IDENTIFYING COMMON TREND DETERMINANTS IN PANEL DATA

By

Yoonseok Lee, Peter C. B. Phillips, Suyong Song and Donggyu Sul

March 2026

COWLES FOUNDATION DISCUSSION PAPER NO. 2504

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Identifying Common Trend Determinants in Panel Data*

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

March 10, 2026

Abstract

This paper develops a novel method for identifying observable determinants of latent common trends in nonstationary panel data, which are typically removed or controlled in two-way fixed effects regressions. By examining cross sectional dispersion processes, we assess whether panel series exhibit distributional convergence toward specific observed time series, revealing them as long run determinants of the underlying latent trend. The approach 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 demonstrates 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 standard two-way fixed effects analysis results, 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 a Kelly Fellowship at the University of Auckland and a KLC Fellowship at Singapore Management University.

†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. Research interest often lies in estimating the marginal effects of control variables or predictors, and traditional panel data regression analysis relies on two-way fixed effect (TWFE) models, where such a common trend is regarded as a nuisance parameter that can be controlled by the time fixed effect. For this reason, when the research question centers on identifying the drivers of the temporal evolution (i.e., the leading trend itself) of the dependent variable, the use of TWFE fails to address the issue or answer the relevant question. Instead, a time fixed effect controls for leading trends in both 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.

The present paper focuses on discovering the latent dynamics in panel data and begins with the question: "what are the limitations of TWFE regression in such analysis?" We investigate the unobserved common trends among panel data series, seeking to unveil the underlying drivers of leading dynamics that are often neglected in standard TWFE analysis. In particular, we propose a novel regression based method to verify if observed variables determine the underlying common trend of a large number of panel series, especially for nonstationary micro panel data where the underlying stochastic trend is often hard to identify. This approach avoids estimating the latent trend or latent nonstationary factors and is computationally easy to implement. In these ways the method is distinct from Bai and Ng (2006) and Parker and Sul (2016) which dealt with stationary panel data and employed common factor estimation.

The key insight in the approach is the observation that many panel data series (or a subgroup of them), even those with stochastic trends, tend to show convergent patterns towards some common time path in the long run. This behavior typically results in a negative association between the cross sectional dispersion of panel data and time, as studied by Phillips and Sul (2007) and Kong, Phillips, and Sul (2019). This concept of distributional convergence is leveraged to identify observed time series that co-move with a latent trend: candidate time series are designated as common trend determinants if the cross sectional dispersion of the panel series from a combination of these candidate series shows a negative association with time in the long run. Although we consider panel data with large time series (T) and cross section (n) dimensions for asymptotic analysis, our conditions only require that $n/T \to \infty$. This means the proposed methodology can be implemented even for relatively short panel data, as it benefits from the information carried in the large cross section.

This study contributes also to empirical work examining comovement in panel data and the common factors driving such comovement. 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 (e.g., Levitt (2004); Moody and Marvell (2010)). To demonstrate the effectiveness of our methodology, we provide an empirical analysis of state-level crime rates in the U.S. Unlike existing studies that rely on TWFE regression analysis, we identify a demographic factor of the young adult population in the U.S. as a key determinant of the latent common trend in violent crime rates. The results thereby highlight a previously overlooked factor compared to traditional research which focuses on factors like the size of the police force or income disparity. For property crime, on the other hand, it is incarceration rates that are found to be a key determinant of the latent common trend.

The rest of the paper is organized as follows. Section 2 motivates our approach, addressing a limitation of standard TWFE regression and empirically illustrating distributional convergence in panel data. Section 3 formally defines long run trend determinants using the concept of distributional convergence and outlines the procedure for identifying its key components. Section 4 provides asymptotic theory for the primary test statistic, which justifies the proposed method. In doing so, the analysis connects the key ideas in the present approach to panel cointegration, thereby introducing a novel perspective on cointegration between time series and panel data. Section 5 addresses those cases where only a subgroup of the series exhibits distributional convergence, proposing a pre-screening method to identify the subgroup associated with the underlying common trend. Section 6 revisits the crime rate example and provides a detailed empirical analysis, yielding results that differ from those obtained through standard TWFE estimation. Section 7 concludes with some final remarks. The Appendix gives a summary of the main procedure, proofs of the main theorems, and tables of critical value. The Online Supplement provides proofs of relevant technical lemmas, further results with linear trends, simulation findings, and additional tables relating to the empirical analysis in Section 6.

2 Distributional Convergence of Panel Data

In the literature on crime rate determinants a major focus has been explaining the decline in U.S. national crime rates since the early 1990s. Figure 1 exhibits time series of log national crime rates from 1985 to 2022 and shows that sharp declines occurred across all types of crime. Many studies seek key economic variables associated with crime rate trends in search of the main drivers of such declines. See Levitt (2004), Moody and Marvell (2010), and the references therein for discussion. To find such key determinants, existing studies often consider TWFE regression models of the following form

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

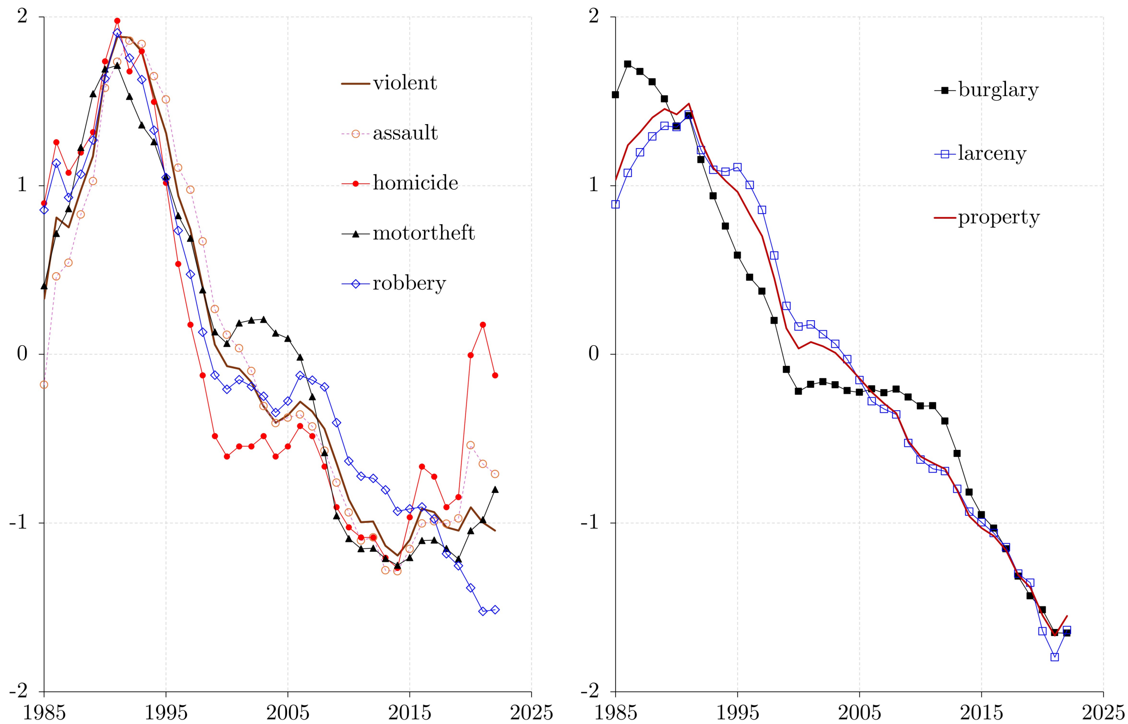
for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where y_{it} is the logarithm of the crime rate in state i and year t, η_i is a state fixed effect, ϱ_t is a year fixed effect, and z_{it} is a vector of explanatory variables possibly including a lagged dependent variable. Instrumental variables regression is often employed to control for potential simultaneity. However, when y_{it} or z_{it} have stochastic trends, this regression approach should be interpreted with caution, particularly when y_{it} and z_{it} are not cointegrated. In such cases, first-differenced variables are frequently used, but the interpretation of β then differs from that in level regression.

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, a 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 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 explain this concern 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; and

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



(a) Violent Crimes and Motor Vehicle Theft

(b) Property Crimes

Note: The figure shows the logarithm of the number of offenses per 100,000 inhabitants from 1985 to 2022. All series are standardized to place them in a single 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 rape crimes as the definition by the FBI changed in 2013. Data source: the *Uniform Crime Report.*

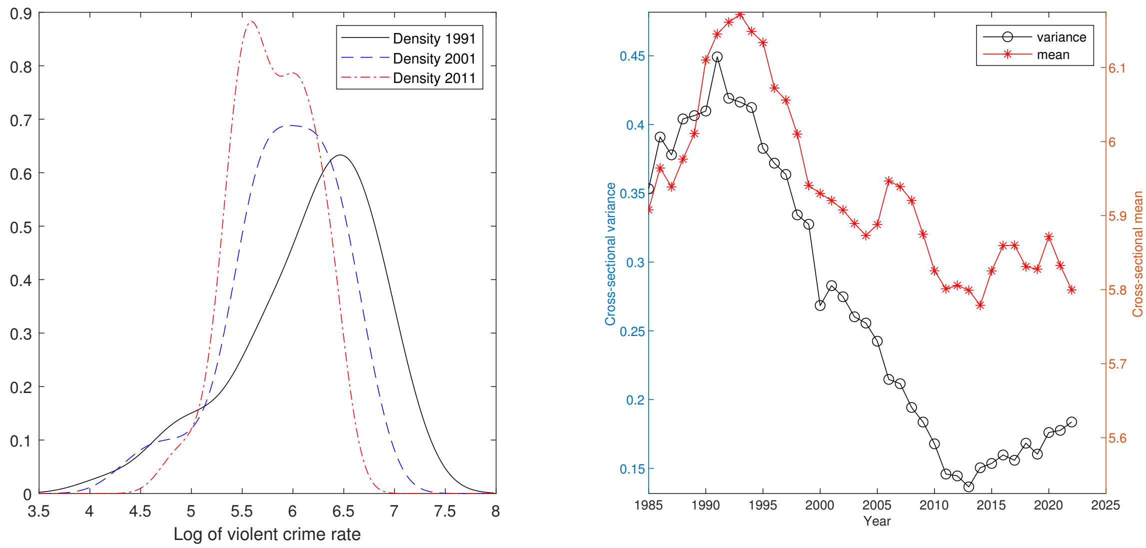
x_{it}^* satisfies the mean-reversion property though it can be heteroskedastic over i and t. The panel process y_{it} can be nonstationary and accommodates stochastic trends. In the representation (2), the component τ_t describes latent common trends among the individual panel series y_{it}, encompassing both deterministic and stochastic components. Suppose

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

where v_{it} is uncorrelated with $(\alpha_{z,i}, \tau_{z,t}, z_{it}^*)$. Then (2) can be rewritten as

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

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



(a) Changes in Densities over time

(b) Cross-sectional Mean and Variance

Note: Violent crime comprises 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 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 the early 1990s.

which yields the TWFE regression in (1), where the fixed effects, $\eta_i = (\alpha_i - \beta'\alpha_{z,i})$ and $\varrho_t = (\tau_t - \beta'\tau_{z,t})$, are arbitrarily correlated with the explanatory variables z_{it}. The coefficient β in this TWFE regression captures 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 example shows that finding significant elements of β in the standard TWFE regression model (1) cannot identify key variables or factors explaining the crime rate trend. Our approach is instead 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 observe a decline in cross section 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 relates to the concept of $\sigma-$convergence in economic growth theory, whereby the cross sectional distribution monotonically shrinks over time t. Rather than using this strong concept of distributional convergence we adopt here the concept of weak $\sigma-$convergence (Kong, Phillips, and Sul, 2019) in which distributional convergence is defined when the cross sectional dispersion measure is negatively associated with time t in the limit. The formal definition in the present settings follows,[1] where the notation $\widetilde{z}_t = z_t - T^{-1}\sum_{t=1}^{T} z_t$ is employed.

Definition 1 (Weak $\sigma-$Convergence towards a Common Trend) *A panel series y_{it} is said to be 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* $T\to\infty$*.*

According to this definition the cross section variation of the idiosyncratic component in (2)

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

is negatively associated with t. The panel y_{it} then shares a homogeneous trend τ_t and the cross sectional distribution of y_{it} weakly converges towards τ_t as t increases.

The common trend τ_t is unobserved and unknown in most cases. The question then arises, as in the aforementioned crime rate study, how to detect observed time series variables θ_t related to τ_t, which can be identified as key determinants of the dynamics of the panel series y_{it}. If y_{it} were stationary, regression of y_{it} on candidate variables θ_t to assess their explanatory power as trend determinants would be an option, similar to Chen, Roll, and Ross (1986). But as noted in Bai and Ng (2006) this approach is valid only when

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

the error variance is small enough for the correlation between y_{it} and θ_t to dominate. For this reason, Bai and Ng (2006) and Parker and Sul (2016) proposed studying directly the correlation between an estimated trend $\widehat{\tau}_t$ and θ_t.[2]

Our approach is closer to Chen, Roll, and Ross (1986), does not require estimating any latent common trend or factors, and is better suited for nonstationary panel processes. To this end, the cross sectional variation of x_{it} (or equivalently that of x_{it}^*) is permitted to vary over t but the temporal variation is required to negatively associate with t (so that it satisfies the weak $\sigma-$convergence condition of Kong, Phillips, and Sul, 2019). This is equivalent to assuming that cross sectional variation of y_{it} around a nonstationary time series τ_t is negatively associated with t as defined in Definition 1.[3]

If this negative association remains when τ_t is replaced by some linear combination of the observed time series variables θ_t, then we conclude that θ_t is 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, a genuinely difficult task in practice (see, e.g., Bai, 2004; Onatski and Wang, 2021). We formalize these ideas in the next section.

Remark 1 *Our approach presumes y_{it} satisfies Definition 1 or that x_{it} is weakly $\sigma-$convergent. Because $y_{it} - \overline{y}_t = x_{it} - \overline{x}_t$ from (4), this property can be checked by examining whether y_{it} itself is weakly $\sigma-$convergent. To achieve this end, Kong, Phillips, and Sul (2019) suggested use of a suitable t-test (say $\mathcal{T}^0_{\psi_K}$) of the significance of the coefficient ψ_K in the*

[2]More precisely, these authors 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 influences 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 may be empirically more relevant to discuss common trend or comovement in the framework of (2).

[3]Such distributional convergence behavior is often found in panel data series and arises because panel data typically share some common attributes that lead to interactions and commonalities as the series evolve. Even when all the panel series do not converge to a common trend series, certain subgroups often emerge that reveal temporal convergence. See Section 5 for discussion of this type of club convergence.

following auxiliary trend regression:

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

3 Evaluating Long Run Trend Determinants

Let θ_t be an $m \times 1$ vector of observed time series and δ be an m-vector of parameters.

Definition 2 (Long Run Trend Determinants) *θ_t is a vector of long run trend determinants of a panel series y_{it} if there exists some $\delta \neq 0$ such that y_{it} is weakly $\sigma-$convergent towards $\delta'\theta_t$ as defined in Definition 1.*

To develop a procedure assessing whether certain nonstationary time series θ_t are long run trend determinants of the panel y_{it}, allowing for the latent common trend τ_t to have a stochastic trend, we first suppose that τ_t and θ_t themselves have a long run association. A standard method might be to specify a potential cointegrating relationship between them. More precisely, suppose that there exists a mean-zero stationary process ξ_t for which

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

where there is no cointegration among the components of θ_t. When τ_t is observed and T is large enough, a natural way to assess θ_t as trend determinants would be by cointegration testing. But since τ_t is unobserved, we proceed instead by using y_t and combine (2), (4), and (6), giving

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

Definition 2 implies that the components of θ_t may be interpreted as long run trend determinants of y_{it} when the cross sectional variation of $y_{it} - \delta'\theta_t = \xi_t + x_{it}$, is negatively associated with t, provided that the vector $\delta \neq 0$.

The cross section sample variation of y_{it} from $\delta'\theta_t$ can be decomposed as follows:

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

where ξ_t and x_{it} are assumed to have mean zero and be 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 in the limit. If x_{it} is weakly $\sigma-$convergent, the association between ξ_t^2 and t must be negative or 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 remains. 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), thereby allowing 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; and ξ_t need not necessarily be 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 (as in the case of segmented cointegration, cf. Kim, 2003). In such a case, standard cointegration tests are likely to fail to detect a long run relation (even when τ_t is observed), whereas the present approach based on Definition 2 may successfully capture the trend relation, irrespective of whether τ_t is observed or not.

In practice, δ is unknown and the criterion

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

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

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

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 estimate of δ. Further, 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 $\widehat{\delta}$ can be checked using a standard t-test (Phillips and Park, 1988). When a particular $\theta_{j,t}$ $(j=1,\ldots,m)$ has an insignificant

[4]This approach can be viewed as a cointegrating regression with τ_t replaced by $\overline{y}_t$, and is akin to use of a cross sectional average as a common factor (Pesaran, 2006) or Mundlak (1978)'s approach for time fixed effects. In practice, other (weighted) mean series of y_{it} could be used instead of the cross section sample mean series $\overline{y}_t$. For instance, in the crime rate application of Section 6, the national crime rate could be used instead of the cross-state average crime rate.

[5]When they are correlated, FM-OLS (Phillips and Hansen, 1990), CCR (Park, 1992) or IVX (Phillips and Magdalinos, 2009) may be employed.

coefficient in this regression it is not considered a potential long run trend determinant. In that case, θ_t is reconfigured to include only significant elements and $S_{n,t}$ is redefined in (9) using this reconstructed θ_t.[6]

The association between $S_{n,t}$ and t can be assessed by a t-test of the estimate of the coefficient ϕ in the auxiliary trend regression

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

Even though the trend regression (11) is likely misspecified, the sign of ϕ is still informative of the direction of association between $S_{n,t}$ and t, or between $n^{-1}\sum_{i=1}^{n}\left(y_{it} - \delta'\theta_t\right)^2$ 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 a HAR (heteroskedasticity autocorrelation robust) long run variance estimator for some fixed-b coefficient $b \in (0,1]$ (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 $K(0) = 1$ and $\int K(\nu)\,d\nu = 1$.[7] If u_t is homoskedastic, $\mathcal{T}_\phi(b)$ may be simplified

[6] If e_t has a stochastic trend and the regression (10) is spurious, $\widehat{\delta}$ may appear significant when T is large so that all θ_t would be included in $S_{n,t}$. But then, due to the nonstationarity of e_t the auxiliary trend regression (11) would reveal that $S_{n,t}$ is not negatively associated with t, so that θ_t would be deselected as long run determinants.

[7] The required kernel function conditions are given in the Appendix. For the Bartlett kernel, the HAR estimator is given as

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

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-sided 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 $\delta'\theta_t$. This leads to the conclusion that the elements of θ_t are long run trend determinants of y_{it}. On the other hand, when the t-test fails to reject $\phi \geq 0$, $S_{n,t}$ is either positively associated or unassociated with t, implying that the cross section variation of $y_{it} - \delta'\theta_t$ either increases or has negligible association with t, and hence θ_t is not a common trend determinant vector of y_{it}. Section 4 derives the limiting distribution of $\mathcal{T}_\phi(b)$ in this auxiliary trend regression and provides critical values obtained by simulation.

Remark 2 *The interpretation of the decomposition in (8) offers an interesting perspective on cointegration between the panel 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). By comparison, standard cointegration only accounts for the first aspect of association. In this respect, our approach can be viewed as revealing more in the time series and panel context than a standard cointegration test.*

4 Asymptotics for Testing Trend Determinants in Panels

This section provides the asymptotic theory for the t-statistic $\mathcal{T}_\phi(b)$ given in (12) for testing long run trend determinants in panel data. This statistic is a directional t-test on the sign of ϕ in the trend regression (11). The directional restriction on ϕ is not used in deriving

with $L = \lfloor bT \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function.

the asymptotics for $\mathcal{T}_\phi(b)$ because the auxiliary trend regression (11) is misspecified by construction and there is no true value of the coefficient ϕ. To address this complication and distinct from standard analysis, the limit distribution of $\mathcal{T}_\phi(b)$ is obtained by using a proxy data generating process designed to characterize the sign behavior under the assumption of a 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 behavior of $n^{-1}\sum_{i=1}^{n}(x_{it}-\overline{x}_t)^2$ and ξ_t^2. Let

$$R_t^x = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(x_{it}-\overline{x}_t)^2,$$

and as in Definition 1-(c) denote the limit of the linear association between a time series z_t and t by the quantity

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

for some $c_T \to \infty$. Under the assumption that x_{it} is weakly $\sigma-$convergent we have $\gamma(R_t^x,t)<0$ and consider the following three cases:

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

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

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

Under (i), $\gamma(\xi_t^2,t)$ is positive and dominates $\gamma(R_t^x,t)$, so the common trend of y_{it} typically diverges from $\delta'\theta_t$. Under (ii), y_{it} shows neither weak $\sigma-$convergence toward $\delta'\theta_t$ nor divergence from $\delta'\theta_t$ because ξ_t is stationary. But under (iii), with $\gamma(R_t^x,t)<0$, y_{it} reveals weak $\sigma-$convergence toward $\delta'\theta_t$. So case (iii) occures when the statistic $\mathcal{T}_\phi(b)$ rejects $\phi\geq 0$ in favor of $\phi<0$, whereas cases (i) and (ii) arise when $\mathcal{T}_\phi(b)$ cannot reject $\phi\geq 0$.

These different cases are formally distinguished for testing purposes by supposing that x_{it} is a trend-stationary process for each i with shrinking variance over t and is generated by a system of the form

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

for some constants $\kappa_1,\kappa_2 \in (0,1/2)$. This specification of weakly $\sigma-$convergent x_{it} is similar to that used in the models of Kong, Phillips, and Sul (2019) and Kong, Phillips, and Sul (2020) but the present framework is more general. The current setup in effect 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 then determine the decay rate of the cross section variance of x_{it}, enabling flexible control over the relative behavior of $\gamma(\xi_t^2, t)$ and $\gamma(R_t^x, t)$, especially for cases (ii) and (iii) above. In particular, case (ii) is specified with $\kappa_1, \kappa_2 > 1/4$, which ensures that the cross section variance of x_{it} vanishes rapidly so that $\gamma(R_t^x, t)$ is small enough to be dominated by $\gamma(\xi_t^2, t)$. On the other hand, case (iii) can be specified with $\kappa_1, \kappa_2 \leq 1/4$.[8]

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 where the components of θ_t are not long run determinants of $y_{i,t}$. The following notation is used: '$\rightsquigarrow$' denotes weak convergence of the associated probability measures, '$\equiv$' stands for distributional equivalence, '$\xrightarrow{p}$' is convergence in probability, $\lfloor \cdot \rfloor$ is the floor function, b is the fixed-b parameter, $K(\cdot)$ is the kernel function used in HAR estimation in (13), $\kappa_1 \wedge \kappa_2 = \min\{\kappa_1, \kappa_2\} = \kappa_1 1\{\kappa_1 \leq \kappa_2\} + \kappa_2 1\{\kappa_1 > \kappa_2\}$, and $1\{\cdot\}$ is the binary indicator.

Theorem 1 *Suppose Assumptions 1-3 in the Appendix hold and $n/T \to \infty$ 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 (15) or (16)), $\mathcal{T}_\phi(b)$ in (12) satisfies*

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

where

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

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

[8] The case with $\kappa_1, \kappa_2 = 0$ is excluded because x_{it} is presumed to satisfy weak $\sigma-$convergence.

[9] This process is a 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); Park and Phillips (1988) for further details and other applications.

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

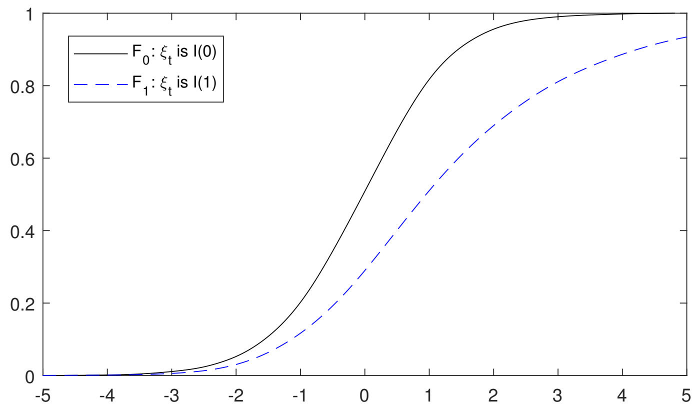
Note: The values are the simulated percentiles of the limiting distributions of $\mathcal{T}_\phi(b)$ and $\mathcal{T}_\phi^0(b)$ given in (19) and (20), obtained from 2 million replications. Brownian motion is approximated by normalized sums of standard normal random variables using 10,000 steps and 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 in (14). Critical values for other $b \in (0,1]$ are available in the Appendix.

Theorem **1** gives the limit distribution of $\mathcal{T}_\phi(b)$ under the two scenarios for ξ_t in (15) and (16). The first (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 (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), and the null limit 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). Importantly, both $F_1(b)$ and $F_0(b)$ in Theorem **1** (as well as $F_0^0(b)$) are free of nuisance parameters. As depicted in Figure 3, $F_1(b)$ 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 $F_0(b)$ and $F_0^0(b)$ 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 case (iii) in (17), and yields the alternative limit distribution of $\mathcal{T}_\phi(b)$.

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 $T = 5,000$ using 10,000 replications.

Theorem 2 *Assume 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., (17) holds). When $\kappa_1 \wedge \kappa_2 = 1/4$,*

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

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

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

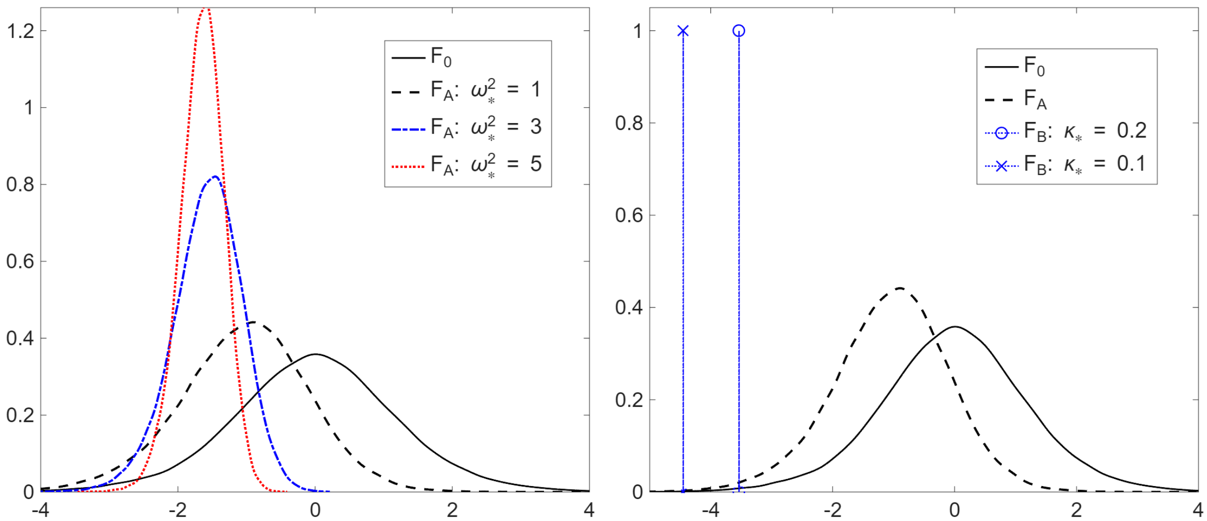
with $\sigma_\mu^2 = \mathbb{E}\mu_i^2$, $\sigma_\varepsilon^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathbb{E}\varepsilon_{it}^2$, and $\omega_{\xi\xi}^2 = \sum_{j=-\infty}^{\infty} \mathbb{E}(\xi_t^2 - \mathbb{E}\xi_t^2)(\xi_{t+j}^2 - \mathbb{E}\xi_t^2)$. When $\kappa_1 \wedge \kappa_2 < 1/4$,

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

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

When $\kappa_1 \wedge \kappa_2 = 1/4$ under $\xi_t \sim I(0)$, neither $\gamma(R_t^x, t)$ nor $\gamma(\xi_t^2, t)$ is dominant. Recall 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, and hence $R_t^x = \text{plim}_{n\to\infty} R_{n,t}$. In this case, the ratio between the cross section

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) shows how the limit distribution $F_A(b; \omega_*^2)$ in (21) changes with ω_*^2. As ω_*^2 increases, the distribution shifts to the left and variance reduces, as evident in the black dashed ($\omega_*^2 = 1$), blue dash-dotted ($\omega_*^2 = 3$), then red dotted ($\omega_*^2 = 5$) lines. Figure 4(b) shows 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 density $F_0(b)$. All limit distributions were simulated with $b = 0.1$, $T = 5,000$, and were based on 10,000 replications.

variance of the panel series $y_{i,t}$ (i.e., R_t^x) 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 limit distribution of $\mathcal{T}_\phi(b)$ as in (21). As depicted in Figure 4(a), as ω_*^2 gets large the limit distribution shifts in the negative direction and shrinks toward its new shifted center. When ω_*^2 is small, on the other hand, it can hardly be distinguished from $F_0(b)$. This case hence can be understood as a local alternative in the present context and demonstrates how 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 components of θ_t are long run determinants of the common trend of the panel series y_{it} only if the cross section variation of y_{it} outweighs that of the cointegration error ξ_t, as earlier discussed in Remark 2.

When $\kappa_1 \wedge \kappa_2 < 1/4$, $\gamma(R_t^x, t)$ dominates $\gamma(\xi_t^2, t)$ and the limit 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^x

becomes important, solely depends on $\kappa_* = \kappa_1 \wedge \kappa_2$, and 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. Further, simulations show that the point of degeneration is always below the 5% critical value given in Table 1 at each b value. From these two cases, under $\kappa_1 \wedge \kappa_2 < 1/4$, the limit distribution of $\mathcal{T}_\phi(b)$ shrinks toward its negative mode as T grows and eventually degenerates as in (23), giving test consistency. The online Appendix provides further simulation evidence of this behavior.

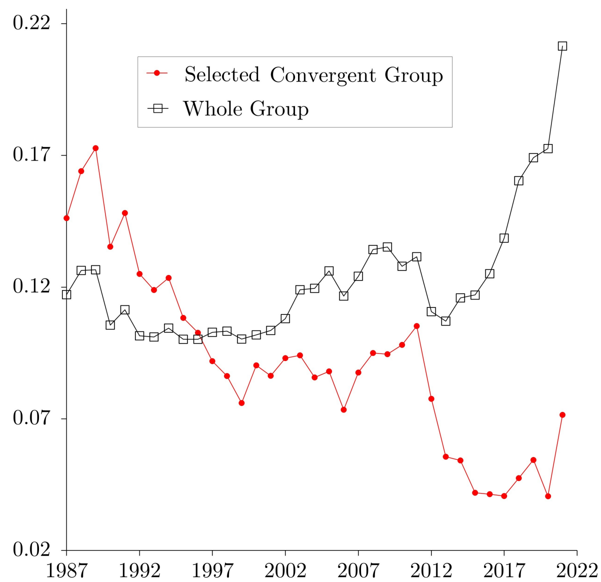
Remark 3 *Nonstationary variables often exhibit linear trends. For instance, suppose the nonstationary common trend τ_t imposes a linear trend (e.g., a random walk with drift). If we properly choose a variable θ_t that also follows a random walk with drift and the regression of τ_t on θ_t successfully 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 be $I(1)$ or, at best, $I(0)$ with a linear trend, under which the test $\mathcal{T}_\phi(b)$ fails to 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 therefore does not affect the test, and the same critical values can be used as in the no-drift case in Table 1.*

5 The Case of Partial Distributional Convergence

The focus of the approach explained earlier is whether the entire panel y_{it} weakly $\sigma-$converges to a common trend τ_t in (2). But in practice it is not unusual for panels to include some outlier series with their own specific trend behavior. In such cases it is unlikely for the full panel to manifest weak $\sigma-$convergence to a common trend. For example, Figure 5 shows cross section 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 series in the panel have different trend behavior from τ_t or their idiosyncratic terms x_{it}^* do not satisfy weak $\sigma-$convergence.

Nonetheless, there is a subgroup of 25 series that show weak $\sigma-$convergence towards the national burglary rates (red circle line). In fact, using the methods of Phillips and

Figure 5: Partial Distributional Convergence of Burglary Rates



Note: The figure plots cross section 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.

Sul (2007) such subgroup of the panel can be determined within which there is weak $\sigma-$convergence towards a common trend. Once this convergent subgroup is found, we can identify the long run trend determinants θ_t of the leading trend τ_t within this subgroup using the methods developed in the previous section. This approach enables outlier series to be eliminated from the analysis and the components of θ_t are identified as the main drivers of the trend determinants of y_{it}, especially when the convergent subgroup comprises a majority, i.e., more than half, of the sample.

To simplify implementation in the case of a wide panel the procedure of Phillips and Sul (2007) can be employed, leading to the following approach to determine a subgroup of weakly convergent series in the panel, given chosen time series θ_t. Start by running individual-specific auxiliary trend regressions

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

for each i and apply t-tests on the fitted coefficients of φ_i, where Δ_{it} is an approximate distance measure between y_{it} and $\delta'\theta_t$ defined as

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

The estimate $\widehat{\delta}$ is from the time series regression of $\overline{y}_t$ on θ_t given in (10), which is statistically significant. 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 below some threshold, we conclude that Δ_{it} is not positively associated with t, implying that the trend of y_{it} is unlikely to depart from $\delta'\theta_t$; consequently, we assign i to the estimated subgroup $\widehat{\mathcal{G}}(\theta)$ for the chosen θ_t.

This subgroup selection process can be regarded as a pre-test and choice of critical values in testing is therefore important. In fact, the limiting behavior of $\mathcal{T}_{\varphi_i}(b)$ is similar to that of $\mathcal{T}_\phi(b)$, and critical values from Table 1 can be used. But it is helpful to choose a first-step critical value so that with high probability true convergent members will be selected to the subgroup estimate $\widehat{\mathcal{G}}(\theta)$, which helps 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 $\Pr(\mathcal{T}_{\varphi_i}(b) < c_1 | i \in \mathcal{G}(\theta))$ remains high, where $\mathcal{G}(\theta)$ is the true convergent group towards the trend $\delta'\theta_t$. The following corollary provides a useful guideline for this process and is derived in the online supplement.

Corollary 1 *Suppose conditions in Theorems* **1** *and* **2** *hold for each i with $T \to \infty$. $F_1(b)$ and $F_0(b)$ are defined in Theorems* **1** *and* **2**, *respectively. If $\xi_t \sim I(1)$, $\mathcal{T}_{\varphi_i}(b) \rightsquigarrow F_1(b)$. 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) for each i.*

When $\xi_t \sim I(1)$, which is the case that Δ_{it} is positively associated with t, the 10th percentile 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 at least a 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 $\widehat{\mathcal{G}}(\theta)$.

The subgroup selection process can be improved by implementing the following iterative algorithm. The procedure described above defines Δ_{it} in (25) using the $\widehat{\delta}$ that is obtained from the time series regression of $\overline{y}_t$ on θ_t (i.e., $\overline{y}_t = \alpha_0 + \delta'\theta_t + e_t$), where $\overline{y}_t = n^{-1}\sum_{i=1}^{n} y_{it}$ is the average of y_{it} for all i. Once we find the subgroup, say $\widehat{\mathcal{G}}^{(1)}(\theta)$, we can update $\overline{y}_t$ such that $\overline{y}_t^{(1)} = |\widehat{\mathcal{G}}^{(1)}(\theta)|^{-1}\sum_{i\in\widehat{\mathcal{G}}^{(1)}(\theta)} y_{it}$ and obtain $\widehat{\delta}^{(1)}$ from the time series regression of $\overline{y}_t^{(1)}$ 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 $\mathcal{T}_{\varphi_i}(b)$ to update the subgroup estimate, yielding $\widehat{\mathcal{G}}^{(2)}(\theta)$. This procedure is repeated 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 the subgroup, this iteration will yield a strictly positive $|\widehat{\mathcal{G}}(\theta)|$ at the end of the 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 the iteration will continue to decrease the subgroup size and eventually yield an empty $\widehat{\mathcal{G}}(\theta)$.

Once the subgroup $\widehat{\mathcal{G}}(\theta)$ is determined, the regression t-test $\mathcal{T}_\phi(b)$ is conducted using only those members of $\widehat{\mathcal{G}}(\theta)$. We may instead consider this subgroup as a core subgroup of convergent members, and enrich it by adding potentially missing panel members in the selection step in a similar manner to Phillips and Sul (2007), before the regression t-test. One way of doing so is to use a distance measure from y_{it} to $\widehat{\delta}'\theta_t$ for each $i \in \widehat{\mathcal{G}}(\theta)^c$, the complement of the subgroup, over some periods and include those elements in the panel 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. Measures such as the forecast depth (Lee and Sul, 2023a) or more general versions (Lee and Sul, 2023b) can be used for this purpose. The Appendix provides details.

6 Trend Determinants of U.S. Crime Rates

This section uses the methodology described above to explore trend determinants of national crime rates in the U.S. Of particular interest is the sharp decline in these crime rate in the 1990s. As explained in Section 2 there are major difficulties using standard TWFE regression methods to investigate such issues. Instead, the approach developed in the present paper is used to check whether given time series are long run trend determinants of the crime rates y_{it} as defined in Definition 2, thereby identifying explanatory factors of the sharp decline in national crime in the 1990s. In particular, we revisit the following four variables that Levitt (2004) considered and study which can be identified as trend determinants: the number of sworn police officers, the incarceration rate, real GDP, and demographics (i.e., the proportion of population in some specific age group). Levitt (2004) found the first two variables were determinants of the crime rate decline in the 1990s but the latter two were not.[10] Unlike these findings, our results show demographics as the key long run trend determinant of violent crime rates and the incarceration rate is the key long run trend determinant of property crime rates, which together well describe crime rate dynamics including its decline in the 1990s. Details of the findings are given in the following two subsections.

6.1 Violent Crimes

We considered logarithms 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 was excluded from this analysis due to data limitations. Instead, we included the aggregate violent crime rate, which encompasses all four categories (i.e., assault, homicide, robbery, and rape). All crime rates were collected from the Uniform Crime Report on the FBI Crime Data Explorer, covering the 50 U.S. states for the period from 1987 to 2021.[11]

Prior to the analysis we checked two items: (i) whether x^*_{it} in (2) satisfied the weak $\sigma-$convergence criterion; and (ii) if y_{it} imposed stochastic trend nonstationarity. As noted in Remark 1, (i) was assessed by studying weak $\sigma-$convergence in the panel y_{it} using the t-test $\mathcal{T}^0_{\psi_K}$ of ψ_K from the trend regression (5). Regarding (ii), instead of directly testing for a unit root in each of the panel series, we indirectly checked if y_{it} satisfied weak

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

[11]https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/crime-trend. The period was chosen to match the available sample period of the potential determinants considered in our analysis.

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 lag length $\lfloor T^{1/3} \rfloor$; $\mathcal{T}^0_{\psi_K} < -1.65$ 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. (* indicates significance at 5%.) (ii) Violent crime includes homicide, rape, robbery, and assault. (iii) The sample period is 1987–2021, which matches the sample period of candidate determinants θ_t that are lagged by one period.

$\sigma-$convergence toward some deterministic trend.[12] This was done by replacing τ_t with a deterministic trend function and checking weak $\sigma-$convergence of the detrended y_{it}. Specifically, a linear trend function $\tau_t = \delta t$ was used, and the weak $\sigma-$convergence test of Kong, Phillips, and Sul (2019) was conducted as in Remark 1 using $R_{n,t} = n^{-1} \sum_{i=1}^{n} (y_{it} - \widehat{\alpha}_0 - \widehat{\delta} t)^2$, where $(\widehat{\alpha}_0, \widehat{\delta})$ were obtained from the trend regression: $\overline{y}_t = \alpha_0 + \delta t + e_t$. If y_{it} is nonstationary, its cross sectional variation should be positively associated with t even after this detrending. When weak $\sigma-$convergence holds in this case, we can 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 a lag length $\lfloor T^{1/3} \rfloor$. The specific form of $\mathcal{T}^0_{\psi_K}$ is given in Appendix A.1. Following Kong, Phillips, and Sul (2019), we rejected the null of no weak $\sigma-$convergence if $\mathcal{T}^0_{\psi_K} < -1.65$. All series satisfied weak $\sigma-$convergence, indicating that the panel series y_{it} share a common long run trend. In addition to the sample average process $\overline{y}_t$, we also considered the logarithm of the national crime rate and reported the t-statistics in the second column, finding similar results to those with the sample average

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

process. The third column reports the t-statistics of the weak $\sigma-$convergence test toward a linear deterministic trend, indicating that y_{it} is most likely nonstationary.[13]

Based on these preliminary results, we applied the ideas in Section 3 to the following time series as candidates for long run determinants θ_t: the fraction of the young adult population with ages between 10 and 39 (Demog), the number of non-civilian police officers (Police), the local incarceration rate (Prison), and real GDP per capita (RGDP).[14] All variables were log-transformed and lagged by one period to minimize potential simultaneity.[15] Augmented Dickey–Fuller, KPSS, and Phillips-Perron tests showed that all these time series were unit root processes (with linear trends). For each of the candidate time series, we conducted two types of t-tests, $\mathcal{T}_\phi(b)$ in (12) and $\mathcal{T}_\phi^0(b)$ 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 logarithms of the four crime rates: violent crime, assault, homicide, and robbery. Among the candidate determinants, only 'Demog' is identified as a long run trend determinant; it yields an estimate $\widehat{\delta}$ that is significantly different from zero, and the trend regression t-statistics $\mathcal{T}_\phi(b)$ and $\mathcal{T}_\phi^0(b)$ are below the 5% critical values (i.e., $\mathcal{T}_\phi(0.1) < -2.04$ and $\mathcal{T}_\phi^0(0.1) < -1.96$). In contrast, 'Police' yields an insignificant $\widehat{\delta}$ and thus its trend regression t-statistics are not obtained.[16]

Knowledge that the fraction of the young adult population is a long run trend deter-

[13] Though $T = 35$ is relatively small for standard unit root testing, we also conducted ADF, KPSS, and Phillips-Perron tests to check whether the $\overline{y}_t$ series have a unit root, allowing for possible linear trends. All these tests supported the presence of a unit root.

[14] Summary statistics, data sources, and variable details are reported in the online supplement.

[15] As a robustness check, we considered variables lagged by two periods as well, with findings in the line with those presented here with a one-period lag. These additional results are reported in the online supplement. Also note that (i) the choice between non-civilian and civilian police officers does not change the results; and (ii) neither does the choice among the three types of incarceration rates (i.e., federal, state, and local).

[16] We also conducted residual based cointegration tests between the average of each crime rate ($\overline{y}_t$) and potential determinants θ_t. In all cases these tests did not reject the null of no cointegration, although the tests were performed with short time series $T = 35$. The comparison shows that the convergence approach can identify time series that are associated with a latent common trend in the panel or their common trend factor, even when a standard cointegration test may be unable to do so. From this perspective, as also discussed in Remark 2, our approach can be viewed as an alternative to standard (homogeneous) panel cointegration testing (e.g., Kao, 1999), particularly when T is not large.

Table 3: Long-Run Trend Determinants of Violent Crimes

Crime	θ_t	$\hat{\delta}$	$se(\hat{\delta})$	$se^0(\hat{\delta})$	$\mathcal{T}_\phi(0.1)$	$\mathcal{T}^0_\phi(0.1)$
Violent	Demog	3.601*	0.431	0.311	-6.658*	-7.299*
	Police	-0.467	1.224	1.462	n.a.	n.a.
	Prison	-0.813*	0.227	0.240	0.293	0.390
	RGDP	-1.432*	0.171	0.145	62.835	6.376
Assault	Demog	3.063*	0.460	0.338	-3.423*	-7.089*
	Police	-0.285	1.057	1.301	n.a.	n.a.
	Prison	-0.704*	0.216	0.220	0.341	0.578
	RGDP	-1.204*	0.184	0.159	34.889	5.367
Homicide	Demog	3.308*	0.466	0.479	-7.948*	-3.875*
	Police	-1.548	1.216	1.378	n.a.	n.a.
	Prison	-0.971*	0.212	0.184	0.030	0.623
	RGDP	-1.281*	0.224	0.219	13.621	2.912
Robbery	Demog	5.498*	0.645	0.568	-19.629*	-2.964*
	Police	-0.321	1.864	2.164	n.a.	n.a.
	Prison	-1.080*	0.323	0.395	-0.615	-1.218
	RGDP	-2.258*	0.230	0.204	87.087	3.228

Note: (i) $\hat{\delta}$ is the least squares estimate from (10); $se(\hat{\delta})$ and $se^0(\hat{\delta})$ are the standard errors for the heteroskedastic error and the homoskedastic error cases, respectively, from Phillips and Park (1988). $\mathcal{T}_\phi(0.1)$ and $\mathcal{T}^0_\phi(0.1)$ are respectively the t-ratios defined in (12) and (14) with $b = 0.1$ and the Bartlett kernel, which are reported only when $\hat{\delta}$ is significant. (ii) 'Demog' is the logarithm of the fraction of the young adult population between the ages of 10 and 39, 'Police' is the logarithm of the number of non-civilian police officers per capita, 'Prison' is the logarithm of the local incarceration per capita, and 'RGDP' is the logarithm of Real GDP per capita. (iii) 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 $\mathcal{T}^0_\phi(0.1) < -1.96$, 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 shown in the table are $\mathcal{T}_{\phi}^{0}(0.1)$ with the fraction of population in each age group as θ_t. (* and ** indicate significance at 5% and 1%, respectively.)

minant of all violent crime rates suggests further investigation into the age groups that are most responsible for each case. The findings from this investigation are shown in Table 4 and lead to the following conclusions: the fraction of young adult population between ages 10 and 39 best describes the common long run trends in assault and overall violent crimes (which includes rape); ages between 20 and 39 best capture trends in homicide; and ages between 20 and 49 best capture trends in robbery.

6.2 Property Crimes

For the analysis of property crimes the following categories were considered: burglary, larceny, motor vehicle theft, and overall property crime rates. Due to data limitations arson was excluded but the overall property crime category encompassed all four types of property crime and so included arson. As in the analysis of violent crime rates all state-level property crime rates were collected from the Uniform Crime Report on FBI Crime Data Explorer for the 50 U.S. states over the period 1987–2021. A preliminary check was conducted to determine whether the panel series satisfied the weak $\sigma-$convergence criterion and were not trend stationary. Table 5 reports the t-statistics of ψ_K in the auxiliary trend regression (5) for the four panels and the results show that only the motor vehicle theft rate manifests weak $\sigma-$convergence.[17] The preliminary results indicate that

[17]The findings in Table 5 also indicate that the motor vehicle theft rate is not linear trend stationary. Trend stationary checks were not conducted for the other variables because they are not weak $\sigma-$convergent (meaning our approach cannot be directly applied for the entire sample) and their trajectories do not appear trend stationary.

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

	Weak σ-convergence toward:		
	Sample Mean	National Average	Linear Trend
Overall 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. (* indicates significance at 5%.) (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.

the appropriate procedure is to assess evidence of partial convergence as discussed in Section 5, whereby a search is conducted for distributionally convergent subgroups from each of the panel series for property crime, burglary, and larceny rates.

To this end subgroup selection was conducted using $\mathcal{T}_{\varphi_i}(b)$ with $b = 0.1$ for each θ_t variable chosen and the threshold value in the test was set to $c_1 = -1.2$. Panel series i were selected for inclusion in the convergent subgroup $\widehat{\mathcal{G}}(\theta)$ if the statistic $\mathcal{T}_{\varphi_i}(b)$ did not exceed the threshold c_1. Interestingly, as long as it was negative the choice of c_1 had little impact on the final subgroup selection outcomes after the iterations described in Section 5, which were completed quickly within a couple of rounds in all cases. Iterations were stopped at the rth round when $|\widehat{\delta}^{(r)} - \widehat{\delta}^{(r-1)}| < 0.001$, with $\widehat{\delta}^{(r)}$ being the rth round estimate of δ in the time series regression (10), ensuring that the subgroup estimate no longer changed.

Table 6 shows long run trend determinant results based on the final subgroup estimate $\widehat{\mathcal{G}}(\theta)$. The same candidate time series θ_t (viz., Demog, Police, Prison, and RGDP) were considered as in the violent crime case. Interestingly, 'Police' did not yield a significant estimate $\widehat{\delta}$ and was therefore not identified as a long run trend determinant. For this reason, neither its subgroup estimate nor the trend regression t-statistics were obtained. Among the other variables, only 'Prison' yielded large subgroup estimates $\widehat{\mathcal{G}}(\theta)$, with over

80% of the panel series included, whereas 'Demog' and 'RGDP' yielded no convergent subgroups. 'Prison' produced a significant $\widehat{\delta}$ and its test statistics satistied $\mathcal{T}_{\phi}(0.1) < -2.04$ and $\mathcal{T}_{\phi}^{0}(0.1) < -1.96$, identifying it as a strong long run trend determinant for property crime, burglary, and larceny. This outcome aligns with a known finding in empirical sociology of a close link between incarceration rates and property crimes, particularly those for burglary (e.g., Rosenfeld and Messner, 2009; Weatherburn, Hua, and Moffatt, 2006). Unlike the violent crime case, 'Demog' is no longer a long run trend determinant because the associated subgroup $\widehat{\mathcal{G}}(\theta)$ is nearly empty with few panel series manifesting a long run trend with this time series.

Notably, the results for motor vehicle theft reported in Table 6 differ from those of other property crimes. As shown in Table 5 motor vehicle theft satisfies the weak $\sigma-$convergence test, so that the panel series all share a long run common trend and there is no need for a subgroup selection analysis. Furthermore, 'Demog' is identified as a long run trend determinant of the motor vehicle theft rate, whereas 'Prison' is not.[18] Although motor vehicle theft is classified as a property crime, its trajectories are very similar to those of violent crime as depicted in Figure 1(a). This similarity may stem from the fact that motor vehicle theft can be elevated from a misdemeanor to a felony, carrying a potential prison sentence of up to 10 years. Furthermore, unlike other property crimes, grand theft auto is often linked to its use as a mode of transportation for other more serious offenses or as a tool in violent crimes.

[18]For this analysis the young adult population fraction of ages between 10 and 49 was used instead of ages between 10 and 39. Similar to Table 4, the t-statistic $\mathcal{T}_{\phi}^{0}(0.1)$ was calculated for the motor vehicle theft rate for different age groups and is shown below. The findings provide weak support for 'Demog' as a long-run common trend determinant when the older population with ages between 40 and 49 is excluded. A possible explanation is that teenagers are less likely to drive and have less incentive to steal motor vehicles, so that considering only a younger population decreases the explanatory power of this variable.

Age Group	10–29	10–39	10–49	20–49
$\mathcal{T}_{\phi}^{0}(0.1)$	-0.202	-2.005	-2.832	-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	n.a.	n.a.	n.a.
	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	n.a.	n.a.	n.a.
	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	n.a.	n.a.	n.a.
	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	n.a.	n.a.	
	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 $\mathcal{T}^0_{\varphi_i}(0.1)$, obtained only when $\hat{\delta}$ is significant. 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); $se(\hat{\delta})$ and $se^0(\hat{\delta})$ are the standard errors for the heteroskedastic error and the homoskedastic error cases, respectively, from Phillips and Park (1988). $\mathcal{T}_\phi(0.1)$ and $\mathcal{T}^0_\phi(0.1)$ are the t-ratios defined in (12) and (14) with $b = 0.1$ and the Bartlett kernel, which are reported only when $\hat{\delta}$ is significant. 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 $\mathcal{T}^0_\phi(0.1) < -1.96$. 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%.)

7 Concluding Remarks

The methods of this paper assist in identifying common trend determinants in nonstationary panel data. In contrast to TWFE regression where time effects are eliminated, the approach seeks to discover and analyze the underlying common trend variables. The methodology is well suited to research where the goal is to locate and understand observed factors that drive shared latent trends among panel series. The approach also sheds new light on cointegration between panel data and time series, emphasizing the need for analysis of the relative variation between panel data and cointegrating error. The concept of distributional convergence, in particular weak $\sigma-$convergence, is leveraged to develop a practical procedure of discovery for trend determinants in panels. The key advantages of this methodology are its applicability to relatively short panel datasets and its ability to bypass the complex estimation of latent common factors. Importantly, the method relies on simple regression techniques that can be implemented using any standard statistical package. This practical simplicity offers a highly accessible framework for analyzing latent trends in micro panel data across diverse domains, including consumer behavior in business, voter dynamics in political science, longitudinal patient outcomes in medical research, household income mobility in economics, and firm-level responses to environmental policy.

Our application to U.S. crime rate data demonstrates that the percentage of young adults significantly influences violent crime trends and that incarceration rates drive property crime trends. These findings differ from standard TWFE analyses, which often highlight factors like police numbers and income levels. Interestingly, research by Farrell, Tilley, and Tseloni (2014) aligns with our present 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 help provide an opening for new policy thinking in matters of crime control.

References

Bai, J. (2004): "Estimating cross-section common stochastic trends in nonstationary panel data," *Journal of Econometrics*, 122(1), 137–183. 8

Bai, J., and S. Ng (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131(1-2), 507–537. 2, 7, 8

Chen, N.-F., R. Roll, and S. A. Ross (1986): "Economic Forces and the Stock Market," *The Journal of Business*, 59(3), 383–403. 7, 8

Farrell, G., N. Tilley, and A. Tseloni (2014): "Why the crime drop?," *Crime and justice*, 43(1), 421–490. 30

Kao, C. (1999): "Spurious regression and residual-based tests for cointegration in panel data," *Journal of Econometrics*, 90(1), 1–44. 24

Kiefer, N. M., and T. J. Vogelsang (2005): "A new asymptotic theory for heteroskedasticity-autocorrelation robust tests," *Econometric Theory*, 21(6), 1130–1164. 11

Kim, J.-Y. (2003): "Inference on Segmented Cointegration," *Econometric Theory*, 19(4), 620–639. 10

Kong, J., P. C. B. Phillips, and D. Sul (2019): "Weak σ-convergence: Theory and applications," *Journal of Econometrics*, 209(2), 185–207. 2, 7, 8, 14, 23, 33

——— (2020): "Testing convergence using HAR inference," *Advances in Econometrics*, 41, 25–72. 14

Lee, Y., and D. Sul (2023a): "Depth-weighted Forecast Combination: Application to COVID-19 Cases," *Advances in Econometrics*, 45B, 235–260. 21, 36

——— (2023b): "Depth-weighted means of noisy data: An application to estimating the average effect in heterogeneous panels," *Journal of Multivariate Analysis*, 196, 105–165. 21, 36

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. 3, 4, 22

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. 14

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. 3, 4

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85. 10

Onatski, A., and C. Wang (2021): "Spurious factor analysis," *Econometrica*, 89(2), 591–614. 8

Park, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60(1), 119. 10

Park, J. Y., and P. C. B. Phillips (1988): "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4(3), 468–497. 14

Parker, J., and D. Sul (2016): "Identification of unknown common factors: Leaders and followers," *Journal of Business & Economic Statistics*, 34(2), 227–239. 2, 8

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012. 10

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. 10

Phillips, P. C. B., and T. Magdalinos (2009): "Econometric inference in the vicinity of unity," *Singapore Management University, CoFie Working Paper*, 7, 35. 10

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. 10, 25, 29, 35

Phillips, P. C. B., and D. Sul (2007): "Transition modeling and econometric convergence tests," *Econometrica*, 75(6), 1771–1855. 2, 18, 19, 21, 36

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. 28

Weatherburn, D., J. Hua, and S. Moffatt (2006): "How Much Crime Does Prison Stop-The Incapacitation Effect of Prison on Burglary," *International Journal of Punishment and Sentencing*, 2(1), 8–32. 28

Appendix

Appendix A: Summary of Procedures

We outline the procedures that were implemented in the empirical analysis.

A.1 Establishing Weak $\sigma-$Convergence of x_{it}^*

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

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

satisfies weak $\sigma-$convergence (toward its mean). As discussed in Remark 1, this can be done by analyzing the weak $\sigma-$convergence properties of y_{it} using the methods of Kong, Phillips, and Sul (2019). The steps in the procedure are as follows.

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 the coefficient ψ_K as
$$\mathcal{T}_{\psi_K}^0 = \frac{\widehat{\psi}_K}{\left(\widehat{\Omega}_{u_K} / \sum_{t=1}^{T} (\widetilde{t})^2\right)^{1/2}},$$
where $\widehat{\Omega}_{u_K}$ is a long run variance estimator of $u_{K,t}$ and $\widetilde{t} = t - T^{-1} \sum_{r=1}^{T} r$.

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

Remark A1 (Trend Stationarity) When T is too small and n is large, it is well known that standard panel unit root tests do not perform well. In such a case, the panel process y_{it} can be analyzed to assess whether weak $\sigma-$convergence holds towards some deterministic trend function, such as a linear trend. If the test fails to reject the null of no weak $\sigma-$convergence, then the outcome can be interpreted as evidence that y_{it} is not (linear) trend stationary in the long run. The test can be conducted by using $R_{n,t} = n^{-1}\sum_{i=1}^{n}(y_{it} - \widehat{\alpha}_0 - \widehat{\delta}t)^2$ in step 1, where the estimates $(\widehat{\alpha}_0, \widehat{\delta})$ are obtained from the regression $\overline{y}_t = \alpha_0 + \delta t + e_t$, followed by steps 2 and 3 as described above.

A.2 Assessing Long-Run Trend Determinants

When x_{it}^* satisfies weak $\sigma-$convergence, candidate time series θ_t can be assessed as potential long run trend determinants as follows. 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 obtained by fitting the time series regression $\overline{y}_t = \alpha_0 + \delta'\theta_t + e_t$.

2. Run the trend regression
$$S_{n,t} = \phi_0 + \phi t + u_t,$$
as given in (11) and compute the t-statistic of ϕ (using here 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 $L = \lfloor bT \rfloor$ for some fixed-b coefficient $b \in (0,1)$, where $\widehat{u}_t = S_{n,t} - \widehat{\phi}_0 - \widehat{\phi}t$.

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

Remark A2 For vector θ_t, we can assess each element individually. If we want to evaluate the elements jointly, we can perform the following preliminary steps before estimating $\overline{y}_t = \alpha_0 + \delta'\theta_t + e_t$. First: test for cointegration within the $m \times 1$ vector θ_t to identify redundancies, retaining only a subset of elements equal to the number of common stochastic trends (i.e., m minus the number of cointegrating relations). Second: conduct element-by-element t-tests following Phillips and Park (1988), including only the significant elements in the final vector θ_t. This applies to Section A.3 below as well.

A.3 Finding Convergent Subgroups

When x^*_{it} does not satisfy the weak $\sigma-$convergence criterion in Section A.1, we seek potential convergent subgroups for a given θ_t. Let $\widehat{\mathcal{G}}(\theta)$ be the largest such subgroup estimate; we then verify whether θ_t is indeed a long run trend determinant within this estimated subgroup by implementing the steps in A.2 using only the panel series $i \in \widehat{\mathcal{G}}(\theta)$. The following steps detail the procedure for obtaining $\widehat{\mathcal{G}}(\theta)$. For simplicity, we again employ the simplified t-statistic, denoting $\mathcal{T}^0_{\varphi_i}(b)$ as $\mathcal{T}^0_{\phi}(b)$ similar to (14).

1. Given θ_t, compute $\Delta_{it} = (y_{it} - \widehat{\delta}'\theta_t)^2$ for each i, where $\widehat{\delta}$ is obtained by fitting the time series regression $\overline{y}_t = \alpha_0 + \delta'\theta_t + e_t$, which is statistically significant. (If $\widehat{\delta}$ is not significant, then seek different θ_t.)

2. For each i, run the trend regression

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

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

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

   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 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 $\overline{y}^{(1)}_t = |\widehat{\mathcal{G}}^{(1)}(\theta)|^{-1}\sum_{i\in\widehat{\mathcal{G}}^{(1)}(\theta)} y_{it}$ and obtain $\widehat{\delta}^{(1)}$ from the regression $\overline{y}^{(1)}_t = \alpha^{(1)}_0 + \delta^{(1)\prime}\theta_t + e^{(1)}_t$.

5. Update $\Delta_{it}^{(1)} = (y_{it} - \widehat{\delta}^{(1)\prime}\theta_t)^2$ for each i 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 $\widehat{\mathcal{G}}(\theta)$.

A.4 Enriching Convergent Subgroups

As noted in Section 5, the subgroup estimate $\widehat{\mathcal{G}}(\theta)$ can be enriched by identifying potentially missing convergent members through the following automated procedure. The idea is to include individuals that exhibit the smallest distance from y_{it} to the common trend during the most recent sampling periods. This is motivated by the finding that divergent series often manifest signs of departure toward the end of the sample period (Phillips and Sul, 2007). To quantify this distance, the following procedure utilizes the 'forecast depth' concept developed in Lee and Sul (2023a).

1. For all $i \in \widehat{\mathcal{G}}(\theta)^c$, the complement of the convergent subgroup estimate $\widehat{\mathcal{G}}(\theta)$, let $d_{it} = y_{it} - \widehat{\delta}'\theta_t$ and $d_i^\epsilon = (d_{iT^\epsilon}, \ldots, d_{iT})'$ over the most recent sampling periods $t = T^\epsilon, \ldots, T$, where $T^\epsilon = T - [\epsilon T]$ for some small $\epsilon > 0$ and $\widehat{\delta}$ is from Section A.3. Using this vector of recent deviations d_i^ϵ, calculate the Mahalanobis forecast depth as
$$\mathcal{D}_i = \frac{1}{1 + d_i^{\epsilon\prime} V_\epsilon^{-1} d_i^\epsilon} \quad \text{with} \quad V_\epsilon = \frac{1}{|\widehat{\mathcal{G}}(\theta)^c|} \sum\nolimits_{i \in \widehat{\mathcal{G}}(\theta)^c} d_i^\epsilon d_i^{\epsilon\prime}.$$
(Refer to Lee and Sul (2023b) for alternative depth measures.)

2. Rank the individuals by their forecast depths in descending order. This ranking reflects the proximity of y_{it} to the common trend $\widehat{\delta}'\theta_t$ over $t = T^\epsilon, \ldots, T$ (i.e., the largest depth indicates highest degree of proximity).

3. Holding $\widehat{\delta}$ constant, conduct the sequential t-test $\mathcal{T}_\phi^0(b)$ using either of the following approaches to obtain the enriched subgroup estimate:

   (i) Top-Down Approach: Start with all $i = 1, \ldots, n$, sequentially remove individuals in $\widehat{\mathcal{G}}(\theta)^c$ with the smallest depth, until $\mathcal{T}_\phi^0(b)$ of the extended subgroup falls below the critical value (e.g., -1.96 for $b = 0.1$ and 5% significance level).

   (ii) Bottom-Up Approach: Sequentially add individuals in $\widehat{\mathcal{G}}(\theta)^c$ with the largest depth to $\widehat{\mathcal{G}}(\theta)$ until $\mathcal{T}_\phi^0(b)$ of the extended subgroup exceeds the critical value.

Appendix B: Proofs of the Main Theorems

We assume the following conditions and use technical lemmas whose proofs 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 moments, 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: $T^{-1/2}\sum_{t=1}^{[Tr]} J_{it} \rightsquigarrow B_i^*(r)$ 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 are not cointegrated.

(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: $T^{-1/2}\sum_{t=1}^{[Tr]} J_{0,t} \rightsquigarrow B_0^*(r) = (B_\xi(r), B_{\xi\xi}(r), B_\theta'(r))'$ as $T \to \infty$ for $r \in [0,1]$, where $B_0^*(\cdot)$ is $(2+m)\times 1$ vector Brownian motion with covariance matrix

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

and 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: $T^{-1/2}\sum_{t=1}^{[Tr]} J_{1,t} \rightsquigarrow B_1^*(r) = (B_\xi(r), B_\theta'(r))'$ 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 $t = 1, \ldots, T$, 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 $n, T \to \infty$,

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

Define the partial sum process

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

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

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

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 $n, T \to \infty$,

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

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 a standard 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 $\kappa_* = \kappa_1 \wedge \kappa_2$. As $n, T \to \infty$, the following results hold.

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

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

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 $\kappa_* = 1/4$,

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

(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 following hold.

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

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

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

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

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 $\widehat{u}_t = \widetilde{S}_{n,t} - \widehat{\phi}\widetilde{t}$ is the regression residual in (11).

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

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

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 $\kappa_* = \kappa_1 \wedge \kappa_2$. As $n, T \to \infty$, the following results hold.

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

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

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

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

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

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,

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

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

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

□

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

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

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

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

$$\begin{aligned}
\mathcal{T}_\phi(b) &= \frac{T^{1+2\kappa_*}\widehat{\phi}}{\left\{\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1} T^{-(4-4\kappa_*)}\Psi_{nT}(b)\left(T^{-3}\sum_{t=1}^{T}(\widetilde{t})^2\right)^{-1}\right\}^{1/2}} \\
&\xrightarrow{p} \frac{-\beta_\phi^*}{\left\{(1/12)^{-1}\int_0^1\int_0^1 K\left(\frac{t-s}{b}\right)\widetilde{rs}\lambda^*(r)\lambda^*(s)\,drds\,(1/12)^{-1}\right\}^{1/2}},
\end{aligned}$$

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

Appendix C: Asymptotic Critical Values

Tables 7 and 8 provide one-sided asymptotic critical values for a grid of values of $b \in (0, 1]$, adding to Table 1. The values are the simulated percentiles of the limiting distribution of $\mathcal{T}_\phi(b)$ and $\mathcal{T}^0_\phi(b)$ given in (19) and (20), respectively, from 2 million replications. 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.

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 $\mathcal{T}^0_\phi(b)$ with the Bartlett kernel, under the homoskedasticity restriction.