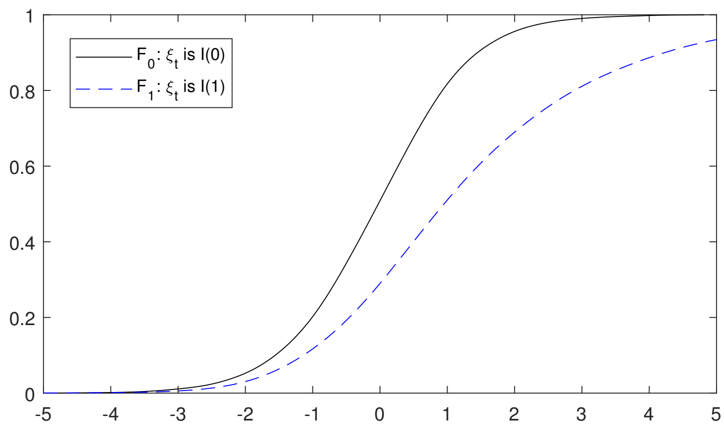
$$F_0^0(b) \equiv \frac{\mathcal{Z}}{\left\{\int_0^1 \int_0^1 K\left(\frac{t-s}{b}\right) dW^\tau(r)\, dW^\tau(s)\right\}^{1/2}} \tag{20}$$

instead of (19).

Interestingly, both $F_1(b)$ and $F_0(b)$ in Theorem **1** (and $F_0^0(b)$ as well) are free from nuisance parameters. As depicted in Figure 3, simulations exhibit that $F_1(b)$ stochastically dominates $F_0(b)$. Thus, in conjunction with Theorem **2** below, we can construct critical values for this one-sided test using $F_0(b)$. The asymptotic critical values for both $F_0(b)$ and

[9]It is equivalent to linearly $L_2[0,1]$ demeaned and detrended standard Brownian motion $W(r)$. More precisely, $W^\tau(r) = W(r) - a^* - b^* r$ and the coefficients (a^*, b^*) are solutions of $\min_{(a,b)} \int_0^1 \{W(r) - a^* - b^* r\}^2 dr$. See MacNeill (1978) and Park and Phillips (1988) for the details.

Figure 3: Limiting null distributions of $\mathcal{T}_\phi(b)$



Note: The black solid line is $F_0(b)$ and the blue dashed line is $F_1(b)$ with $b = 0.1$. The limiting distributions are simulated with $T = 5,000$ and 10,000 replications.

$F_0^0(b)$ are provided in Table 1.

The next theorem gives the limit of $\mathcal{T}_\phi(b)$ when $S_{n,t}$ is negatively associated with t, which corresponds to the case in (17). It hence yields the limiting alternative distribution of $\mathcal{T}_\phi(b)$. '$\xrightarrow{p}$' denotes the convergence in probability.

Theorem 2 *Suppose the conditions in Theorem 1 hold and $\xi_t \sim I(0)$. For given $b \in (0, 1]$, let $S_{n,t}$ be negatively associated with t (i.e., the case (17) holds). When $\min\{\kappa_1, \kappa_2\} = 1/4$,*

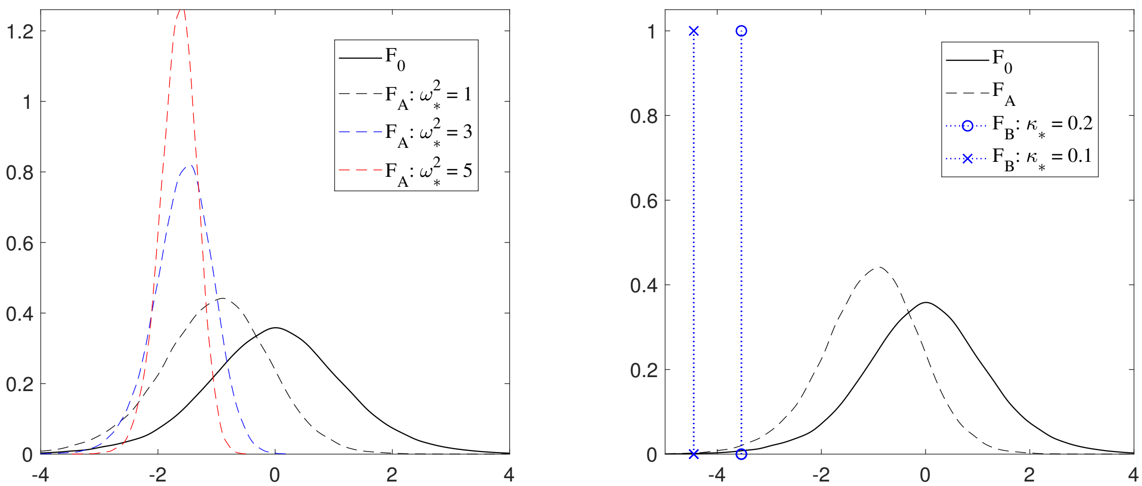
$$\mathcal{T}_\phi(b) \Rightarrow \frac{\mathcal{Z} - (2/\sqrt{3})\omega_*^2}{\left\{12 \int_0^1 \int_0^1 K\left(\frac{t-s}{b}\right)\left(r - \frac{1}{2}\right)\left(s - \frac{1}{2}\right)\left(dW^\tau(r) + \omega_*^2 h(r)dr\right)\left(dW^\tau(s) + \omega_*^2 h(s)ds\right)\right\}^{1/2}}, \tag{21}$$

where $h(r) = 4r + r^{-1/2} - 4$ and

$$\omega_*^2 = \begin{cases} \sigma_\mu^2/\omega_{\xi\xi} & \text{if } \kappa_1 < \kappa_2 \\ \sigma_\varepsilon^2/\omega_{\xi\xi} & \text{if } \kappa_1 > \kappa_2 \\ \left(\sigma_\mu^2 + \sigma_\varepsilon^2\right)/\omega_{\xi\xi} & \text{if } \kappa_1 = \kappa_2, \end{cases} \tag{22}$$

with $\sigma_\mu^2 = \mathbb{E}\mu_i^2$, $\sigma_\varepsilon^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathbb{E}\varepsilon_{it}^2$, and $\omega_{\xi\xi}^2 = \sum_{j=-\infty}^{\infty} \mathbb{E}(\xi_t^2 - \mathbb{E}\xi_t^2)(\xi_{t+j}^2 - \mathbb{E}\xi_t^2)$. When $\min\{\kappa_1, \kappa_2\} < 1/4$,

Figure 4: Limiting alternative distributions of $\mathcal{T}_\phi(b)$



(a) When $\min\{\kappa_1, \kappa_2\} = 1/4$ (b) When $\min\{\kappa_1, \kappa_2\} < 1/4$

Note: Figure 4(a) illustrates how the limiting distribution $F_A(b; \omega_*^2)$ in (21) changes with ω_*^2. As ω_*^2 increases, the distribution shifts to the left and its variance reduces, as shown by the black ($\omega_*^2 = 1$), blue ($\omega_*^2 = 3$), then red ($\omega_*^2 = 5$) dashed lines. Figure 4(b) illustrates how the limiting point $F_B(b; \kappa_*)$ in (23) changes with κ_*, where $\omega_*^2 = 1$. As κ_* decreases, the limiting point shifts to the left, from "circle" ($\kappa_* = 0.2$) to "cross" ($\kappa_* = 0.1$). In both figures, the black solid line is the limiting null distribution $F_0(b)$. All the limiting distributions are simulated with $b = 0.1$, $T = 5,000$, and 10,000 replications.

$$\mathcal{T}_\phi(b) \xrightarrow{p} \frac{-\kappa_*/2}{\left\{\int_0^1 \int_0^1 K\left(\frac{t-s}{b}\right)\left(r - \frac{1}{2}\right)\left(s - \frac{1}{2}\right) g\left(r; \kappa_*\right) g\left(s; \kappa_*\right) drds\right\}^{1/2}} < 0, \tag{23}$$

where $\kappa_* = \min\{\kappa_1, \kappa_2\}$ *and* $g\left(r; \kappa_*\right) = 6\kappa_* r + (1 - \kappa_*)(1 - 2\kappa_*)r^{-2\kappa_*} - (1 + 2\kappa_*)$.

When $\min\{\kappa_1, \kappa_2\} = 1/4$ under $\xi_t \sim I(0)$, neither $\gamma(R_t, t)$ nor $\gamma(\xi_t^2, t)$ is dominant. In this case, the ratio between the cross-sectional variance of the panel series $y_{i,t}$ (i.e., R_t) and the variance of the disequilibrium error ξ_t (i.e., ξ_t^2) becomes crucial. This variance ratio is measured by ω_*^2 and it influences the limiting distribution of $\mathcal{T}_\phi(b)$ as in (21). As depicted in Figure 4(a), as ω_*^2 gets large, the limiting distribution shifts to the negative direction and shrinks toward the shifted center. When ω_*^2 is small, on the other hand, it could be hardly distinguished from $F_0(b)$. This hence can be understood as a local alternative in this context and demonstrates that the power of our one-sided test depends on ω_*^2. This implies that, even when θ_t is cointegrated with the latent common trend τ_t, the test can conclude that θ_t are the long-run determinants of the common trend of the panel series y_{it} only if the cross-sectional variation of y_{it} outweighs that of the cointegration error ξ_t, as we have also discussed in Remark 2.

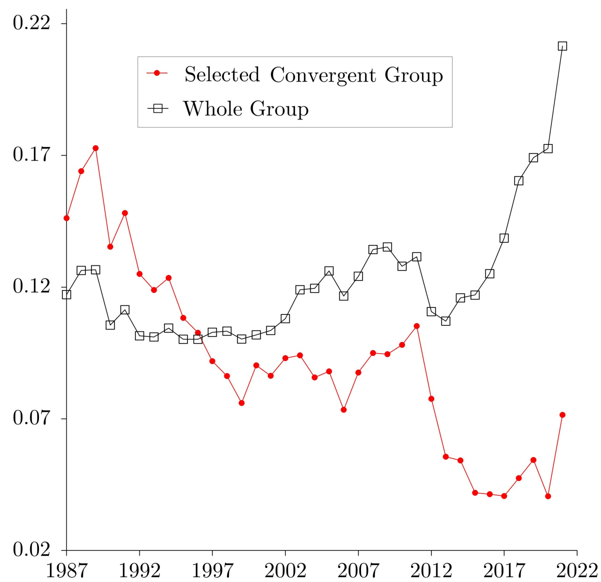
When $\min\{\kappa_1, \kappa_2\} < 1/4$, $\gamma(R_t, t)$ dominates $\gamma(\xi_t^2, t)$ and the limiting distribution of $\mathcal{T}_\phi(b)$ no longer depends on the variance ratio ω_*^2. In this case, the rate of diminishing variance R_t becomes the important factor, which solely depends on $\kappa_* = \min\{\kappa_1, \kappa_2\}$. The limit of $\mathcal{T}_\phi(b)$ degenerates to a negative value as given in (23). Figure 4(b) shows that the probability limit moves to the left as κ_* gets smaller. Furthermore, simulations show that the degenerating point is always below the 5% critical value given in Table 1 at each b value. From these two cases, under $\min\{\kappa_1, \kappa_2\} \leq 1/4$, we can find that the limiting distribution of $\mathcal{T}_\phi(b)$ shrinks toward its negative mode as T grows and eventually degenerates as in (23), which gives the consistency of the test. The online Appendix provides further simulation evidence.

Remark 3 *Nonstationary variables often exhibit linear trends. For instance, suppose the nonstationary common trend τ_t imposes a linear trend (e.g., a random walk with drift). If we properly choose a variable θ_t that also follows a random walk with drift and the regression of τ_t on θ_t successfully accounts for both the linear and stochastic trends, the cointegration error $\xi_t = \tau_t - \delta'\theta_t$ will be a stationary process without a linear trend. In this case, the limiting null distribution $F_0(b)$ of $\mathcal{T}_\phi(b)$ is unaffected by the drift terms. However, if θ_t is not correctly chosen or the linear trend is not controlled for, the cointegration error ξ_t will either $I(1)$ or, at best, $I(0)$ with a linear trend, under which the test $\mathcal{T}_\phi(b)$ would not reject $\phi \geq 0$. In the online supplement, we derive the limiting distributions for such cases and show that this leads to pivotal limiting distributions similar to $F_1(b)$ in Theorem* **1***, which have much thinner left tails than that of $F_1(b)$. The drift term in τ_t hence do not affect our test, and we can use the same critical values as in the no-drift case in Table 1.*

5 Partial Distributional Convergence Case

The important presumption of our approach developed in the previous sections is that the panel data series y_{it} is weakly $\sigma-$convergent towards its common trend τ_t in (2), or equivalently x_{it}^* is weakly $\sigma-$convergent. However, it is not rare that the panel series include

Figure 5: Partial Distributional Convergence of Burglary Rates



Note: The figure plots the cross-sectional dispersion trajectories of burglary rates across all the 50 states (black square line) from 1987 to 2021 and among the selected 25 states (red circle line) that exhibit convergence toward the national burglary rates.

some outlying individual series that impose their unique trends. In such a case, it is unlikely that all the individual series in the panel sample yield the weak $\sigma-$convergence towards the common trend τ_t. For example, Figure 5 depicts the cross-sectional dispersion of the burglary rates across all 50 states from the national rate (black square line), which shows a divergent trajectory, unlike the violent crime rates in Figure 2. This is because some of the panel series have different trends from τ_t or their idiosyncratic terms x_{it}^* do not satisfy the weak $\sigma-$convergence.

However, we can find a subgroup of 25 panel series that show weak $\sigma-$convergence towards the national burglary rates (red circle line). In fact, we can seek a subgroup of the panel series within which the the weak $\sigma-$convergence towards its common trend holds. Once we find this convergent subgroup, we can identify the long-run trend determinants θ_t of the leading trend τ_t within this subgroup using the method developed in the previous section. In this way, we can exclude some outlying individuals that could (incorrectly) influence the leading common trend and result in incorrect trend determinant identification. When such a subgroup is the majority (or becomes the core group), we conclude this θ_t as the main drivers of the overall dynamics of y_{it}.

For any given reference time series θ_t, selecting a convergent subgroup can be done by repeatedly applying the procedure developed in the previous section across all the different subsets among y_{it}. In practice, however, this approach can be infeasible when the cross-section size n is large and hence the number of different combinations becomes massive. We instead take a similar approach as Phillips and Sul (2007) and suggest the following procedure. For time series θ_t chosen, we use the t-test of φ_i in the individual-specific auxiliary trend regression,

$$\Delta_{it} = \varphi_{0i} + \varphi_i t + u_{it} \tag{24}$$

for each i, where Δ_{it} is a distance measure between y_{it} and $\delta'\theta_t$ defined as

$$\Delta_{it} = (y_{it} - \widehat{\delta}'\theta_t)^2 \tag{25}$$

and $\widehat{\delta}$ is from the time series regression of $\bar{y}_t$ on θ_t given in (10). For given $b \in (0,1]$, we construct the t-statistic $\mathcal{T}_{\varphi_i}(b)$ of φ_i as in (12) or (14), where $\widehat{u}_t$ is replaced with $\widehat{u}_{it} = \Delta_{it} - \widehat{\varphi}_{0i} + \widehat{\varphi}_i t$ for each i. When $\mathcal{T}_{\varphi_i}(b)$ is smaller than some threshold, we conclude that Δ_{it} is not positively associated with t, which implies y_{it} shares a common trend with $\delta'\theta_t$, and we select i as a member of the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ for this chosen θ_t. This is because, when τ_t and θ_t are cointegrated, Δ_{it} would not grow with t if the distance between y_{it} and the common trend τ_t decreases over t.

It is worthy to discuss the choice of critical values in this subgroup selection test, which can be regarded as a pre-test. In fact, the limit of $\mathcal{T}_{\varphi_i}(b)$ behaves quite similarly as that of $\mathcal{T}_\phi(b)$, and one could use the critical values from Table 1. However, we want to choose a first-step critical value so that the true convergent members will be selected into the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ with high probability and hence can contribute to improve the power of the test $\mathcal{T}_\phi(b)$ in the second-step. This can be achieved by choosing a larger critical value in this first-step one-sided test, say c_1, with which the nominal size becomes higher than the usual practice and $\Pr(\mathcal{T}_{\varphi_i}(\theta; b) < c_1 | i \in \mathcal{G}(\theta))$ stays high, where $\mathcal{G}(\theta)$ is the true convergent group towards $\delta'\theta_t$. The following corollary provides a useful guideline for this. Derivation of this corollary is given in the online supplement.

Corollary 1 *Suppose conditions in Theorems 1 and 2 hold for each i, but now $T/n = T \to \infty$. $F_1(b)$ and $F_0(b)$ are defined in Theorems 1 and 2, respectively. If $\xi_t \sim I(1)$, $\mathcal{T}_{\varphi_i}(b) \Rightarrow F_1(b)$ as $T \to \infty$ for each i. If $\xi_t \sim I(0)$ and for $\kappa_1, \kappa_2 \in (0, 1/2)$, $\mathcal{T}_{\varphi_i}(b)$ converges to $F_0(b)$, negatively shifted $F_0(b)$, or a negative point not larger than (23) as $T \to \infty$ for each i.*

When $\xi_t \sim I(1)$, which is the case that Δ_{it} is positively associated with t, the 10%-quantile of $F_1(b)$ is about -1.13 with $b = 0.1$. From Corollary 1, when Δ_{it} is not positively associated with t, the limit of $\mathcal{T}_{\varphi_i}(b)$ will be stochastically dominated by $F_1(b)$. Based on this finding, we use $c_1 = -1.2$ for our empirical analysis in the next section. This choice allows for at least 10% type I error in the first-step selection test with $b = 0.1$, while it is more likely to select the non-diverging members (i.e., the case with $\xi_t \sim I(0)$) into the subgroup estimate $\widehat{\mathcal{G}}(\theta)$.

We can improve the subgroup selection by implementing the following iterative algorithm. The procedure described above defines Δ_{it} in (25) using $\widehat{\delta}$ that is obtained from the time series regression of $\bar{y}_t$ on θ_t (i.e., $\bar{y}_t = a_0 + \delta'\theta_t + e_t$), where $\bar{y}_t = n^{-1}\sum_{i=1}^n y_{it}$ is the average of y_{it} for all i. Once we find the subgroup, say $\widehat{\mathcal{G}}^{(1)}(\theta)$, we can update $\bar{y}_t$ such that $\bar{y}_t^{(1)} = |\widehat{\mathcal{G}}^{(1)}(\theta)|^{-1}\sum_{i\in\widehat{\mathcal{G}}^{(1)}(\theta)} y_{it}$ and obtain $\widehat{\delta}^{(1)}$ from the time series regression of $\bar{y}_t^{(1)}$ on θ_t, where $|\widehat{\mathcal{G}}^{(1)}(\theta)|$ is the cardinality of $\widehat{\mathcal{G}}^{(1)}(\theta)$. Then we run the trend regression in (24) using $\Delta_{it}^{(1)} = (y_{it} - \widehat{\delta}^{(1)\prime}\theta_t)^2$ and conduct the t-test $\mathcal{T}_{\varphi_i}(b)$ only for $i \in \widehat{\mathcal{G}}^{(1)}(\theta)$ to update the subgroup estimate to have $\widehat{\mathcal{G}}^{(2)}(\theta)$. We repeat this procedure until the subgroup membership is not further updated or the estimate of δ does not change. If θ_t is indeed a long-run trend determinant of this subgroup, this iteration will yield a strictly positive $|\widehat{\mathcal{G}}(\theta)|$ at the end of this iteration procedure. If θ_t is not a long-run trend determinant (i.e., none of the panel series share the same long-run trend with θ_t), we can expect that this iteration will decrease the subgroup size each round and eventually yield an empty $\widehat{\mathcal{G}}(\theta)$.

Once we conclude the subgroup $\widehat{\mathcal{G}}(\theta)$, we conduct the t-test $\mathcal{T}_{\phi}(b)$ only using members in $\widehat{\mathcal{G}}(\theta)$. We want to have a large enough $|\widehat{\mathcal{G}}(\theta)|$ to claim this subgroup is the majority (i.e., becomes the core group) and hence this θ_t well describes the overall common long-run trend in the panel series.[10]

[10]We can enrich the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ by adding potentially missing agents in this selection

6 Trend Determinants of U.S. Crime Rates

We now look into the examples introduced in the previous sections and identify the key trend determinants yielding sharp decline of the national crime rate in the U.S. in the 1990s. As we discussed in Section 2, we claim that the desired answer to this research question cannot be obtained from the standard TWFE regression. Instead, we suggest using our approach to check if a given time series is a long-run trend determinant of the crime rate y_{it} as defined in Definition 2, which then should be regarded as one of the determinants that govern the evolution of the national crime rate and hence explains the sharp crime rate decline in the 1990s. In particular, we revisit the following four variables that Levitt (2004) considered and study which one can be identified as trend determinants of our definition: the number of sworn police officers, the incarceration rate, the real GDP, and demographics (i.e., the proportion of population in some specific age group). Levitt (2004) found the first two variables were the determinants of the crime rate decline in the 1990s but the two latter were not.[11]. Unlike such findings, we find that demographics is the key long-run trend determinant of the violent crime rates and the incarceration rate is the key long-run trend determinant of the property crime rates, which well describe the crime rate dynamics including its decline in the 1990s. We report the detailed results in the following two subsections.

6.1 Violent Crimes

We consider the log of the following state-level violent crime rates: aggravated assault, homicide (murder and nonnegligent manslaughter), robbery, and the overall violent crime. As noted in Figure 1, forcible rape is excluded in this analysis due to data limitation; instead we include the violent crime rate that encompasses all four types of violent crime (i.e., assault,

step. One idea is to measure some distance from y_{it} toward $\widehat{\delta}'\theta_t$ for each $i \in \widehat{\mathcal{G}}(\theta)^c$ over some periods and include the agents with the smallest distance to $\widehat{\mathcal{G}}(\theta)$ as long as the t-test $\mathcal{T}_\phi(b)$ stays below the desired critical value. The forecast depth by Lee and Sul (2023a) or more generally Lee and Sul (2023b) can be used for such distance measure. See the online supplement for the details.

[11] Levitt (2004) concluded that the following four variables are the main determinants of the crime rate decline – *increases in the number of police officers, rising prison population, receding crack epidemic, and legalization of abortion* – but the following variables are not the determinants – *strong economy of the 1990s, demographics, policing strategies, gun control laws, carrying of concealed weapons, and capital punishment.*

homicide, robbery, and rape). We collect all the crime rates from the Uniform Crime Report on FBI Crime Data Explorer, and study the period from 1987 to 2021 over 50 states in the U.S.[12]

Prior to our analysis, we need to check if x_{it}^* in (2) satisfies the weak $\sigma-$convergence and if y_{it} imposes stochastic trends (i.e., nonstationary process). As noted in Remark 1, the first point can be readily checked by studying the weak $\sigma-$convergence of the panel series y_{it} using the t-test of ψ_K, $\mathcal{T}^0_{\psi_K}$, from the trend regression (5). For the second point, instead of directly testing for the unit root of all the panel series, we indirectly check if y_{it} satisfies the weak $\sigma-$convergence toward some deterministic trend.[13] This can be done by replacing τ_t by some deterministic trend function and check the weak $\sigma-$convergence of detrended y_{it}. More specifically, we consider a linear trend function $\tau_t = \delta t$ in this analysis, and conduct the weak $\sigma-$convergence test of Kong, Phillips, and Sul (2019) as in Remark 1 using $R_{n,t} = n^{-1}\sum_{i=1}^n (y_{it} - \widehat{\alpha}_0 - \widehat{\delta}t)^2$, where $(\widehat{\alpha}_0, \widehat{\delta})$ is from the trend regression: $\overline{y}_t = \alpha_0 + \delta t + e_t$. If y_{it} is nonstationary, its cross-sectional variation would be positively associated with t even after this detrending. Thus, when the weak $\sigma-$convergence holds in this case, we conclude y_{it} to be trend stationary.

The first column in Table 2 reports these t-statistics $\mathcal{T}^0_{\psi_K}$ using the HAC long-run variance estimator with the lag length of $[T^{1/3}]$. The specific form of $\mathcal{T}^0_{\psi_K}$ is given in Appendix A.1. As suggested in Kong, Phillips, and Sul (2019), we reject the null of no weak $\sigma-$convergence if $\mathcal{T}^0_{\psi_K} < -1.65$. All satisfy weak $\sigma-$convergence and hence the panel series y_{it} share a common long-run trend. Instead of the sample average process $\overline{y}_t$, we also consider the log of national crime rate and report the t-statistics in the second column, but the results are quire similar to those with the sample average process. Next, the third column in Table 2 reports the t-statistics of the weak $\sigma-$convergence test toward a linear deterministic trend, which show y_{it} is most likely nonstationary.[14]

Based on these preliminary results, we now apply the idea in Section 2 for the following

[12] https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/crime-trend. The period is chosen to match the available sample period of potential determinant variables we consider.

[13] This approach can be a useful alternative to the standard panel unit root tests especially when T is small (but n is large) as our case.

[14] Though $T = 35$ is rather small in this analysis, we also conduct Augmented Dickey–Fuller, KPSS, and Phillips-Perron tests to check whether $\overline{y}_t$'s are unit root (with linear trends). All tests cannot reject the null of unit root.

Table 2: Preliminary Weak $\sigma-$Convergence Tests of Violent Crime Rates

	Weak $\sigma-$convergence toward:		
	Sample Mean	National Average	Linear Trend
Violent	-10.01	-11.14	6.36
Assault	-6.23	-8.44	5.07
Homicide	-3.06	-4.83	1.96
Robbery	-13.27	-9.47	4.79

Note: (i) The first two columns report the t-ratio $\mathcal{T}^0_{\psi_K}$ of the weak $\sigma-$convergence test from (5) using the HAC long-run variance estimator with the lag length of $[T^{1/3}]$; $\mathcal{T}^0_{\psi_K} < -1.65$ implies weak $\sigma-$convergence toward the sample mean (1st column) or national average (2nd column). The last column reports the t-ratio of the weak $\sigma-$convergence test toward a linear trend, which is $\mathcal{T}^0_{\psi_K}$ from (5) with $R_{n,t} = n^{-1}\sum_{i=1}^{n}(y_{it} - \widehat{\alpha}_0 - \widehat{\delta}t)^2$, where $(\widehat{\alpha}_0, \widehat{\delta})$ is obtained from $\overline{y}_t = \alpha_0 + \delta t + e_t$; a value larger than -1.65 implies y_{it} is not trend stationary. The specific form of $\mathcal{T}^0_{\psi_K}$ is given in Appendix A.1. (ii) Violent crime includes homicide, rape, robbery, and assault. (iii) The sample period is from 1987 to 2021, which matches with the sample period of candidate determinants θ_t that are lagged by one period.

time series as candidates of long-run determinants θ_t: the fraction of young adult population of age between 10 and 39 (Demog), the number of non-civilian police officers (Police), the local incarceration rate (Prison), and the real GDP per capita (RGDP).[15] All variables are log-transformed and lagged by one period to minimize potential simultaneity.[16] Augmented Dickey–Fuller, KPSS, and Phillips-Perron tests show that all these time series are unit root (with linear trends). For each of those candidate time series, we conduct two-types of the t-tests, $\mathcal{T}_\phi(b)$ in (12) and $\mathcal{T}^0_\phi(b)$ in (14), using the HAR long-run variance estimator with $b = 0.1$ and the Bartlett kernel.

Table 3 reports the test results for potential common long-run trend determinants θ_t of the log of four crime rates: violent crime, assault, homicide, and robbery. Among the candidate determinants, only 'Demog' can be identified as a long-run trend determinant because it shows that δ is significantly different from zero and the trend regression t-statistics $\mathcal{T}_\phi(b)$

[15]The summary statistics, the source of each data and the data details are reported in Table S3 in the online supplement.

[16]As a robustness check, we considered variables lagged by two periods as well, but the findings are in the line with those presented here with one-period lag. The additional results are reported in the online supplement. Also note that choice between non-civilian and civilian police officers does not change the results. Choice among the three types of incarceration rates (i.e., federal, state, and local) does not change the results, either.

Table 3: Long-Run Trend Determinants for Violent Crimes

Crime	θ_t	$\hat{\delta}$	$se(\hat{\delta})$	$\mathcal{T}_\phi(0.1)$	$\mathcal{T}_\phi^0(0.1)$
Violent	Demog	3.601*	0.431	-6.658*	-7.299*
	Police	-0.467	1.224	-2.748*	-4.915*
	Prison	-0.813*	0.227	0.293	0.390
	RGDP	-1.432*	0.171	62.835	6.376
Assault	Demog	3.063*	0.460	-3.423*	-7.089*
	Police	-0.285	1.057	-1.440	-3.681*
	Prison	-0.704*	0.216	0.341	0.578
	RGDP	-1.204*	0.184	34.889	5.367
Homicide	Demog	3.308*	0.466	-7.948*	-3.875*
	Police	-1.548	1.216	-0.262	-3.931*
	Prison	-0.971*	0.212	0.030	0.623
	RGDP	-1.281*	0.224	13.621	2.912
Robbery	Demog	5.498*	0.645	-19.629*	-2.964*
	Police	-0.321	1.864	-2.278*	-8.331*
	Prison	-1.080*	0.323	-0.615	-1.218
	RGDP	-2.258*	0.230	87.087	3.228

Note: (i) $\hat{\delta}$ is the least squares estimate from (10) and $se(\hat{\delta})$ is its standard errors from Phillips and Park (1988). $\mathcal{T}_\phi(0.1)$ and $\mathcal{T}_\phi^0(0.1)$ are respectively the t-ratios defined in (12) and (14) with $b = 0.1$ and the Bartlett kernel. (ii) 'Demog' is the log of the fraction of young adult population between age 10 and 39, 'Police' is the log of the number of non-civilian police officers per capita, 'Prison' is the log of the local incarceration per capita, and 'RGDP' is the log of the Real GDP per capita. (iii) From Definition 1, θ_t becomes a long-run trend determinant if $\widehat{\delta}$ is significantly different from zero and $\mathcal{T}_\phi(0.1) < -2.04$ or $\mathcal{T}_\phi^0(0.1) < -1.96$, where the 5% critical values are from Table 1; only 'Demog' satisfies all the conditions required. (* indicates significance at 5%.)

Table 4: Which Ages Are More Violent?

Age Group	Violent	Assault	Homicide	Robbery
10 to 29	-2.312	-2.135	-1.433	-1.320
10 to 39	-7.299	-7.089	-3.875	-2.964
10 to 49	-4.045	-3.793	-1.078	-3.500
20 to 39	-5.067	-5.755	-5.853	-1.695
20 to 49	-6.379	-5.848	-1.655	-4.464

Note: The values are $\mathcal{T}_\phi^0(0.1)$ with the fraction of population in each age group as θ_t.

and $\mathcal{T}_\phi^0(b)$ are less than the 5% critical values (i.e., $\mathcal{T}_\phi(0.1) < -2.04$ and $\mathcal{T}_\phi^0(0.1) < -1.96$). In comparison, 'Police' shows that δ is insignificant though some trend regression t-statistics $\mathcal{T}_\phi(b)$ or $\mathcal{T}_\phi^0(b)$ are less than the critical values, which hence does not satisfy Definition 2.[17]

Knowing that the fraction of young adult population is a long-run trend determinant of all the violent crime rates, we further investigate which age groups are more responsible for each case in Table 4. The fraction of young adult population of age between 10 and 39 best describes the common long-run trends of assault and overall violent crimes (which includes rape); age between 20 and 39 for homicide; age between 20 and 49 for robbery.

6.2 Property Crimes

For the property crimes, we consider burglary, larceny-theft, and motor vehicle theft, as well as the overall property crime rates. Due to data limitation, arson is excluded, but the overall property crime encompasses all these four types of property crimes. Like the violent crime rates, all these state-level property crime rates are collected from the Uniform Crime Report on FBI Crime Data Explorer from 1987 to 2021 over 50 states in the U.S. We first check if these panel series satisfy the weak $\sigma-$convergence and they are not trend stationary.

[17]We also conducted residual-based cointegration test between the average of each crime rate ($\overline{y}_t$) and the potential determinant θ_t, which cannot reject the null of no cointegration for all the cases. It should be noted that, however, this cointegration test would not perform properly with small T, which is $T = 35$ in this analysis. This comparison shows that our approach can identify time series that is associated with the latent common trend of panel series or their common trend factor, even when the standard cointegration test may not be able to do so. From this perspective, as also discussed in Remark 2, our approach can seen as an alternative to the standard (homogeneous) panel cointegration tests (e.g., Kao (1999)), particularly when T is not large enough.

Table 5: Preliminary Weak $\sigma-$Convergence Tests of Property Crime Rates

	Weak σ-convergence toward:		
	Sample Mean	National Average	Linear Trend
Property	0.33	-0.83	n.a.
Burglary	3.53	1.43	n.a.
Larceny	-1.11	-1.44	n.a.
Motor Vehicle Theft	-1.80	-3.78	2.58

Note: (i) The first two columns report the t-ratio $\mathcal{T}^0_{\psi_K}$ of the weak $\sigma-$convergence test from (5) as in Table 2; $\mathcal{T}^0_{\psi_K} < -1.65$ implies weak $\sigma-$convergence toward the sample mean (1st column) or national average (2nd column). The last column reports the t-ratio of the weak $\sigma-$convergence test toward a linear trend as in Table 2; a value larger than -1.65 implies y_{it} is not trend stationary. (ii) Property crime includes burglary, larceny-theft, motor vehicle theft, and arson. (iii) The sample period is from 1987 to 2021, which matches with the sample period of candidate determinants θ_t that are lagged by one period.

Table 5 reports the t-statistics of ψ_K in the auxiliary trend regression (5) for these four panel time series. Only motor vehicle theft rate satisfies the weak $\sigma-$convergence.[18] This result implies that we need to apply the procedure in Section 5 to first find the subgroup $\mathcal{G}$ from each of the property crime, burglary, and larceny rates panel series that yields distributional convergence.

To this end, we conduct the subgroup selection procedure using $\mathcal{T}_{\varphi_i}(b)$ with $b = 0.1$ as described in Section 5 for each θ_t variable chosen. We use the threshold value $c_1 = -1.2$. We elect the panel series i to be included in the convergent subgroup $\widehat{\mathcal{G}}(\theta)$ if its $\mathcal{T}_{\varphi_i}(b)$ does not exceed this threshold c_1. One interesting finding is that the choice of c_1 does not much affect the final subgroup selection outcomes after the iterations we described in Section 5, as long as it is negative. For the iterations, we stop at the rth round if $|\widehat{\delta}^{(r)} - \widehat{\delta}^{(r-1)}| < 0.001$, where $\widehat{\delta}^{(r)}$ is the estimate of δ in the time series regression of $\bar{y}_t$ on θ_t given in (10), and hence the distance measure $\Delta_{it} = (y_{it} - \widehat{\delta}'\theta_t)^2$ in (25) and the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ no longer change.[19]

[18] The table also suggests that motor vehicle theft rate is not linear trend stationary. We do not consider this trend stationary check for the other variables because they are not weak $\sigma-$convergent and hence our approach cannot be applied. However, we checked their trajectories and concluded they are not trend stationary.

[19] This iteration concludes very quickly within a couple of rounds for all the cases.

Table 6 shows the long-run trend determinant identification results using the final subgroup estimate $\widehat{\mathcal{G}}(\theta)$. We consider the same candidate time series θ_t (i.e., Demog, Police, Prison, and RGDP) as we considered for the violent crime case. We find that only 'Prison' and 'Police' yield large size of the subgroup estimates $\widehat{\mathcal{G}}(\theta)$ (i.e., over 80% of the panel series are included), whereas 'Demog' and 'RGDP' have no such convergent subgroups. For 'Prison', it gives significant δ and satisfies $\mathcal{T}_\phi(0.1) < -2.04$ and $\mathcal{T}^0_\phi(0.1) < -1.96$. Therefore, it can be identified as a strong long-run trend determinant for the property crime, burglary, and larceny. In fact, it is well known in sociology that there is a closed link between incarceration rate and property crime, particularly for burglary crime (e.g., Rosenfeld and Messner (2009); Weatherburn, Hua, and Moffatt (2006)). Interestingly, 'Police' yields both t-ratios are far below the critical values but its δ is not significantly different from zero for all the cases, and hence it cannot be identified as a long-run trend determinant. Unlike the violent crime case, 'Demog' is no longer a long-run trend determinant because $\widehat{\mathcal{G}}(\theta)$ becomes (nearly) empty, showing almost none of the panel series seem to share a long-run trend with this time series.

It is worthy to note that the motor vehicle theft shows quite a different result from the rest of the property crimes in Table 6. As we found in Table 5, it satisfies the weak $\sigma-$convergence (i.e., all the panel series share a long-run common trend) and hence it does not require to go through the subgroup selection. 'Demog' is identified as a long-run trend determinant for the motor vehicle theft rate, whereas 'Prison' is not.[20] Though the motor vehicle theft is not classified into violent crime, its trajectories are very similar to violent crime as already depicted in Figure 1(a). In fact, a motor vehicle theft can be elevated from a misdemeanor to a felony, carrying a potential prison sentence of up to 10 years. Unlike other property crimes, grand theft auto is often linked to its use as a mode of transportation

[20]We use the fraction of young adult population of age between 10 and 49, instead of age between 10 and 39, in this case. Similarly as Table 4, we obtained the t-statistic $\mathcal{T}^0_\phi(0.1)$ of the motor vehicle theft rate for different age groups below; it is found that the age group between 10 and 39 has rather a weak result supporting 'Demog' as a long-run common trend determinant. This is because too young teenagers are less likely to drive and hence less incentive to steal motor vehicles. For this reason, adding more older population increases the explanatory power in this case.

Age Group	10–29	10–39	10–49	20–39	20–49
$\mathcal{T}^0_\phi(0.1)$	-0.202	-2.005	-2.832	-0.918	-4.429

Table 6: Long-run Trend Determinants for Property Crimes

Crime	θ_t	$\hat{\delta}$	$se(\hat{\delta})$	$se^0(\hat{\delta})$	$\mathcal{T}_\phi(0.1)$	$\mathcal{T}^0_\phi(0.1)$	Group
Property	Demog	n.a.	n.a.	n.a.	n.a.	n.a.	0
	Police	0.300	1.400	1.680	-19.276*	-24.490*	50
	Prison	-0.850*	0.220	0.350	-3.087*	-3.581*	40
	RGDP	-2.620*	0.170	0.180	-1.745	-1.991*	1
Burglary	Demog	3.210*	0.480	0.530	-10.056*	-10.701*	1
	Police	0.250	1.990	2.250	-12.450*	-11.397*	49
	Prison	-1.080*	0.310	0.460	-3.077*	-3.360*	42
	RGDP	-3.810*	0.280	0.300	-2.082*	-1.848*	1
Larceny	Demog	n.a.	n.a.	n.a.	n.a.	n.a.	0
	Police	0.360	1.220	1.510	-15.541*	-20.334*	50
	Prison	-0.760*	0.200	–	-2.993*	–	40
		-0.760*	–	0.310	–	-3.596*	41
	RGDP	-2.180*	0.170	0.140	-1.419	-1.493	1
Motor Vehicle Theft	Demog	5.816*	0.587	0.576	-15.308*	-2.832*	
	Police	-0.107	2.166	3.390	-3.275*	-6.816*	
	Prison	-1.247*	0.340	0.429	-1.283	-1.248	
	RGDP	-2.270*	0.229	0.300	51.261	1.419	

Note: (i) 'Group' is the number of states selected in the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ using $\mathcal{T}_{\varphi_i}(0.1)$ and $\mathcal{T}^0_{\varphi_i}(0.1)$. When they are different, each of $\hat{\delta}$ and $\mathcal{T}_\phi(0.1)$ are reported in separate lines. (ii) $\hat{\delta}$ is the least squares estimate from (10), and $se(\hat{\delta})$ and $se^0(\hat{\delta})$ are the standard error from Phillips and Park (1988) using $\mathcal{T}_{\varphi_i}(0.1)$ and $\mathcal{T}^0_{\varphi_i}(0.1)$, respectively. $\mathcal{T}_\phi(0.1)$ and $\mathcal{T}^0_\phi(0.1)$ are the t-ratios defined in (12) and (14) with $b = 0.1$ and the Bartlett kernel. When $\widehat{\mathcal{G}}(\theta)$ is empty, $\hat{\delta}$ and $\mathcal{T}_\phi(b)$ cannot be obtained and marked as 'n.a.'. (iii) 'Demog' is the log of the fraction of young adult population between age 10 and 39, 'Police' is the log of the number of non-civilian police officers per capita, 'Prison' is the log of the local incarceration per capita, and 'RGDP' is the log of the Real GDP per capita. (iv) From Definition 1, θ_t becomes a long-run trend determinant if $\hat{\delta}$ is significantly different from zero and $\mathcal{T}_\phi(0.1) < -2.04$ or $\mathcal{T}^0_\phi(0.1) < -1.96$. Furthermore, the size of the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ should be large enough to include most of the states. For 'Property', 'Burglary', and 'Larceny', only 'Prison' satisfies all these three conditions. (v) 'Motor Vehicle Theft' satisfies the weak σ-convergence and hence does not require to get the subgroup, so no group size is given. For this case, 'Demog' is identified as a long-run trend determinant, which is defined as the population fraction of age from 10 to 49. (* indicates significance at 5%.)

for other offenses or as a tool in violent crimes.

7 Concluding Remarks

We develop a novel method to identify common trend determinants in nonstationary panel data. Unlike the two-way fixed effects analysis, which eliminates time effects, our method directly analyzes the underlying common trends. This approach is particularly valuable when researchers seek to understand the underlying observed factors that drive the shared latent stochastic trend among panel series. This approach also sheds new light on cointegration between panel data and time series, emphasizing the needs for analysis of relative variation between panel data and the cointegration error. We leverage the concept of distributional convergence (i.e., weak $\sigma-$convergence) to formalize this idea. The key advantages of our approach are applicability to relatively short panel datasets and circumvention of the need to estimate latent common factors.

Our application to crime rates demonstrates that the percentage of young adults significantly influences violent crime trends, while incarceration rates drive property crime trends. These findings differ from the standard two-way fixed effect analysis, which often highlights factors like police numbers and income levels. Interestingly, research by Farrell, Tilley, and Tseloni (2014) aligns with our findings, suggesting a connection between declining violent crime rates in Canada and the U.K. with a decrease in the young adult population. Additionally, our analysis reveals a similar trend between motor vehicle theft and violent crime, distinct from property crime. These insights can open doors for novel policy considerations in crime control.

References

Bai, J., and S. Ng (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131(1-2), 507–537.

Chang, Y., J. Y. Park, and P. C. B. Phillips (2001): "Nonlinear econometric models with cointegrated and deterministically trending regressors," *The Econometrics Journal*, 4(1), 1–36.

Chen, N.-F., R. Roll, and S. A. Ross (1986): "Economic Forces and the Stock Market," *The Journal of Business*, 59(3), 383–403.

Farrell, G., N. Tilley, and A. Tseloni (2014): "Why the crime drop?," *Crime and justice*, 43(1), 421–490.

Kao, C. (1999): "Spurious regression and residual-based tests for cointegration in panel data," *Journal of Econometrics*, 90(1), 1–44.

Kiefer, N. M., and T. J. Vogelsang (2005): "A new asymptotic theory for heteroskedasticity-autocorrelation robust tests," *Econometric Theory*, 21(06), 1130–1164.

Kim, J.-Y. (2003): "Inference on Segmented Cointegration," *Econometric Theory*, 19(4), 620–639.

Kong, J., P. C. B. Phillips, and D. Sul (2019): "Weak σ-convergence: Theory and applications," *Journal of Econometrics*, 209(2), 185–207.

——— (2020): "Testing convergence using HAR inference," *Advances in Econometrics*, 41, 25–72.

Lee, Y., and D. Sul (2023a): "Depth-weighted Forecast Combination: Application to COVID-19 Cases," *Advances in Econometrics*, 45B, 235–260.

——— (2023b): "Depth-weighted means of noisy data: An application to estimating the average effect in heterogeneous panels," *Journal of Multivariate Analysis*, 196, 105–165.

Levitt, S. D. (2004): "Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not," *Journal of Economic Perspectives*, 18(1), 163–190.

MacNeill, I. B. (1978): "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times," *The Annals of Statistics*, 6(2), 422–433.

MOODY, C. E., AND T. B. MARVELL (2010): "On the choice of control variables in the crime equation," *Oxford Bulletin of Economics and Statistics*, 72(5), 696–715.

MUNDLAK, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85.

PARK, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60(1), 119.

PARK, J. Y., AND P. C. B. PHILLIPS (1988): "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4(3), 468–497.

PARKER, J., AND D. SUL (2016): "Identification of unknown common factors: Leaders and followers," *Journal of Business & Economic Statistics*, 34(2), 227–239.

PESARAN, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *The Review of Economic Studies*, 57(1), 99–125.

PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1111.

PHILLIPS, P. C. B., AND J. Y. PARK (1988): "Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors," *Journal of the American Statistical Association*, 83(401), 111–115.

PHILLIPS, P. C. B., AND D. SUL (2007): "Transition modeling and econometric convergence tests," *Econometrica*, 75(6), 1771–1855.

ROSENFELD, R., AND S. F. MESSNER (2009): "The crime drop in comparative perspective: the impact of the economy and imprisonment on American and European burglary rates," *The British Journal of Sociology*, 60(3), 445–471.

SUN, Y. (2004): "A Convergent t-Statistic in Spurious Regressions," *Econometric Theory*, 20(05), 943–962.

WEATHERBURN, D., J. HUA, AND S. MOFFATT (2006): "How Much Crime Does Prison Stop-The Incapacitation Effect of Prison on Burglary," *IJPS*, 2, 8.

Appendix

A Summary of Procedures

We outline the procedures we implemented in the empirical analysis.

A.1 Weak $\sigma-$Convergence of x^*_{it}

The key presumption of our procedure is that the idiosyncratic term x^*_{it} in

$$y_{it} = \alpha_i + \tau_t + x^*_{it}$$

satisfies the weak $\sigma-$convergence (toward its mean). As discussed in Remark 1, this can be done by studying the weak $\sigma-$convergence of y_{it} by Kong, Phillips, and Sul (2019).

1. Obtain $R_{n,t} = n^{-1}\sum_{i=1}^{n}(y_{it} - \overline{y}_t)^2$, where $\overline{y}_t = n^{-1}\sum_{i=1}^{n} y_{it}$.

2. Run a trend regression:
$$R_{n,t} = \psi_{K0} + \psi_K t + u_{K,t}$$
as given in (5) and obtain the t-statistic of ψ_K as
$$\mathcal{T}^0_{\psi_K} = \frac{\widehat{\psi}_K}{\left(\widehat{\omega}^2_{u_K} / \sum_{t=1}^{T}(\widetilde{t})^2\right)^{1/2}},$$
where $\widehat{\omega}^2_{u_K}$ is a long-run variance estimator of $u_{K,t}$ and $\widetilde{t} = t - T^{-1}\sum_{r=1}^{T} r$.

3. If $\mathcal{T}^0_{\psi_K}$ is less than a critical value (e.g., -1.65 for 5% test when $\widehat{\omega}^2_{u_K}$ is a HAC estimator), then we conclude y_{it} satisfies the weak $\sigma-$convergence and hence so does x^*_{it}. (More critical values are available in Kong, Phillips, and Sul (2019).)

Remark A (Trend Stationarity) When T is too small and n is large, the standard panel unit root test would not perform properly. In such a case, we could see if the panel process y_{it} satisfies the weak $\sigma-$convergence toward some deterministic trend function, such as a linear trend. If this weak $\sigma-$convergence is rejected, then it gives a supporting evidence that y_{it} is not (linear) trend stationary in the long run. We can conduct this test using $R_{n,t} = n^{-1}\sum_{i=1}^{n}(y_{it} - \widehat{\alpha}_0 - \widehat{\delta}t)^2$, where $(\widehat{\alpha}_0, \widehat{\delta})$ is from the trend regression: $\overline{y}_t = \alpha_0 + \delta t + e_t$. Then follow the steps 2 and 3 as above.

A.2 Long-Run Trend Determinant

When x_{it}^* satisfies the weak $\sigma-$convergence from Section A.1, we can tell if a candidate time series θ_t is a long-run trend determinant as follow. For simplicity, we employ the t-statistic $\mathcal{T}_\phi^0(b)$ given in (14).

1. Obtain $S_{n,t} = n^{-1}\sum_{i=1}^{n}(y_{it} - \widehat{\delta}\theta_t)^2$, where $\widehat{\delta}$ is from the time series regression: $\overline{y}_t = \alpha_0 + \delta\theta_t + e_t$.

2. Run a trend regression:
$$S_{n,t} = \phi_0 + \phi t + u_t$$
as given in (11) and obtain the t-statistic of ϕ as (using the Bartlett kernel)
$$\mathcal{T}_\phi^0(b) = \frac{\widehat{\phi}}{\left(\widehat{\Omega}_u(b)/\sum_{t=1}^{T}(\widetilde{t})^2\right)^{1/2}}$$
where
$$\widehat{\Omega}_u(b) = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 + \frac{2}{T}\sum_{\ell=1}^{L}\sum_{t=1}^{T-\ell}\left(1 - \frac{\ell}{L+1}\right)\widehat{u}_t\widehat{u}_{t+\ell}$$
with $L = [bT]$ for some fixed-b coefficient $b \in (0,1)$, where $\widehat{u}_t = S_{n,t} - \widehat{\phi}_0 - \widehat{\phi}t$ and $[c]$ is the largest integer smaller than or equal to c.

3. If $\mathcal{T}_\phi^0(b)$ is less than a critical value (e.g., -1.96 for 5% test with $b = 0.1$), then we conclude θ_t as a long-run (convergent) trend determinant of y_{it}. More critical values are available in Tables 7 and 8 at the end of the Appndix.

A.3 Convergent Subgroup

When x_{it}^* does not satisfy the weak $\sigma-$convergence from Section A.1, we need to estimate the convergent subgroup $\mathcal{G}(\theta)$ for a given θ_t. If this subgroup estimate $\widehat{\mathcal{G}}(\theta)$ is majority of the entire panel sample, then we check if θ_t is indeed a long-run trend determinant within this estimated subgroup by implementing the steps in A.2 above only using the panel series $i \in \widehat{\mathcal{G}}(\theta)$. For simplicity, we here employ the simplified t-statistic, denoting $\mathcal{T}^0_{\varphi_i}(b)$, similarly as $\mathcal{T}^0_{\phi}(b)$ given in (14).

1. Given θ_t, obtain $\Delta_{it} = (y_{it} - \widehat{\delta}\theta_t)^2$ for each i, where $\widehat{\delta}$ is from the time series regression: $\overline{y}_t = \alpha_0 + \delta\theta_t + e_t$.

2. For each i, run a trend regression:

$$\Delta_{it} = \varphi_{0i} + \varphi_i t + u_{it}$$

   as given in (24) and obtain the t-statistic of φ_i as

$$\mathcal{T}^0_{\varphi_i}(b) = \frac{\widehat{\varphi}_i}{\left(\widehat{\Omega}_{u,i}(b)/\sum_{t=1}^{T}(\widetilde{t})^2\right)^{1/2}},$$

   where the HAR long-run variance estimator $\widehat{\Omega}_{u,i}(b)$ is obtained as $\widehat{\Omega}_u(b)$ above by replacing $\widehat{u}_t$ with $\widehat{u}_{it} = \Delta_{it} - \widehat{\varphi}_{0i} - \widehat{\varphi}_i t$.

3. If $\mathcal{T}^0_{\varphi_i}(b)$ is less than a threshold $c_1 < 0$ (e.g., we chose $c_1 = -1.2$ in the empirical study), then we conclude this member i belongs to the subgroup estimate. Do this step for all i to get the subgroup estimate $\widehat{\mathcal{G}}^{(1)}(\theta)$.

4. Update $\overline{y}_t$ in step 1 as $\overline{y}_t^{(1)} = |\widehat{\mathcal{G}}^{(1)}(\theta)|^{-1}\sum_{i\in\widehat{\mathcal{G}}^{(1)}(\theta)} y_{it}$ and obtain $\widehat{\delta}^{(1)}$ from $\overline{y}_t^{(1)} = \alpha_0^{(1)} + \delta^{(1)}\theta_t + e_t^{(1)}$.

5. Update $\Delta_{it}^{(1)} = (y_{it} - \widehat{\delta}^{(1)}\theta_t)^2$ for each $i \in \widehat{\mathcal{G}}^{(1)}(\theta)$ and repeat the steps 2 to 4 until the subgroup membership is not further updated or $|\widehat{\delta}^{(r)} - \widehat{\delta}^{(r-1)}|$ falls below some threshold at the rth iteration. This resulting subgroup yields $\widehat{\mathcal{G}}(\theta)$.

B Proof of Main Theorems

We assume the following conditions. '$\Rightarrow$' denotes weak convergence of the associated probability measures and $[c]$ is the largest integer smaller than or equal to c. The proofs of all the technical lemmas are provided in the online supplement.

Assumption 1

(i) $(\alpha_i, \mu_i)'$ is i.i.d. with mean zero and finite second moment, satisfying $\mathbb{E}\alpha_i\mu_i = 0$. Let $\sigma_\mu^2 = \mathbb{E}\mu_i^2$.

(ii) $(\varepsilon_{it}, \epsilon_{it})'$ is independent across i with mean zero and uniformly finite fourth moment, satisfying $\mathbb{E}\varepsilon_{it}\epsilon_{it} = 0$. Let $\sigma_{\varepsilon,i}^2 = \mathbb{E}\varepsilon_{it}^2$ and $\sigma_\varepsilon^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_{\varepsilon,i}^2 < \infty$.

(iii) For each i, $J_{it} = \left(\varepsilon_{it}, \epsilon_{it}, , \varepsilon_{it}\epsilon_{it}, \varepsilon_{it}^2 - \sigma_{\varepsilon,i}^2\right)'$ satisfies a multivariate invariance principle: $T^{-1/2}\sum_{t=1}^{[Tr]} J_{it} \Rightarrow B_i^*(r)$ as $T \to \infty$ for $r \in [0,1]$, where $B_i^*(\cdot)$ is 4×1 vector Brownian motion.

(iv) The elements of θ_t do not have cointegration among them.

(v) $\kappa_1, \kappa_2 \in (0, 1/2)$.

(vi) $n/T \to \infty$ as $(n,T) \to \infty$.

(vii) The kernel function $K : \mathbb{R} \mapsto [0,1]$ satisfies $K(0) = 1$, $K(-\nu) = K(\nu)$, $\int K(\nu)\, d\nu = 1$, and $\int K^2(\nu)\, d\nu < \infty$.

Assumption 2 When $\xi_t \sim I(0)$, $(\xi_t, \Delta\theta_t')'$ is mean zero and $\sigma_\xi^2 = \mathbb{E}\xi_t^2 < \infty$. Moreover, $J_{0,t} = \left(\xi_t, \xi_t^2 - \sigma_\xi^2, \Delta\theta_t'\right)'$ satisfies a multivariate invariance principle: $T^{-1/2}\sum_{t=1}^{[Tr]} J_{0,t} \Rightarrow B_0^*(r) = (B_\xi(r), B_{\xi\xi}(r), B_\theta'(r))'$ as $T \to \infty$ for $r \in [0,1]$, where $B_0^*(\cdot)$ is $(2+m) \times 1$ vector Brownian motion with covariance matrix

$$\Omega_0 = \sum_{j=-\infty}^{\infty} \mathbb{E}(J_{0,t}J_{0,t+j}') = \begin{pmatrix} \omega_\xi^2 & \omega_{\xi,\xi\xi} & \Omega_{\theta\xi}' \\ \omega_{\xi,\xi\xi} & \omega_{\xi\xi}^2 & \Omega_{\theta\xi\xi}' \\ \Omega_{\theta\xi} & \Omega_{\theta\xi\xi} & \Omega_\theta \end{pmatrix} < \infty,$$

that is uncorrelated with $B_i^*(\cdot)$ defined in Assumption 1.

Assumption 3 When $\xi_t \sim I(1)$, $J_{1,t} = (\Delta\xi_t, \Delta\theta_t')'$ is mean zero and satisfies a multivariate invariance principle: $T^{-1/2}\sum_{t=1}^{[Tr]} J_{1,t} \Rightarrow B_1^*(r) = (B_\xi(r), B_\theta'(r))'$ as $T \to \infty$ for $r \in [0,1]$, where $B_1^*(\cdot)$ is $(1+m)\times 1$ vector Brownian motion with covariance matrix

$$\Omega_1 = \sum_{j=-\infty}^{\infty} \mathbb{E}(J_{1,t}J_{1,t+j}') = \begin{pmatrix} \omega_\xi^2 & \Omega_{\theta\xi}' \\ \Omega_{\theta\xi} & \Omega_\theta \end{pmatrix} < \infty,$$

that is uncorrelated with $B_i^*(\cdot)$ defined in Assumption 1.

For any process z_t in discrete time $t = 1, \dots, T$, we denote the demeaned process as $\widetilde{z}_t = z_t - T^{-1}\sum_{s=1}^{T} z_s$. Similarly, for any process $z(r)$ in continuous time $r \in [0,1]$, we denote the demeaned process as $\widetilde{z}(r) = z(r) - \int_0^1 z(s)\,ds$.

Lemma B1 Let $\widehat{\delta}$ be the least squares estimator of δ in (10). Under Assumptions 1-3, as $n, T \to \infty$,

$$\begin{cases} T(\widehat{\delta}-\delta) \Rightarrow \left(\int_0^1 \widetilde{B}_\theta(r)\widetilde{B}_\theta(r)'\,dr\right)^{-1}\int_0^1 \widetilde{B}_\theta(r)\,dB_\xi(r) & \text{if } \xi_t \sim I(0) \\ \widehat{\delta}-\delta \Rightarrow \left(\int_0^1 \widetilde{B}_\theta(r)\widetilde{B}_\theta(r)'\,dr\right)^{-1}\int_0^1 \widetilde{B}_\theta(r)\,B_\xi(r)\,dr & \text{if } \xi_t \sim I(1). \end{cases}$$

Define the partial sum process

$$Z_{nT}(r) = \sum_{t=1}^{[Tr]} \widetilde{t}\widetilde{S}_{n,t}$$

for $r \in [0,1]$ and

$$V(r) = \left\{ W_1(r) - \int_0^1 W_1(s)\widetilde{W}_m(s)'\,ds \left(\int_0^1 \widetilde{W}_m(s)\widetilde{W}_m(s)'\,ds\right)^{-1} W_m(r) \right\}^2, \tag{B.1}$$

where W_1 and W_m are standard vector Brownian motions such that $B_\xi(\cdot) = \omega_\xi W_1(\cdot)$ and $B_\theta(\cdot) = \Omega_\theta^{1/2} W_m(\cdot)$.

Lemma B2 Suppose $\xi_t \sim I(1)$ and Assumptions 1 and 3 hold. As $n, T \to \infty$,

$$T^{-3} Z_{nT}(r) \Rightarrow \omega_\xi^2 \int_0^r \widetilde{s}\widetilde{V}(s)\,ds$$

for $r \in [0,1]$, where $\omega_\xi^2 = \sum_{j=-\infty}^{\infty} \mathbb{E}(\xi_t \xi_{t+j})$.

Lemma B3 Suppose $\xi_t \sim I(0)$ and Assumptions 1 and 2 hold. Define $\mathcal{B}(r)$ as the Brownian bridge and

$$q(\kappa; r) = \int_0^r \left(s - \frac{1}{2}\right)\left(s^{-2\kappa} - \frac{1}{1-2\kappa}\right) ds$$

for $0 < \kappa < 1/2$. Also let $\kappa_* = \min\{\kappa_1, \kappa_2\}$. As $n, T \to \infty$, the following results hold.

(i) When $\kappa_* > 1/4$,

$$T^{-3/2} Z_{nT}(r) \Rightarrow \omega_{\xi\xi} \int_0^r \widetilde{s} d\mathcal{B}(s),$$

where $\omega_{\xi\xi}^2 = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(\xi_t^2 - \sigma_\xi^2\right)\left(\xi_{t+j}^2 - \sigma_\xi^2\right)$ and $\sigma_\xi^2 = \mathbb{E}\xi_t^2$.

(ii) When $\kappa_* = 1/4$,

$$T^{-3/2} Z_{nT}(r) \Rightarrow \omega_{\xi\xi} \int_0^r \widetilde{s} d\mathcal{B}(s) + \begin{cases} q(1/4; r)\,\sigma_\mu^2 & \text{if } \kappa_1 < \kappa_2 \\ q(1/4; r)\,\sigma_\varepsilon^2 & \text{if } \kappa_1 > \kappa_2 \\ q(1/4; r)\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if } \kappa_1 = \kappa_2. \end{cases}$$

(iii) When $\kappa_* < 1/4$,

$$T^{-(2-2\kappa_*)} Z_{nT}(r) \xrightarrow{p} \begin{cases} q(\kappa_*; r)\,\sigma_\mu^2 & \text{if } \kappa_1 < \kappa_2 \\ q(\kappa_*; r)\,\sigma_\varepsilon^2 & \text{if } \kappa_1 > \kappa_2 \\ q(\kappa_*; r)\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if } \kappa_1 = \kappa_2. \end{cases}$$

Lemma B4 Let $\widehat{\phi}$ be the least squares estimator of ϕ in (11) and Assumptions 1-3 hold. As $n, T \to \infty$, the followings hold.

(i) Suppose $\xi_t \sim I(1)$. Then,

$$\widehat{\phi} \Rightarrow 12\omega_\xi^2 \int_0^1 \widetilde{s}\widetilde{V}(s)\, ds.$$

(ii) Suppose $\xi_t \sim I(0)$ and let $\kappa_* = \min\{\kappa_1, \kappa_2\}$. Then,

$$\begin{cases} T^{3/2}\widehat{\phi} \Rightarrow 12\omega_{\xi\xi} \int_0^1 \widetilde{s} d\mathcal{B}(s) \sim \mathcal{N}\left(0, 12\omega_{\xi\xi}^2\right) & \text{when } \kappa_* > 1/4 \\ T^{3/2}\widehat{\phi} \Rightarrow 12\omega_{\xi\xi} \int_0^1 \widetilde{s} d\mathcal{B}(s) - \beta_\phi \sim \mathcal{N}\left(-\beta_\phi, 12\omega_{\xi\xi}^2\right) & \text{when } \kappa_* = 1/4 \\ T^{1+2\kappa_*}\widehat{\phi} \xrightarrow{p} -\beta_\phi^* & \text{when } \kappa_* < 1/4, \end{cases}$$

where

$$\beta_\phi = \begin{cases} 4\sigma_\mu^2 & \text{if } \kappa_1 < \kappa_2 \\ 4\sigma_\varepsilon^2 & \text{if } \kappa_1 > \kappa_2 \\ 4\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if } \kappa_1 = \kappa_2 \end{cases}$$

and

$$\beta_\phi^* = \begin{cases} \frac{6\kappa_*}{(1-\kappa_*)(1-2\kappa_*)}\sigma_\mu^2 & \text{if } \kappa_1 < \kappa_2 \\ \frac{6\kappa_*}{(1-\kappa_*)(1-2\kappa_*)}\sigma_\varepsilon^2 & \text{if } \kappa_1 > \kappa_2 \\ \frac{6\kappa_*}{(1-\kappa_*)(1-2\kappa_*)}\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if } \kappa_1 = \kappa_2. \end{cases}$$

Define

$$\Psi_{nT}(b) = \sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{t-s}{Tb}\right)\left(\widetilde{t}\widehat{u}_t\right)\left(\widetilde{s}\widehat{u}_s\right) \tag{B.2}$$

for some $b \in (0,1]$ and a kernel function $K(\cdot)$ given in Assumption 1, where $\widehat{u}_t = \widetilde{S}_{n,t} - \widetilde{\widehat{\phi} t}$ is the regression residual in (11).

Lemma B5 Suppose $\xi_t \sim I(1)$ and Assumptions 1 and 3 hold. As $n, T \to \infty$,

$$T^{-6}\Psi_{nT}(b) \Rightarrow \omega_\xi^4 \int_0^1\int_0^1 K\left(\frac{r-s}{b}\right)\widetilde{r}\widetilde{s}V^\tau(r)\,V^\tau(s)\,drds,$$

where

$$V^\tau(r) = \widetilde{V}(r) - \widetilde{r}\left(\int_0^1 (\widetilde{\nu})^2 d\nu\right)^{-1}\int_0^1 \widetilde{\nu}\widetilde{V}(\nu)\,d\nu \tag{B.3}$$

and $\widetilde{V}(r)$ is the demeaned $V(r)$ in (B.1).

Lemma B6 Suppose $\xi_t \sim I(0)$ and Assumptions 1 and 2 hold. Let $\kappa_* = \min\{\kappa_1, \kappa_2\}$. As $n, T \to \infty$, the following results hold.

(i) When $\kappa_* > 1/4$,

$$T^{-3}\Psi_{nT}(b) \Rightarrow \omega_{\xi\xi}^2 \int_0^1\int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{r}\widetilde{s}dW^\tau(r)\,dW^\tau(s),$$

where $W^\tau(r)$ is the second-level Brownian bridge.

(ii) When $\kappa_* = 1/4$,

$$T^{-3}\Psi_{nT}(b) \Rightarrow \omega_{\xi\xi}^2 \int_0^1 \int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{rs}\left\{dW^{\tau}(r) + \frac{\lambda(r)}{\omega_{\xi\xi}}dr\right\}\left\{dW^{\tau}(s) + \frac{\lambda(s)}{\omega_{\xi\xi}}ds\right\},$$

where

$$\lambda(r) = \begin{cases} \left(4r + r^{-1/2} - 4\right)\sigma_\mu^2 & \text{if } \kappa_1 < \kappa_2 \\ \left(4r + r^{-1/2} - 4\right)\sigma_\varepsilon^2 & \text{if } \kappa_1 > \kappa_2 \\ \left(4r + r^{-1/2} - 4\right)\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if } \kappa_1 = \kappa_2. \end{cases}$$

(iii) When $\kappa_* < 1/4$,

$$T^{-(4-4\kappa_*)}\Psi_{nT}(b) \xrightarrow{p} \int_0^1 \int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{rs}\lambda^*(r)\lambda^*(s)\,drds,$$

where

$$\lambda^*(r) = \begin{cases} c(\kappa_*; r)\sigma_\mu^2 & \text{if } \kappa_1 < \kappa_2 \\ c(\kappa_*; r)\sigma_\varepsilon^2 & \text{if } \kappa_1 > \kappa_2 \\ c(\kappa_*; r)\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if } \kappa_1 = \kappa_2 \end{cases}$$

with

$$c(\kappa; r) = r^{-2\kappa} + \frac{6\kappa r - (1 + 2\kappa)}{(1-\kappa)(1-2\kappa)}.$$

Proof of Theorem 1 Note that, for some fixed-b coefficient $b \in (0, 1]$,

$$T\widehat{\Omega}(b) = \sum_{\ell=-(T-1)}^{T-1} K\left(\frac{\ell}{Tb}\right)T\widehat{\Gamma}_\ell = \sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{t-s}{Tb}\right)\varkappa_t\varkappa_s = \Psi_{nT}(b)$$

in (B.2), where $\varkappa_t = \widehat{u}_t\widetilde{t}$. Then, when $\xi_t \sim I(1)$, by Lemmas B4-(i) and B5,

$$\begin{aligned} \mathcal{T}_\phi(b) &= \frac{\widehat{\phi}}{\left\{\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1} T^{-6}\Psi_{nT}(b)\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}\right\}^{1/2}} \\ &\Rightarrow \frac{12\omega_\xi^2 \int_0^1 \widetilde{s}\widetilde{V}(s)\,ds}{\left\{(1/12)^{-1}\omega_\xi^4 \int_0^1\int_0^1 K\left(\frac{r-s}{b}\right)\widetilde{rs}V^{\tau}(r)V^{\tau}(s)\,drds\,(1/12)^{-1}\right\}^{1/2}} \end{aligned}$$

$$= \frac{\int_0^1 \widetilde{s}\widetilde{V}(s)\,ds}{\left\{\int_0^1\int_0^1 K\left(\frac{r-s}{b}\right)\widetilde{r}\widetilde{s}V^{\tau}(r)\,V^{\tau}(s)\,drds\right\}^{1/2}}, \tag{B.4}$$

where V^{τ} is defined in (B.3). When $\xi_t \sim I(0)$ and $\min\{\kappa_1, \kappa_2\} > 1/4$, Lemmas B4-(ii) and B6-(i) yield

$$\begin{aligned}\mathcal{T}_{\phi}(b) &= \frac{T^{3/2}\widehat{\phi}}{\left\{\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}T^{-3}\Psi_{nT}(b)\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}\right\}^{1/2}}\\ &\Rightarrow \frac{\mathcal{N}\left(0, 12\omega_{\xi\xi}^2\right)}{\left\{(1/12)^{-1}\omega_{\xi\xi}^2\int_0^1\int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{r}\widetilde{s}dW^{\tau}(r)\,dW^{\tau}(s)\,(1/12)^{-1}\right\}^{1/2}}\\ &= \frac{\mathcal{N}(0,1)}{\left\{12\int_0^1\int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{r}\widetilde{s}dW^{\tau}(r)\,dW^{\tau}(s)\right\}^{1/2}},\end{aligned}$$

where $W^{\tau}(r)$ is the second-level Brownian bridge. $\square$

Proof of Theorem 2 When $\xi_t \sim I(0)$ and $\min\{\kappa_1, \kappa_2\} = 1/4$, by Lemmas B4-(ii) and B6-(ii), we have

$$\begin{aligned}\mathcal{T}_{\phi}(b) &= \frac{T^{3/2}\widehat{\phi}}{\left\{\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}T^{-3}\Psi_{nT}(b)\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}\right\}^{1/2}}\\ &\Rightarrow \frac{\mathcal{N}\left(0, 12\omega_{\xi\xi}^2\right) - \beta_{\phi}}{\left\{(1/12)^{-1}\omega_{\xi\xi}^2\int_0^1\int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{r}\widetilde{s}\left\{dW^{\tau}(r) + \frac{\lambda(r)}{\omega_{\xi\xi}}dr\right\}\left\{dW^{\tau}(s) + \frac{\lambda(s)}{\omega_{\xi\xi}}ds\right\}(1/12)^{-1}\right\}^{1/2}},\end{aligned}$$

which yields the desired result by multiplying $(12\omega_{\xi\xi}^2)^{-1/2}$ to both the numerator and the denominator.

When $\kappa_* = \min\{\kappa_1, \kappa_2\} < 1/4$, Lemmas B4-(ii) and B6-(iii) yield

$$\mathcal{T}_{\phi}(b) = \frac{T^{1+2\kappa_*}\widehat{\phi}}{\left\{\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}T^{-(4-4\kappa_*)}\Psi_{nT}(b)\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}\right\}^{1/2}}$$

Table 7: One-Sided Asymptotic Critical Values (Bartlett kernel): Heteroskedastic Case

$b=$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1%	-3.037	-3.758	-4.350	-4.861	-5.391	-5.838	-6.280	-6.641	-6.891	-7.220
2.5%	-2.488	-3.045	-3.500	-3.895	-4.286	-4.622	-4.942	-5.227	-5.423	-5.682
5%	-2.040	-2.467	-2.826	-3.135	-3.429	-3.679	-3.918	-4.131	-4.289	-4.493
10%	-1.554	-1.861	-2.117	-2.340	-2.543	-2.710	-2.866	-3.013	-3.133	-3.284
20%	-0.999	-1.181	-1.336	-1.472	-1.591	-1.683	-1.767	-1.847	-1.923	-2.016

Note: The values are the simulated percentiles of the limiting distribution $F_0(b)$ in (19) of $\mathcal{T}_\phi(b)$ with the Bartlett kernel, which allows for heteroskedasticity.

Table 8: One-Sided Asymptotic Critical Values (Bartlett kernel): Homoskedastic Case

$b=$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1%	-2.914	-3.598	-4.268	-4.988	-5.540	-6.087	-6.596	-7.046	-7.579	-8.020
2.5%	-2.385	-2.890	-3.407	-3.974	-4.428	-4.872	-5.301	-5.685	-6.111	-6.467
5%	-1.961	-2.340	-2.735	-3.181	-3.556	-3.921	-4.279	-4.608	-4.950	-5.238
10%	-1.501	-1.759	-2.035	-2.354	-2.639	-2.924	-3.206	-3.463	-3.721	-3.935
20%	-0.968	-1.117	-1.278	-1.469	-1.650	-1.836	-2.021	-2.193	-2.356	-2.491

Note: The values are the simulated percentiles of the limiting distribution $F_0^0(b)$ in (20) of $\mathcal{T}_\phi^0(b)$ with the Bartlett kernel, under the homoskedasticity restriction.

$$\xrightarrow{p} \frac{-\beta_\phi^*}{\left\{(1/12)^{-1}\int_0^1\int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{rs}\lambda^*(r)\,\lambda^*(s)\,drds\,(1/12)^{-1}\right\}^{1/2}}$$

and the desired result follows by multiplying $(1-\kappa_*)(1-2\kappa_*)/12$ to both the numerator and the denominator, where either σ_μ^2 or σ_ε^2 are all canceled out. $\square$

C Asymptotic Critical Values

Tables 7 and 8 provide one-sided asymptotic critical values for $b \in (0, 1]$, which adds to Table 1. The values are the simulated percentiles of the limiting distribution of $\mathcal{T}_\phi(b)$ and $\mathcal{T}_\phi^0(b)$ given in (19) and (20), respectively, from 2 million replications. The Brownian motion is approximated by normalized sums of standard normal random variables using 10,000 steps and the Bartlett kernel is used for HAR estimation.