A Test for Convergence Horizon

Yong Bao[a], Dustin Chambers[b], Xiaotian Liu[c], and Xuewen Yu *[d]

[a]*Department of Economics, Purdue University, USA (ybao@purdue.edu)*
[b]*Department of Economics, Salisbury University, USA (dlchambers@salisbury.edu)*
[c]*Department of Agricultural Economics, Huazhong Agricultural University, China (xiaotianliu@mail.hzau.edu.cn)*
[d]*Department of Applied Economics, Fudan University, China (xuewenyu@fudan.edu.cn)*

July 29, 2024

Abstract

Dynamic panel data models have been extensively employed to test for convergence in economic time series between cross-sectional units of interest (e.g., nations or states). However, an important issue often ignored by empirical researchers is the appropriate choice of the convergence horizon. Rather than employing an objective, data-driven test to determine the appropriate horizon, researchers choose an arbitrary time horizon popular to the literature (e.g., 5, 10, or 20 years). This paper proposes a two-step procedure for determining the convergence horizon. In the first step, we use a recentered method of moments estimator to construct a supreme t test statistic, up to the longest horizon possible, to determine whether there is evidence of convergence. Upon rejection of the null (of no convergence), then in the second step we search over horizons of varying length to determine the convergence horizon. We derive the limiting distribution of the supreme t test and show that the horizon chosen is consistent. Monte Carlo simulations provide evidence that our procedure works well in finite samples. We employ our procedure to study one century of price data (1917–2017) across 18 major U.S. cities and find that price shocks may be longer-lived than expected.

Keywords: dynamic panel; convergence horizon; supreme t

JEL Classification: C23, C13, C26

*Corresponding author.

1. Introduction

Economists have long been interested in the short and long-run changes that occur within markets and economies. Concepts like long-run equilibria, structural change, and transitional dynamics are abundant in the economics literature (Smith, 1776; Marshall, 1890; Keynes, 1936; Solow, 1956). Often, economists are not interested in the evolution of these economic time series in isolation (e.g., for a single country), but rather their parallel changes across multiple units of interest (e.g., a panel of countries). This interest in the parallel co-movement of economic time series has sparked interest in whether or not said data converge toward a common equilibrium or distinctly different equilibria (Chatterji, 1992). Topics for which convergence is widely studied include per capita income levels (Barro, 1991), income inequality (Ravallion, 2003), and prices (Parsley and Wei, 1996), among others.

Despite the diversity of topics considered by the literature, a common drawback to these empirical studies is the lack of an appropriately chosen time horizon to test convergence. Studies that rely upon cointegration methods implicitly assume a pair of time series grow at a common rate in the long-run if a cointegrating relationship exists (Bernard and Durlauf, 1995). However, if one wishes to test for joint convergence among a large number of cross-sectional units, panel unit root tests are employed (Fleissig and Strauss, 2001). Within this dynamic panel framework, the choice of the autoregressive order (τ) and hence the horizon over which convergence is being tested is arbitrarily chosen by the researcher. Popular choices for τ in the literature include 1, 5, 10, and 20 years for annual data.

The purpose of this paper is to introduce a data-driven, consistent procedure for determining the convergence horizon. It involves two steps. In the first step, a supreme t statistic is constructed to test whether there is evidence of convergence, up to the longest horizon possible. This statistic is based on a recentered method of moments (RMM) estimator of the autoregressive parameter in the convergence model for a given τ. The RMM estimator is motivated by the fact that the convergence model is a special case of a general higher-order dynamic panel data model. We follow Bao (2024) to derive the asymptotic properties of the RMM estimator under different asymptotic regimes (large cross-sectional size or long time span or both), stable conditions (whether there is a unit root), and error characteristics (whether there is heteroskedasticity). Relative to other methods in the literature, the RMM estimator performs much better in terms of bias and root mean squared error (RMSE) and its associated inference has better size and power performance, as demonstrated in our Monte Carlo experiments. If the RMM-based supreme t rejects the null of no convergence, then in the second step, we search over convergence horizons of varying length and select the horizon that yields the lowest sum of squared residuals (SSR) (adjusted for the degrees of freedom). We show that this procedure is statistically consistent in selecting the true convergence horizon. Although other statistics like the mean t or supreme Wald statistic may also be used in the first step, our Monte Carlo experiments demonstrate that the supreme t is most reliable.

To illustrate the use of our convergence horizon selection procedure, we investigate consumer price index (CPI) data across 18 major U.S. cities between 1917 and 2017. Our two-step procedure finds strong evidence of convergence in both the price level and inflation rate. While the single-year convergence horizon ($\tau = 1$) chosen by our data-driven approach for price levels is consistent with the existing literature, for the inflation series, the chosen horizon of 32 is *significantly* longer than the typical horizon chosen in existing studies (Kočenda and Papell, 1997; Lopez and Papell, 2012).

The remainder of this paper is organized as follows. We first present the convergence model and assumptions in Section 2, followed by the details of the RMM estimation method and the associated inference, assuming that the convergence horizon is given. Section 3 discusses the RMM-based supreme t test and two-step horizon selection procedure. Monte Carlo simulations and comparisons are reported in Section 4 to assess the performances of the RMM estimator and the horizon selection procedure under various scenarios. Section 5 contains the empirical application, and Section 6 concludes. The proofs of our main theoretical results and other technical details are provided in the appendix.

2. Estimation and Inference with Test Horizon Given

We consider the typical dynamic panel (DP) model used in convergence studies such as Caselli et al. (1996):

$$y_{it} = \alpha_i + \phi y_{i,t-\tau} + u_{it}, \tag{1}$$

where the true parameter value $\phi_0 \in (-1, 1]$, y_{it} is the variable of interest (e.g., per capita income), and τ is the time span over which researchers want to test convergence.[1] If $\phi_0 - 1 < 0$, and hence $\phi_0 < 1$, then the literature concludes there is convergence in y_{it}.[2]

For a given τ, to estimate the convergence model (1) and conduct inference, Caselli et al. (1996) take a τ-th order difference approach and propose using the generalized method of moments (GMM) estimator based on the differenced data (GMM-τ for short):

$$y_{it} - y_{i,t-\tau} = \phi(y_{i,t-\tau} - y_{i,t-2\tau}) + u_{it} - u_{it-\tau}. \tag{2}$$

Bao and Dhongde (2009) instead consider a simple first-order difference ordinary least squares (OLS) estimator (OLS1 for short) which eliminates the need for lagged instruments and greatly preserves the usable size of the sample data:

$$y_{it} - y_{i,t-1} = \phi(y_{i,t-\tau} - y_{i,t-\tau-1}) + u_{it} - u_{it-1}. \tag{3}$$

Recently, Bao and Yu (2023) revisit this strand of literature and propose an estimation strategy that is based on the idea of indirect inference (II). They provide Monte Carlo evidence showing that their II estimator performs better than GMM-τ and OLS1 in finite samples. Note that the convergence model is a special case of a general higher-order DP and Bao and Yu (2023) re-cast the estimation of (1) and associated inference in a more general framework. Their II approach matches the biased within-group (WG) estimator

[1] This could be in terms of deviations from period mean or alternatively, one could add time-fixed effects in the model with unadjusted measures of y_{it}.

[2] When $\phi_0 = 0$, it is obviously true that $\phi_0 - 1 < 0$. However, this should not be interpreted as convergence since $y_{it} = \alpha_i + 0 \cdot y_{i,t-\tau_1} + u_{it}$ and $y_{it} = \alpha_i + 0 \cdot y_{i,t-\tau_2} + u_{it}$ are equivalent for any $\tau_1 \neq \tau_2$. It does not imply divergence either, so we define $\tau = 0$ when $\phi_0 = 0$.

with its analytical approximate expectation for the fixed-effects DP models when N is large but T is finite. Bao (2024) further explores this idea by exploiting correlation between the WG transformed lagged dependent variables and the error term, and develops a recentered method of moments (RMM) estimator. It is shown in Bao (2024) that the RMM estimator has very good finite-sample performance under different asymptotic regimes (either $N \to \infty$ or $T \to \infty$), stable conditions (with or without a unit root), and idiosyncratic error distributions (homoskedasticity or heteroskedasticity of different forms). Since the convergence model is a special case of higher-order DP models, we can apply the RMM approach in Bao (2024) to estimate model (1) for a given horizon τ, which to our knowledge has not been done before. In what follows, we use $\mathrm{tr}(\cdot)$ to denote the matrix trace operator. $\mathrm{Dg}(\cdot)$ creates a diagonal matrix such that it either stacks diagonally all its scalar arguments in order or copies the diagonal elements of a matrix argument.

2.1. Assumptions and Estimation

For each cross-sectional unit i, denote $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})'$, $\boldsymbol{y}_{i,(-\tau)} = (y_{i,1-\tau}, \cdots, y_{i,T-\tau})'$, and $\boldsymbol{u}_i = (u_{i1}, \cdots, u_{iT})'$. Stacking over i, we put $\boldsymbol{y} = (\boldsymbol{y}_1', \cdots, \boldsymbol{y}_N')'$, $\boldsymbol{y}_{(-\tau)} = (\boldsymbol{y}_{1,(-\tau)}', \cdots, \boldsymbol{y}_{N,(-\tau)}')'$, $\boldsymbol{u} = (\boldsymbol{u}_1', \cdots, \boldsymbol{u}_N')'$, and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_N)'$. In matrix notation, (1) can be written as

$$\boldsymbol{y} = (\boldsymbol{I}_N \otimes \mathbf{1}_T)\boldsymbol{\alpha} + \phi \boldsymbol{y}_{(-\tau)} + \boldsymbol{u}, \tag{4}$$

where $\boldsymbol{I}_N$ denotes the identity matrix of size N and $\mathbf{1}_T$ is a $T \times 1$ vector consisting of 1's.

Suppose we use the WG transformation to eliminate the fixed effects, then (4) becomes

$$\boldsymbol{A}\boldsymbol{y} = \phi \boldsymbol{A}\boldsymbol{y}_{(-\tau)} + \boldsymbol{A}\boldsymbol{u}, \tag{5}$$

where $\boldsymbol{A} = \boldsymbol{I}_N \otimes \boldsymbol{M}_T$, $\boldsymbol{M}_T = \boldsymbol{I}_T - T^{-1}\mathbf{1}_T\mathbf{1}_T'$. Applying the OLS procedure to (5) yields the WG estimator $\tilde{\phi}_{WG} = (\boldsymbol{y}_{(-\tau)}'\boldsymbol{A}\boldsymbol{y}_{(-\tau)})^{-1}(\boldsymbol{y}_{(-\tau)}'\boldsymbol{A}\boldsymbol{y})$. Because the regressor $\boldsymbol{A}\boldsymbol{y}_{(-\tau)}$ in (5) is endogenous, i.e., $\mathrm{E}[(\boldsymbol{A}\boldsymbol{y}_{(-\tau)})'(\boldsymbol{A}\boldsymbol{u})] = \mathrm{E}(\boldsymbol{y}_{(-\tau)}'\boldsymbol{A}\boldsymbol{u}) \neq 0$, the WG estimator is inconsistent when T is finite.

We can follow the RMM approach of Bao (2024) by recentering $\boldsymbol{y}_{(-\tau)}'\boldsymbol{A}\boldsymbol{u}$ by its non-zero expectation $\mathrm{E}(\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{u})$ to build a moment condition for estimating ϕ_0. In particular, we first write

$$\boldsymbol{y}_{(-\tau)} = (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u} + (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau\mathbf{1})\boldsymbol{\alpha} + \sum_{j=0}^{\tau-1}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^{\tau-1-j}\boldsymbol{e}_1)\boldsymbol{y}_{-j}, \tag{6}$$

where $\boldsymbol{\Phi}_\tau = \boldsymbol{\Phi}_\tau(\phi_0)$, $\boldsymbol{\Phi}_\tau(\phi) = \boldsymbol{I}_T - \phi\boldsymbol{L}^\tau$, $\boldsymbol{L} = \boldsymbol{L}_T$ is a $T \times T$ strictly lower triangular matrix with 1's on the first sub-diagonals and zero elsewhere, $\boldsymbol{y}_{-j} = (y_{1,-j}, \cdots, y_{N,-j})'$, $j = 0, \cdots, \tau - 1$, collects the cross-sectional initial observations, and $\boldsymbol{e}_1$ is the first column of $\boldsymbol{I}_T$. Under homoskedasticity (see Assumption 1(i) to be introduced) and regular conditions, it follows that

$$\mathrm{E}(\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{u}) = \mathrm{E}[\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u}] = \mathrm{E}[\boldsymbol{u}'(\boldsymbol{I}_N \otimes \boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u}] = -\frac{N\sigma^2}{T}\mathbf{1}'_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau\mathbf{1}_T, \tag{7}$$

where σ^2 is the variance of u_{it}, $\boldsymbol{\Phi}_\tau = \boldsymbol{\Phi}_\tau(\phi_0)$, $\boldsymbol{\Phi}_\tau(\phi) = \boldsymbol{I} - \phi\boldsymbol{L}^\tau$. In deriving (7), we take the expectation of a quadratic form in $\boldsymbol{u}$ and use the fact that $\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau$ is strictly lower triangular, and thus $\mathrm{E}[\boldsymbol{u}'(\boldsymbol{I}_N \otimes \boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u}] = \sum_{i=1}^N \mathrm{E}(\boldsymbol{u}_i'\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}_T^\tau\boldsymbol{u}_i) = N\sigma^2\mathrm{tr}(\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}_T^\tau\boldsymbol{u}_i) = -T^{-1}N\sigma^2\mathbf{1}'\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau\mathbf{1}$. Since $\mathrm{E}(\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}) = N(T-1)\sigma^2$ under homoskedasticity, a natural moment condition is

$$g_{NT}(\phi) = \frac{1}{NT}\big[\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}(\boldsymbol{y} - \phi\boldsymbol{y}_{(-\tau)}) + (\boldsymbol{y} - \phi\boldsymbol{y}_{(-\tau)})'\boldsymbol{A}(\boldsymbol{y} - \phi\boldsymbol{y}_{(-\tau)})h(\phi)\big], \tag{8}$$

where $h(\phi) = [T(T-1)]^{-1}\mathbf{1}'\boldsymbol{\Phi}_\tau^{-1}(\phi)\boldsymbol{L}^\tau\mathbf{1}$, such that $\mathrm{E}(g_{NT}(\phi_0)) = 0$. Accordingly, the RMM estimator is defined by $\hat{\phi} = \arg_\phi\{g_{NT}(\phi) = 0\}$.

If $\mathrm{Var}(u_{it}) = \sigma_{it}^2$, by defining $\boldsymbol{\Sigma}_i = \mathrm{Dg}(\sigma_{i1}^2, \cdots, \sigma_{iT}^2)$, the expectation of the cross moment of the endogenous regressor and error term in the WG transformed model (5) becomes

$$\mathrm{E}(\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{u}) = \mathrm{E}[\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u}] = \sum_{i=1}^N \mathrm{tr}(\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau\boldsymbol{\Sigma}_i). \tag{9}$$

Instead of seeking an estimator of $\boldsymbol{\Sigma}_i$, we follow Bao (2024) by using another quadratic form in $\boldsymbol{u}$ such that its expectation is the same as that of $\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u}$. Specifically, $\mathrm{E}[\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^\tau)\boldsymbol{u}] = \mathrm{E}[\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{\Psi})\boldsymbol{A}\boldsymbol{u}]$, where $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\phi_0)$, $\boldsymbol{\Psi}(\phi) = T(T-$

$2)^{-1}\mathrm{Dg}(\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}(\phi)\boldsymbol{L}^\tau) - [(T-1)(T-2)]^{-1}\mathrm{tr}(\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}(\phi)\boldsymbol{L}^\tau)\boldsymbol{I}_T$. Correspondingly, the re-centered moment condition under heteroskedasticity becomes

$$g_{NT}(\phi) = \frac{1}{NT}\sum_{i=1}^{N}\left[\boldsymbol{y}_{i,-\tau}'\boldsymbol{M}_T(\boldsymbol{y}_i - \phi\boldsymbol{y}_{i,(-\tau)}) - (\boldsymbol{y}_i - \phi\boldsymbol{y}_{i,(-\tau)})'\boldsymbol{M}_T\boldsymbol{\Psi}(\phi)\boldsymbol{M}_T(\boldsymbol{y}_i - \phi\boldsymbol{y}_{i,(-\tau)})\right], \quad (10)$$

which has the property that $\mathrm{E}(g_{NT}(\phi_0)) = 0$. It should be noted that this moment condition is still valid under homoskedasticity.

2.2. Inference

If τ is known, the asymptotic properties of $\hat{\phi}$ can be established under the following set of assumptions.

Assumption 1. *The series of error terms u_{it}, $i = 1, \cdots, N$, $t = 1, \cdots, T$ satisfies one of the following two conditions: (i) it is independent and identically distributed (i.i.d.) across time and individuals with $\mathrm{E}(u_{it}) = 0$, $\mathrm{Var}(u_{it}) = \sigma^2$, and finite moments up to the fourth order, and (ii) it is independent across time and individuals with $\mathrm{E}(u_{it}) = 0$, $\mathrm{Var}(u_{it}) = \sigma_{it}^2$, and finite moments up to the $(4+\eta)$-th order, $\eta > 0$.*

Assumption 2. *The series of fixed effects α_i, $i = 1, \cdots, N$, is i.i.d. across individuals with finite moments up to the fourth order.*

Assumption 3. *The error terms u_{it} and fixed effects α_i are independent for any $i = 1, \cdots, N$, $t = 1, \cdots, T$.*

Assumption 4. *The initial values $y_{i,-\tau}$, $i = 1, \cdots, N$, are either fixed or random. When they are fixed, $y_{i,-\tau} = O(1)$. When they are random, the following hold: (i) $y_{i,-\tau} = O_P(1)$ with finite moments up to the fourth order, (ii) $\mathrm{E}(\alpha_i^{r_1} y_{i,-\tau}^{r_2}) = O(1)$ and $\mathrm{Cov}(\alpha_i^{r_1} y_{i,-\tau}^{r_2}, \alpha_j^{r_1} y_{j,-\tau}^{r_2}) = 0$, $r_1 + r_2 \leq 4$, $r_1 \geq 0$, $r_2 \geq 0$, $i \neq j$, $i, j = 1, \cdots, N$, and (iii) $\mathrm{Cov}(y_{i,-\tau}, u_{it}) = 0$, $t = 1, \cdots, T$, $i = 1, \cdots, N$.*

Note that Assumption 2 is not needed for estimation, since the WG transformation eliminates the fixed effects and the moment condition $g_{NT}(\phi)$ does not involve them. However, it is used, together with the other assumptions, to ensure that g_{NT} and its derivative have properly defined probability limits. The i.i.d. condition in Assumption 2 is only for convenience and can be relaxed. In Assumption 1(ii), the idiosyncratic error term can be heteroskedastic across units and/or over time. The following theorem establishes the asymptotic properties under various asymptotic regimes, stable conditions, and variance schemes. Let $\hat{\phi} = \arg_{\phi}\{g_{NT}(\phi) = 0\}$, $\hat{\boldsymbol{v}} = \boldsymbol{y} - \hat{\phi}\boldsymbol{y}_{(-\tau)}$, and $\hat{\boldsymbol{v}}_i = \boldsymbol{y}_i - \hat{\phi}\boldsymbol{y}_{i,(-\tau)}$, where $g_{NT}(\phi)$ follows (8) under Assumption 1(i) or (10) under Assumption 1(ii).

Theorem 1. *For model* (1) *with true parameter* ϕ_0*, under Assumptions 1–4 and the condition* $\operatorname{plim}(g_{NT}(\phi)) \neq 0$ *for* $\phi \neq \phi_0$ *in the parameter space* $(-1, 1]$*, when either (i)* $N \to \infty$ *with* T *fixed or (ii)* $T \to \infty$*, the RMM estimator* $\hat{\phi}$ *is consistent. Furthermore,*

$$t_{\hat{\phi}} = \frac{\hat{\phi} - \phi_0}{\text{s.e.}(\hat{\phi})} \xrightarrow{d} \mathrm{N}(0, 1), \tag{11}$$

where $\text{s.e.}(\hat{\phi})$ *is defined according to whether* u_{it} *is homoskedastic or heteroskedastic. If Assumption 1(i) applies, then*

$$\text{s.e.}(\hat{\phi}) = \begin{cases} \sqrt{N^{-1}\hat{G}_T^{-2}\left(N^{-1}\sum_{i=1}^{N}\hat{g}_{Ti}^2\right)} & \text{if } N \to \infty \text{ and } T \text{ is fixed} \\ \sqrt{\hat{\sigma}^2(\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{y}_{(-\tau)})^{-1}} & \text{if } T \to \infty \end{cases}, \tag{12}$$

where $\hat{g}_{Ti} = T^{-1}(\boldsymbol{y}'_{i,(-\tau)}\boldsymbol{M}_T\hat{\boldsymbol{v}}_i + \hat{h}\hat{\boldsymbol{v}}'_i\boldsymbol{M}_T\hat{\boldsymbol{v}}_i)$, $\hat{G}_T = (NT)^{-1}[\hat{H}\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}} - \boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{y}_{(-\tau)} - 2\hat{h}\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\hat{\boldsymbol{v}}]$, $\hat{h} = h(\hat{\phi})$, $\hat{H} = [T(T-1)]^{-1}\boldsymbol{1}'\boldsymbol{\Phi}_\tau^{-1}(\hat{\phi})\boldsymbol{L}^\tau\boldsymbol{\Phi}_\tau^{-1}(\hat{\phi})\boldsymbol{L}^\tau\boldsymbol{1}$*, and* $\hat{\sigma}^2 = \hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}}/[N(T-1)]$*. If Assumption 1(ii) applies, then*

$$\text{s.e.}(\hat{\phi}) = \begin{cases} \sqrt{N^{-1}\hat{G}_T^{-2}\left(N^{-1}\sum_{i=1}^{N}\hat{g}_{Ti}^2\right)} & \text{if } N \to \infty \text{ and } T \text{ is fixed} \\ \sqrt{(\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{y}_{(-\tau)})^{-2}\sum_{i=1}^{N}(\boldsymbol{y}'_{i,(-\tau)}\boldsymbol{M}_T\hat{\boldsymbol{v}}_i)^2} & \text{if } T \to \infty \text{ and } N \to \infty \end{cases}, \tag{13}$$

where $\hat{g}_{Ti} = T^{-1}(\boldsymbol{y}'_{i,(-\tau)}\boldsymbol{M}_T\hat{\boldsymbol{v}}_i + \hat{\boldsymbol{v}}'_i\boldsymbol{M}_T\hat{\boldsymbol{\Psi}}\boldsymbol{M}_T\hat{\boldsymbol{v}}_i)$, $\hat{G}_T = (NT)^{-1}\sum_{i=1}^{N}(2\hat{\boldsymbol{v}}'_i\boldsymbol{M}_T\hat{\boldsymbol{\Psi}}\boldsymbol{M}_T\boldsymbol{y}_{i,(-\tau)} - \boldsymbol{y}'_{i,(-\tau)}\boldsymbol{M}_T\boldsymbol{y}_{i,(-\tau)} - \hat{\boldsymbol{v}}'_i\boldsymbol{M}_T\widehat{\nabla\boldsymbol{\Psi}}\boldsymbol{M}_T\hat{\boldsymbol{v}}_i)$, $\hat{\boldsymbol{\Psi}} = \boldsymbol{\Psi}(\hat{\phi})$*, and* $\widehat{\nabla\boldsymbol{\Psi}} = T(T-2)^{-1}\mathrm{Dg}[\boldsymbol{M}_T(\boldsymbol{\Phi}_\tau^{-1}(\hat{\phi})\boldsymbol{L}^\tau)^2] - [(T-1)(T-2)]^{-1}\mathrm{tr}[\boldsymbol{M}_T(\boldsymbol{\Phi}_\tau^{-1}(\hat{\phi})\boldsymbol{L}^\tau)^2]\boldsymbol{I}_T$*.*

The t-statistic in Theorem 1 follows an asymptotic standard normal distribution, making inference straightforward and simple. Under homoskedasticity, the cross-sectional size N can be either fixed or grow as T grows. But under (unconditional) heteroskedasticity, if $T \to \infty$, we need N to grow as well. On the other hand, following Bao (2024), if $T \to \infty$ and N is fixed, the t-statistic with standard error of the form $\sqrt{(\boldsymbol{y}'_{(-\tau)}\boldsymbol{A}\boldsymbol{y}_{(-\tau)})^{-2}\sum_{i=1}^{N}(\boldsymbol{y}'_{i,(-\tau)}\boldsymbol{M}_T\hat{\boldsymbol{v}}_i)^2}$ may still be useful under homoskedasticity (with possible conditional temporal heteroskedasticity) or temporal heteroskedasticity. In this case, $t_{\hat{\phi}} \xrightarrow{d} \sqrt{N/(N-1)}\mathrm{t}_{N-1}$, where t_{N-1} denotes the t distribution with $N-1$ degrees of freedom.

Note that in Theorem 1, when both N and T grow, we do not address whether they grow simultaneously or sequentially. Following Bao (2024, endnote 13), we can show that such a concern is irrelevant for Theorem 1. In the next section, we present the horizon selection procedure and assume a joint asymptotic regime, namely, N and T approach infinity simultaneously.[3]

3. Horizon Determination

The preceding analysis shows how to consistently estimate the convergence measure and conduct statistical inference when the test horizon τ is known and given. However, generally no formal economic theory specifies a definitive value for τ, and empirical researchers often resort to ad hoc choices like 5 or 10 years. In this section, we propose a consistent selection procedure in which the underlying data determines the appropriate test horizon.

Since the convergence model is embodied in a general pth-order dynamic panel model (DP(p)),

$$y_{it} = \alpha_i + \phi_1 y_{i,t-1} + \ldots + \phi_p y_{i,t-p} + u_{it}, \tag{14}$$

if the convergence horizon is truly τ, then the appropriate null hypothesis is,

$$H_0 : \phi_s = 0, \quad s \neq \tau,\ s = 1, \cdots, p. \tag{15}$$

Under the null, y_{it} is related to $y_{i,t-\tau}$ only, subject to the individual effect and idiosyncratic shock. With the true convergence horizon unknown, we can naturally try $\tau = 1, \cdots, p$, where

[3] In the next section, Theorems 2–4 are expected to hold as well under a sequential regime, as all the relevant statistics are essentially in terms of linear and quadratic forms in $\boldsymbol{u}$. However, it is beyond the scope of this paper to provide a rigorous treatment of this topic.

the order p can be interpreted as the maximum horizon supported by economic reasoning. For example, when studying city-level aggregate prices in Section 5, we set $p = 50$, since it is unlikely that monetary policy decisions or supply and demand shocks more than half a century ago have any meaningful direct impact on current price/inflation deviations. As such, we first study the asymptotic properties of the RMM estimator $\hat{\phi}_s$ from the DP-s regression model ($y_{it} = \alpha_i + \phi_s y_{i,t-s} + u_{it}$) when we choose a horizon of s, but the correctly specified model is $y_{it} = \alpha_i + \phi_{0\tau} y_{i,t-\tau} + u_{it}$. In this section, we denote the true parameter $\phi_{0\tau}$ to emphasize that the DP-τ model is a special case of the DP(p) model. This is stated in the following theorem.

Theorem 2. *For model (14), under H_0 and Assumptions 1–4, as $N, T \to \infty$, the RMM estimator $\hat{\phi}_s$ from a DP-s regression has the following properties : (i) if $|\phi_{0\tau}| < 1$, then $\hat{\phi}_s \xrightarrow{p} 0$ when s is not a multiple of τ and $\hat{\phi}_s \xrightarrow{p} \phi_{0\tau}^m$ when $s/\tau = m$ for some positive integer m; (ii) if $\phi_{0\tau} = 1$, then $\hat{\phi}_s \xrightarrow{p} 1$.*

In Theorem 2, we do not specify which recentered moment condition, (8) or (10), is used for $\hat{\phi}_s$. Recall that even if the idiosyncratic errors are homoskedastic, we can still use (10). Next, for a given $\hat{\phi}_s$, we define the sum of squared residuals (SSR) $SSR(s) = (\boldsymbol{y} - \hat{\phi}_s \boldsymbol{y}_{(-s)})' \boldsymbol{A} (\boldsymbol{y} - \hat{\phi}_s \boldsymbol{y}_{(-s)})$ and its adjusted version $SSR^*(s) = SSR(s)/[N(T-s)]$. If the chosen model coincides with the true model, then $\hat{\phi}_s \xrightarrow{p} \phi_{0\tau}$ in view of Theorem 2, and thus we would expect that it best fits the data. Accordingly, we consider the following procedure to select the horizon:

$$\hat{s} = \arg \min_{s \in \{1, \cdots, p\}} SSR^*(s). \tag{16}$$

Theorem 3. *For model (14), as $N, T \to \infty$, under H_0 and Assumptions 1–4, $\hat{s} \xrightarrow{p} \tau$.*

Thus, the selection procedure (16) ensures that with large samples, we will obtain the true test horizon. Of course, a properly defined convergence model only makes sense when $\phi_{0\tau} \neq 0$. If under H_0, it is further the case that if $\phi_{0\tau} = 0$, and thus $y_{it} = \alpha_i + u_{it}$, then a regression model $y_{it} = \alpha_i + \phi_s y_{i,t-\ell} + u_{it}$ gives $\hat{\phi}_\ell \xrightarrow{p} 0$ for any ℓ according to Theorem 2.

Therefore, prior to proceeding with the decision rule based on (16), we need to first reject

$$H_0' : \phi_\ell = 0, \quad \ell = 1, \cdots, p. \tag{17}$$

for ϕ_ℓ in $y_{it} = \alpha_i + \phi_\ell y_{i,t-\ell} + u_{it}$. For this purpose, we can consider a sup-$|t|$ test:

$$\text{sup-}|t| = \sup_\ell |t(\ell)|, \; t(\ell) = t_{\hat{\phi}_\ell}, \quad \ell = 1, \cdots, p, \tag{18}$$

where $t_{\hat{\phi}_\ell}$ is the t-statistic associated with ϕ_ℓ in the regression $y_{it} = \alpha_i + \phi_\ell y_{i,t-\ell} + u_{it}$ estimated by the RMM method.

Theorem 4. *For model (14), as $N, T \to \infty$, and $N/T \to 0$, under H_0' and Assumptions 1–4, $\text{sup-}|t| \xrightarrow{d} \sup_{\ell \in \{1,\ldots,p\}} |Z_\ell|$, where $\{Z_\ell\}_{\ell=1}^p$ is i.i.d. $\mathrm{N}(0,1)$.*

Alternatively, we can adopt a conventional Wald-type test for testing joint significance of all the ϕ parameters in model (14) by using the statistic

$$W = NT\hat{\boldsymbol{\phi}}'\hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\phi}}, \tag{19}$$

where $\hat{\boldsymbol{\phi}}$ is the RMM estimator of $\boldsymbol{\phi}_0$ in Bao (2024) for the higher-order DP model (14) and $\hat{\boldsymbol{V}}$ is the estimated variance of $\sqrt{NT}\hat{\boldsymbol{\phi}}$. It follows an asymptotic chi-squared distribution with p degrees of freedom under H_0' and Assumptions 1–4.

In summary, our horizon selection procedure involves two steps. In the first step, we need to make a decision on H_0' (the null of no convergence for any horizon). If it is rejected, then in the second step we use (16) to determine the appropriate convergence horizon, where each $SSR^*(s)$ is constructed under H_0 (the null of a given horizon).

4. Monte Carlo Simulations

This section presents a set of Monte Carlo experiments that assesses the finite-sample performance of the proposed estimator and the resulting inference procedure against existing methods for estimating the convergence parameter $\rho_0 = (\phi_0 - 1)/\tau$ when the test horizon is given. It also assesses the performance of the test horizon selection procedure. Through-out, the individual effects are simulated as i.i.d. $\mathrm{N}(0,1)$ and independent of the idiosyncratic errors. All the results are based on 10,000 simulations. For the idiosyncratic errors, they may be simulated as i.i.d. $\mathrm{N}(0,1)$ (across i and t). For ease of reference, we call this the homoskedastic case. We also consider $u_{it} = \sqrt{z_{it}} e_{it}$, $e_{it} \sim$ i.i.d. $\mathrm{N}(0,1)$ (across i and t), $z_{it} = t\sqrt{N/i}/T$ if $t\sqrt{N/i}/T \leq 100$, and otherwise z_{it} is randomly drawn from a chi-squared distribution with 10 degrees of freedom. We refer to this case as heteroskedasticity.[4]

4.1. Bias, RMSE, Size, and Power Performances: τ Given

For comparison, we include the WG, GMM-τ, and OLS1 estimators. We use RMM and RMM$_r$ to denote the RMM estimator based on model (8) and its robust version that is based on model (10), respectively. In addition to the bias and RMSE results (scaled up by 100 for each estimator), we also calculate the empirical rejection rate of a 5% two-sided t-test for whether the convergence parameter is equal to its true value. For the WG and GMM-τ estimators, the White standard errors may also be used, and the corresponding entries in the tables are denoted by WG(h) and GMM-τ(h) respectively. Similarly, for the OLS1 estimator of Bao and Dhongde (2009), we use OLS1 and OLS1(h) to distinguish if White standard errors are used in conducting the t-test. The entry RMM(N) signifies the empirical size when the large-N-fixed-T standard error is used and RMM(T) corresponds to the case when the large-T standard error is used. Similar notation is used for RMM$_r$(N) (large-N-fixed-T) and RMM$_r$(NT) (large-N-large-T), whereas RMM$_r$(T) signifies the empirical size when the $\sqrt{N/(N-1)}t_{N-1}$ approximation is used.

Table 1 reports the simulation results when $\tau = 5$ under different combinations of N and T for the homoskedastic case, whereas Table 2 reports results for different combinations of τ and T when $N = 50$. Results in both tables are related to the convergence parameter $\rho_0 = (\phi_0 - 1)/\tau$. We experiment with $\phi_0 = 1, 0.8, 0.5, -0.2$ in Table 1, where the first parameter configuration corresponds to a model of no convergence and the other three imply convergence. In Table 2, we again include a non-convergent data generating process (DGP).

[4] For all the (N,T) combinations in this section, $t\sqrt{N/i}/T$ never exceeds 100.

Table 1. Simulation Results: DP-5 under Homoskedasticity

	(N,T)	(25,50)	(25,100)	(25,200)	(50,50)	(50,100)	(50,200)	(25,50)	(25,100)	(25,200)	(50,50)	(50,100)	(50,200)
		$\phi_0 = 1$						$\phi_0 = 0.5$					
Bias($\times$100)	WG	−0.10	−0.04	−0.01	−0.10	−0.04	−0.01	−0.55	−0.30	−0.15	−0.56	−0.29	−0.15
	GMM-τ	−0.25	−0.13	−0.06	−0.12	−0.06	−0.03	−2.21	−1.53	−0.89	−1.29	−0.89	−0.70
	OLS1	−0.01	−0.01	−0.01	−0.01	0.00	0.00	−0.01	−0.01	0.00	−0.02	0.00	−0.01
	RMM	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	−0.01	0.00	−0.02	0.00	−0.01
	RMM$_r$	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	−0.01	0.00	−0.02	0.00	−0.01
RMSE($\times$100)	WG	0.18	0.08	0.03	0.14	0.06	0.02	0.77	0.47	0.29	0.67	0.38	0.23
	GMM-τ	0.54	0.21	0.08	0.35	0.13	0.05	2.96	1.85	1.08	1.97	1.18	0.84
	OLS1	0.20	0.12	0.07	0.14	0.08	0.05	0.63	0.43	0.30	0.45	0.31	0.21
	RMM	0.14	0.06	0.02	0.10	0.04	0.02	0.54	0.36	0.25	0.38	0.26	0.18
	RMM$_r$	0.14	0.06	0.02	0.10	0.04	0.02	0.54	0.36	0.25	0.38	0.26	0.18
Size(5%)	WG	11.81	11.08	8.89	17.29	16.22	12.41	17.77	12.67	8.85	32.01	19.88	13.81
	WG(h)	14.01	13.95	12.95	17.92	17.90	14.34	20.72	14.74	10.65	33.10	21.19	14.55
	GMM-τ	7.85	11.92	16.55	5.89	8.58	12.05	18.23	30.21	29.68	12.87	19.85	32.51
	GMM-$\tau(h)$	12.84	16.70	20.93	8.55	11.15	14.56	22.87	33.41	32.41	14.44	21.20	33.52
	OLS1	5.30	4.85	5.25	5.25	4.79	5.19	5.34	4.76	4.84	5.05	4.88	4.69
	OLS1(h)	7.85	7.43	8.08	6.38	6.25	6.51	7.33	6.65	6.94	6.15	5.84	5.40
	RMM(N)	7.23	7.55	8.07	5.79	5.80	6.41	6.72	6.17	6.18	5.71	5.79	5.36
	RMM(T)	5.67	5.14	5.19	5.06	5.07	5.18	5.83	5.27	5.03	5.50	5.51	5.03
	RMM$_r$(N)	7.15	7.54	8.01	5.71	5.83	6.40	6.69	6.15	6.19	5.62	5.66	5.37
	RMM$_r$(NT)	7.67	8.07	8.54	5.94	6.17	6.71	7.70	6.87	6.92	6.49	6.29	5.65
	RMM$_r$(T)	5.25	5.74	6.32	4.96	5.12	5.53	5.37	4.77	4.75	5.30	5.22	4.68
		$\phi_0 = 0.8$						$\phi_0 = -0.2$					
Bias($\times$100)	WG	−0.48	−0.31	−0.17	−0.48	−0.31	−0.17	−0.31	−0.16	−0.08	−0.32	−0.16	−0.08
	GMM-τ	−5.18	−2.65	−1.27	−4.07	−1.94	−1.09	−0.63	−0.60	−0.40	−0.32	−0.32	−0.31
	OLS1	−0.03	−0.02	−0.01	−0.02	−0.01	0.00	0.02	0.01	0.01	0.00	0.00	0.00
	RMM	−0.02	−0.02	−0.01	−0.01	−0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	RMM$_r$	−0.02	−0.02	−0.01	−0.01	−0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
RMSE($\times$100)	WG	0.63	0.41	0.25	0.55	0.36	0.21	0.65	0.42	0.29	0.52	0.32	0.21
	GMM-τ	5.85	2.88	1.38	4.72	2.16	1.17	1.58	1.10	0.75	1.10	0.76	0.55
	OLS1	0.44	0.31	0.21	0.31	0.21	0.15	0.70	0.49	0.34	0.50	0.35	0.24
	RMM	0.39	0.26	0.18	0.28	0.19	0.13	0.58	0.40	0.28	0.41	0.29	0.20
	RMM$_r$	0.39	0.26	0.18	0.28	0.19	0.13	0.58	0.40	0.28	0.41	0.29	0.20
Size(5%)	WG	25.11	23.61	16.38	42.81	39.65	27.22	7.78	6.48	5.72	12.24	8.61	7.05
	WG(h)	25.76	24.71	17.97	40.58	39.09	27.62	10.29	8.52	7.90	13.74	9.89	7.79
	GMM-τ	47.26	73.14	72.38	39.20	59.59	76.48	6.60	9.09	9.61	5.95	7.77	10.17
	GMM-$\tau(h)$	53.06	69.74	70.39	42.06	55.66	72.71	10.59	12.87	12.89	7.50	9.38	11.64
	OLS1	4.73	5.20	5.00	4.93	5.35	4.88	4.93	4.83	4.84	4.88	5.31	5.34
	OLS1(h)	7.37	7.59	6.68	6.23	6.27	6.02	6.79	6.70	6.75	5.78	6.13	5.93
	RMM(N)	6.46	6.48	6.23	5.99	6.00	6.14	6.28	6.01	6.34	5.82	5.69	5.65
	RMM(T)	5.74	5.68	5.48	6.10	5.91	6.05	5.04	4.90	4.94	5.14	5.20	5.12
	RMM$_r$(N)	6.26	6.34	6.26	5.86	5.89	6.15	6.29	6.00	6.35	5.80	5.78	5.65
	RMM$_r$(NT)	7.19	7.38	7.13	6.71	6.67	6.61	6.96	6.48	6.80	6.19	6.14	5.95
	RMM$_r$(T)	4.95	5.25	4.76	5.57	5.68	5.79	4.85	4.50	4.80	5.18	5.10	4.92

Table 2. Simulation Results: DP-τ under Homoskedasticity, $N = 50$

	(τ, T)	(5,50)	(10,50)	(15,50)	(5,100)	(10,100)	(15,100)	(5,50)	(10,50)	(15,50)	(5,100)	(10,100)	(15,100)
		$(\phi_0-1)/\tau = 0$						$(\phi_0-1)/\tau = -0.01$					
Bias(×100)	WG	−0.10	−0.03	−0.01	−0.04	−0.02	−0.01	−0.20	−0.08	−0.04	−0.12	−0.07	−0.05
	GMM-τ	−0.11	−0.09	−0.06	−0.06	−0.06	−0.05	−0.52	−1.62	−3.38	−0.38	−1.14	−2.19
	OLS1	0.00	0.00	0.00	−0.01	0.00	0.00	−0.01	0.00	0.00	−0.01	0.00	0.00
	RMM	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00	0.00	0.00	0.00	0.00
	RMM$_r$	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00	0.00	0.00	0.00	0.00
RMSE(×100)	WG	0.14	0.07	0.04	0.06	0.04	0.03	0.25	0.14	0.09	0.15	0.10	0.07
	GMM-τ	0.35	0.48	0.46	0.13	0.18	0.20	0.89	2.65	4.49	0.50	1.46	2.59
	OLS1	0.14	0.08	0.05	0.08	0.05	0.03	0.19	0.12	0.09	0.12	0.08	0.06
	RMM	0.10	0.06	0.04	0.04	0.03	0.03	0.15	0.11	0.08	0.08	0.07	0.05
	RMM$_r$	0.10	0.06	0.04	0.04	0.03	0.03	0.15	0.11	0.08	0.08	0.07	0.05
Size(5%)	WG	16.52	8.16	6.40	16.45	10.63	8.41	28.00	12.49	9.23	30.48	19.64	13.19
	WG(h)	17.42	8.73	7.27	17.68	11.35	9.17	27.20	13.10	9.88	30.08	19.70	13.53
	GMM-τ	6.33	4.17	2.98	8.28	6.42	5.88	10.85	2.71	0.06	22.19	26.06	12.64
	GMM-τ(h)	8.81	7.55	7.42	10.50	8.66	8.88	13.20	14.73	23.83	22.34	26.82	36.73
	OLS1	4.81	4.83	4.89	4.49	5.25	4.83	5.07	4.80	5.22	5.01	5.00	4.91
	OLS1(h)	6.15	6.31	6.08	6.05	6.02	5.86	6.57	5.76	6.32	6.48	5.94	5.52
	RMM(N)	5.71	5.43	5.65	6.49	5.98	5.86	5.84	5.59	6.07	5.95	5.36	5.59
	RMM(T)	5.15	4.80	4.89	5.20	5.26	5.25	5.61	4.92	5.41	5.60	4.93	5.08
	RMM$_r$(N)	5.66	5.43	5.67	6.45	6.00	5.89	5.74	5.54	6.12	5.94	5.28	5.56
	RMM$_r$(NT)	5.91	5.68	5.88	6.74	6.21	6.08	6.11	5.80	6.38	6.46	5.63	5.89
	RMM$_r$(T)	4.86	4.72	4.83	5.62	5.11	5.17	5.19	4.82	5.35	5.31	4.68	4.75
		$(\phi_0-1)/\tau = -0.005$						$(\phi_0-1)/\tau = -0.05$					
Bias(×100)	WG	−0.14	−0.05	−0.03	−0.08	−0.05	−0.03	−0.51	−0.21	−0.11	−0.31	−0.13	−0.08
	GMM-τ	−0.25	−0.34	−0.50	−0.15	−0.26	−0.44	−3.02	−0.95	−0.31	−1.56	−0.62	−0.29
	OLS1	−0.01	0.00	0.00	−0.01	0.00	0.00	−0.01	−0.01	0.00	−0.01	0.00	0.00
	RMM	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00	0.00	−0.01	0.00	0.00
	RMM$_r$	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00	0.00	−0.01	0.00	0.00
RMSE(×100)	WG	0.19	0.10	0.07	0.10	0.07	0.05	0.60	0.29	0.17	0.37	0.19	0.12
	GMM-τ	0.54	0.98	1.38	0.25	0.45	0.73	3.66	1.90	1.13	1.79	0.97	0.65
	OLS1	0.16	0.10	0.07	0.10	0.07	0.05	0.34	0.24	0.17	0.23	0.16	0.12
	RMM	0.12	0.08	0.06	0.06	0.05	0.04	0.30	0.20	0.14	0.20	0.13	0.10
	RMM$_r$	0.12	0.08	0.06	0.06	0.05	0.04	0.30	0.20	0.14	0.20	0.13	0.10
Size(5%)	WG	21.88	9.51	7.52	23.48	15.59	11.31	42.40	19.28	12.28	35.66	17.30	12.04
	WG(h)	22.02	10.28	8.20	23.50	16.25	12.09	40.40	20.12	13.14	36.02	18.36	13.32
	GMM-τ	7.88	3.82	1.10	12.75	10.28	8.87	29.64	6.24	4.68	45.75	11.66	6.81
	GMM-τ(h)	10.40	8.97	9.76	14.45	12.78	13.78	31.38	10.89	8.54	43.48	13.45	9.05
	OLS1	5.40	4.65	5.20	5.14	5.21	4.78	5.12	5.02	4.87	4.69	5.06	4.64
	OLS1(h)	6.29	5.86	6.07	6.49	6.42	5.80	6.47	5.94	5.62	5.87	6.09	5.59
	RMM(N)	6.24	5.28	5.33	5.95	6.26	5.75	5.57	5.57	5.79	5.87	6.19	5.63
	RMM(T)	5.79	4.73	4.76	5.34	5.48	5.24	5.75	5.09	5.12	5.64	5.74	5.10
	RMM$_r$(N)	6.16	5.23	5.33	5.91	6.19	5.78	5.56	5.55	5.79	5.80	6.14	5.60
	RMM$_r$(NT)	6.59	5.52	5.67	6.35	6.51	6.00	6.26	6.02	6.06	6.54	6.75	5.97
	RMM$_r$(T)	5.25	4.53	4.67	5.07	5.52	4.97	5.25	4.87	5.02	5.50	5.58	4.82

Table 3. Simulation Results: DP-5 under Heteroskedasticity

	(N,T)	(25,50)	(25,100)	(25,200)	(50,50)	(50,100)	(50,200)	(25,50)	(25,100)	(25,200)	(50,50)	(50,100)	(50,200)
		$\phi_0 = 1$						$\phi_0 = 0.5$					
Bias(×100)	WG	−0.15	−0.05	−0.01	−0.15	−0.05	−0.01	−0.58	−0.30	−0.16	−0.57	−0.30	−0.15
	GMM-τ	−0.36	−0.16	−0.06	−0.20	−0.09	−0.04	−2.60	−1.83	−1.02	−1.65	−1.17	−0.88
	OLS1	−0.03	−0.02	−0.01	−0.01	−0.01	−0.01	−0.03	−0.01	−0.02	−0.02	−0.01	0.00
	RMM	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	−0.01	0.02	0.01	0.00
	RMM$_r$	−0.01	0.00	0.00	0.00	0.00	0.00	−0.02	0.00	−0.02	−0.01	−0.01	0.00
RMSE(×100)	WG	0.24	0.09	0.03	0.21	0.07	0.02	0.86	0.53	0.35	0.73	0.44	0.27
	GMM-τ	0.66	0.25	0.08	0.44	0.16	0.06	3.42	2.20	1.26	2.41	1.51	1.05
	OLS1	0.30	0.19	0.11	0.23	0.14	0.08	0.75	0.52	0.37	0.56	0.39	0.27
	RMM	0.19	0.08	0.03	0.14	0.06	0.02	0.65	0.44	0.31	0.47	0.32	0.23
	RMM$_r$	0.19	0.08	0.03	0.14	0.05	0.02	0.64	0.44	0.31	0.47	0.32	0.23
Size(5%)	WG	15.36	12.41	8.26	24.98	18.10	11.38	21.91	16.37	14.98	33.70	25.56	19.62
	WG(h)	15.58	14.15	11.00	21.09	17.53	12.28	19.91	14.90	11.72	27.45	19.07	13.05
	GMM-τ	10.37	15.12	16.95	8.44	11.40	15.05	26.65	43.18	40.07	22.58	34.47	49.50
	GMM-$\tau(h)$	14.21	18.88	20.64	10.20	13.35	17.11	27.21	39.70	34.73	18.48	26.99	38.25
	OLS1	8.62	10.32	11.50	10.97	11.10	13.52	8.60	9.14	10.66	9.76	10.49	11.72
	OLS1(h)	8.65	9.19	8.88	7.74	7.62	7.61	8.56	8.13	8.17	7.15	6.84	6.79
	RMM(N)	7.06	7.95	7.64	6.40	6.67	6.75	7.67	7.31	7.22	6.73	6.56	6.53
	RMM(T)	7.44	7.04	5.73	9.10	7.48	6.86	9.83	9.81	10.69	10.95	10.94	12.28
	RMM$_r$(N)	6.65	7.40	7.34	5.70	6.09	6.24	7.13	7.15	7.25	5.90	6.15	6.31
	RMM$_r$(NT)	7.55	8.15	8.07	6.60	6.67	6.64	8.65	8.07	7.94	7.37	7.06	6.94
	RMM$_r$(T)	5.34	6.08	5.89	5.38	5.63	5.68	6.22	5.80	5.88	6.04	5.95	5.87
		$\phi_0 = 0.8$						$\phi_0 = -0.2$					
Bias(×100)	WG	−0.54	−0.32	−0.18	−0.53	−0.32	−0.18	−0.34	−0.15	−0.08	−0.33	−0.16	−0.08
	GMM-τ	−5.28	−2.76	−1.34	−4.19	−2.13	−1.23	−0.83	−0.77	−0.47	−0.45	−0.46	−0.43
	OLS1	−0.04	−0.02	−0.01	−0.02	−0.02	−0.01	0.01	0.02	0.00	0.00	0.01	0.00
	RMM	0.02	0.02	0.00	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.00
	RMM$_r$	−0.03	−0.01	−0.01	−0.02	−0.01	−0.01	0.00	0.01	0.00	0.00	0.01	0.00
RMSE(×100)	WG	0.73	0.46	0.29	0.64	0.40	0.24	0.76	0.50	0.35	0.60	0.39	0.27
	GMM-τ	6.01	3.03	1.48	4.90	2.39	1.33	1.86	1.34	0.90	1.32	0.96	0.71
	OLS1	0.55	0.38	0.27	0.40	0.28	0.20	0.84	0.58	0.42	0.62	0.44	0.31
	RMM	0.49	0.33	0.22	0.36	0.24	0.17	0.70	0.48	0.34	0.51	0.36	0.26
	RMM$_r$	0.49	0.33	0.22	0.36	0.24	0.17	0.70	0.48	0.34	0.51	0.36	0.26
Size(5%)	WG	28.13	26.25	21.16	44.88	41.70	32.25	12.34	10.90	11.35	16.71	14.52	14.04
	WG(h)	24.89	22.94	17.72	35.66	32.31	23.41	12.15	9.65	8.88	13.15	9.96	8.34
	GMM-τ	55.56	77.36	74.99	50.71	71.24	84.18	11.10	16.31	16.25	10.49	15.52	21.19
	GMM-$\tau(h)$	54.37	67.96	65.70	43.67	58.19	71.61	13.18	16.19	14.11	8.72	11.63	14.67
	OLS1	8.45	9.41	10.62	9.55	10.77	11.99	9.05	8.97	10.70	9.84	11.06	11.88
	OLS1(h)	8.69	8.54	8.85	6.83	7.12	7.04	8.56	7.95	8.07	6.85	7.01	6.63
	RMM(N)	7.37	7.60	7.91	6.59	7.18	6.70	7.72	7.48	7.54	6.76	6.90	6.43
	RMM(T)	9.85	10.86	11.13	10.95	12.57	12.91	8.82	9.72	10.32	10.33	11.49	11.95
	RMM$_r$(N)	6.45	6.96	7.37	5.42	5.99	6.23	7.62	7.35	7.47	6.68	6.97	6.54
	RMM$_r$(NT)	8.11	8.60	8.67	7.03	7.77	7.34	8.45	8.06	8.24	7.13	7.39	6.76
	RMM$_r$(T)	5.85	6.03	6.35	5.86	6.33	6.27	6.04	5.78	5.59	6.13	6.11	5.67

Table 4. Simulation Results: DP-τ under Heteroskedasticity, $N = 50$

	(τ, T)	(5,50)	(10,50)	(15,50)	(5,100)	(10,100)	(15,100)	(5,50)	(10,50)	(15,50)	(5,100)	(10,100)	(15,100)
		$(\phi_0 - 1)/\tau = 0$						$(\phi_0 - 1)/\tau = -0.01$					
Bias(×100)	WG	−0.15	−0.06	−0.02	−0.05	−0.03	−0.02	−0.27	−0.11	−0.06	−0.14	−0.09	−0.06
	GMM-τ	−0.20	−0.18	−0.14	−0.09	−0.10	−0.09	−0.80	−2.34	−3.65	−0.49	−1.37	−2.19
	OLS1	−0.02	0.00	0.00	−0.01	0.00	0.00	−0.02	−0.01	0.00	−0.01	0.00	0.00
	RMM	0.01	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.02	0.01	0.00
	RMM$_r$	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	−0.01	0.00	0.00	0.00	0.00
RMSE(×100)	WG	0.20	0.12	0.07	0.07	0.06	0.05	0.35	0.19	0.12	0.18	0.13	0.09
	GMM-τ	0.44	0.62	0.60	0.16	0.22	0.25	1.16	3.42	4.84	0.60	1.67	2.62
	OLS1	0.23	0.13	0.08	0.14	0.08	0.06	0.28	0.18	0.12	0.18	0.12	0.09
	RMM	0.14	0.10	0.07	0.06	0.05	0.04	0.21	0.15	0.10	0.11	0.09	0.07
	RMM$_r$	0.14	0.10	0.07	0.05	0.05	0.04	0.21	0.15	0.10	0.11	0.09	0.07
Size(5%)	WG	24.24	14.30	11.50	17.68	15.69	13.44	35.20	18.94	14.53	33.55	25.54	20.29
	WG(h)	21.07	10.85	8.59	16.51	13.58	10.47	28.12	14.61	10.54	28.04	19.43	15.07
	GMM-τ	8.66	5.32	4.47	11.01	7.63	6.75	17.15	6.45	0.88	33.77	40.25	30.94
	GMM-$\tau(h)$	10.46	8.02	8.20	12.96	9.65	8.87	17.37	20.15	27.45	31.76	34.38	40.18
	OLS1	10.45	10.15	9.42	11.40	10.55	10.12	10.73	9.72	9.41	11.01	10.33	10.24
	OLS1(h)	7.62	7.10	7.01	7.33	7.39	6.82	7.87	6.76	6.83	7.29	7.27	6.79
	RMM(N)	6.38	6.58	6.32	6.47	6.26	6.53	6.55	6.86	6.25	6.90	6.51	6.73
	RMM(T)	8.77	9.50	9.42	7.43	8.80	9.31	10.04	10.03	9.29	9.50	10.65	10.94
	RMM$_r$(N)	5.76	6.32	6.28	5.99	5.98	6.33	5.62	6.59	6.06	5.87	6.19	6.52
	RMM$_r$(NT)	6.45	6.80	6.65	6.59	6.45	6.70	6.86	7.14	6.41	6.93	6.94	7.11
	RMM$_r$(T)	5.32	5.59	5.36	5.52	5.47	5.48	5.64	5.96	5.36	5.72	5.72	5.93
		$(\phi_0 - 1)/\tau = -0.005$						$(\phi_0 - 1)/\tau = -0.05$					
Bias(×100)	WG	−0.21	−0.09	−0.04	−0.09	−0.06	−0.04	−0.56	−0.22	−0.11	−0.32	−0.14	−0.08
	GMM-τ	−0.40	−0.63	−0.89	−0.22	−0.40	−0.67	−3.44	−1.18	−0.39	−1.85	−0.83	−0.38
	OLS1	−0.02	−0.01	0.00	−0.01	−0.01	0.00	−0.03	0.00	0.00	−0.01	0.00	0.00
	RMM	0.02	0.00	0.00	0.01	0.00	0.00	0.03	0.01	0.00	0.02	0.01	0.00
	RMM$_r$	−0.01	−0.01	0.00	0.00	0.00	0.00	−0.02	0.00	0.00	−0.01	0.00	0.00
RMSE(×100)	WG	0.27	0.15	0.10	0.12	0.10	0.07	0.68	0.33	0.20	0.41	0.21	0.14
	GMM-τ	0.70	1.31	1.87	0.30	0.58	0.96	4.15	2.20	1.26	2.12	1.20	0.79
	OLS1	0.25	0.15	0.11	0.16	0.10	0.07	0.44	0.29	0.20	0.30	0.20	0.14
	RMM	0.18	0.13	0.09	0.08	0.07	0.06	0.38	0.25	0.17	0.26	0.16	0.12
	RMM$_r$	0.18	0.13	0.09	0.08	0.07	0.06	0.38	0.25	0.17	0.26	0.16	0.12
Size(5%)	WG	30.65	16.84	13.12	25.38	22.29	16.97	43.60	22.65	16.24	37.31	21.25	16.21
	WG(h)	25.21	13.10	10.05	22.14	17.83	13.13	35.86	17.94	12.78	28.68	16.41	12.07
	GMM-τ	11.45	6.74	2.09	18.65	16.79	15.64	43.20	11.48	7.61	62.00	22.59	12.21
	GMM-$\tau(h)$	12.84	11.53	12.15	19.89	16.97	17.49	36.51	14.11	9.45	49.25	18.76	11.91
	OLS1	10.40	9.34	10.21	11.33	10.77	10.08	9.45	9.39	9.21	10.99	9.73	9.42
	OLS1(h)	7.66	7.10	7.63	7.62	7.78	7.07	7.21	6.84	6.68	7.46	6.53	6.83
	RMM(N)	6.36	6.66	6.80	6.76	6.62	6.94	6.77	6.58	6.40	7.27	6.63	6.21
	RMM(T)	9.28	9.27	9.90	8.57	9.40	10.10	11.11	10.04	9.74	12.02	10.54	9.89
	RMM$_r$(N)	5.80	6.28	6.74	6.12	6.07	6.62	5.82	6.01	6.12	6.33	6.31	6.17
	RMM$_r$(NT)	6.52	6.86	7.01	6.96	6.80	7.13	7.36	6.84	6.59	7.84	7.04	6.53
	RMM$_r$(T)	5.54	5.51	5.89	5.70	5.66	6.01	6.33	5.63	5.49	6.68	5.88	5.52

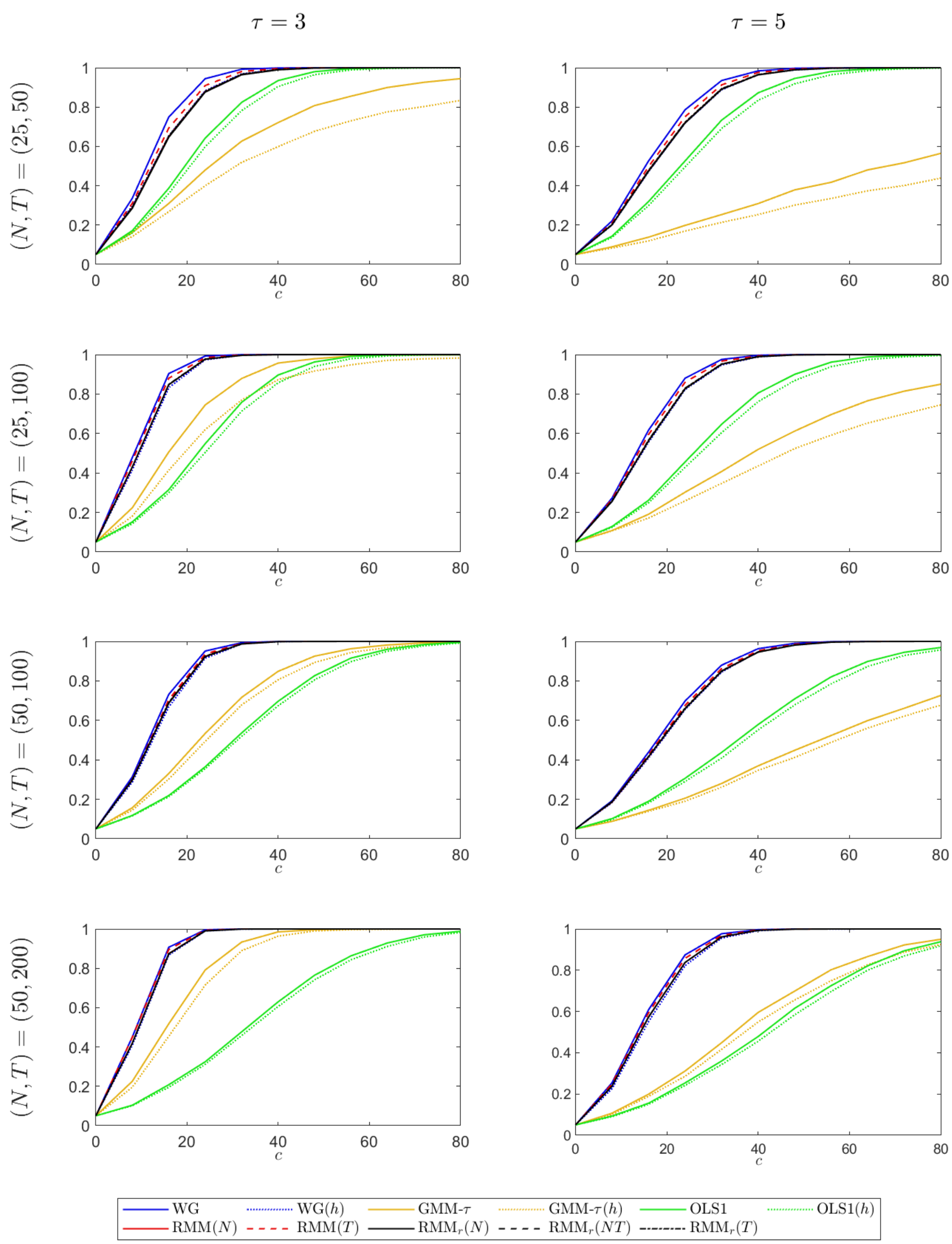
We observe that, in terms of bias and RMSE, the RMM-based methods uniformly dominate the other estimators in all cases. The GMM-τ estimator has the worst performance with respect to bias and RMSE. The WG estimator shows smaller bias and RMSE than GMM-τ but still exhibits much worse performance than the other three, especially in terms of bias. The OLS1 estimator performs reasonably well, usually slightly worse than RMM and RMM$_r$. In terms of hypothesis testing, OLS1, RMM, and RMM$_r$ rank at the top, with their empirical sizes very close to the nominal size, across different parameter configurations and combinations of N,T, and τ. Keep in mind that the robust version RMM$_r$ is still valid under homoskedasticity. In contrast, inference based on WG and GMM-τ perform very poorly, with the t-test from GMM-τ usually exhibiting the most severe upward size distortion.

Tables 3 and 4 present parallel results under heteroskedasticity. Overall, we observe similar patterns of performance across the various estimators as described above. Notably, the robust RMM$_r$ has better overall size performance than its non-robust version RMM. This is consistent with the fact that the DGP contains heteroskedastic errors. Its size performance is also better than that of OLS1.

We further investigate the local power performance of each method when testing the convergence parameter. We set $\rho_0 = -c/(NT)$, $c = 0, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80$, $\tau = 3, 5, 10, 15$, and $(N, T) = (25, 50), (25, 100), (50, 100), (50, 200)$. Specifically, the power comparison results are obtained from testing for the null of $\rho_0 = 0$ (no convergence) against the one-sided alternative $\rho_0 < 0$ (convergence). Figure 1 presents the results when $\tau = 3$ or 5, and Figure 2 reports the results when $\tau = 10$ or 15, both under homoskedasticity. Figures 3 and 4 display results analogous to Figures 1 and 2 under heteroskedasticity. To make a fair comparison, all the finite-sample power results are size-adjusted with the actual rejection rate at 5% when $c = 0$.

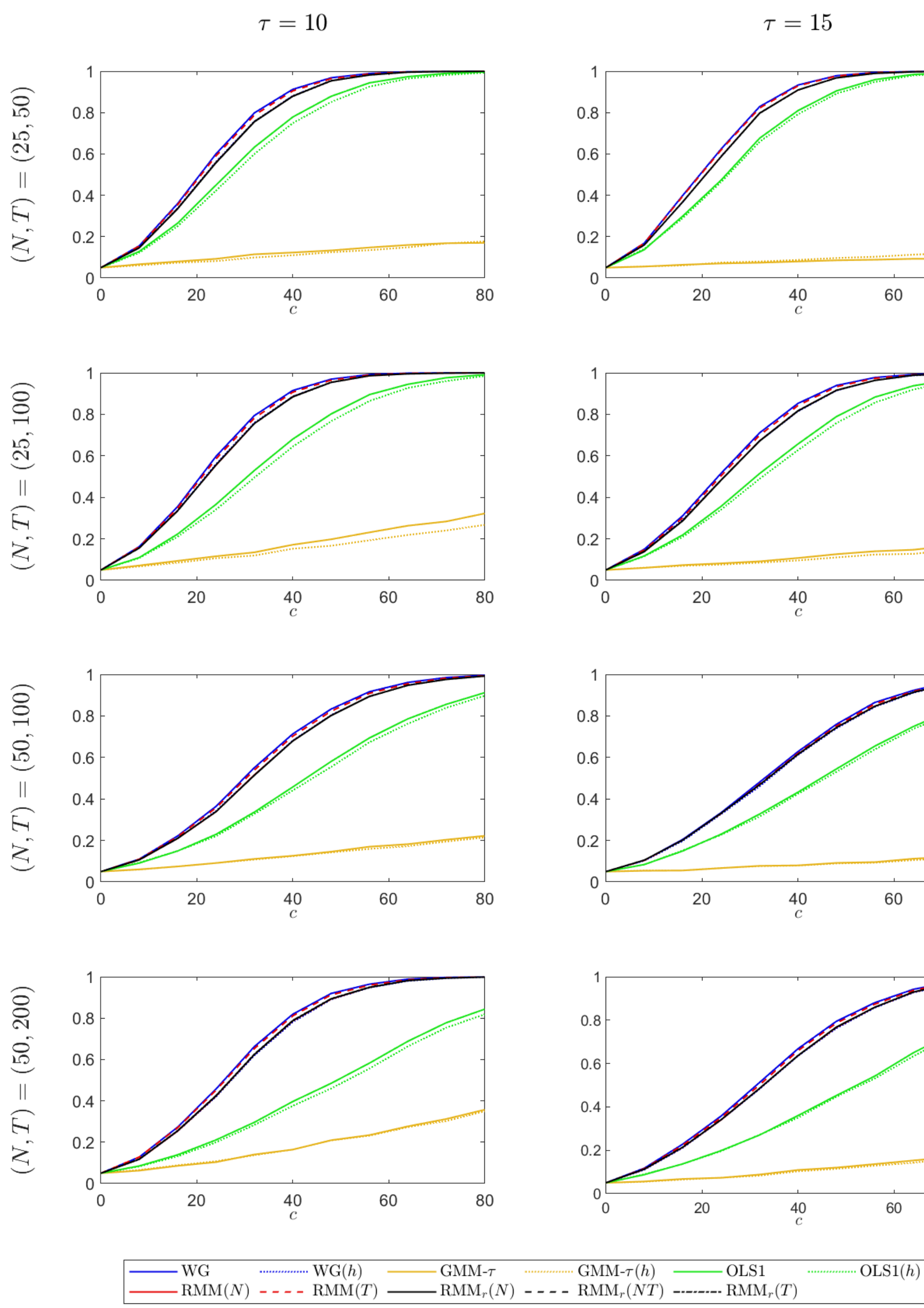
It is clear from Figure 1 that the RMM- and WG-based inference procedures have the highest size-adjusted power in every scenario, usually quite close to each other. There is no definitive ranking when comparing OLS1-based and GMM-based inferences in terms of size-

Figure 1. Size-Adjusted Power: $\tau = 3, 5$, Homoskedastic Case



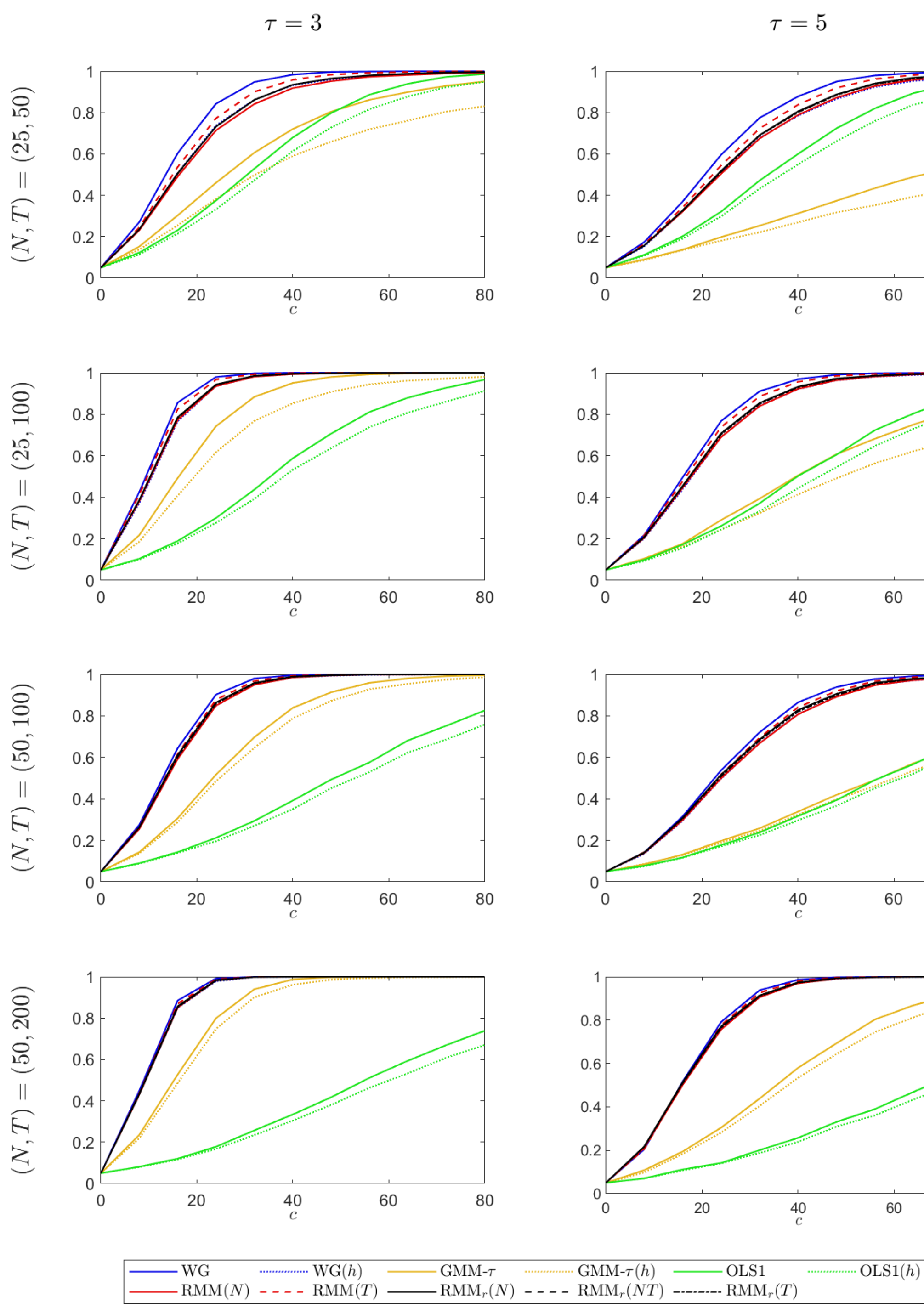
adjusted power under the shorter horizons of 3 or 5 periods, but OLS1 dominates GMM-τ under the longer horizons of 10 or 15 periods. Furthermore, the power disparities between

Figure 2. Size-Adjusted Power: $\tau = 10, 15$, Homoskedastic Case



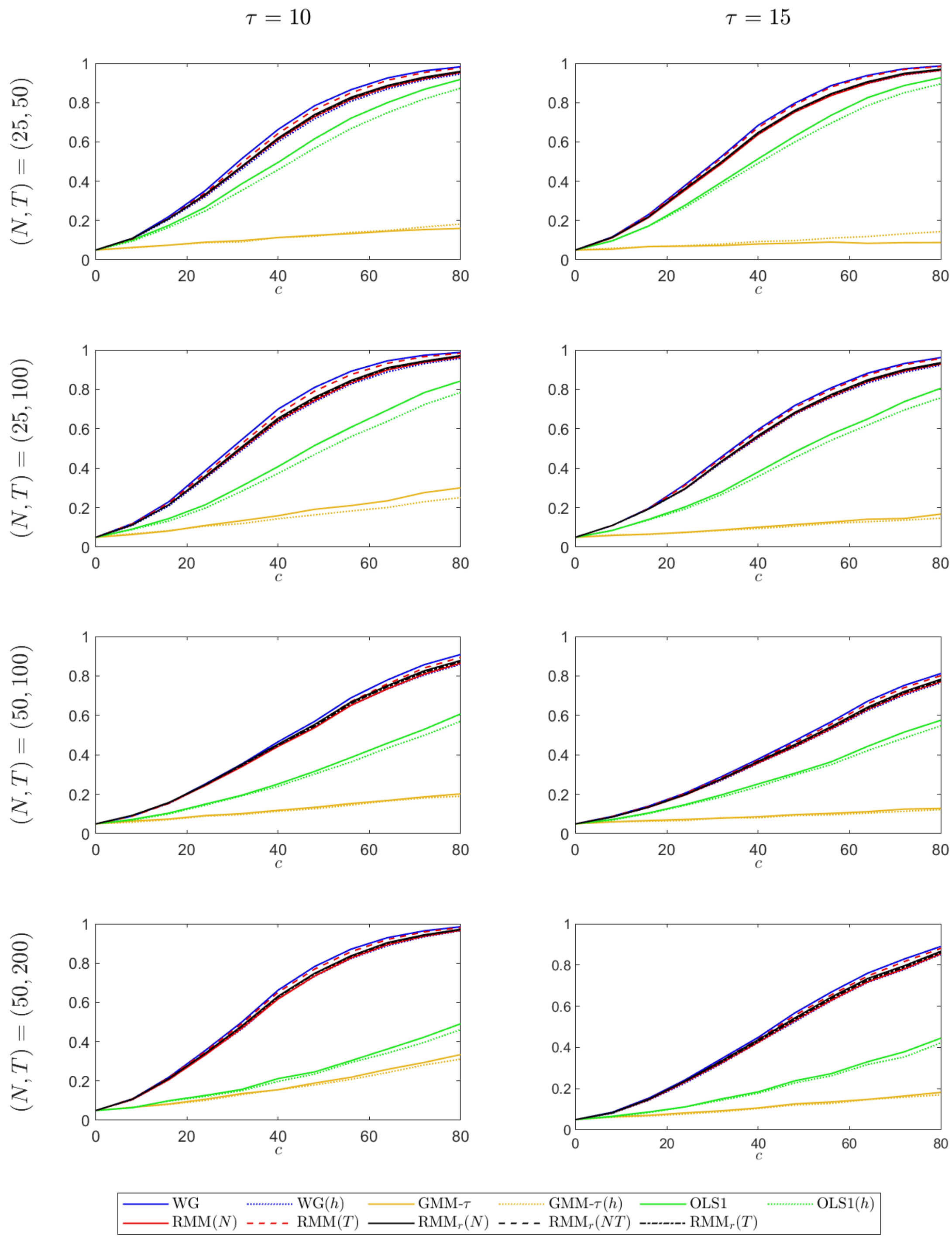
the top performers (RMM and WG) and worst (GMM-τ) are more pronounced under the longer horizons.

Figure 3. Size-Adjusted Power: $\tau = 3, 5$, Heteroskedastic Case



Comparing Figure 2 with Figure 1, notice that the power is lower under longer horizons than under shorter horizons. This is to be expected, as increasing τ obviously reduces the

Figure 4. Size-Adjusted Power: $\tau = 10, 15$, Heteroskedastic Case



effective sample size over time. When the idiosyncratic error is heteroskedastic, Figures 3 and 4 provide largely similar findings. Again, the RMM- and WG-based inference procedures possess the highest power. Recall that the WG approach delivers severe size distortions.

4.2. Test Horizon Selection

In this subsection, we present the finite-sample performance of our RMM-based test horizon selection procedure as described in Section 3 for a variety of DGPs. Specifically, the proposed procedure works as follows. First, we conduct a 5% sup-$|t|$ test, as defined in (18), up to a maximum horizon set by p. Second, upon rejection, we search through a continuous range of discrete horizons to find the horizon that possesses the smallest adjusted SSR as defined in (16); otherwise, the procedure stops with a conclusion that there is neither convergence nor divergence and thus the true horizon is 0. In this spirit, Tables 5 and 6 report the results under homoskedasticity and heteroskedasticity, respectively. As in the previous subsection, RMM(T) and RMM$_r$(NT) signal what corresponding estimated standard errors are used in constructing the statistic in the first step, namely, large-T for RMM and large-N-large-T for RMM$_r$. Recall that Theorem 1 indicates the large-T inference is valid for RMM if N is also large and Theorems 2–4 require both N and T to be large. In the second step, RMM(T) and RMM$_r$(NT) signal that the corresponding adjusted SSR are constructed from the RMM and RMM$_r$ estimators, respectively.

In Table 5, the horizon selection results are reported (with horizon ranging from 0 to 10) when the true horizon is 5 (i.e., DP-5). We consider combinations of the following parameter values and sample sizes: $\phi_0 = 0, 0.5, 0.8, 1, -0.2$, $(N,T) = (25,50), (25,100), (25,200), (50,50), (50,100), (50,200)$. We first examine the results when $\phi_0 = 0$. The selection rate from RMM(T) concentrates around 95% at horizon 0 for all cases. This corroborates the asymptotic result derived in Theorem 4: the rejection rate under the null, namely, the probability of type-I error, approaches the nominal level 5%. In comparison, the robust version RMM$_r$(NT) slightly under-selects in this case, which is in line with the slight upward size distortion for the t-test from RMM$_r$(NT) under homoskedasticity in Tables 1 and 2. Next, when $\phi_0 \neq 0$, the selection frequency of choosing the true horizon of 5 virtually hits 100%, with the exception of slight under-selection when $\phi_0 = -0.2$. The correct selection frequencies of RMM(T)

and $\text{RMM}_r(NT)$ are very close to each other across the board when $\phi_0 \neq 0$.

Table 6 presents parallel results for the heteroskedastic case. When $\phi_0 \neq 0$, the findings from this table are the same as the homoskedastic counterparts shown in Table 5. When $\phi_0 = 0$, both $\text{RMM}(T)$ and $\text{RMM}_r(NT)$ under-select a horizon of 0, but under relatively larger N and T, $\text{RMM}_r(NT)$ performs better than $\text{RMM}(T)$. Recall that for inference, $\text{RMM}_r(NT)$ needs both N and T to be large.

Note that the sup-$|t|$ test in the first step of our selection algorithm may be replaced with a simple Wald-type test as described in (19). Moreover, as is common in the structural breaks literature (e.g., Bai and Perron, 1998), the mean-type of the joint $|t|$ tests may also be used.[5] So we include both variants for comparison in Table 7, where the frequencies of rejecting $H_0' : \phi_\ell = 0, \ \ell = 1, \cdots, p$ at 5% in the first step of our selection procedure are reported, out of the same experiments that produce the results in Tables 5 and 6. In total, based on $\text{RMM}(T)$ and $\text{RMM}_r(NT)$, we have six tests under the three variants. We use, for instance, sup-$|t|$-$\text{RMM}(T)$ to denote the sup-$|t|$ statistic when the large-T standard errors from RMM are used in constructing the t-ratios. We first look at the case $\phi_0 = 0$ under homoskedasticity. It is evident that all three of the $\text{RMM}(T)$-based tests yield very good empirical rejection rates, each being near 5%. For the $\text{RMM}_r(NT)$-based tests, while the empirical rejection rates from sup-$|t|$ and mean-$|t|$ are very close to each other, somewhat above 5%, the Wald test over-rejects substantially. Turning to the heteroskedastic case, we again find that the Wald test undesirably rejects much more than the other tests. The $\text{RMM}_r(NT)$-based sup-$|t|$ and mean-$|t|$ tests give empirical rejection rates closer to the nominal 5% when N and T are relatively larger. For the cases when $\phi_0 \neq 0$, both the sup-$|t|$ and mean-$|t|$ tests reject at a frequency of 100% or close to 100%. However, the rejection rates from Wald-$\text{RMM}(T)$ are much smaller than those from the sup-$|t|$ and mean-$|t|$ tests, especially for the homoskedastic case. These results strongly suggest that the simple Wald-type test would not be very reliable if used in the first step of our horizon selection procedure.

[5] To be specific, the mean-$|t|$ test is defined by: mean-$|t| = p^{-1} \sum_{\ell=1}^{p} |t(\ell)|$, where p is the longest horizon under consideration.

Table 5. Selection Frequencies ($\times 100\%$) in DP-5: Homoskedastic Case

	s	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
ϕ_0	(N,T)	RMM(T)											RMM$_r$(NT)										
0	(25,50)	94.57	0.73	0.61	0.50	0.46	0.27	0.46	0.35	0.40	0.71	0.94	88.26	2.07	1.15	0.81	0.94	0.52	1.09	0.98	0.81	1.36	2.00
	(25,100)	94.75	0.77	0.66	0.42	0.31	0.33	0.35	0.39	0.48	0.54	1.00	89.04	1.74	1.18	0.91	0.73	0.73	0.93	0.77	1.12	0.98	1.87
	(25,200)	94.86	0.93	0.54	0.39	0.37	0.37	0.41	0.31	0.43	0.47	0.91	89.36	2.31	1.07	0.66	0.85	0.87	0.83	0.60	0.83	0.89	1.75
	(50,50)	94.78	0.72	0.53	0.44	0.40	0.25	0.35	0.55	0.53	0.49	0.97	91.29	1.48	0.76	0.69	0.62	0.53	0.56	0.95	0.56	0.83	1.74
	(50,100)	95.01	0.82	0.47	0.47	0.35	0.26	0.33	0.30	0.53	0.65	0.79	92.52	1.37	0.63	0.61	0.58	0.37	0.63	0.56	0.56	0.91	1.25
	(50,200)	95.37	0.90	0.50	0.33	0.28	0.40	0.31	0.29	0.38	0.57	0.67	92.55	1.47	0.66	0.64	0.48	0.59	0.52	0.59	0.62	0.73	1.16
0.5	(25,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
0.8	(25,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
1	(25,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
-0.2	(25,50)	0.00	0.08	0.14	0.00	0.00	98.66	0.02	0.06	0.14	0.37	0.54	0.00	0.08	0.14	0.00	0.00	98.62	0.02	0.06	0.14	0.37	0.58
	(25,100)	0.00	0.00	0.00	0.00	0.00	99.98	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	99.98	0.00	0.00	0.02	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	99.98	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	99.98	0.00	0.00	0.00	0.00	0.02
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00

Table 6. Selection Frequencies ($\times 100\%$) in DP-5: Heteroskedastic Case

	s	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
ϕ_0	(N,T)	RMM(T)											RMM$_r$(NT)										
0	(25,50)	85.64	11.71	1.90	0.46	0.11	0.09	0.04	0.02	0.00	0.02	0.00	84.71	12.70	1.93	0.44	0.15	0.04	0.00	0.00	0.00	0.02	0.00
	(25,100)	84.81	14.34	0.79	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00	84.76	14.54	0.63	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	81.09	18.67	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	85.23	14.58	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	81.79	16.90	0.96	0.23	0.10	0.02	0.00	0.00	0.00	0.00	0.00	88.71	10.48	0.62	0.15	0.02	0.02	0.00	0.00	0.00	0.00	0.00
	(50,100)	79.97	19.78	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	89.40	10.43	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	77.48	22.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	89.92	10.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	(25,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
0.8	(25,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
1	(25,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
−0.2	(25,50)	0.08	12.27	0.78	0.00	0.02	86.85	0.00	0.00	0.00	0.00	0.00	0.66	11.86	0.76	0.00	0.02	86.71	0.00	0.00	0.00	0.00	0.00
	(25,100)	0.00	0.71	0.02	0.00	0.00	99.27	0.00	0.00	0.00	0.00	0.00	0.00	0.71	0.02	0.00	0.00	99.27	0.00	0.00	0.00	0.00	0.00
	(25,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	(50,50)	0.00	5.81	0.11	0.00	0.00	94.08	0.00	0.00	0.00	0.00	0.00	0.00	5.81	0.11	0.00	0.00	94.08	0.00	0.00	0.00	0.00	0.00
	(50,100)	0.00	0.05	0.00	0.00	0.00	99.95	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	99.95	0.00	0.00	0.00	0.00	0.00
	(50,200)	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00

Table 7. Rejection Frequencies ($\times 100\%$) of Alternative Tests in DP-5

	Statistic	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
ϕ_0	(N,T)	Homoskedastic Case						Heteroskedastic Case					
0	(25,50)	5.43	11.74	5.24	9.02	21.15	53.75	14.36	15.29	16.38	13.68	32.40	66.82
	(25,100)	5.25	10.96	3.92	7.49	20.83	52.44	15.19	15.24	17.73	12.25	34.03	67.95
	(25,200)	5.14	10.64	4.27	7.67	21.91	53.51	18.91	14.77	19.56	12.07	34.60	68.85
	(50,50)	5.22	8.71	4.94	7.05	29.41	35.42	18.21	11.29	20.37	9.29	43.14	48.46
	(50,100)	4.99	7.48	4.51	6.00	29.37	35.40	20.03	10.60	21.08	8.28	44.09	48.69
	(50,200)	4.63	7.45	3.92	5.52	28.92	33.92	22.52	10.08	23.05	7.22	43.31	47.20
0.5	(25,50)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(25,100)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.91
	(25,200)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.95
	(50,50)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,100)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,200)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.8	(25,50)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(25,100)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(25,200)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,50)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,100)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,200)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1	(25,50)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(25,100)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(25,200)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,50)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,100)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(50,200)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
-0.2	(25,50)	100.00	100.00	93.83	94.95	67.59	89.40	99.92	99.34	94.63	90.54	72.88	92.11
	(25,100)	100.00	100.00	99.92	99.85	66.89	88.53	100.00	100.00	99.83	98.53	73.28	92.61
	(25,200)	100.00	100.00	100.00	100.00	67.98	89.91	100.00	100.00	100.00	99.96	73.04	92.13
	(50,50)	100.00	100.00	99.89	99.82	91.97	95.62	100.00	100.00	99.86	98.55	92.09	95.46
	(50,100)	100.00	100.00	100.00	100.00	92.54	95.68	100.00	100.00	100.00	99.98	93.22	95.33
	(50,200)	100.00	100.00	100.00	100.00	92.16	95.41	100.00	100.00	100.00	100.00	93.64	95.77

Note: (1)–(6) stand for the sup-$|t|$-RMM(T), sup-$|t|$-RMM$_r$(NT), mean-$|t|$-RMM(T), mean-$|t|$-RMM$_r$(NT), Wald-RMM(T), and Wald-RMM$_r$(NT) statistics, respectively.

5. Empirical Application: U.S. City-Level Prices

Studies of purchasing power parity (PPP) among countries using national price indices (adjusted for exchange rate differences) typically find evidence of moderately slow convergence in aggregate prices, wherein a deviation from PPP typically has an estimated half-life of about four years (Frankel and Rose, 1996). This slow rate of PPP convergence across countries despite the high volatility of short-term real exchange rates has been labeled the "PPP Puzzle" by Rogoff (1996). Researchers wanting to better understand the dynamics of price adjustments without the confounding influence of exchange rate volatility focus on the convergence of regional price indices in common currency zones. Given the relatively young age of the Eurozone, most of these studies focus on city-level aggregate prices in the United States because of better availability of long-run data (Cecchetti et al., 2002).

In their seminal study, Cecchetti et al. (2002) find evidence of convergence in price indices for 19 major U.S. cities over the period 1918 to 1995. However, the speed of convergence was even *slower* than international panel studies, with an estimated convergence half-life of nine years. This result is puzzling given the elimination of exchange rate volatility in a common currency zone, which should have accelerated the convergence process. Various explanations provided by the literature for this slow pace of convergence between cities include high transportation costs, geographical barriers, local monopoly power, differences in the speed of adjustment among cities, sticky prices, and the inclusion of non-traded goods in the price indices (Engel and Rogers, 2001; Cecchetti et al., 2002; Busetti et al., 2006). In a subsequent study, Chen and Devereux (2003) find strong evidence of price level convergence among 19 U.S. cities from 1918 to 2000, with an estimated half-life of five years. Huang et al. (2012) find similar results when applying the OLS approach developed by Bao and Dhongde (2009) to study convergence in 17 major U.S. cities from 1918 to 2008. They find a half-life of convergence between five to seven years depending upon the choice of time horizon τ in their dynamic panel model. In a recent study, Belaire-Franch (2020) utilize the same data but examine it differently, looking for persistent changes using the structural break tests developed by Kejriwal et al. (2013). The paper, which only considers time lags of up to four years, finds weak evidence of change in persistence (between stationary and unit root processes) in the deviation of city level prices measured relative to Chicago, but acknowledges that a panel framework would improve the power of the underlying tests. Given the mixed findings of the preceding studies and the ad hoc approach to time horizon selection, this topic is an ideal application of our horizon selection method.

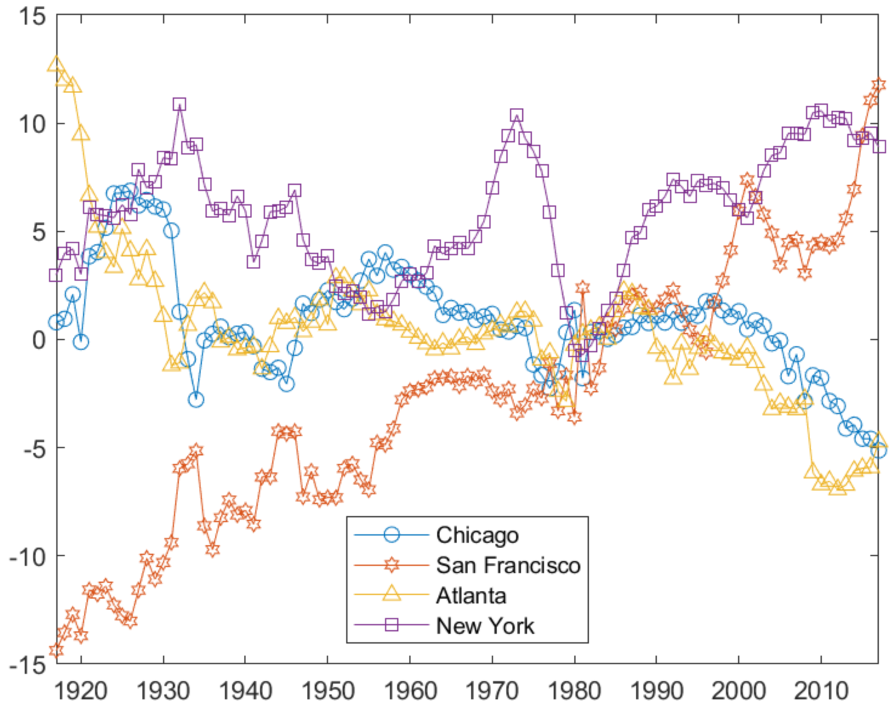
The dataset used by Cecchetti et al. (2002) originally consisted of 19 major U.S. cities' data from 1918 to 1995. We directly downloaded the latest version of this dataset from the Bureau of Labor Statistics.[6] The original 19 major cities are Atlanta, Baltimore, Boston, Chicago, Cincinnati, Cleveland, Detroit, Houston, Los Angeles, Milwaukee, Minneapolis, New York, Philadelphia, Pittsburgh, Portland, San Francisco, Saint Louis, Seattle, and Washington, D.C. However, since the Bureau of Labor Statistics dropped coverage for Cincinnati, Cleveland, Milwaukee, Pittsburgh, and Portland starting in 2017, we limit our focus to the period 1917–2017. Furthermore, we exclude Milwaukee from the analysis since coverage for that city begins in 1935. To summarize, we have a sample consisting of 101 observations on CPI from 18 cities.

We first investigate the price level in each city relative to the cross-sectional average for each year. Following Cecchetti et al. (2002), we plot deviations in the $(100\times)$ log CPI series for Chicago, San Francisco, Atlanta, and New York from the annual cross-sectional mean of all cities in Figure 5. The patterns in this plot are very similar to those reported in Figure 1 of Cecchetti et al. (2002) and suggest that shocks to U.S. city prices can persist for many years. However, the range of price movements at the end of our sample in 2017 is about twice as wide as that of the 1995 terminal values in Cecchetti et al. (2002). As such, we expect our expanded data set will yield a longer estimated half-life than Cecchetti et al. (2002).

We can show that the DP-τ model used in Caselli et al. (1996) is equivalent to a restricted version of the augmented Dicky-Fuller regression used in Cecchetti et al. (2002), but has

[6] While searching for the data on the Bureau's website, we occasionally came across two vintages that contained data on overlapping time periods. In such cases, we merged the two vintages to create a comprehensive dataset for the analysis in this paper.

Figure 5. Deviations in Log CPI from the Cross-Sectional Average



different interpretations.[7] The dynamic panel model in Caselli et al. (1996) is designed to test for β convergence – i.e., if a change in the variable of interest over a given horizon is inversely related to its starting value, while the unit-root test in the augmented Dicky-Fuller regression in Cecchetti et al. (2002) tests for stationarity in the variable of interest. While stationarity is a necessary condition for β convergence, confirming the stationarity of a time series does not reveal the appropriate horizon over which to test for β convergence, hence the need for a data-driven method to select appropriate convergence horizons.

Fitting the DP-τ model (1) to the log CPI data (measured in deviations from cross-sectional mean) and conducting the two-step procedure described in Section 3, we strongly reject the null hypothesis of no convergence (see Panel A in Table 8). Searching over $SSR^*(s)$ for $s = 1, \cdots, 50$ to determine the appropriate horizon, we find that $\hat{s} = 1$. Panel B in Table 8 reports the estimated convergence parameter, standard error, and the corresponding half-life

[7] To illustrate, consider a time series model $y_t = \phi_4 y_{t-4} + u_t$. It is equivalent to $\Delta y_t = (\phi_4 - 1)y_{t-1} - \phi_4 \Delta y_{t-1} - \phi_4 \Delta y_{t-2} - \phi_4 \Delta y_{t-3} + u_t$.

from each method under the chosen horizon 1.

Table 8. CPI Levels in 18 U.S. Cities: Empirical Results

Panel A: Test Statistics (sup-$|t|$ and mean-$|t|$)			under Possible Horizon up to 50			
	RMM(N)	RMM(T)	RMM$_r$(N)	RMM$_r$(NT)	RMM$_r$(T)	95% C.V.
sup-$|t|$	54.95	108.54	55.31	80.13	77.87	3.29
mean-$|t|$	4.53	15.49	5.19	5.22	5.07	0.94

Panel B: Convergence Measure Estimates and Standard Errors under Horizon 1					
	WG	GMM	OLS1	RMM	RMM$_r$
$\hat{\rho}$	−0.071125	−0.062490	—	−0.048407	−0.045626
s.e.	0.008750	0.009004	—		
s.e.(White)	0.013871	0.012343	—		
s.e.(large N)				0.017319	0.020241
s.e.(large N, T)				0.008767	0.011911
s.e.(large T, fixed N)					0.012256
Half-Life	9.39	10.74	—	13.97	14.84

Note: OLS1 is not applicable when the convergence horizon is 1.

The DP-1 convergence model selected by our procedure implies that a city's current relative CPI is affected by all of its past shocks up to the current year since a first-order autoregressive process can be alternatively specified as an infinite-order moving average process. Our estimation results stand in sharp contrast to those in Huang et al. (2012), where ad-hoc τ values of 5, 10, and 15 years are considered. Furthermore, our RMM- and RMMr-based estimation results reveal that it takes about 14 years for the impact of a current shock to a city's price level to decline by 50%, which is much longer than the half-life estimate reported in Cecchetti et al. (2002), and much longer than one would anticipate. This slow pace of convergence is likely due to multiple factors including transportation costs, differences in the speed of adjustment among cities, and especially the inclusion of non-traded goods in the price indices (Cecchetti et al., 2002). Given long-run differences in development patterns across U.S. regions and cities (e.g., San Francisco versus Detroit), and the resulting differences in new job formation, average wages, in and out migration, and ultimately demand for non-tradable goods and services like housing, it is not surprising that we observe persistent differences in aggregate prices. This result is consistent with

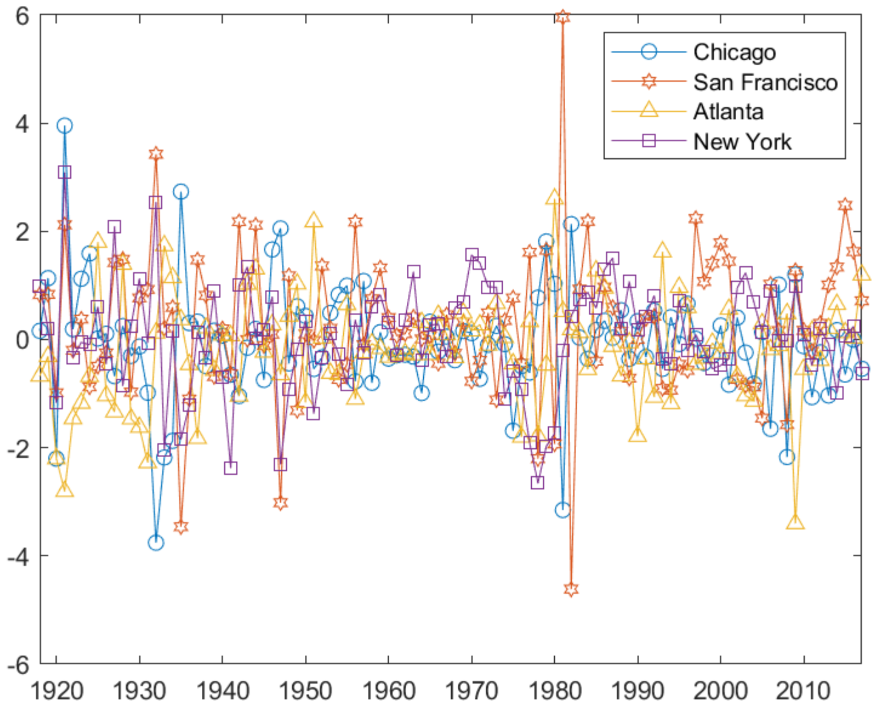
Engel and Rogers (2001) who find high and persistent variability in disaggregated relative consumer prices in 29 U.S. cities between 1986 and 1996. They attribute their finding to high transportation costs and sticky nominal prices.

According to Busetti et al. (2006), to determine if price levels have converged or are in the process of converging, there must be convergence in both the price level and inflation series (vis-a-vis unit root tests).[8] This underscores that the state of this complex and dynamic system of prices is jointly determined by both the price levels and their rates of change. However, even if inflation is stationary, we still know nothing about the dynamics of the β convergence process. This fact is clearly displayed in Figure 6, which plots deviations in cities' CPI growth from the cross-sectional average rate of inflation. There are no obvious cycles in the data and deviations from the cross-sectional inflation rate are more fleeting than deviations in the price level data.

By itself, inflation convergence has been an important topic of interest for those studying the European Monetary Union (EMU) because a common currency and hence a common nominal interest rate will lead to differences in real interest rates among member countries if their inflation rates differ. In such a situation, slower growing countries with excess production capacity and low inflation rates will face higher real interest rates, thus increasing the cost of capital and further slowing growth. At the same time, better performing countries with less economic slack and hence higher inflationary pressure will face lower real interest rates, further stimulating growth. This could lead to a sharp and persistent economic divergence within the common currency zone (Busetti et al., 2007). The consensus among researchers is that there is strong evidence of inflation convergence among EMU members, but the range of half-life estimates varies widely from four months to eight years (Kočenda and Papell, 1997; Busetti et al., 2007; Lopez and Papell, 2012). Although similar studies have been conducted for OECD countries (Lee and Wu, 2001), Italian provinces (Busetti et al., 2006), and East African Community nations (Dridi and Nguyen, 2019), we are unaware of

[8]To determine if price levels *have* converged *or* are in the *process* of converging, additional stationarity tests on the price levels are required. See Figure 2 in Busetti et al. (2006) for more details.

Figure 6. Deviation in CPI Growth from Cross-Sectional Average



any similar studies for U.S. cities.

Therefore, we conduct the same two-step horizon selection exercise over U.S. city inflation rates, expressed in deviations from cross-sectional averages. Panel A of Table 9 strongly rejects the null hypothesis of no convergence. This is not surprising since stationarity in price necessarily implies stationarity in inflation. When we search for the convergence horizon, $SSR^*(s)$ reaches its lowest value at $s = 32$ for $s = 1, \cdots, 50$. While the horizon is quite long, implying that inflation contains a very low-frequency covariance structure, it paradoxically hints at a relatively short half-life of shocks since contemporaneous shocks do not persist in the short-run. Viewed as an impulse response, current shocks rapidly dissipate but lead to long-run echoes in the distant future. Panel B of Table 9 supports this intuition, with a reported half-life around 6 years. This estimated half-life is consistent with the range of values reported by Lopez and Papell (2012) for member states of the European Union (EU).

Table 9. CPI Growth in 18 U.S. Cities: Empirical Results

Panel A: Test Statistics (sup-$\|t\|$ and mean-$\|t\|$)						
	RMM(N)	RMM(T)	RMM$_r$(N)	RMM$_r$(NT)	RMM$_r$(T)	95% C.V.
sup-$\|t\|$	5.31	4.42	5.11	5.36	5.21	3.29
mean-$\|t\|$	1.23	1.15	1.23	1.27	1.23	0.94

Panel B: Convergence Measure Estimates and Standard Errors under Horizon 32					
	WG	GMM	OLS1	RMM	RMM$_r$
$\hat{\rho}$	−0.030523	−0.032087	−0.030965	−0.030373	−0.030338
s.e.	0.000699	0.003808	0.000859		
s.e.(White)	0.000679	0.004479	0.000774		
s.e.(large N)				0.000700	0.000696
s.e.(large N, T)				0.000699	0.000680
s.e.(large T, fixed N)					0.000700
Half-Life	5.90	6.13	4.72	6.21	6.28

6. Conclusions

Dynamic panel models have been the empirical workhorse of convergence studies going back to the 1990s. Unfortunately, the order of these dynamic panel models and hence their convergence horizon were often chosen arbitrarily, thus drawing into question the validity of their underlying convergence test results. To address this issue, we propose a two-step procedure for determining if there is (i) evidence of convergence and if so, (ii) the appropriate time horizon length.

In the first step, we use a recentered method of moments estimator to construct a supreme t test statistic, up to the longest horizon possible, to determine whether there is evidence of convergence. In the second step, we search over convergence horizons of varying length and select the horizon that yields the lowest sum of squared residuals (adjusted for the degrees of freedom). We show that this procedure is consistent in selecting the true convergence horizon and also performs well in finite samples vis-a-vis Monte Carlo simulations.

We use our procedure to study aggregate price indices for 18 major U.S. cities from 1917 to 2017 and find strong evidence of convergence in the price level and inflation rate. Our price level half-life estimate of 14 years is greater than the largest prior estimate of nine years in Cecchetti et al. (2002), which relied on a sample that ended in 1995. Although Huang et al. (2012) estimate a smaller half-life (5–7 years), they rely on longer time horizons (τ) in their empirical DP-τ model, which are not supported by our data-driven horizon selection procedure. With regard to inflation convergence, we find that the convergence horizon is quite long (32 years) *but* the corresponding half-life of 6 years is quite short and in-line with reported inflation convergence results in the EU.

For future research, our method for estimating the DP-τ model should be extended to handle serial correlation and/or cross-sectional dependence in the idiosyncratic shocks. For example, Panizza (2001) and Lin and Huang (2011, 2012a,b) find strong evidence of convergence in income inequality among U.S. states. However, Ho (2015) finds evidence of cross-sectional dependence and claims that this short-run covariance in income distribution is being mistaken for long-run convergence. An extended version of our convergence horizon estimator would be well suited to investigate this open empirical question.

Appendix

The appendix collects technical details and proofs of the four theorems in the main text. For ease of presentation, we introduce some further notation: $a_1 \equiv a_2(\text{mod}, c)$ means that two integers a_1 and a_2 are congruent modulo c, while $a_1 \not\equiv a_2(\text{ mod }, c)$ indicates that two integers a_1 and a_2 are not congruent modulo c. For a variable δ, we write $\delta = \delta_0 + o(\delta_0)$ (when it is nonrandom) or $\delta = \delta_0 + o_p(\delta_0)$ (when it is random) as $\delta = \delta_0 + s.o.$, where $s.o.$ represents a term of smaller order (in probability). For a fixed quantity a, denote $a^+ = \max(a, 0)$. Define $\underline{\sigma}_i^2 = \min_{1\le t\le T}\{\sigma_{it}^2\}$ and $\bar{\sigma}_i^2 = \max_{1\le t\le T}\{\sigma_{it}^2\}$. The element-by-element product (Hadarmard) operator is $\odot$, and $\text{dg}(\cdot)$ collects in order the diagonal elements of its matrix argument as a column vector.

A. Some Useful Lemmas

Lemma 1. *For any integers k_1, k_2 that satisfy $0 \le k_1 \le T$, $0 \le k_2 \le T$, we have: (i) $\text{tr}(\boldsymbol{L}^{k_1}\boldsymbol{L}^{k_2}) = 0$; (ii) $\text{tr}(\boldsymbol{L}^{k_1\prime}\boldsymbol{L}^{k_2}) = T - k$ if $k_1 = k_2 = k$, and equal to 0 otherwise; (iii)*

$\mathbf{1}'\boldsymbol{L}^{k_1}\mathbf{1} = T - k_1$; *(iv)* $\mathbf{1}'\boldsymbol{L}^{k_1}\boldsymbol{L}^{k_2}\mathbf{1} = T - \min\{k_1 + k_2, T\}$; *(v)* $\mathbf{1}'\boldsymbol{L}^{k_1\prime}\boldsymbol{L}^{k_2}\mathbf{1} = \mathbf{1}'\boldsymbol{L}^{k_2}\boldsymbol{L}^{k_1\prime}\mathbf{1} = T - \max\{k_1, k_2\}$.

Proof. These results follow directly from the definition of $\boldsymbol{L}$. □

Lemma 2. *Suppose* $|\phi_0| < 1$, $0 \le l \le T$, *and* $0 \le s \le T$. *Then* $\mathbf{1}'\boldsymbol{\Phi}_s^{-1}\boldsymbol{L}^s\mathbf{1} = O(T)$, $\mathbf{1}'\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\mathbf{1} = O(T)$, $\mathrm{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) = O(T)$ *if* $l \equiv s(\mathrm{mod}, \tau)$ *and* $|l - s|/\tau = m$, *where* m *is some positive integer, and* $\mathrm{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) = 0$ *if* $l \not\equiv s(\mathrm{mod}, \tau)$. *Moreover, for* $0 \equiv s(\mathrm{~mod~}, \tau)$ *and* $s/\tau = m$, $\phi_0^m\mathrm{tr}(\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Sigma}_i)/T - \mathrm{tr}(\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i)/T = O(T^{-1})$ *and* $\mathrm{tr}(\boldsymbol{L}^{s\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i)/T - \mathrm{tr}(\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Sigma}_i)/T = O(T^{-1})$.

Proof. Both $\mathbf{1}'\boldsymbol{\Phi}_s^{-1}\boldsymbol{L}^s\mathbf{1} = O(T)$ and $\mathbf{1}'\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\mathbf{1} = O(T)$ follow from Lemma A2 and the proof of Lemma A8 in Bao (2024) if we treat the convergence model as a special case of the DP(τ) model. Next, $\mathrm{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) = [\mathrm{dg}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s)]'[\mathrm{dg}(\boldsymbol{\Sigma}_i)]$ by noting that $\boldsymbol{\Sigma}_i$ is diagonal. Thus, in what follows, we focus on the properties of the diagonal elements of $\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s$. Using $\boldsymbol{\Phi}_\tau^{-1} = \sum_{i=0}^\infty \phi_0^i\boldsymbol{L}^{i\tau}$, we first write $\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s = \sum_{i=0}^\infty\sum_{j=0}^\infty \phi_0^{i+j}\boldsymbol{L}^{i\tau+l\prime}\boldsymbol{L}^{j\tau+s}$. From Lemma 1, the diagonal elements of $\boldsymbol{L}^{i\tau+l\prime}\boldsymbol{L}^{j\tau+s}$ are nonzeros if and only if $i\tau + l = j\tau + s$. Thus, if $l \not\equiv s(\mathrm{~mod~}, \tau)$, the resulting $\mathrm{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i)$ is 0. If $l \equiv s(\mathrm{mod}, \tau)$ and $|l - s|/\tau = m$, without loss of generality, assuming $s \ge l$, we have $s = l + \tau m$ and

$$
\begin{aligned}
\mathrm{tr}\left(\sum_{i=0}^\infty\sum_{j=0}^\infty \phi_0^{i+j}\boldsymbol{L}^{i\tau+l\prime}\boldsymbol{L}^{j\tau+s} \times \boldsymbol{\Sigma}_i\right) &= \mathrm{tr}\left(\sum_{i=0}^\infty\sum_{j=0}^\infty \phi_0^{i+j}\boldsymbol{L}^{i\tau+l\prime}\boldsymbol{L}^{(j+m)\tau+l} \times \boldsymbol{\Sigma}_i\right) \\
&= \mathrm{tr}\left(\sum_{j=0}^\infty \phi_0^{2j+m}\boldsymbol{L}^{(m+j)\tau+l\prime}\boldsymbol{L}^{(m+j)\tau+l} \times \boldsymbol{\Sigma}_i\right) \\
&= \sum_{j=0}^\infty \phi_0^{2j+m}(\mathbf{1}'_{T-(m+j)\tau-l}, \mathbf{0}'_{(m+j)\tau+l})'[\mathrm{dg}(\boldsymbol{\Sigma}_i)] \\
&= \sum_{j=0}^\infty \phi_0^{2j+m}\sum_{t=1}^{T-(m+j)\tau-l}\sigma_{it}^2 \\
&\ge \underline{\sigma}_i^2\sum_{j=0}^\infty \phi_0^{2j+m}[T - (m+j)\tau - l]^+ \\
&= \underline{\sigma}_i^2\left[\frac{\phi^m(T - l - \tau m)}{1 - \phi_0^2} - \frac{\tau\phi_0^{m+2}}{(1 - \phi_0^2)^2}\right] + s.o. = O(T).
\end{aligned}
$$

Obviously, we can bound $\sum_{j=0}^{\infty}\phi_0^{2j+m}\sum_{t=1}^{T-(m+j)\tau-l}\sigma_{it}^2$ by $\sum_{j=0}^{\infty}\phi_0^{2j+m}\sum_{t=1}^{T-(m+j)\tau-l}\bar{\sigma}_i^2$, which is the same as $\bar{\sigma}_i^2\sum_{j=0}^{\infty}\phi_0^{2j+m}[T-(m+j)\tau-l]^+ = O(T)$. Finally, again using $\boldsymbol{\Phi}_\tau^{-1} = \sum_{i=0}^{\infty}\phi_0^i\boldsymbol{L}^{i\tau}$ and Lemma 1, we have

$$\begin{aligned}
&\left|\frac{\phi_0^m}{T}\mathrm{tr}(\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Sigma}_i) - \frac{1}{T}\mathrm{tr}(\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i)\right| \\
&= \left|\frac{1}{T}\left[\phi_0^m\sum_{j=0}^{\infty}\phi_0^{2j}\sum_{t=1}^{T-j\tau}\sigma_{it}^2 - \sum_{j=0}^{\infty}\phi_0^{2j+m}\sum_{t=1}^{T-(m+j)\tau}\sigma_{it}^2\right]\right| \\
&= \frac{1}{T}\sum_{j=0}^{\infty}\phi_0^{2j+m}\sum_{t=T-(m+j)\tau}^{T-j\tau}\sigma_{it}^2 \\
&\le \frac{\bar{\sigma}_i^2}{T}\sum_{j=0}^{\infty}\phi_0^{2j+m}m\tau = \frac{m\tau\phi_0^m\bar{\sigma}_i^2}{T(1-\phi_0^2)} = O(T^{-1})
\end{aligned}$$

and

$$\begin{aligned}
&\left|\frac{1}{T}\mathrm{tr}(\boldsymbol{L}^{s\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) - \frac{1}{T}\mathrm{tr}(\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Sigma}_i)\right| \\
&= \left|\frac{1}{T}\left[\sum_{j=0}^{\infty}\phi_0^{2j}\sum_{t=1}^{T-(m+j)\tau}\sigma_{it}^2 - \sum_{j=0}^{\infty}\phi_0^{2j}\sum_{t=1}^{T-j\tau}\sigma_{it}^2\right]\right| \\
&= \left|-\frac{1}{T}\sum_{j=0}^{\infty}\phi_0^{2j}\sum_{t=T-(m+j)\tau}^{T-j\tau}\sigma_{it}^2\right| \\
&\le \frac{\bar{\sigma}_i^2}{T}\sum_{j=0}^{\infty}\phi_0^{2j}m\tau = \frac{m\tau\bar{\sigma}_i^2}{T(1-\phi_0^2)} = O(T^{-1})
\end{aligned}$$

hold for $0 \equiv s(\mathrm{mod}, \tau)$ and $s/\tau = m$. □

Lemma 3. *Suppose* $\phi_0 = 1$, $0 \le l \le T$, *and* $0 \le s \le T$. *Then* $\boldsymbol{1}'\boldsymbol{\Phi}_s^{-1}\boldsymbol{L}^s\boldsymbol{1} = O(T^2)$, $\boldsymbol{1}'\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{1} = O(T^3)$, $\boldsymbol{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Sigma}_i\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{1} = O(T^3)$, $\mathrm{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) = O(T^2)$ *if* $l \equiv s(\mathrm{mod}, \tau)$ *and* $|l-s|/\tau = m$, *and* $\mathrm{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) = 0$ *if* $l \not\equiv s(\mathrm{mod}, \tau)$.

Proof. Lemma A3 and the proof of A16 in Bao (2024) give straightforwardly the results $\boldsymbol{1}'\boldsymbol{\Phi}_s^{-1}\boldsymbol{L}^s\boldsymbol{1} = O(T^2)$, $\boldsymbol{1}'\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{1} = O(T^3)$, and $\boldsymbol{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Sigma}_i\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{1} = O(T^3)$ if we interpret the convergence model as a special case of DP(τ) model. Similar to the proof of

Lemma 2, when $l \equiv s(\text{mod}, \tau)$ and $|l-s|/\tau = m$, we have, assuming $s \geq l$,

$$\begin{aligned}
\text{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i) &= \text{tr}\left(\sum_{j=0}^{\infty}\boldsymbol{L}^{(m+j)\tau+l\prime}\boldsymbol{L}^{(m+j)\tau+l}\boldsymbol{\Sigma}_i\right) \\
&= \sum_{j=0}^{\infty}(\mathbf{1}'_{T-(m+j)\tau-l}, \mathbf{0}'_{(m+j)\tau+l})'[\text{dg}(\boldsymbol{\Sigma}_i)] \\
&= \sum_{j=0}^{\infty}\sum_{t=1}^{T-(m+j)\tau-l}\sigma_{it}^2 \\
&\geq \underline{\sigma}_i^2\sum_{j=0}^{\infty}[T-(m+j)\tau-l]^+ \\
&= \frac{1}{2}[2(T-l-\tau m)-k\tau](k+1)\underline{\sigma}_i^2 + s.o. = O(T^2),
\end{aligned}$$

where $k = \lfloor (T-m\tau-l)/\tau \rfloor$. Likewise, given $\sum_{j=0}^{\infty}\sum_{t=1}^{T-(m+j)\tau-l}\sigma_{it}^2 \leq \sum_{j=0}^{\infty}\sum_{t=1}^{T-(m+j)\tau-l}\bar{\sigma}_i^2$, we can show the upper bound of $\sum_{j=0}^{\infty}\sum_{t=1}^{T-(m+j)\tau-l}\sigma_{it}^2$ is also $O(T^2)$. If $l \not\equiv s(\text{mod}, \tau)$, $\text{tr}(\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{\Sigma}_i)$ is obviously equal to 0 in view of Lemma 1. □

Lemma 4. *When $|\phi_0| < 1$, under Assumptions 1–4, (i) $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} = O_p(NT)$; (ii) for $0 \not\equiv s(\bmod, \tau)$, $\boldsymbol{y}'_{(-s)}\boldsymbol{A}\boldsymbol{y} = O(N)+O_p(\sqrt{NT})$; (iii) for $0 \equiv s(\bmod, \tau)$ and $s/\tau = m$, $\boldsymbol{y}'_{(-s)}\boldsymbol{A}(\boldsymbol{y}-\phi_0^m\boldsymbol{y}_{(-s)}) = O(N)+O_p(\sqrt{NT})$, $(\boldsymbol{y}-\phi_0^m\boldsymbol{y}_{(-s)})'\boldsymbol{A}(\boldsymbol{y}-\phi_0^m\boldsymbol{y}_{(-s)}) = O_p(NT)$.*

Proof. (i) follows from Lemma A5 of Bao (2024) if we treat DP-τ as a special case of DP(τ). To prove (ii), we note that for $0 \leq s \leq \tau$, $\boldsymbol{y}_{(-s)} = (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s)\boldsymbol{u} + (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\mathbf{1})\boldsymbol{\alpha} + \sum_{j=0}^{s-1}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^{s-1-j}\boldsymbol{e}_1)\boldsymbol{\mathcal{y}}_{(-j)} + \phi_0\sum_{j=s}^{\tau-1}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^{\tau+s-1-j}\boldsymbol{e}_1)\boldsymbol{\mathcal{y}}_{(-j)}$, and for $s > \tau$, $\boldsymbol{y}_{(-s)} = (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s)\boldsymbol{u} + (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\mathbf{1})\boldsymbol{\alpha} + \sum_{j=0}^{\tau-1}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^{s-1-j}\boldsymbol{e}_1)\boldsymbol{\mathcal{y}}_{(-j)} + \sum_{j=\tau}^{s-1}(\boldsymbol{I}_N \otimes \boldsymbol{L}^{s-1-j}\boldsymbol{e}_1)\boldsymbol{\mathcal{y}}_{(-j)}$, where $\boldsymbol{\mathcal{y}}_{(-j)} = (y_{1,-j}, \cdots, y_{N,-j})'$, $j = 0, \cdots, \tau-1$, collects the cross-sectional initial observations.[9] Setting $s = 0$ yields $\boldsymbol{y} = (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1})\boldsymbol{u} + (\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\mathbf{1})\boldsymbol{\alpha} + \phi_0\sum_{j=0}^{\tau-1}(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^{\tau-1-j}\boldsymbol{e}_1)\boldsymbol{\mathcal{y}}_{(-j)}$. With the expressions of $\boldsymbol{y}_{(-s)}$ and $\boldsymbol{y}$, the leading term in $\boldsymbol{y}'_{(-s)}\boldsymbol{A}\boldsymbol{y}$ is $[(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s)\boldsymbol{u}]'\boldsymbol{A}[(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1})\boldsymbol{u}]$, with its expectation

$$\text{E}\left\{[(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s)\boldsymbol{u}]'\boldsymbol{A}[(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1})\boldsymbol{u}]\right\}$$

[9] We can derive $\boldsymbol{y}_{(-s)}$ from Equation (5) of Bao (2024) by treating the convergence model as a special case of the DP(τ) model.

$$
\begin{aligned}
&= \sum_{i=1}^{N} \mathrm{E}[\boldsymbol{u}_i' \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{u}_i] \\
&= \sum_{i=1}^{N} \left[\mathrm{tr}(\boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{\Sigma}_i) - \frac{1}{T} \mathbf{1}' \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \mathbf{1} \right] \\
&= \sum_{i=1}^{N} [0 - O(1)] = O(N),
\end{aligned}
$$

where we have used the results in Lemma 2. Further,

$$
\begin{aligned}
&\mathrm{Var}([(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{L}^s)\boldsymbol{u}]' \boldsymbol{A} [(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1})\boldsymbol{u}]) \\
&= \sum_{i=1}^{N} \mathrm{Var}(\boldsymbol{u}_i' \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{u}_i) \\
&= \sum_{i=1}^{N} \Big[\mathrm{tr}(\boldsymbol{\Sigma}_i^{(4)} \odot \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \odot \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1}) \\
&\quad + \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1}) \\
&\quad + \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{L}^s) \Big],
\end{aligned}
$$

where $\boldsymbol{\Sigma}_i^{(4)} = \mathrm{Dg}(\mathrm{E}(u_{i1}^4) - 3\sigma_{i1}^4, \cdots, \mathrm{E}(u_{iT}^4) - 3\sigma_{iT}^4))$. We can verify that both $\mathrm{tr}(\boldsymbol{\Sigma}_i^{(4)} \odot \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \odot \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1})$ and $\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1}$ are $O(1)$, but $\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{L}^s)$ is $O(T)$, by using (ix), (x), (xi) of Lemma A2 in Bao (2024). So $\boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y} = \mathrm{E}(\boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y}) + \boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y} - \mathrm{E}(\boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y}) = O(N) + O_p(\sqrt{NT})$. For (iii), if $0 \equiv s(\mathrm{mod}, \tau)$ and $s/\tau = m$, using Lemma 2 again, we have $\mathrm{E}(\boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y}_{(-s)}) = \sum_{i=1}^{N} \mathrm{tr}(\boldsymbol{L}^{s\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{L}^s \boldsymbol{\Sigma}_i) + O(N)$ and $\mathrm{Var}(\boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y}_{(-s)}) = O(NT)$ from Lemma A5 of Bao (2024). Based on these results and (i) and (ii), we can claim that $\boldsymbol{y}_{(-s)}' \boldsymbol{A}(\boldsymbol{y} - \phi_0^m \boldsymbol{y}_{(-s)}) = O(N) + O_p(\sqrt{NT})$ and $(\boldsymbol{y} - \phi_0^m \boldsymbol{y}_{(-s)})' \boldsymbol{A}(\boldsymbol{y} - \phi_0^m \boldsymbol{y}_{(-s)}) = \boldsymbol{y}' \boldsymbol{A} \boldsymbol{y} - 2\phi_0^m \boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y} + \phi_0^{2m} \boldsymbol{y}_{(-s)}' \boldsymbol{A} \boldsymbol{y}_{(-s)} = O_p(NT)$. □

Lemma 5. *When $\phi_0 = 1$, under Assumptions 1–4, $\boldsymbol{y}_{(-l)}' \boldsymbol{A} \boldsymbol{y}_{(-s)} = O_p(NT^3)$.*

Proof. Following the proof of Lemma A12 in Bao (2024), we note that, regardless of $l \equiv s(\mathrm{mod}, \tau)$ or not, $\boldsymbol{y}_{(-l)}' \boldsymbol{A} \boldsymbol{y}_{(-s)}$ is dominated by $(\mathbf{1}' \boldsymbol{L}^{l\prime} \boldsymbol{\Phi}_\tau^{-1\prime} \boldsymbol{M}_T \boldsymbol{\Phi}_\tau^{-1} \boldsymbol{L}^s \mathbf{1})(\bar{\boldsymbol{\alpha}}' \bar{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\alpha}}' \tilde{\boldsymbol{\alpha}})$, where $\bar{\boldsymbol{\alpha}} = \mathrm{E}(\boldsymbol{\alpha})$ and $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}$. Furthermore, $(\bar{\boldsymbol{\alpha}}' \bar{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\alpha}}' \tilde{\boldsymbol{\alpha}}) = O_p(N)$, and in view of Lemma 3,

$\mathbf{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_{\tau}^{-1\prime}\boldsymbol{M}_T\boldsymbol{\Phi}_{\tau}^{-1}\boldsymbol{L}^s\mathbf{1} = \mathbf{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_{\tau}^{-1\prime}\boldsymbol{\Phi}_{\tau}^{-1}\boldsymbol{L}^s\mathbf{1} - \mathbf{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_{\tau}^{-1\prime}\mathbf{1}\mathbf{1}'\boldsymbol{\Phi}_{\tau}^{-1}\boldsymbol{L}^s\mathbf{1}/T = O(T^3)$. In conclusion, $\boldsymbol{y}'_{(-l)}\boldsymbol{A}\boldsymbol{y}_{(-s)}$ is $O_p(NT^3)$. □

B. Proof of Main Results

Proof of Theorem 1. The proof follows from Theorems 1–5 and Corollaries 1,2,4, and 5 in Bao (2024) by treating the convergence model as a special case of DP(τ) model without exogenous variables $\boldsymbol{X}$, and the true parameter values $\phi_{01} = \phi_{02} = \cdots = \phi_{0,\tau-1} = 0$. We omit the details here to save space.

By definition, the estimator $\hat{\phi}_s = \arg_{\phi_s}\{g_{NT}(\phi_s) = 0\}$. When $|\phi_{0\tau}| < 1$, under homoskedasticity,

$$g_{NT}(\phi_s) = \frac{1}{NT}\big[\boldsymbol{y}'_{(-s)}\boldsymbol{A}(\boldsymbol{y} - \phi_s\boldsymbol{y}_{(-s)}) + (\boldsymbol{y} - \phi_s\boldsymbol{y}_{(-s)})'\boldsymbol{A}(\boldsymbol{y} - \phi_s\boldsymbol{y}_{(-s)})h(\phi_s)\big],$$

where $h(\phi_s) = [T(T-1)]^{-1}\mathbf{1}'\boldsymbol{\Phi}_s^{-1}(\phi_s)\boldsymbol{L}^s\mathbf{1}$. We first consider the case $0 \not\equiv s(\mathrm{mod}, \tau)$. By a first-order expansion, $g_{NT}(\hat{\phi}_s) = g_{NT}(0) + g'_{NT}(0)(\hat{\phi}_s - 0) + s.o.$ Then it follows that $\hat{\phi}_s = -[g'_{NT}(0)]^{-1}g_{NT}(0) + s.o.$, where $[g'_{NT}(0)]^{-1}$ has a bounded probability limit from Lemma A5 in Bao (2024). Thus, $\hat{\phi}_s \xrightarrow{p} 0$ holds if we prove $g_{NT}(0) = o_p(1)$. Specifically,

$$g_{NT}(0) = \frac{1}{NT}\big[\boldsymbol{y}'_{(-s)}\boldsymbol{A}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}h(\phi_s)\big].$$

By Lemma 4, we have $\boldsymbol{y}'_{(-s)}\boldsymbol{A}\boldsymbol{y} = O(N) + O_p(\sqrt{NT})$, $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} = O_p(NT)$, $h(\phi_s) = [T(T-1)]^{-1} \times O(T) = O(T^{-1})$, so $g_{NT}(0) = [O(N) + O_p(\sqrt{NT}) + O_p(NT) \times O(T^{-1})]/(NT) = o_p(1)$. Next, for the case $0 \equiv s(\bmod, \tau)$, we expand $g_{NT}(\hat{\phi}_s) = g_{NT}(\phi_{0\tau}^m) + g'_{NT}(\phi_{0\tau}^m)(\hat{\phi}_s - \phi_{0\tau}^m) + s.o.$, and thus $\hat{\phi}_s - \phi_{0\tau}^m = -[g'_{NT}(\phi_{0\tau}^m)]^{-1}g_{NT}(\phi_{0\tau}^m) + o_p(1)$. Again by Lemma 4, $g_{NT}(\phi_{0\tau}^m) = \big[\boldsymbol{y}'_{(-s)}\boldsymbol{A}(\boldsymbol{y} - \phi_0^m\boldsymbol{y}_{(-s)}) + (\boldsymbol{y} - \phi_{0\tau}^m\boldsymbol{y}_{(-s)})'\boldsymbol{A}(\boldsymbol{y} - \phi_{0\tau}^m\boldsymbol{y}_{(-s)})h(\phi_{0\tau}^m)\big]/(NT) = \big[O(N) + O_p(\sqrt{NT}) + O_p(NT) \times O(T^{-1})\big]/(NT) = o_p(1)$. Therefore, $\hat{\phi}_s \xrightarrow{p} \phi_{0\tau}^m$ when $0 \equiv s(\mathrm{mod}, \tau)$. Under heteroskedasticity,

$$g_{NT}(\phi_s) = \frac{1}{NT}\sum_{i=1}^{N}\big[\boldsymbol{y}'_{i,(-s)}\boldsymbol{M}_T(\boldsymbol{y}_i - \phi_s\boldsymbol{y}_{i,(-s)}) - (\boldsymbol{y}_i - \phi_s\boldsymbol{y}_{i,(-s)})'\boldsymbol{M}_T\boldsymbol{\Psi}(\phi_s)\boldsymbol{M}_T(\boldsymbol{y}_i - \phi_s\boldsymbol{y}_{i,(-s)})\big],$$

where $\boldsymbol{\Psi}(\phi_s) = T(T-2)^{-1}\mathrm{Dg}(\boldsymbol{M}_T\boldsymbol{\Phi}_s^{-1}(\phi_s)\boldsymbol{L}^s) - [(T-1)(T-2)]^{-1}\mathrm{tr}(\boldsymbol{M}_T\boldsymbol{\Phi}_s^{-1}(\phi_s)\boldsymbol{L}^s)\boldsymbol{I}$. For the case $0 \not\equiv s(\mathrm{mod}, \tau)$. Similar to the homoskedastic case, we essentially want to prove

$g_{NT}(0) = \sum_{i=1}^{N} \left[\boldsymbol{y}_{i,(-s)}' \boldsymbol{M}_T \boldsymbol{y}_i - \boldsymbol{y}_i' \boldsymbol{M}_T \boldsymbol{\Psi} \boldsymbol{M}_T \boldsymbol{y}_i\right]/(NT) = o_p(1)$. Using Lemma 2, we can prove a similar result to Lemma 4 that claims $\boldsymbol{y}_{i,(-s)}' \boldsymbol{M}_T \boldsymbol{y}_i = O(1) + O_p(\sqrt{T})$. Moreover, notice that $\boldsymbol{\Psi}(\phi_s)$ is diagonal with $O(T^{-1})$ elements, so $\boldsymbol{y}_i' \boldsymbol{M}_T \boldsymbol{\Psi}(\phi_s) \boldsymbol{M}_T \boldsymbol{y}_i = O(1)$, following the lines of Lemma A8 in Bao (2024). Thus, $g_{NT}(0) = o_p(1)$. For the case $0 \equiv s(\bmod, \tau)$, we can derive a similar result to Lemma 4 that claims $\boldsymbol{y}_{i,(-s)}' \boldsymbol{M}_T (\boldsymbol{y}_i - \phi_{0\tau}^m \boldsymbol{y}_{i,(-s)}) = O(1) + O_p(\sqrt{T})$ by Lemma 2. Furthermore, $(\boldsymbol{y}_i - \phi_{0\tau}^m \boldsymbol{y}_{i,(-s)})' \boldsymbol{M}_T \boldsymbol{\Psi} \boldsymbol{M}_T (\boldsymbol{y}_i - \phi_{0\tau}^m \boldsymbol{y}_{i,(-s)}) = O(1)$ by noting that $2\phi_{0\tau}^m \boldsymbol{y}_i' \boldsymbol{M}_T \boldsymbol{y}_{i,(-s)}$ and $\phi_{0\tau}^{2m} \boldsymbol{y}_{i,(-s)}' \boldsymbol{M}_T \boldsymbol{y}_{i,(-s)}$ are both $O_p(1)$. Thus, $g_{NT}(\phi_{0\tau}^m) = o_p(1)$.

When $\phi_{0\tau} = 1$, under homoskedasticity, the analysis is similar to that of the homoskedastic scenario in the stable case. By a first-order expansion, $g_{NT}(\hat{\phi}_s) = g_{NT}(1) + g_{NT}'(1)(\hat{\phi}_s - 1) + s.o.$, and thus $\hat{\phi}_s - 1 = -[T^{-2} g_{NT}'(1)]^{-1} T^{-2} g_{NT}(1) + s.o.$, where

$$\frac{1}{T^2} g_{NT}(1) = \frac{1}{NT^3} \left[\boldsymbol{y}_{(-s)}' \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{y}_{(-s)}) + (\boldsymbol{y} - \boldsymbol{y}_{(-s)})' \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{y}_{(-s)}) h(\phi_{0\tau}^m)\right].$$

By Lemma 5, $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$, $\boldsymbol{y}_{(-s)}'\boldsymbol{A}\boldsymbol{y}$, and $\boldsymbol{y}_{(-s)}'\boldsymbol{A}\boldsymbol{y}_{-s}$ are $O_p(NT^3)$. More specifically, from the proof of Lemma 5, we know $(\boldsymbol{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{1})(\bar{\boldsymbol{\alpha}}'\bar{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\alpha}}'\tilde{\boldsymbol{\alpha}})$ dominates $\boldsymbol{y}_{(-l)}'\boldsymbol{A}\boldsymbol{y}_{(-s)}$. By Lemma A3(iii) of Bao (2024), $\boldsymbol{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{1} = T^3/12 + O(T)$, for any $l, s = 0, 1, \cdots$.[10] Moreover, $\boldsymbol{y}_{(-l)}'\boldsymbol{A}\boldsymbol{y}_{(-s)}$ contains $[(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^l)\boldsymbol{u}]'\boldsymbol{A}[(\boldsymbol{I}_N \otimes \boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s)\boldsymbol{u}]$, which is $O_p(NT^2)$. Also, $h(\phi_{0\tau}^m) = h(1) = [T(T-1)]^{-1} \times O(T^2) = O(1)$. Thus, $T^{-2}g_{NT}(1) = (NT^3)^{-1}O_p(NT^2) = O_p(T^{-1}) = o_p(1)$. The result $T^{-2}g_{NT}(1) = o_p(1)$ for the heteroskedastic case can be established similarly and is therefore omitted.

To show the consistency, it suffices to show that $SSR(s)/[N(T-s)]$ achieves the minimum at $s = \tau$ when $N, T \to \infty$. For ease of representation, we denote

$$\eta(s) = \frac{T-s}{T} \times \frac{SSR(s)}{N(T-s)} = \frac{SSR(s)}{NT}.$$

Obviously, the proof is then equivalent to showing that $\eta(s)$ achieves the minimum at $s = \tau$ when $N, T \to \infty$.

When $|\phi_{0\tau}| < 1$, by Theorem 2, $\hat{\phi}_s \xrightarrow{p} \phi_s$, where for ease of presentation, we have defined a pseudo parameter ϕ_s that is equal to $\phi_{0\tau}^m$ when $0 \equiv s(\bmod, \tau)$ and equal to 0 otherwise,

[10] The proof of result (iii) of Lemma A3 in Bao (2024) states that for a unit-root higher-order DP model, $\boldsymbol{1}'\boldsymbol{L}^{l\prime}\boldsymbol{\Phi}_\tau^{-1\prime}\boldsymbol{M}_T\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{L}^s\boldsymbol{1} = a_1 T^3/12 + O(T)$, where a_1 is a term in the solution of the homogeneous equation associated with the autoregressive parameters in the DP model. When it applies to our context, $a_1 = 1$.

$\lfloor s/\tau \rfloor = m$. It then follows that

$$\begin{aligned}\eta(s) &= \frac{1}{NT}(\boldsymbol{y} - \phi_s \boldsymbol{y}_{(-s)})'\boldsymbol{A}(\boldsymbol{y} - \phi_s \boldsymbol{y}_{(-s)}) + s.o. \\ &= \frac{1}{NT}(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} + \phi_s^2 \boldsymbol{y}_{(-s)}'\boldsymbol{A}\boldsymbol{y}_{(-s)} - 2\phi_s \boldsymbol{y}_{(-s)}'\boldsymbol{A}\boldsymbol{y}) + s.o.\end{aligned}$$

When $0 \not\equiv s(\bmod, \tau)$, $\phi_s = 0$, $\eta(s) = (NT)^{-1}\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} + s.o. \xrightarrow{p} \text{plim}_{N,T\to\infty}(NT)^{-1}\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$. When $0 \equiv s(\text{mod}, \tau)$, $s = m\tau$, we have $\eta(s) = (NT)^{-1}(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} + \phi_{0\tau}^{2m}\boldsymbol{y}_{(-s)}'\boldsymbol{A}\boldsymbol{y}_{(-s)} - 2\phi_{0\tau}^{m}\boldsymbol{y}_{(-s)}'\boldsymbol{A}\boldsymbol{y}) + s.o \xrightarrow{p} (1 - \phi_{0\tau}^{2m})\text{plim}_{N,T\to\infty}(NT)^{-1}\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$, in light of Lemmas 2 and 4. Thus, $\eta(s)$ achieves the minimum at $m = 1$. This result holds under both homoskedasticity and heteroskedasticity. Interestingly, under homoskedasticity, the limit of $\eta(s)$ can be further simplified.[11] Specifically,

$$\eta(s) \xrightarrow{p} \begin{cases} \frac{1-\phi_{0\tau}^{2m}}{1-\phi_{0\tau}^{2}}\sigma^2 & 0 \equiv s(\text{mod}, \tau), s \neq \tau, s/\tau = m \\ \sigma^2 & s = \tau \\ \frac{1}{1-\phi_{0\tau}^2}\sigma^2 & 0 \not\equiv s(\text{mod}, \tau) \end{cases}.$$

Obviously, for $m > 1$, $(1 - \phi_{0\tau}^2)^{-1} > (1 - \phi_{0\tau}^{2m})/(1 - \phi_{0\tau}^2) > 1$, and thus $\eta(s)$ achieves the minimum at $s = \tau$.

When $\phi_{0\tau} = 1$, in view of the proofs of Theorem 2 and Lemma 3, we have

$$\eta(s) \xrightarrow{p} \begin{cases} \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^N \sum_{j=0}^\infty \sum_{t=T-(m+j)\tau}^{T-j\tau} \sigma_{it}^2 & 0 \equiv s(\text{mod}, \tau), s/\tau = m \\ \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^N \sum_{j=0}^\infty (\sum_{t=1}^{T-j\tau} \sigma_{it}^2 + \sum_{t=1}^{T-j\tau-s} \sigma_{it}^2) & 0 \not\equiv s(\text{mod}, \tau) \end{cases}.$$

Note that

$$\begin{aligned}\lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^N \sum_{j=0}^\infty \sum_{t=T-(m+j)\tau}^{T-j\tau} \sigma_{it}^2 &\leq \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^N \bar{\sigma}_i^2 \sum_{j=0}^\infty \sum_{t=T-(m+j)\tau}^{T-j\tau} \\ &\leq \frac{1}{N}\sum_{i=1}^N \bar{\sigma}_i^2 \lim_{N,T\to\infty} \frac{1}{T}\sum_{j=0}^{\lfloor T/\tau \rfloor} m\tau = \frac{m}{N}\sum_{i=1}^N \bar{\sigma}_i^2 = O(1).\end{aligned}$$

In a similar way, we can show that $\lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N \sum_{j=0}^\infty \sum_{t=T-(m+j)\tau}^{T-j\tau} \sigma_{it}^2$ is bounded below by $(m/N)\sum_{i=1}^N \underline{\sigma}_i^2 = O(1)$. Therefore, $\lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N \sum_{j=0}^\infty \sum_{t=T-(m+j)\tau}^{T-j\tau} \sigma_{it}^2 = O(1)$. In comparison, $\lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N \sum_{j=0}^\infty (\sum_{t=1}^{T-j\tau} \sigma_{it}^2 + \sum_{t=1}^{T-j\tau-s} \sigma_{it}^2) = O(T)$, in

[11]The result is thus analogous to the conclusions that we can get from pure time series regressions.

light of the proof of Lemma 3. Furthermore, $\lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{j=0}^{\infty}\sum_{t=T-(m+j)\tau}^{T-j\tau}\sigma_{it}^2$ is strictly increasing in m. Based on all these results, we can conclude that $\eta(s)$ achieves the minimum at $m=1$, which is the case of $s=\tau$. From the preceding derivations, this result holds under both homoskedasticity and heteroskedasticity. Interestingly, with some further calculations, we notice that under homoskedasticity, the above result can be presented as follows,

$$\eta(s) \xrightarrow{p} \begin{cases} m\sigma^2 & 0 \equiv s(\mathrm{mod}, \tau), s \neq \tau, s/\tau = m \\ \sigma^2 & s = \tau \\ \infty & 0 \not\equiv s(\mathrm{mod}, \tau) \end{cases},$$

which again mimics the pure time series regression scenarios. □

Proof of Theorem 4. From Theorem 1, we know that under $H_0' : \phi_\ell = 0, \ell = 1, \ldots, p$, using proper standard errors, each t-statistic follows a standard normal distribution asymptotically. Thus, to show the supreme of these p t-statistics follows the supreme of a set of independent standard normal distributions, it is sufficient to show that the t-statistics are asymptotically uncorrelated. According to Theorem 1, as $T \to \infty$, the t-statistics have the following two forms depending on error heteroskedasticities:

$$t(\ell) = \begin{cases} \dfrac{\sqrt{NT}\hat{\phi}_\ell}{\sqrt{\hat{\sigma}^2 NT(\boldsymbol{y}_{(-\ell)}'\boldsymbol{A}\boldsymbol{y}_{(-\ell)})^{-1}}} & \text{homoskedasticity} \\ \dfrac{\sqrt{NT}\hat{\phi}_s}{NT(\boldsymbol{y}_{(-\ell)}'\boldsymbol{A}\boldsymbol{y}_{(-\ell)})^{-1}\sqrt{(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{y}_{i,(-\ell)}'\boldsymbol{M}_T\hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}}_i'\boldsymbol{M}_T\boldsymbol{y}_{i,(-\ell)}}} & \text{heteroskedasticity} \end{cases}.$$

We first claim that after certain normalization (as shown in the above formula), the standard errors in the t-statistics have well-defined probability limits, which is a common property shared by the conventional t-statistics. First, under homoskedasticity, it is obvious that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, $(NT)^{-1}\boldsymbol{y}_{(-\ell)}'\boldsymbol{A}\boldsymbol{y}_{(-\ell)} \xrightarrow{p} \sigma^2$, and thus $\sqrt{\hat{\sigma}^2 NT(\boldsymbol{y}_{(-\ell)}'\boldsymbol{A}\boldsymbol{y}_{(-\ell)})^{-1}} \xrightarrow{p} 1$. Next, for the heteroskedastic case, Bao (2024) shows the above standard error converges in probability to a term that does not possess an analytical expression. Next, for the numerator in $t(\ell)$, by a first-order expansion, $\sqrt{NT}\hat{\phi}_\ell = -\sqrt{NT}[g_{NT}'(\phi_\ell)^{-1}]g_{NT}(\phi_\ell) + o_p(1)$, where $g_{NT}'(\phi_\ell) \xrightarrow{p} -\mathrm{plim}_{T\to\infty}(NT)^{-1}\boldsymbol{y}_{(-\ell)}'\boldsymbol{A}\boldsymbol{y}_{(-\ell)}$ holds under both homoskedasticity (Corollary 1 in Bao (2024)) and heteroskedasticity (Lemma A9 in Bao (2024)).

With these results in hand, we now investigate $\text{plim}_{N,T\to\infty} t(\ell_1)t(\ell_2)$. Subject to some scaling constants, we obtain, for the homoskedastic case,

$$\begin{aligned}\underset{N,T\to\infty}{\text{plim}}\; t(\ell_1)t(\ell_2) &= c_1 \underset{N,T\to\infty}{\text{plim}}\; NT\hat{\phi}_{\ell_1}\hat{\phi}_{\ell_2} = c_2 \underset{N,T\to\infty}{\text{plim}}\; NT g_{NT}(\phi_{\ell_1}) g_{NT}(\phi_{\ell_2}) \\ &= c_2 \underset{N,T\to\infty}{\text{plim}} \frac{1}{NT}(\boldsymbol{y}'_{(-\ell_1)}\boldsymbol{A}\boldsymbol{u} + \boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}h_{\ell_1})(\boldsymbol{y}'_{(-\ell_2)}\boldsymbol{A}\boldsymbol{u} + \boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}h_{\ell_2})],\end{aligned}$$

where c_1 and c_2 are some constant numbers that arise when we replace the relevant random terms with their probability limits, $h_{\ell_1} = h(\phi_{\ell_1})$, and $h_{\ell_2} = h(\phi_{\ell_2})$. Using the results from Bao (2024) and the results derived in this paper, for a given ℓ, $\boldsymbol{y}'_{(-\ell)}\boldsymbol{A}\boldsymbol{u} = O(N) + O_p(\sqrt{NT})$, $h_\ell = O(T^{-1})$, $\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u} = O(NT) + O_p(\sqrt{NT})$. Then $N/T \to 0$, $\text{plim}_{N,T\to\infty} t(\ell_1)t(\ell_2)$

$$\underset{N,T\to\infty}{\text{plim}}\; t(\ell_1)t(\ell_2) = c_2 \underset{N,T\to\infty}{\text{plim}} \frac{1}{NT}\boldsymbol{y}'_{(-\ell_1)}\boldsymbol{A}\boldsymbol{u}\boldsymbol{y}'_{(-\ell_2)}\boldsymbol{A}\boldsymbol{u}.$$

Furthermore, for a certain ℓ, the leading term in $\boldsymbol{y}'_{(-\ell)}\boldsymbol{A}\boldsymbol{u}$ is $\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{L}^\ell)\boldsymbol{u}$, so

$$\begin{aligned}\underset{N,T\to\infty}{\text{plim}}\; t(\ell_1)t(\ell_2) &= c_2 \underset{N,T\to\infty}{\text{plim}} \left[\frac{1}{NT}\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{L}^{\ell_1})\boldsymbol{u}\boldsymbol{u}'\boldsymbol{A}(\boldsymbol{I}_N \otimes \boldsymbol{L}^{\ell_2})\boldsymbol{u}\right] \\ &= c_2 \underset{N,T\to\infty}{\text{plim}} \left[\frac{1}{NT}\left(\sum_{i=1}^N \boldsymbol{u}'_i\boldsymbol{M}_T\boldsymbol{L}^{\ell_1}\boldsymbol{u}_i\right)\left(\sum_{i=1}^N \boldsymbol{u}'_i\boldsymbol{M}_T\boldsymbol{L}^{\ell_2}\boldsymbol{u}_i\right)\right].\end{aligned}$$

We can easily show that $\boldsymbol{u}'_i\boldsymbol{M}_T\boldsymbol{L}^\ell\boldsymbol{u}_i = O_p(1)$. With $N/T \to 0$, we have $\text{plim}_{N,T\to\infty} t(\ell_1)t(\ell_2) = 0$. For the heteroskedastic case, we can similarly show the leading term in $\text{plim}_{N,T\to\infty} t(\ell_1)t(\ell_2)$ is proportional to $\text{plim}_{N,T\to\infty}[(NT)^{-1}(\sum_{i=1}^N \boldsymbol{u}'_i\boldsymbol{M}_T\boldsymbol{L}^{\ell_1}\boldsymbol{u}_i)(\sum_{i=1}^N \boldsymbol{u}'_i\boldsymbol{M}_T\boldsymbol{L}^{\ell_2}\boldsymbol{u}_i)]$. The property $\boldsymbol{u}'_i\boldsymbol{M}_T\boldsymbol{L}^\ell\boldsymbol{u}_i = O_p(1)$ also holds under heteroskedasticity. It follows that $\text{plim}_{N,T\to\infty} t(\ell_1)t(\ell_2) = 0$ as $N/T \to 0$. $\square$

References

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Bao, Y. (2024). Estimating linear dynamic panels with recentered moments. *Econometrics 12*(1), 3.

Bao, Y. and S. Dhongde (2009). Testing convergence in income distribution. *Oxford Bulletin of Economics and Statistics 71*(2), 295–302.

Bao, Y. and X. Yu (2023). Indirect inference estimation of dynamic panel data models. *Journal of Econometrics 235*(2), 1027–1053.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics 106*(2), 407–443.

Belaire-Franch, J. (2020). A note on change in persistence of US city prices. *Studies in Nonlinear Dynamics & Econometrics 26*(5), 649–653.

Bernard, A. B. and S. N. Durlauf (1995). Convergence in international output. *Journal of Applied Econometrics 10*(2), 97–108.

Busetti, F., S. Fabiani, and A. Harvey (2006). Convergence of prices and rates of inflation. *Oxford Bulletin of Economics and Statistics 68*, 863–877.

Busetti, F., L. Forni, A. Harvey, and F. Venditti (2007). Inflation convergence and divergence within the European monetary union. *International Journal of Central Banking 3*(2), 95–121.

Caselli, F., G. Esquivel, and F. Lefort (1996). Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth 1*, 363–389.

Cecchetti, S. G., N. C. Mark, and R. J. Sonora (2002). Price index convergence among United States cities. *International Economic Review 43*(4), 1081–1099.

Chatterji, M. (1992). Convergence clubs and endogenous growth. *Oxford Review of Economic Policy 8*(4), 57–69.

Chen, L. and J. Devereux (2003). What can US city price data tell us about purchasing power parity? *Journal of International Money and Finance 22*, 213–222.

Dridi, J. and A. D. Nguyen (2019). Assessing inflation convergence in the east African community. *Journal of International Development 31*, 119–136.

Engel, C. and J. H. Rogers (2001). Violating the law of one price: Should we make a federal case out of it? *Journal of Money, Credit, and Banking 33*(1), 1–15.

Fleissig, A. and J. Strauss (2001). Panel unit-root tests of OECD stochastic convergence. *Review of International Economics 9*(1), 153–162.

Frankel, J. A. and A. K. Rose (1996). A panel project on purchasing power parity: Mean reversion within and between countries. *Journal of International Economics 40*(1–2), 209–224.

Ho, T.-w. (2015). Income inequality may not converge after all: Testing panel unit roots in the presence of cross-section cointegration. *The Quarterly Review of Economics and Finance 56*, 68–79.

Huang, H.-C. R., W.-H. Liu, and C.-C. Yeh (2012). Convergence in price levels across US cities. *Economics Letters 114*(3), 245–248.

Kejriwal, M., P. Perron, and J. Zhou (2013). Wald tests for detecting multiple structural changes in persistence. *Econometric Theory 29*(2), 289–323.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* Palgrave Macmillan.

Kočenda, E. and D. H. Papell (1997). Inflation convergence within the European union: A panel data analysis. *International Journal of Finance and Economics 2*, 189–198.

Lee, H.-Y. and J.-L. Wu (2001). Mean reversion of inflation rates: Evidence from 13 OECD countries. *Journal of Macroeconomics 23*(3), 477–487.

Lin, P.-C. and H.-C. Huang (2011). Inequality convergence in a panel of states. *Journal of Economic Inequality 9*, 195–206.

Lin, P.-C. and H.-C. Huang (2012a). Convergence in income inequality? Evidence from panel unit root tests with structural breaks. *Empirical Economics 43*, 153–174.

Lin, P.-C. and H.-C. Huang (2012b). Inequality convergence revisited: Evidence from stationarity panel tests with breaks and cross correlation. *Economic Modelling 29*(2), 316–325.

Lopez, C. and D. H. Papell (2012). Convergence of Euro area inflation rates. *Journal of International Money and Finance 31*, 1440–1458.

Marshall, A. (1890). *Principles of Economics.* Macmillan.

Panizza, U. (2001). Convergence in income inequality. *Journal of Income Distribution 10*(1–2), 5–12.

Parsley, D. C. and S.-J. Wei (1996). Convergence to the law of one price without trade barriers or currency fluctuations. *Quarterly Journal of Economics 111*(4), 1211–1236.

Ravallion, M. (2003). Inequality convergence. *Economics Letters 80*(3), 351–356.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature 34*, 647–668.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* W. Strahan and T. Cadell.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics 70*, 65—-94.