A ROBUST TEST FOR PARAMETER HOMOGENEITY IN PANEL DATA MODELS

Chang-Ching Lin*

Institute of Economics

Academia Sinica

February 26, 2007

Abstract

This paper proposes a robust test for parameter homogeneity in panel data models. Under the assumption of parameter homogeneity, the sample estimates across individuals should be less dispersed than the estimates under parameter heterogeneity. Conventional dispersion tests based on sample variances, however, can be highly distorted when extreme observations are present in small samples. This paper utilizes Hosking's (1990) second L-moment to measure the dispersion of the distribution of the sample individual parameter estimates. Due to the robustness of the sample second L-moment against the presence of outliers, the proposed test tends to provide more accurate size in smaller samples. The proposed test is shown to have a standard normal limiting distribution. Monte Carlo simulations also show that the proposed test delivers more accurate finite sample sizes than existing tests for various combinations of N and T in linear and nonlinear panel data models. As an illustration, the proposed test is applied to examine whether or not the rates of convergence of economic growth and the rates of technology growth are the same across countries. The results show that neither of them are the same across countries.

Keywords: L-moments, Gini's Mean Difference, Nonlinear Regressions, Economic Growth.
JEL classification: C12, C23

*I would like to thank Serena Ng, Shinichi Sakata, and Lutz Kilian for encouragement and valuable advice. I am grateful to the seminar participants at the 2006 Midwest Economics Association's 70^{th} Annual Meeting. Any remaining errors are mine. Contact address: Institute of Economics, Academia Sinica, 128 Academia Road, Sec 2, Taipei 115, Taiwan. Tel: 886-2-27822791 ext. 301. E-mail: *lincc@econ.sinica.edu.tw.*

1 Introduction

The assumption of parameter homogeneity is widely used in panel data regression models. For example, in the literature on income dynamics, the autoregressive parameters in the mean are usually assumed to be the same across individuals in the same demographic or geographic cohort. However, if the true parameters are not the same across individuals, then imposing that assumption will bias estimation and inference. In the literature, only a handful of tests have been developed to test slope homogeneity in linear panel data models. Swamy (1970) proposed a dispersion test for panel data sets with a large number of time periods (T) but a small number of individuals (N). In contrast, in many empirical studies, N is usually (much) larger than T. For the latter panel data sets, Pesaran, Smith, and Im (1996) proposed a Hausman type test, but that test is invalid for a simple dynamic panel model or a panel data model with only strictly exogenous explanatory variables. In view of this, Pesaran and Yamagata (2005) modified Swamy's test for large N panel data sets and in their simulations that dispersion test has more accurate size than the other tests. See Pesaran and Yamagata (2005) for a survey of tests for slope homogeneity in linear panel data models.

To make use of a dispersion test, it is necessary to estimate the parameter of interest in a panel data model for each cross-sectional unit using T observations. When T is not large enough, the presence of outliers might bias inference. Moreover, even though current dispersion tests can potentially be extended to nonlinear panel data models, there may be more outliers when estimating the parameter for each individual in nonlinear models than in linear models. It is essential to develop a general and robust test for parameter homogeneity not only in the linear but also in the nonlinear panel data models.

The proposed robust dispersion test is designed for parameter homogeneity in linear and nonlinear panel data models. Like other standard dispersion tests, the main idea of the proposed test is that sample estimates of the parameter of interest across individuals tend to be more tightly bound under parameter homogeneity than sample estimates under parameter heterogeneity. To develop a new dispersion test based on this idea, the following three steps should be considered: 1) standardize the estimator of the parameter of interest for each cross-sectional unit, 2) choose a measure to capture the dispersion of (standardized) sample estimates across individual units, and 3) find critical values to determine whether or not the estimate of the chosen dispersion measure is significantly different from the value under parameter homogeneity.

In Step 1, it is natural to standardize the estimator using its asymptotic standard error. After suitably controlling for the rate of convergence, the standardized estimator tends to have a standard normal limiting distribution under the null hypothesis, as long as some mild conditions are met. Then, given a dispersion measure, we usually can obtain the critical values analytically or numerically by using samples drawn from the standard normal distribution. But what dispersion measure should be chosen? Although the standard deviation and the mean absolute deviation are commonly used dispersion measures, the standard deviation is very sensitive to outliers. Also, taking the absolute value to compute the mean absolute deviation will introduce extra complexity even though it is robust against the presence of outliers.

For Step 2 this paper utilizes the second L-moment developed by Hosking (1990) to measure the dispersion of the sample standardized estimates across individual units. The estimator of the second L-moment is a linear combination of order statistics, and has been shown to be more robust against the presence of extreme observations than the standard deviation. Then, for Step 3 the distribution of the estimator of second L-moment when a sample is drawn from the standard normal distribution is studied. We approximate the asymptotic and finite-sample variances numerically and show that the proposed test statistic has a standard normal limiting distribution under parameter homogeneity, as long as some mild conditions are met. The simulation results demonstrate that this test has accurate size and satisfactory power for various pairs of N and T commonly used in empirical applications.

The rest of this paper is organized as follows. In Section 2, the second L-moment is introduced and its properties are studied. Based on these properties, we propose a dispersion test in Section 3. We provide three simple examples that demonstrate how to apply the test to linear and nonlinear panel data models. Section 4 includes Monte Carlo simulations to evaluate the finite sample properties of the proposed test and to compare the finite sample size and power of the proposed test with those of the tests that are commonly used in the literature. As an illustration, we apply the proposed test to examine whether or not the rates of convergence of economic growth and the rates of technology growth are the same across countries in Section 5. Section 6 concludes this paper, and the proofs are in the Appendix A.

2 The Second L-moment

Suppose that we randomly draw a sample of size 2, X_i and X_j, $i, j = 1, 2, i \neq j$, from a population distribution. The distance between the two observations, $|X_i - X_j|$, reveals the information about the dispersion of the distribution. If the distribution is widely dispersed, then the distance between X_i and X_j tends to be large. If the distribution is tightly bunched around the central value, then these two observations are closer. Thus, the dispersion of the distribution can be measured by the expectation of $|X_i - X_j|$:

$$\lambda_2 = \frac{1}{2}\mathrm{E}[|X_i - X_j|].$$

The above expectation can be expressed as a function of order statistics. Consider a sample $X_1, X_2, \cdots X_m$ of m independent and identically distributed real-valued random variables from a distribution $F(x)$. The order statistics of this sample can be obtained by arranging these X_i in ascending order of magnitude such that

$$X_{1:m} \leq X_{2:m} \leq \cdots \leq X_{m:m},$$

where $X_{j:m}$ denotes the j^{th} smallest observation from a sample of m. Then,

$$\lambda_2 = \frac{1}{2}\mathrm{E}[|X_i - X_j|] = \frac{1}{2}\mathrm{E}[X_{2:2} - X_{1:2}].$$

Hosking (1990) calls λ_2 the second (order) L-moment. Also, let $x(F)$ denote the quantile function with respect to $F(x)$. Since

$$\mathrm{E}[X_{r:n}] = \frac{n!}{(r-1)!(n-r)!}\int x(F(x))F(x)^{r-1}(1-F(x))^{n-r}dF(x),$$

the second L-moment can be rewritten as

$$\begin{aligned}
\lambda_2 &= \frac{1}{2}\mathrm{E}[|X_i - X_j|] = \frac{1}{2}\mathrm{E}[X_{2:2} - X_{1:2}] \\
&= \int x(F(x))F(x)dF(x) - \int x(F(x))(1-F(x))dF(x) \\
&= 2\int x(F(x))F(x)dF(x) - \int x(F(x))dF(x) \\
&= 2\beta_1 - \beta_0,
\end{aligned}$$

where $\beta_r = E[x(F(x))F(x)^r]$.

As the classical moments of a distribution are usually estimated by a sample from the population, so can the second L-moment be estimated from an ordered sample of size N: $x_{1:N} \leq x_{2:N} \leq \cdots \leq x_{N:N}$. Hosking, Wallis, and Wood (1985) proposed the following unbiased estimator of λ_2:

$$l_2 = 2b_1 - b_0 = \frac{1}{N}\sum_{j=1}^{N}\frac{2j-N-1}{N-1}x_{j:N} = \frac{1}{N}\sum_{j=1}^{N} w_{j:N}x_{j:N},$$

where

$$b_1 = N^{-1}\sum_{j=2}^{N}\frac{j-1}{N-1}x_{j:N},$$
$$b_0 = N^{-1}\sum_{j=1}^{N}x_{j:N},$$

respectively are the unbiased estimators of β_1 and β_0 originally proposed by Landwehr, Matalas, and Wallis (1979). The estimator l_2 is a linear combination of order statistics and identical to one-half of Gini's mean difference statistic (Gini (1912)).[1]

Analogous to the standard deviation, l_2 is a measure of the dispersion of a distribution. However, l_2 is less sensitive than the sample standard deviation to the presence of outliers because l_2 assigns relatively less weight to the observations far from the sample mean. Moreover, λ_2 can provide better identification of the underlying population distribution even if two different distributions have the same variance. For example, a standard normal distribution has the same variance as an exponential distribution with $\theta = 1$. However, $\lambda_2 = 1/2$ for the exponential distribution and $\lambda_2 = 1/\sqrt{\pi}$ for the standard normal distribution. Hosking (1990) lists the values of λ_2 associated with the distributions commonly used in empirical applications.

In this paper, we are particularly interested in whether l_2 is significantly different from the λ_2 associated with a specific distribution F. If so, then it is likely that the sample is not drawn from the distribution considered. The following lemma forms the basis of this idea:

Lemma 1: *Let X be a real-valued random variable with distribution function F and the second L-moment λ_2. Suppose that $\mathrm{E}[X^2]$ is finite. Let l_2 denote the sample second L-moment from a random sample of size N drawn from the distribution of X. Then, as $N \to \infty$:*

$$\sqrt{N}(l_2 - \lambda_2) \xrightarrow{d} N(0, \Lambda),$$

where the asymptotic variance $\Lambda = 2\int\int_{y<z}[1 + F(y)F(z)][F(y)(1 - F(z)]\mathrm{d}y\mathrm{d}z$.

Given λ_2 and Λ, an asymptotic 95% confidence interval of l_2 can be constructed by

$$CI_\infty = [\lambda_2 - 1.96\sqrt{\Lambda/N}, \lambda_2 + 1.96\sqrt{\Lambda/N}].$$

[1] Gini's mean difference statistic can be expressed as

$$\frac{1}{N(N-1)}\sum_{i}^{N}\sum_{j,j\neq i}^{N}|x_i - x_j|.$$

Similarly to the conventional testing procedure, if λ_2 is located outside the confidence interval above, then the hypothesis that X is drawn from the distribution F tends to be rejected.

Notice that λ_2 and Λ are fully determined by the distribution of X. For a standard normal distribution, $\lambda_2 = 1/\sqrt{\pi} \approx 0.5642$. While Λ has no explicit form,

$$\Lambda \approx 0.1628$$

by numerical approximation. Table 1 shows the sample mean, variance, and 2.5% and 97.5% quantiles of $\sqrt{N}(l_2 - 1/\sqrt{\pi})$ when a sample of size N is drawn from the standard normal distribution. We find that l_2 is asymptotically symmetric around $\lambda_2 = 1/\sqrt{\pi}$ and its variance for given N is a bit larger than Λ when N is small, but converges to Λ as $N \to \infty$. Additionally, the 2.5% and 97.5% quantiles are roughly equal to $-1.96\sqrt{\Lambda_N}$ and $1.96\sqrt{\Lambda_N}$, respectively. Together with Lemma 1, we have

$$L_N(z) = \frac{\sqrt{N}[l_2(z) - 1/\sqrt{\pi}]}{\sqrt{\Lambda_N}} \xrightarrow{d} N(0,1), \text{ as } N \to \infty, \tag{1}$$

where z is a random sample of size N from the standard normal distribution and Λ_N can easily be approximated by a simulation with an arbitrarily large number of repetitions for a given N.

3 A Test for Parameter Homogeneity

In practice, the true distribution of the estimator for each cross-sectional unit is usually unknown. However, after suitable standardization, it often has a normal limiting distribution. Therefore, given the number of cross-sectional units N, we can apply equation (1) to test for parameter homogeneity. In this section, the case in which the standardized estimator of the parameter of interest has a standard normal distribution is considered first. Then the case in which the distribution of the standardized estimator is asymptotically normally distributed will be studied.

Consider a panel data model:

$$y_{it} = f(x_{it}; \theta_i, \Theta_i) + e_{it}, \quad i = 1, \cdots, N, \ t = 1, \cdots, T,$$

where $f(x_{it}; \cdot)$ can be a linear or nonlinear function in θ_i; x_{it} denotes the regressors; θ_i is the parameter of interest for cross-sectional unit i, which can be an intercept, slope coefficient, or a parameter in $f(x_{it}; \cdot)$; Θ_i denotes all other parameters.

Let $\hat{\theta}_i$ denote the estimator of θ_i, $\sigma^2_{\hat{\theta}_i}$ the variance of $\sqrt{T}(\hat{\theta}_i - \theta_i)$ and $\widetilde{\sigma}^2_{\hat{\theta}_i}$ a consistent estimator of $\sigma^2_{\hat{\theta}_i}$. Here, we allow for cross-sectionally heteroskedastic errors in the model above. The hypothesis of interest is

$$H_0 : \theta_i = \theta \quad \forall i,$$

against the alternative hypothesis

$$H_A : \theta_i \neq \theta \quad \text{for some } i.$$

It is assumed that θ is unknown and $\widetilde{\theta}$ denotes a pooled estimator of θ. However, if θ equals a specific value θ_0, we can replace $\widetilde{\theta}$ with θ_0 and the following analysis is still valid.

We assume the following:

Assumptions:

A1. $s_{iT} = \sqrt{T}(\hat{\theta}_i - \theta_i)/\sigma_{\hat{\theta}_i}$ *is independent across* i *and has a standard normal distribution for all* i.

A2. $\sigma^2_{\hat{\theta}_i}$ *is strictly positive and* $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|\frac{1}{\sigma_{\hat{\theta}_i}} - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_{\hat{\theta}_i}}\right| = O_p(1)$.

A3. $\sigma^2_{\hat{\theta}_i}/\widetilde{\sigma}^2_{\hat{\theta}_i} = 1 + O_p(T^{-1})$

A4. Under the null hypothesis, $\widetilde{\theta}$ *is* $\sqrt{NT}$*-consistent.*

Because s_{iT} is infeasible, under the null hypothesis it is natural to replace θ_i with $\widetilde{\theta}$ for all i and estimate $\sigma_{\hat{\theta}_i}$ using $\widetilde{\sigma}_{\hat{\theta}_i}$. Particularly, $\widetilde{\sigma}_{\hat{\theta}_i}$ can be the restricted estimator of $\sigma^2_{\hat{\theta}_i}$ under the null hypothesis, which is shown by Pesaran and Yamagata (2005) to improve the size accuracy for large N panel data sets. Therefore, under the null hypothesis, s_{iT} can be estimated by

$$\widetilde{s}_{iT} = \frac{\sqrt{T}(\hat{\theta}_i - \widetilde{\theta})}{\widetilde{\sigma}_{\hat{\theta}_i}}.$$

The proposed test statistic for parameter homogeneity is then defined as

$$L_N(\widetilde{s}^N_T) = \frac{\sqrt{N}[l_2(\widetilde{s}^N_T) - 1/\sqrt{\pi}]}{\sqrt{\Lambda_N}}, \tag{2}$$

where $\widetilde{s}^N_T = (\widetilde{s}_{1T}, \cdots \widetilde{s}_{NT})'$, Λ_N denotes the approximated variance of $\sqrt{N}[l_2(z) - 1/\sqrt{\pi}]$ from a simulation with an arbitrarily large number of repetitions, and z is a random sample of size N from a standard normal distribution.

Next, the asymptotic distribution of $L_N(\widetilde{s}^N_T)$ under the null hypothesis will be established. Under Assumptions A2 and A3 and assuming that $\theta_i = \theta$ for all i,

$$\widetilde{s}_{iT} = s_{iT} - \frac{\sqrt{T}(\widetilde{\theta} - \theta)}{\sigma_{\hat{\theta}_i}} + O_p(T^{-1}). \tag{3}$$

Moreover, by Assumption A3 and the fact that $l_2(z)$ is a linear combination of z, it is shown in Appendix A that

$$L_N(\widetilde{s}_T^N) = \frac{\sqrt{N}\left[l_2(s_T^N) - 1/\sqrt{\pi}\right]}{\sqrt{\Lambda_N}} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\sqrt{N}/T\right), \tag{4}$$

where $s_T^N = (s_{1T}, \cdots s_{NT})'$. Therefore, by Lemma 1, $L_N(\widetilde{s}_T^N)$ is asymptotically normally distributed with variance Λ under the null hypothesis that $\theta_i = \theta$ for all i if Assumptions A1–A4 hold and $\sqrt{N}/T \to 0$ as N and T grow to infinity jointly. The following theorem provides a formal statement of the results.

Theorem 1: *Suppose that Assumptions A1–A4 hold and N and T tend to infinity jointly such that $\sqrt{N}/T \to 0$. Then, under the null hypothesis that $\theta_i = \theta$ for all i,*

$$L_N(\widetilde{s}_T^N) = \frac{\sqrt{N}[l_2(\widetilde{s}_T^N) - 1/\sqrt{\pi}]}{\sqrt{\Lambda_N}} \xrightarrow{d} N(0, 1).$$

Based on Theorem 1, the null is rejected if $|L_N(\widetilde{s}_T^N)| > 1.96$. This dispersion test can be interpreted as a test for normality because $1/\sqrt{\pi}$ and Λ_N are associated with the standard normal distribution. On the other hand, it is expected that the variance of s_{iT} would become larger under the alternative hypothesis because extra variation is introduced through $(\theta_i - \bar{\theta})$, where $\bar{\theta}$ is a weighted average of θ_i. Therefore, the distribution of s_{iT} across i is no longer standard normal under the alternative hypothesis. Similarly, provided that $s_{iT} \sim (0, 1)$ for all i, $L_N(\widetilde{s}_T^N)$ generally would be far from zero if s_{iT} is not normally distributed.

Next, consider the case in which s_{iT} is not normally distributed but asymptotically has a standard normal distribution. We replace Assumption A1 with:

Assumption *A1'* : $s_{iT} = \sqrt{T}(\hat{\theta}_i - \theta_i)/\sigma_{\hat{\theta}_i}$ is independent across i and $s_{iT} \xrightarrow{d} N(0, 1)$ as $T \to \infty$.

Given the assumption above, to make our test valid, T must increase to infinity faster than N to eliminate the distance of the distribution of s_{iT} from a standard normal distribution. This result is summarized in the following corollary to Theorem 1.

Corollary 1: *Suppose that Assumptions A1', A2– A4 hold and $T \to \infty$ and then $N \to \infty$. Then, under the null hypothesis that $\theta_i = \theta$ for all i,*

$$L_N(\widetilde{s}_T^N) = \frac{\sqrt{N}[l_2(\widetilde{s}_T^N) - 1/\sqrt{\pi}]}{\sqrt{\Lambda_N}} \xrightarrow{d} N(0, 1).$$

Corollary 1 also implies that the proposed test in general has no power against the alternative that $H_A' : \theta_i = \theta + \delta_i/T^a$ for some $a \geq 1/2$. When $a \geq 1/2$, $\widetilde{s}_{iT}$ is still $\sqrt{T}$-consistent and has an asymptotical standard normal distribution. On the other hand, if $a < 1/2$, then $\widetilde{s}_{iT}$ no longer has a standard normal asymptotic distribution and therefore the proposed test statistic would diverge as $T \to \infty$.

The proposed test can be generalized to investigate whether $K(\geq 2)$ parameters of interest are jointly cross-sectionally homogenous or not. For this purpose, $\widetilde{s}_{iT}$ should be modified as

$$\widetilde{s}'_{iT} = \frac{\sqrt{T}R(\hat{\theta}_i - \tilde{\theta})}{\sqrt{R\widetilde{\text{avar}}(\hat{\theta}_i)R'}},$$

where $R = (1, \ldots, 1)$ is a $1 \times K$ vector, $\widetilde{\text{avar}}(\hat{\theta}_i)$ is the sample counter part of asymptotic variance of $\hat{\theta}_i = (\hat{\theta}_{i1}, \ldots, \hat{\theta}_{iK})'$ under the null hypothesis, and $\hat{\theta}_{ik}$ is the estimator of the k^{th} parameter of interest for the i^{th} individual. Suppose that the cross-sectional independence and asymptotic normality of $\widetilde{s}'_{iT}$ hold for all individuals. Corollary 1 can be directly applied and the test becomes

$$L_N(\widetilde{s}^{N'}_T) = \frac{\sqrt{N}[l_2(\widetilde{s}^{N'}_T) - 1/\sqrt{\pi}]}{\sqrt{\Lambda_N}} \xrightarrow{d} N(0,1),$$

where $\widetilde{s}^{N'}_T = (\widetilde{s}'_{1T}, \ldots, \widetilde{s}'_{NT})'$.[2] Alternatively, we can also consider a Wald type version of the $L_N(\widetilde{\tau}^N_T)$ test, in which $\widetilde{\tau}^N_T = (\widetilde{\tau}_{1T}, \ldots, \widetilde{\tau}_{NT})'$ and

$$\widetilde{\tau}_{iT} = T(\hat{\theta}_i - \tilde{\theta})'[\widetilde{\text{avar}}(\hat{\theta}_i)]^{-1}(\hat{\theta}_i - \tilde{\theta}), \quad i = 1, \ldots N,$$

tends to have a $\chi^2(K)$ limiting distribution under very mild conditions. Let $\bar{\lambda}_K$ denote the mean of $l_2(\tau)$ when τ is a random sample of size N from $\chi^2(K)$. In Appendix B, we list $\bar{\lambda}_K$ and the asymptotic variance of $\sqrt{N}(l_2(\tau) - \bar{\lambda}_K)$ for various combinations of N and K, and then briefly describe the properties of this modified test.

Our test can be also extended in some applications where the estimator of the parameter of interest is not $\sqrt{T}$-consistent. For example, the estimator for a deterministic time trend is $T^{3/2}$-consistent. In this case, the proposed test is still valid because its estimator still has an asymptotical standard normal distribution after suitable standardization. In addition to testing for intercept or slope homogeneity, the proposed test potentially can be applied to nonlinear models because the estimator of the parameter of interest for each cross-sectional unit usually has an asymptotic standard normal distribution under quite general conditions. In this case, $\widetilde{s}_{iT}$ usually cannot approximate a standard distribution well in finite samples and the classical dispersion tests tend to be over-sized. Despite that, our test can provide more secure inference due to its robustness against the presence of outliers.

[2]The proposed test works in most of cases. The only exception is when $R\theta_1 = R\theta_2 = \ldots = R\theta_N$, where $R\theta_i = \sum_{k=1}^{K} \theta_{ik}$. However, this situation seems rare.

To illustrate how to make use of the proposed test in panel data models, the following three examples are considered. The first is a linear panel model with fixed effects in which the individual slope estimators are $\sqrt{T}$-consistent. The second example considers processes with deterministic time trends in which the individual slope estimators are $T^{3/2}$-consistent. The third example is a nonlinear panel model with fixed effects.

3.1 Example 1: A Linear Panel Model with Fixed Effects

Consider the following panel data model:

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it}, \quad i = 1, \cdots, N, \; t = 1, \cdots, T, \tag{5}$$

where α_i denotes the fixed effects, x_{it} can be a lagged dependent variable or a strictly exogenous regressor, and β_i denotes the unknown slope coefficient for individual i. The hypothesis of interest in this example is

$$H_0 : \beta_i = \beta \quad \forall i,$$

against the alternative hypothesis

$$H_A : \beta_i \neq \beta \quad \text{for some } i.$$

Under H_A, we further assume that the fraction of β_i that are the same does not tend to one as $N \to \infty$.

Under the null hypothesis, the infeasible standardized estimator can be expressed as

$$s_{iT} = \frac{\sqrt{T}(\hat{\beta}_i - \beta)}{\sigma_{\hat{\beta}_i}}$$

where $\hat{\beta}_i = \frac{\frac{1}{T}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)y_{it}}{Q_{iT}}$ is the OLS estimator of β_i, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, and $Q_{iT} = \frac{1}{T}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)^2$. Additionally, β can be estimated by

$$\widetilde{\beta} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{\sigma}_i^{-2}(y_{it} - \bar{y}_i)(x_{it} - \bar{x}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{\sigma}_i^{-2}(x_{it} - \bar{x}_i)^2}.$$

Under some mild conditions, s_{iT} has a standard normal limiting distribution and $\widetilde{\beta}$ is $\sqrt{NT}$-consistent under the null hypothesis.[3] This weighted FE-estimator of β is used by Swamy (1970) to test for slope homogeneity in large-T small-N panel data models and by Pesaran and Yamagata (2005) for large-N panel data models.

[3] For example, Assumptions A1-A4 are met under the assumptions used in Pesaran and Yamagata (2005).

Suppose that Q_{iT} is positive for all i and $Q_{iT} \xrightarrow{p} Q_i > 0$. If $\xi_{iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{it} - \bar{x}_i) e_{it} \xrightarrow{d} N(0, \sigma_i^2 Q_i)$, then $\sigma_{\hat{\beta}_i}^2 = \sigma_i^2 Q_i^{-1}$. Typically, σ_i^2 can be estimated using the individual OLS residuals. Under the null hypothesis, however, Pesaran and Yamagata (2005) suggest a more efficient estimator established from the pooled fixed effects residuals:

$$\widetilde{\sigma}_i^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left[(y_{it} - \bar{y}_i) - \hat{\beta}_{FE}(x_{it} - \bar{x}_i) \right]^2,$$

where

$$\hat{\beta}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)(x_{it} - \bar{x}_i)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2}.$$

Also, $\sigma_{\hat{\beta}_i}^2$ can be estimated by $\widetilde{\sigma}_{\hat{\beta}_i}^2 = \widetilde{\sigma}_i^2 Q_{iT}^{-1}$. Note that if $\sigma_{\hat{\beta}_i}^2 = Q_i^{-2}\Omega$, we can estimate it by $\widetilde{\sigma}_{\hat{\beta}_i}^2 = Q_{iT}^{-2} \left(\frac{1}{T} \sum_{t=1}^{T} \left[(y_{it} - \bar{y}_i) - \hat{\beta}_{FE}(x_{it} - \bar{x}_i) \right]^2 (x_{it} - \bar{x}_i)^2 \right)$ when errors have conditional heteroskedasticity of unknown forms; and we can use the automatic method developed by Andrews (1991) to consistently estimate $\sigma_{\hat{\beta}_i}^2$ when e_{it} is conditionally heteroskedastic and serially correlated over t for some i. Thus, for each i, s_{iT} can be estimated by

$$\widetilde{s}_{iT} = \frac{\sqrt{T} \left(\hat{\beta}_i - \widetilde{\beta} \right)}{\widetilde{\sigma}_{\hat{\beta}_i}}, \tag{6}$$

and the proposed statistic is

$$L_N(\widetilde{s}_T^N) = \frac{\sqrt{N} \left(l_2(\widetilde{s}_T^N) - 1/\sqrt{\pi} \right)}{\sqrt{\Lambda_N}},$$

where $\widetilde{s}_T^N = (\widetilde{s}_{1T}, \widetilde{s}_{2T}, \cdots, \widetilde{s}_{NT})'$. If $|L_N(\widetilde{s}_T^N)| > 1.96$, then the null that $\beta_i = \beta$ for all i is rejected at the 5% significance level.

3.2 Example 2: Processes with Deterministic Time Trends

Next, consider the following process:

$$y_{it} = \alpha_i + \phi_i t + e_{it}, \quad i = 1, \cdots, N, t = 1, \cdots, T, \tag{7}$$

where $e_{it} \sim (0, \sigma_i^2)$ is a white noise process over t and is cross-sectionally independent. This kind of regression is usually used in the study of economic growth. Here, we focus on whether or not the time trend coefficients are the same across i (i.e. $\phi_i = \phi$ for all i). Notice that $\hat{\phi}_i$, the OLS estimator of ϕ_i, is superconsistent in T for each i and after proper standardization,

$$D \begin{bmatrix} (\hat{\alpha}_i - \alpha_i) \\ (\hat{\phi}_i - \phi_i) \end{bmatrix} \xrightarrow{d} N(0, \sigma_i^2 Q^{-1}),$$

where $D = \text{diag}(\sqrt{T}, \sqrt{T^3})$, $x_t = [1, t]$, and $Q = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}$.

Under the null hypothesis that $\phi_i = \phi$ for all i, σ_i^2 can be estimated by

$$\widetilde{\sigma}_i^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left[(y_{it} - \bar{y}_i) - \hat{\phi}_{FE}(t - \bar{t}) \right]^2,$$

where $\bar{t} = T^{-1} \sum_{t=1}^{T} t$,

$$\hat{\phi}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)(t - \bar{t})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (t - \bar{t})^2}.$$

Moreover, ϕ can be estimated by the weighted-FE estimator

$$\widetilde{\phi} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{\sigma}_i^{-2} (y_{it} - \bar{y}_i)(t - \bar{t})}{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{\sigma}_i^{-2} (t - \bar{t})^2}.$$

Therefore, s_{iT} now can be estimated by

$$\widetilde{s}_{iT} = \frac{T^{3/2} \left(\hat{\phi}_i - \widetilde{\phi} \right)}{\widetilde{\sigma}_i \sqrt{M_{T,22}^{-1}}}, \tag{8}$$

where $M_{T,22}$ denotes the $(2,2)$ element of $(D^{-1} \left(\sum_{i=1}^{T} x_t' x_t \right) D^{-1})^{-1}$. The proposed statistic still takes the following form

$$L_N(\widetilde{s}_T^N) = \frac{\sqrt{N} \left(l_2(\widetilde{s}_T^N) - 1/\sqrt{\pi} \right)}{\sqrt{\Lambda_N}},$$

where $\widetilde{s}_T^N = (\widetilde{s}_{1T}, \widetilde{s}_{2T}, \cdots, \widetilde{s}_{NT})'$.

3.3 Example 3: A Nonlinear Panel Data Model with Fixed Effects

This test can be applied to nonlinear panel data models when s_{iT} has a standard normal limiting distribution. Here, we consider a general form for the panel data regression model with fixed effects and cross-sectionally heteroskedastic errors:

$$y_{it} = \alpha_i + f(x_{it}, \theta_i) + e_{it},$$

where α_i denotes the fixed effects for unit i, θ_i denotes the parameter of interest, and $e_{it} \sim (0, \sigma_i^2)$ and is cross-sectionally independent.

As an illustration, we here apply nonlinear least squares (NLS) estimation. Then, σ_i^2 can be estimated by

$$\widetilde{\sigma}_i^2 = \frac{\sum_{t=1}^{T} [\ddot{y}_{it} - \ddot{f}(x_{it}, \widetilde{\theta}_p)]^2}{T-2},$$

where $\ddot{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\ddot{f}(x_{it}, \cdot) = f(x_{it}, \cdot) - \frac{1}{T}\sum_{t=1}^{T} f(x_{it}, \cdot)$, and

$$\widetilde{\theta}_p = \arg\min_{\theta} \sum_{i=1}^{N}\sum_{t=1}^{T} [\ddot{y}_{it} - \ddot{f}(x_{it}, \theta)]^2.$$

θ can be estimated by

$$\widetilde{\theta}_w = \arg\min_{\theta} \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{[\ddot{y}_{it} - \ddot{f}(x_{it}, \theta)]^2}{\widetilde{\sigma}_i^2}.$$

Also, when the model has heteroskedastic errors of unknown forms over t for some i, $\sigma^2_{\hat{\theta}_i}$ can be estimated by

$$\widetilde{\sigma}^2_{\hat{\theta}_i} = \left(\frac{1}{T}\sum_{t=1}^{T} X_i' X_i\right)^{-2} \left(\frac{1}{T}\sum_{t=1}^{T} [\ddot{y}_{it} - \ddot{f}(x_{it}, \widetilde{\theta}_p)]^2 X_{it}^2\right),$$

where $X_i = (X_{i1}, \cdots, X_{iT})'$, and $X_{it} = \left.\frac{\partial \ddot{f}(x_{it}, \theta_i)}{\partial \theta_i}\right|_{\tilde{\theta}_p}$. We can apply the automatic method used by Andrews (1991) to consistently estimate $\sigma^2_{\hat{\theta}}$ when the model has heteroskedastic and serially correlated errors of unknown forms over t for some i. In the special case of homoskedasticity,

$$\widetilde{\sigma}^2_{\hat{\theta}_i} = \widetilde{\sigma}_i^2 \left(\frac{1}{T}\sum_{t=1}^{T} X_i' X_i\right)^{-1}.$$

Now, s_{iT} can be estimated by

$$\widetilde{s}_{iT} = \frac{\sqrt{T}(\hat{\theta}_i - \widetilde{\theta}_w)}{\widetilde{\sigma}_{\hat{\theta}_i}}, \tag{9}$$

where

$$\hat{\theta}_i = \arg\min_{\theta} \sum_{t=1}^{T} [\ddot{y}_{it} - \ddot{f}(x_{it}, \theta)]^2.$$

Under Assumptions A1', A2–A4, $L_N(\widetilde{s}_T^N)$ can be used to test for parameter homogeneity in the nonlinear parameter model, where $\widetilde{s}_T^N = (\widetilde{s}_{1T}, \cdots, \widetilde{s}_{NT})'$.

Alternatively, it is feasible to use the (quasi) maximum likelihood method or generalized method of moments to estimate the parameter of interest with and without restrictions in the model above. For these estimation methods, $L_N(\widetilde{s}_T^N)$ can still be used to test for parameter homogeneity under very general conditions if $\widetilde{s}_{iT}$ can be modified appropriately. The performance of the proposed test in terms of size and power highly depends on how closely $\widetilde{s}_T^N$ approximates a standard normal distribution. The proposed test might have size distortion in the nonlinear panel data models, especially when $\widetilde{s}_T^N$ is far from a standard distribution. However, as shown in the next section, our test potentially outperforms other tests due to its robustness against the presence of outliers.

4 Finite Sample Properties

In this section, Monte Carlo simulations are used to evaluate the finite sample properties of the proposed test (henceforth, the L_N test) in linear and nonlinear panel data models. We report size and power at the 5% nominal level for various pairs of N and T, which are often considered in panel data. We also compare the finite sample performance of the proposed test with those of the (quasi) likelihood ratio (LR) test, the modified Swamy (MS) test, and the test by Pesaran and Yamagata (2005), (henceforth, the PY test).

Based on the data generating processes considered, the LR test can be expressed as

$$(LR \text{ test}) \qquad LR = T\left(\sum_{i=1}^{N} \ln(\widetilde{\sigma}_i^2) - \ln(\hat{\sigma}_i^2)\right) \xrightarrow{d} \chi^2_{N-1}, \tag{10}$$

where $\widetilde{\sigma}_i^2$ denotes the estimator of σ_i^2 under the null hypothesis and $\hat{\sigma}_i^2$ is calculated without restricting the parameters of interest. This test is used by Lee, Pesaran, and Smith (1997) in a study of cross-country economic growth. The MS test can be expressed as

$$(MS \text{ test}) \qquad MS = \sum_{i=1}^{N} \widetilde{s}_{iT}^2 \xrightarrow{d} \chi^2_{N-1}. \tag{11}$$

Unlike the original test proposed by Swamy (1970), which is designed for data with small N and large T, we calculate $\widetilde{s}_{iT}$ using the restricted estimator $\widetilde{\sigma}_i^2$ instead of the unrestricted estimator $\hat{\sigma}_i^2$. The PY test can be rewritten as

$$(PY \text{ test}) \qquad PY = \sqrt{N}\left(\frac{N^{-1}\sum_{i=1}^{N} \widetilde{s}_{iT}^2 - 1}{\sqrt{2}}\right). \tag{12}$$

Notice that the MS and PY tests were originally developed to test the slope homogeneity in a linear panel data regression. However, these tests can be extended to nonlinear panel data models when $\widetilde{s}_{iT}$ has a standard normal asymptotic distribution.

4.1 Linear Panel Data Models

Two linear panel data models are considered. Following Pesaran and Yamagata (2005), we first consider a simple dynamic panel model:

$$y_{it} = (1-\rho_i)\alpha_i + \rho_i y_{i,t-1} + e_{it}, \tag{13}$$

where $\alpha_i \overset{iid}{\sim} N(1,1)$, $e_{it} \overset{iid}{\sim} N(0,\sigma_i^2)$, and $\sigma_i^2 \overset{iid}{\sim} \chi^2_{(2)}/2$. This model is a special case of (5) with $x_{it} = y_{i,t-1}$. Therefore, $\widetilde{s}_{iT}$ can be computed by (6) under the null hypothesis that $\rho_i = \rho$ for all i. Here, $\rho = \{0.2, 0.8\}$. Under the alternative hypothesis, $\rho_i \overset{iid}{\sim} U(\rho - 0.2, \rho + 0.2)$. For this experiment, $T + 50$ observations are generated but the first 49 observations are discarded. Five thousand replications are generated for each combination of (N, T). α_i, ρ_i, and σ_i^2 are fixed across replications.

Table 2 summarizes the results when data are generated from (13). First, when $\rho = 0.2$, the LR test has an accurate size for each combination of (N, T). However, the LR test tends to over-reject when $\rho_i = \rho = 0.8$. It is evident that the LR test is not appropriate when $N > T$, even if the finite sample size of the LR test declines to the nominal level as T increases. The PY and L_N tests tend to over-reject the null when N is much larger than T. When T is larger than 20, however, both tests tend to have reasonable sizes. It can also be found that their finite sample sizes converge to the nominal level as T increases. Additionally, the L_N test delivers a more accurate size than the PY and MS tests. This result can be traced to the fact that for a given N, the proposed statistic is standardized by using Λ_N instead of its asymptotic approximation Λ. Moreover, the LR and MS tests tend to have greater power than the PY and L_N tests, while the power of the L_N test seems satisfactory in this experiment and is quite comparable to that of the PY test.

The other linear panel data model considered is a process with deterministic time trends, as described in (7):

$$y_{it} = \alpha_i + \phi_i t + e_{it},$$

where $\alpha_i \overset{iid}{\sim} N(1, 1)$, $e_{it} \overset{iid}{\sim} \sigma_i \xi_{it}$, $\sigma_i^2 \overset{iid}{\sim} \chi^2_{(2)}/2$, and ξ_{it} is randomly drawn from $(\chi^2_{(2)}/2 - 1)$. $\phi_i = \phi = 0.03$ under the null hypothesis and $\phi_i \overset{iid}{\sim} U(0, 0.06)$ under the alternative hypothesis. Note that ϕ_i is $T^{3/2}$-consistent. $\widetilde{s}_{iT}$ is defined in (8). Five thousand replications are generated for each combination of (N, T). α_i, ϕ_i, and σ_i^2 are fixed across replications.

The simulation results are reported in Table 3. The main features of this table are similar to those shown in Table 2. The LR test suffers from size distortion when $N > T$. On the other hand, when N and T are small, the MS, PY test, and L_N tests all under-reject the null. The sizes of these tests converge to the nominal level when T increases. Moreover, the size delivered by L_N tends to be closer to the nominal level than those generated from the other tests.

In short, the LR test is not appropriate for the panel data models with $N > T$. The modified Swamy test is more sensitive to outliers than the proposed test. Also, the proposed test tends to have more accurate size than the PY test.

4.2 Nonlinear Panel Data Models

For nonlinear panel models, we consider the following two data generating processes:

$$y_{it} = \alpha_i + x_{it}^{\theta_i} + e_{it}, \quad (14)$$

$$y_{it} = \alpha_i + \exp(x_{it}\theta_i) + e_{it}. \quad (15)$$

The null hypothesis for the above equations is that $\theta_i = \theta$ for all i. The simulation results for the models above are reported in Tables 4 and 5, respectively. The data generating processes are described in detail in the footnote of each table. We apply NLS for these two regressions and, therefore, $\widetilde{s}_{iT}$ is estimated by (9).

Table 4 reports the results when data are generated from (14). As we can see, the finite sample size for each test is very similar to those in the linear panel data models considered in Section 4.1. It seems that all the tests considered can be successfully applied in nonlinear panel data models. In general, however, these tests may have severe size distortion when we consider other nonlinear models, such as (15).

In Table 5, we find that all of the tests tend to over-reject the null even though their size accuracy improves with T. On the other hand, the size of the proposed test is much closer to the nominal level than the others. These results are not surprising. Recall that whether the MS test or the L_N test has an accurate size depends on how well $\widetilde{s}_T^N$ approximates a standard normal distribution. Additionally, the performance of the PY test depends on whether the variance of $\widetilde{s}_T^N$ is close enough to one under the null hypothesis. For nonlinear panel data models, it might not be able to approximate the asymptotic variance (and distribution) well, particularly when T is not large enough and there tend to be more outliers in these cases. As expected, the MS, PY and L_N tests tend to over-reject the null. However, the L_N test can deliver more accurate inference in these cases due to its robustness in the presence of outliers.

5 Application: Cross-country Homogeneity in the Rates of Technology Growth and Speeds of Convergence

Consider the following empirical regression from the neoclassical Solow exogenous growth model, for $i = 1, \cdots, N$, $1 \leq t_1 < t_2 \leq T$:

$$y_{it_2} = \gamma_i y_{it_1} + x_{it_2}\beta_i + \phi_i t_2 + c_i + e_{it}. \quad (16)$$

Here y_{it} is the logarithm of per capita output; x_{it} is a vector consisting of $\ln(s_{it})$ and $\ln(n_{it} + g_i + \delta)$, where s_{it} is the savings rate at time t, n_{it} is the rate of growth of population, g_i is the rate of technology growth, and δ is the rate of depreciation, all of which are assumed to be exogenous. The coefficient for y_{it_1} is a function of the rate of convergence to steady state per capita output for country i, λ_i, i.e $\gamma_i = e^{-\lambda_i \tau}$ and $\tau = t_2 - t_1$. The coefficient β_i is a 2×1 vector, which can be expressed as a function of λ_i and α_i, the exponent of capital in the Cobb-Douglas production function. $\phi_i = g_i(1 - \gamma_i)$ and $c_i = (1 - e^{-\lambda_i \tau})A_i(0) + \phi_i \tau$, where $A_i(0)$ denotes the initial endowments for country i and c_i captures the country-specific effects.

Many empirical studies of cross-county economic growth are based on (16). For example, Mankiw, Romer, and Weil (1992) conduct cross-sectional analysis with $t_1 = 1960$, $t_2 = 1985$. Islam (1995) considers country-specific effects (i.e. the fixed effects) in panel data analysis under the assumption of homogeneity in γ_i and g_i. Lee, Pesaran, and Smith (1997) further reexamine the assumption of homogeneity in γ_i and g_i under the assumption that the steady state savings rate and population growth rate are constant at a country-specific level over time for each country and then (16) can be rewritten as the following linear panel regression with fixed effects μ_i and time effects η_t

$$y_{it} = \mu_i + \gamma_i y_{i,t-1} + g_i(1 - \gamma_i)t + \eta_t + e_{it}. \tag{17}$$

See Durlauf and Quah (1998) for a survey of growth empirics. Lee, Pesaran, and Smith (1997) use the LR test which tends to be over-sized as shown in the simulations in the previous section.

This empirical study is reconsidered as an illustration not only because the new test can deliver more accurate size but also because a longer cross country panel data set has become available. The data comes from the Penn World Tables (PWT v6.1) by Heston, Summers, and Aten (2002). We use a country list similar to that used in Mankiw, Romer, and Weil (1992) in order to compare the results with Islam (1995) and Lee, Pesaran, and Smith (1997). As a result, we examine the full set of 93 non-oil-producing countries, a group of 73 intermediate countries, and a group of 21 OECD countries.[4] The measure of real output per capita used is labeled RGDPL in the Penn World Tables. The parameters of interest are estimated over the period 1965-1995 (i.e. $t = 1$ is 1965) with $T = 31$ observations for each country.

[4]The data set considered in Mankiw, Romer, and Weil (1992) consists of 98 non-oil-producing countries (75 intermediate countries and 22 OECD countries) However, PWT6.1 does not have data for Liberia, Myanmar, Somalia, and Sudan. Also, Germany is removed from data set due to consolidation. See Heston, Summers, and Aten (2002) for details.

The estimation used in this paper follows Lee, Pesaran, and Smith (1997). Briefly, to estimate γ_i and g_i separately, (17) can be rewritten as

$$\begin{aligned} y_{it} &= \mu_i' + g_i t + \varepsilon_{it}, \\ \varepsilon_{it} &= \gamma_i \varepsilon_{i,t-1} + \eta_t + e_{it}, \end{aligned} \tag{18}$$

where

$$\mu_i' = A_i(0) + g_i + \frac{\alpha_i}{1-\alpha_i}\left[\log(s_i) + \log(n_i + g_i + \delta)\right].$$

Then, the following demeaned version of (18) can be used to remove the common time effects:

$$\begin{aligned} y_{it} - \overline{y_t} &= \mu_i' - \overline{\mu'} + g_i^d + \varepsilon_{it} - \overline{\varepsilon_t}, \\ \varepsilon_{it} - \overline{\varepsilon_t} &= \gamma_i(\varepsilon_{i,t-1} - \overline{\varepsilon_{t-1}}) + e_{it} - \overline{e_t}, \end{aligned} \tag{19}$$

where $\overline{y_t} = \frac{1}{N}\sum_{i=1}^N y_{it}$, $\overline{\varepsilon_t} = \frac{1}{N}\sum_{i=1}^N \varepsilon_{it}$, $\overline{e_t} = \frac{1}{N}\sum_{i=1}^N e_{it}$, $g_i^d = g_i - \bar{g}$, and $\overline{\mu'}$ and $\bar{g}$ are the averages of μ_i' and g_i, respectively.[5] Assume that $e_{it} \sim (0, \sigma_i^2)$ is cross-sectionally independent. $\hat{g}_i^d$ and $\hat{\gamma}_i$ are the quasi maximum likelihood estimators of g_i^d and γ_i under the restriction that $|\gamma_i| < 1$ for all individuals. Also, $\bar{g}$ can be estimated by $\hat{\bar{g}}$ based on

$$\begin{aligned} \bar{y}_t &= \overline{\mu'} + \bar{g}t + \overline{\varepsilon_t}, \\ \overline{\varepsilon_t} &= \gamma\overline{\varepsilon_{t-1}} + \overline{e_t}, \end{aligned} \tag{20}$$

and then g_i's are recovered by $\widehat{\bar{g}} + \hat{g}_i^d$.

We first estimate the parameters for each country separately. Then, the parameters are estimated either under the null hypothesis that $g_i = g$ or under the null hypothesis that $\gamma_i = \gamma$ for all i. To make use of the proposed test (and the PY test), we standardize $\hat{g}_i^d$ and $\hat{\gamma}_i$ separately by

$$\widetilde{s_{iT,\hat{g}_i^d}} = (\hat{g}_i^d - \hat{g}^d)/\tilde{\sigma}_{\hat{g}_i^d}, \tag{21}$$

$$\widetilde{s_{iT,\hat{\gamma}_i}} = (\hat{\gamma}_i - \hat{\gamma})/\tilde{\sigma}_{\hat{\gamma}_i}, \tag{22}$$

where $\hat{g}^d$ is estimated under the null hypothesis that $g_i = g$, $\tilde{\sigma}^2_{\hat{g}_i^d}$, the restricted variance of $\hat{g}_i^d$ for country i, is estimated under the null hypothesis that $g_i = g$ using the quadratic spectral kernel (QS) estimator with the data-dependent bandwidth developed by Andrews (1991) for

[5] Equation (19) is derived from (18) when $\gamma_i = \gamma$. However, as pointed out by Lee, Pesaran, and Smith (1997) *"even if γ_i's differ across countries, but not markedly so, removing time effects in this way can still partially eliminate cross-sectional correlation."*

each country.[6] Similarly, $\hat{\gamma}$ and $\tilde{\sigma}^2_{\hat{\gamma}_i}$ are estimated under the null hypothesis that $\gamma_i = \gamma$ for each country. Then, the L_N and PY tests are computed by (2) and (12), respectively.

Table 6 reports the results. Without imposing homogeneity in g_i and γ_i, the mean of $\hat{g}_i$ is 0.0136 for the non-oil-producing country sample, 0.0177 for the intermediate country sample, and 0.0276 the OECD country sample. However, under the null hypothesis that $g_i = g$ for all i, the pooled estimates of g are 0.0194, 0.0206, and 0.0274, respectively. The hypothesis of a common rate of technology growth across countries in each group is strongly rejected, with $L_N = 15.288$, 12.509, and 7.244. Also, the hypothesis of a common rate of convergence for each country to its country-specific steady state growth path is also rejected for all three samples, with $L_N = -4.7861$, -5.160 , and -3.090.[7] In contrast, Lee, Pesaran, and Smith (1997) cannot reject this hypothesis for the OECD country sample by using PWT data from 1965 to 1989.

Figure 1 shows the probability density estimates of standardized $\hat{g}_i^d$ and $\hat{\gamma}_i$ as defined in (21) and (22) for all three samples. The standardized estimated growth rates of technology have bimodal distributions for the non-oil-producing sample and the intermediate sample. This shows that the assumption of parameter homogeneity in g_i is inappropriate and it might be worthwhile to cluster countries into two (or more groups) based on standardized $\hat{g}_i^d$ for growth empirics using cross-country panel data. For the speed of convergence across countries, the kernel density plots show that there are many extreme observations of standardized $\hat{\gamma}_i$ on the left tail when the entire non-oil-producing country sample and the intermediate sample are considered, even though many of the standardized $\hat{\gamma}_i$'s are bunched together. In addition, for the OECD sample, the kernel density plot of standardized $\hat{\gamma}_i$'s has two peaks and provides possible evidence of heterogeneity in γ_i.

6 Concluding Remarks

This paper proposes a new dispersion test for parameter homogeneity in panel data models that is robust against outliers in the data. Conventional dispersion tests based on sample variances may have size distortions when extreme observations are present. We utilize the second L-moment to measure the dispersion of the distribution of sample estimates across individual units. The test statistic is a linear combination of order statistics. Because this

[6] To estimate $\widehat{\text{var}}(\hat{g}_i^d)$ and $\widehat{\text{var}}(\hat{\gamma}_i)$, we also use the Truncated Flat (TF) estimator of long run variance introduced in Chapter 2. The testing results based on the TF estimator are similar to those based on the QS estimator.

[7] When the null is false, the heterogeneity bias in a pooled estimate might bias up the estimates of restricted variance and therefore might result in a negative L_N.

test is less influenced by outliers, it can provide more reliable inference not only in linear but also in nonlinear panel data models.

The proposed test is shown to have a standard normal limiting distribution when either of the following is met: 1) the suitably standardized estimator of the parameter of interest for each cross-sectional unit has a standard normal distribution as long as $\sqrt{N}/T \to 0$ as N and T tend to infinity jointly, or 2) the suitably standardized estimator has a standard normal limiting distribution as long as T grows to infinity faster than N.

Using Monte Carlo simulations, we find that the proposed test delivers accurate size and satisfactory power even when N and T are not large. The finite sample performance of this new test also delivers more accurate size than other existing tests for various combinations of N and T particularly when the number of observations for each cross-sectional unit is small or when the model is nonlinear.

The proposed test is valid under the assumption of cross-sectional independence. Once this assumption is relaxed, the asymptotic mean and variance of the sample second L-moments will no longer be $1/\sqrt{\pi}$ and Λ. More work is needed to develop a proper test for parameter homogeneity under cross-sectional dependence.

Appendix A Proofs

Proof of Lemma 1. See Theorem 3 of Hosking (1990) for details. *Q.E.D.*

Proof of Equation (3). By Assumption A3 and Taylor's expansion around 1, we have

$$\sqrt{\sigma^2_{\hat{\theta}_i}/\tilde{\sigma}^2_{\hat{\theta}_i}} = 1 + \frac{1}{2}O_p(T^{-1}) + o_p(T^{-1}) = 1 + O_p(T^{-1}).$$

Therefore,

$$\begin{aligned} \tilde{s}_{iT} &= \frac{\sqrt{T}(\hat{\theta}_i - \tilde{\theta})}{\tilde{\sigma}_{\hat{\theta}_i}} = \frac{\sqrt{T}(\hat{\theta}_i - \theta)}{\sigma_{\hat{\theta}_i}} - \frac{\sqrt{T}(\tilde{\theta} - \theta)}{\sigma_{\hat{\theta}_i}} + O_p(T^{-1}) \\ &= s_{iT} - \frac{\sqrt{T}(\tilde{\theta} - \theta)}{\sigma_{\hat{\theta}_i}} + O_p(T^{-1}). \end{aligned} \tag{23}$$

Q.E.D.

Proof of Equation (4). Let $s_{[j:N],T}$ denote the j^{th} smallest observation from $s^N_T = (s_{1T}, \cdots, s_{NT})'$ and $w_{j:N} = 2j/(N-1) - 1$ the weight associated with $s_{[j:N],T}$ when computing $l_2(s^N_T)$. Analogously, let $\tilde{s}_{[j:N],T}$ denote the j^{th} smallest observation from $\tilde{s}^N_T = (\tilde{s}_{1T}, \cdots, \tilde{s}_{NT})'$ and let $\tilde{w}_{j:N} = (2j/(N-1)-1)$ be the weight associated with $\tilde{s}_{T,[j:N]}$. By (23) and Assumption A2 that $\tilde{\theta}$ is $\sqrt{NT}$ consistent under the null, we have

$$\tilde{w}_{j:N} - w_{j:N} = O_p\left(\frac{1}{\sqrt{N^3}}\right) + O_p\left(\frac{1}{NT}\right). \tag{24}$$

Therefore,

$$\begin{aligned} L_N(\tilde{s}^N_T) &= \frac{\sqrt{N}\left(l_2(\tilde{s}^N_T) - 1/\sqrt{\pi}\right)}{\sqrt{\Lambda_N}} = \frac{\sqrt{N}\left[\frac{1}{N}\left(\sum_{j=1}^N \tilde{w}_{j:N}\tilde{s}_{[j:N],T}\right) - 1/\sqrt{\pi}\right]}{\sqrt{\Lambda_N}} \\ &= \frac{\sqrt{N}\left[\frac{1}{N}\left(\sum_{j=1}^N w_{j:N}\tilde{s}_{[j:N],T}\right) - 1/\sqrt{\pi}\right]}{\sqrt{\Lambda_N}} + \frac{\sqrt{N}\left[\frac{1}{N}\left(\sum_{j=1}^N (\tilde{w}_{j:N} - w_{j:N})\tilde{s}_{[j:N],T}\right)\right]}{\Lambda_N} \\ &= \frac{\sqrt{N}\left[\frac{1}{N}\left(\sum_{j=1}^N w_{j:N} s_{[j:N],T}\right) - 1/\sqrt{\pi}\right]}{\sqrt{\Lambda_N}} - \sqrt{\frac{T}{\Lambda_N}}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^N \frac{w_{j:N}}{\sigma_{\hat{\theta}_i,j:N}}\right)(\tilde{\theta} - \theta) \\ &\quad + \frac{\sqrt{N}\left[\frac{1}{N}\left(\sum_{j=1}^N (\tilde{w}_{j:N} - w_{j:N}) s_{[j:N],T}\right)\right]}{\sqrt{\Lambda_N}} + O_p\left(\frac{\sqrt{N}}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= L_N(s^N_T) - I_1 + I_2 + O_p\left(\frac{\sqrt{N}}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \end{aligned}$$

where $\sigma_{\hat{\theta}_i,j:N}$ is associated with $s_{[j:N],T}$.

To obtain the desired result, it is required to show the convergence rates of I_1 and I_2. We first show that $I_1 = O_p\left(\frac{1}{\sqrt{N}}\right)$. Since $|w_{j:N}| \le 1$,

$$\frac{1}{\sqrt{N}}\sum_{j=1}^N \frac{w_{j:N}}{\sigma_{\hat{\theta}_i,j:N}} = \frac{1}{\sqrt{N}}\sum_{j=1}^N w_{j:N}\left(\frac{1}{\sigma_{\hat{\theta}_i,j:N}} - \overline{\frac{1}{\sigma}}\right) \le \frac{1}{\sqrt{N}}\sum_{j=1}^N \left|\frac{1}{\sigma_{\hat{\theta}_i,j:N}} - \overline{\frac{1}{\sigma}}\right| = O_p(1),$$

where $\overline{\frac{1}{\sigma}} = \frac{1}{T}\sum_{i=1}^N \frac{1}{\sigma_{\hat{\theta}_i}}$. Together with Assumptions A2 and A3, we have

$$I_1 = \sqrt{\frac{T}{\Lambda_N}}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^N \frac{w_{j:N}}{\sigma_{\hat{\theta}_i,j:N}}\right)(\tilde{\theta} - \theta) = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Next, consider I_2. Note that data are cross-sectionally independent and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}|s_{iT}| = O_p(1)$. Together with (24), we obtain

$$\begin{aligned} I_2 &= \frac{\left[\frac{1}{\sqrt{N}}\left(\sum_{j=1}^{N}(\tilde{w}_{j:N} - w_{j:N})s_{[j:N],T}\right)\right]}{\sqrt{\Lambda_N}} \\ &\leq \frac{\left(\sum_{j=1}^{N}|\tilde{w}_{j:N} - w_{j:N}|\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}|s_{[j:N],T}|\right)}{\sqrt{\Lambda_N}} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{T}\right). \end{aligned}$$

Q.E.D.

Therefore, the desired result follows.

Proof of Theorem 1. Together with Lemma 1, Equations (3) and (4), as N and T grow to infinity such that $\sqrt{N}/T \to 0$,

$$L_N(\tilde{s}_T^N) = \frac{\sqrt{N}\left(l_2(\tilde{s}_T^N) - 1/\sqrt{\pi}\right)}{\sqrt{\Lambda_N}} \xrightarrow{d} N(0,1)$$

Q.E.D.

Appendix B Construct a Test Based on $l_2(\tau)$ as $\tau \sim$ i.i.d $\chi^2(K)$

Let $\widetilde{\tau} = (\widetilde{\tau}_1, \widetilde{\tau}_2, \ldots, \widetilde{\tau}_N)'$, where

$$\widetilde{\tau}_i = T(\hat{\theta}_i - \tilde{\theta})'[\widetilde{\text{avar}}(\hat{\theta}_i)]^{-1}(\hat{\theta}_i - \tilde{\theta}),$$

$\widetilde{\text{avar}}(\hat{\theta}_i)$ is the sample counter-part of asymptotic variance of $\hat{\theta}_i = (\hat{\theta}_{i1}, \ldots, \hat{\theta}_{iK})'$ under the null hypothesis, and $\hat{\theta}_{ik}$ is the estimator of the k^{th} parameter of interest for the i^{th} individual. Under very mild conditions, $\widetilde{\tau}_i$ tends to have a $\chi^2(K)$ limiting distribution. The proposed test can be further extended when $\tilde{\tau}_i \xrightarrow{d} \chi^2(K)$, as $T \to \infty$. Let $\bar{\lambda}_K$ denote the sample mean of $l_2(\tau)$ when τ is a random sample of size N from $\chi^2(K)$ and let $\widehat{\Lambda}_{N,K}$ denote the sample variance of $\sqrt{N}(l_2(\tau) - \bar{\lambda}_K)$ for a given N. One can easily generate $\bar{\lambda}_K$ and $\widehat{\Lambda}_{N,K}$ by simulation for given K and N with a large number of repetitions. Table 7 lists $\bar{\lambda}_K$ and $\widehat{\Lambda}_{N,K}$ over one million repetitions for various combination of K and N commonly used in panel data studies. The test statistic becomes

$$L_2(\widetilde{\tau}_T^N) = \frac{\sqrt{N}(l_2(\tilde{\tau}) - \bar{\lambda}_K)}{\sqrt{\widehat{\Lambda}_{N,K}}}.$$

This test should work when $R\theta_1 = R\theta_2 = \ldots = R\theta_N$, and similarly to Theorem 1, it can be shown that

$$L_2(\widetilde{\tau}_T^N) \xrightarrow{d} N(0,1), \text{ as } T, N \to \infty \text{ jointly}, N/T \to 0.$$

References

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

DURLAUF, S. N., AND D. T. QUAH (1998): "The New Empirics of Economic Growth," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, no. 1, chap. 4, pp. 235–308.

GINI, C. (1912): "Variabilita e mutabilitaa contributo allo studio delle distribuzioni e delle relaztion statistiche," *Studi Economico-Giuridici delle Reale Universita di Cagliari*, 3, 3–159.

HESTON, A., R. SUMMERS, AND B. ATEN (2002): "Penn World Table Version 6.1," *Center for International Comparisons at the University of Pennsylvania.*

HOSKING, J. R. M. (1990): "L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," *Journal of Royal Statistical Society. Series B (Methodological)*, 52(1).

HOSKING, J. R. M., J. R. WALLIS, AND E. F. WOOD (1985): "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments," *Technometrics*, 27(3).

ISLAM, N. (1995): "Growth Empirics: A Panel Approach," *The Quarterly Journal of Economics*, 110(4), 1127–1170.

LANDWEHR, J. M., N. C. MATALAS, AND J. R. WALLIS (1979): "Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles," *Water Resources Research*, 15, 1055–1064.

LEE, K., M. H. PESARAN, AND R. SMITH (1997): "Growth and Convergence in a Multi-country Empirical Stochastic Solow Model," *Journal of Applied Econometrics*, 12, 357–392.

MANKIW, N. G., D. ROMER, AND D. N. WEIL (1992): "A Contribution to the Emprics of Economic Growth," *The Quarterly Journal of Economics*, 107(2), 407–437.

PESARAN, H., R. SMITH, AND K. S. IM (1996): "Dynamic linear model for heterogenous panels. In: Matyas, L., Sevestre, P., eds," *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Second Edtition, 145–195.

PESARAN, M. H., AND T. YAMAGATA (2005): "Testing Slope Homogeneity in Large Panels," *IEPR Working Paper, University of South California.*

SWAMY, P. A. V. B. (1970): "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38(2), 311–323.

Table 1: Descriptive Statistics of $\sqrt{N}(l_2(z) - 1/\sqrt{\pi})$.

N	Mean	Λ_N	2.5%	97.5%	$1.96\times\sqrt{\Lambda_N}$
10	0.0002	0.1851	-0.7845	0.8956	0.8433
15	-0.0000	0.1773	-0.7815	0.8647	0.8253
20	0.0000	0.1733	-0.7795	0.8495	0.8159
25	-0.0002	0.1709	-0.7787	0.8396	0.8104
30	0.0001	0.1692	-0.7778	0.8330	0.8062
40	0.0004	0.1682	-0.7788	0.8281	0.8039
50	-0.0000	0.1668	-0.7780	0.8213	0.8005
100	-0.0000	0.1647	-0.7796	0.8094	0.7955
150	-0.0002	0.1645	-0.7801	0.8072	0.7949
200	0.0005	0.1644	-0.7850	0.8057	0.7946
1000	0.0005	0.1628	-0.7846	0.7956	0.7900

(a) Mean denotes the finite sample mean of $\sqrt{N}(l_2(z) - 1/\pi)$, Λ_N denotes the finite sample variance of $\sqrt{N}(l_2(z) - 1/\pi)$, and 2.5% and 97.5% denote the 2.5 percentile and 97.5 percentile of $\sqrt{N}(l_2(z) - 1/\pi)$ respectively. z denotes a sample of size N randomly draw from a standard normal distribution.

(b) One million repetitions repetitions for each N.

Table 2: Size and Power of the Tests for Parameter Homogeneity in Panel Models with Heteroskadastic AR(1) Specifications.

$\rho = 0.2$ v.s $\rho_i \overset{iid}{\sim} U(0, 0.4)$

		Size(5%)					Size-adjusted Power				
	N	T=10	20	30	50	100	T=10	20	30	50	100
LR	20	5.18	4.84	4.72	4.50	5.28	7.42	17.56	39.60	76.38	89.56
	30	5.80	4.18	4.58	5.28	5.06	7.64	22.72	31.32	59.10	98.96
	50	5.20	4.80	5.10	5.40	4.74	10.28	24.80	43.24	74.80	99.96
	100	4.96	4.92	5.04	4.56	4.78	17.18	41.52	71.70	98.54	100.00
MS	20	0.12	1.18	1.92	2.62	4.18	8.00	18.04	40.02	76.70	89.24
	30	0.10	0.80	1.78	2.76	3.68	7.38	22.54	30.94	59.02	98.98
	50	0.00	0.58	1.10	2.66	3.18	9.76	25.34	44.40	74.88	99.96
	100	0.00	0.14	0.76	1.78	3.08	18.04	41.62	72.12	98.56	100.00
PY	20	2.92	2.00	2.42	2.36	3.54	5.62	12.22	29.48	67.82	83.26
	30	5.84	3.48	3.28	3.20	3.52	5.34	15.18	20.52	49.20	98.16
	50	12.78	5.20	3.84	3.90	4.18	6.36	16.62	34.62	64.42	99.86
	100	32.02	9.58	6.32	4.72	4.46	11.58	31.80	61.84	97.26	100.00
L_N	20	5.54	3.88	3.92	3.54	4.60	5.84	11.76	29.96	66.88	83.30
	30	7.98	4.86	4.34	4.28	4.28	5.42	14.86	20.40	48.52	98.18
	50	14.42	6.46	5.10	4.66	4.96	6.52	16.38	32.86	64.40	99.90
	100	29.26	9.64	6.50	5.08	4.80	11.18	30.52	60.92	97.00	100.00

$\rho = 0.8$ v.s $\rho_i \overset{iid}{\sim} U(0.6, 0.99)$

		Size(5%)					Size-adjusted Power				
	N	T=10	20	30	50	100	T=10	20	30	50	100
LR	20	19.28	14.82	12.24	9.56	7.02	8.68	50.56	62.34	100.00	98.60
	30	24.90	17.20	13.46	11.12	8.60	47.02	99.88	99.98	100.00	100.00
	50	34.14	21.64	17.66	13.60	9.42	68.28	99.74	100.00	100.00	100.00
	100	50.30	33.40	24.58	17.84	11.98	24.20	100.00	100.00	100.00	100.00
MS	20	1.38	5.04	5.90	6.30	5.74	9.10	51.64	60.34	100.00	98.64
	30	1.70	5.26	5.62	7.08	6.76	40.66	99.90	99.98	100.00	100.00
	50	1.40	5.34	7.12	7.80	6.70	54.22	99.02	100.00	100.00	100.00
	100	0.74	6.04	7.88	8.84	8.26	22.68	100.00	100.00	100.00	100.00
PY	20	1.12	3.14	3.54	4.44	3.92	7.38	41.20	49.18	100.00	97.56
	30	1.58	3.12	3.84	4.54	4.68	28.96	99.68	99.96	100.00	100.00
	50	1.70	3.54	4.66	5.80	4.70	41.08	97.52	100.00	100.00	100.00
	100	2.36	4.08	5.08	5.40	5.70	15.96	100.00	100.00	100.00	100.00
L_N	20	1.96	3.14	3.50	4.00	4.56	6.86	24.26	35.08	96.28	97.30
	30	2.22	2.84	3.18	4.56	4.78	17.08	48.34	97.34	98.02	100.00
	50	2.28	2.86	3.68	4.78	4.38	19.46	81.42	67.18	98.40	100.00
	100	3.04	3.10	3.80	4.40	5.12	13.18	86.72	98.56	99.82	100.00

(a) Data are generated by $y_{it} = (1-\rho_i)\alpha_i + \rho_i y_{i,t-1} + e_{it}$, where $\alpha_i \overset{iid}{\sim} N(1,1)$, $e_{it} \overset{iid}{\sim} N(0, \sigma_i^2)$, and $\sigma_i^2 \overset{iid}{\sim} \chi_2^2/2$.

(b) $T + 50$ observations are generated but the first 49 observations are discarded. Five thousand replications are generated for each combination of (N, T).

(c) α_i, ρ_i, and σ_i^2 are fixed across replications.

Table 3: Size and Power of the Tests for Parameter Homogeneity in Panel Models with Deterministic Time Trends.

$\phi = 0.03$ v.s. $\phi_i \overset{iid}{\sim} U(0, 0.06)$

		Size (5%)					Size-adjusted Power				
		T=10	20	30	50	100	T=10	20	30	50	100
LR	$N = 20$	23.54	11.02	7.24	6.46	5.46	5.00	11.36	29.12	89.30	100.00
	30	30.18	12.82	9.20	7.00	6.04	5.76	10.52	28.72	97.48	100.00
	50	40.84	16.06	11.04	8.24	6.78	6.50	15.16	44.70	100.00	100.00
	100	64.70	24.72	15.08	9.80	7.32	6.40	22.60	87.64	100.00	100.00
MS	$N = 20$	0.58	1.78	2.40	3.42	4.04	4.84	11.24	29.02	89.36	100.00
	30	0.40	1.58	2.64	3.46	4.46	5.92	10.62	28.46	97.44	100.00
	50	0.22	1.28	2.82	3.04	4.58	6.24	14.94	44.56	100.00	100.00
	100	0.08	0.98	2.20	3.16	3.92	6.50	22.30	87.80	100.00	100.00
PY	$N = 20$	1.04	1.58	2.08	2.52	2.88	4.58	8.46	19.54	83.72	100.00
	30	1.64	2.24	2.86	2.98	3.68	5.32	7.24	18.94	95.66	100.00
	50	2.80	2.42	3.44	3.64	4.28	5.36	10.08	31.12	99.96	100.00
	100	5.78	3.08	3.70	4.02	3.86	5.48	16.26	78.88	100.00	100.00
L_N	$N = 20$	2.26	2.78	3.44	3.82	4.12	4.64	8.22	19.00	82.02	100.00
	30	2.56	3.24	3.92	3.84	4.64	5.26	7.20	19.46	95.24	100.00
	50	3.16	3.14	3.92	4.58	4.90	5.36	9.42	30.84	99.96	100.00
	100	4.64	3.22	4.12	4.58	4.16	5.48	15.68	78.22	100.00	100.00

(a) Data are generated from $y_{it} = \alpha_i + \phi_i t + e_{it}$, where $\alpha_i \overset{iid}{\sim} N(1,1)$, $e_{it} \overset{iid}{\sim} \sigma_i \xi_{it}$, $\sigma_i^2 \overset{iid}{\sim} \chi^2_{(2)}/2$, and ξ_{it} is randomly drawn from $(\chi^2_{(2)}/2 - 1)$.

(b) $\phi_i = \phi = 0.03$ under the null hypothesis and $\phi_i \overset{iid}{\sim} U(0, 0.06)$ under the alternative hypothesis. Note that ϕ_i is $T^{3/2}$-consistent. $\widetilde{s}_{iT}$ is defined in (8).

(c) Five thousand replications are generated for each combination of (N, T). α_i, ϕ_i, and σ_i^2 are fixed across replications.

Table 4: Size and Power of the Tests for Parameter Homogeneity in a Nonlinear Panel Model – (14).

$\theta = 0.4$ v.s. $\theta_i \overset{iid}{\sim} U(0.2, 0.6)$

	Size (5%)					Size-adjusted Power				
		$T = 20$	30	50	100		$T = 20$	30	50	100
MS	$N = 20$	2.90	1.20	1.40	2.40	$N = 20$	41.50	77.80	96.20	100.00
	30	7.40	5.90	7.10	10.30	30	39.00	84.30	99.30	100.00
	50	4.20	1.80	2.70	6.20	50	74.40	99.70	100.00	100.00
	100	2.90	0.60	0.90	2.30	100	95.90	100.00	100.00	100.00
PY	$N = 20$	4.20	1.90	2.70	2.30	$N = 20$	25.50	70.70	94.90	100.00
	30	6.50	4.00	4.70	7.60	30	20.30	76.30	97.80	100.00
	50	5.80	3.60	3.70	6.00	50	52.80	99.60	100.00	100.00
	100	12.40	7.90	5.10	5.20	100	81.60	99.90	100.00	100.00
L_N	$N = 20$	4.60	3.30	4.40	3.10	$N = 20$	32.80	71.00	95.20	100.00
	30	5.50	4.10	5.00	9.20	30	28.90	76.60	98.50	100.00
	50	5.10	3.50	4.70	6.30	50	70.40	99.60	100.00	100.00
	100	10.80	8.80	5.40	5.40	100	96.70	100.00	100.00	100.00

(a) Data are generated from (14): $y_{it} = \alpha_i + x_{it}^{\theta_i} + e_{it}$, where the individual effects $\alpha_i \overset{iid}{\sim}$, $x_{it} \overset{iid}{\sim} \chi^2_{(2)}/2$, $e_{it} \overset{iid}{\sim} N(0, \sigma_i^2)$, and $\sigma_i^2 \overset{iid}{\sim} U(0.5, 1.5)$.

(b) Under the null hypothesis, $\theta_i = 0.4$ for all i; under the alternative hypothesis, $\theta_i \overset{iid}{\sim} U(0.2, 0.6)$.

(c) One thousand replications are generated for each combination of (N, T). α_i, θ_i, and σ_i^2 are fixed across replications.

Table 5: Size and Power of the Tests for Parameter Homogeneity in a Nonlinear Panel Model – (15).

$\theta_i = 0.7$ v.s. $\theta_i \overset{iid}{\sim} U(0.4, 1)$

		Size (5%)					Size-adjusted Power			
		$T = 20$	30	50	100		$T = 20$	30	50	100
MS	$N = 20$	24.00	18.30	14.80	11.70	$N = 20$	27.20	51.90	78.50	99.50
	30	36.20	30.00	20.10	16.30	30	38.00	63.30	93.20	100.00
	50	42.20	31.30	25.00	18.20	50	54.30	81.90	98.90	100.00
	100	56.60	41.60	30.50	22.30	100	72.10	96.40	99.90	100.00
PY	$N = 20$	18.70	13.00	11.50	8.70	$N = 20$	21.00	41.80	71.70	99.20
	30	29.50	23.00	13.90	12.30	30	27.00	59.50	89.20	100.00
	50	34.10	23.40	18.70	13.50	50	42.40	78.40	98.20	100.00
	100	48.20	32.00	21.40	15.50	100	60.40	94.20	99.80	100.00
L_N	$N = 20$	14.40	9.70	10.10	7.00	$N = 20$	19.60	41.20	72.10	98.80
	30	24.70	19.90	11.20	9.80	30	29.10	55.50	87.80	99.90
	50	27.60	18.30	14.80	10.60	50	43.20	73.60	97.80	100.00
	100	40.60	26.00	16.30	11.90	100	62.50	93.30	99.70	100.00

(a) Data are generated from (15): $y_{it} = \alpha_i + \exp(x_{it}\theta_i) + e_{it}$, where the individual effects $\alpha_i \overset{iid}{\sim} U(-0.5, 0.5)$, $x_{it} \overset{iid}{\sim} \chi^2_{(2)}/2$, $e_{it} \overset{iid}{\sim} N(0, \sigma_i^2)$, and $\sigma_i^2 \overset{iid}{\sim} U(0.5, 1.5)$.

(b) Under the null hypothesis, $\theta_i = 0.7$ for all i and under the alternative hypothesis that $\theta_i \overset{iid}{\sim} U(0.69, 0.71)$.

(c) One thousand replications are generated for each combination of (N, T). α_i, θ_i, and σ_i^2 are fixed across replications.

Table 6: Descriptive Statistics of Estimated Coefficients and Test Results for the Cross-country Homogeneity in the Rates of Convergence and Technology Growth.

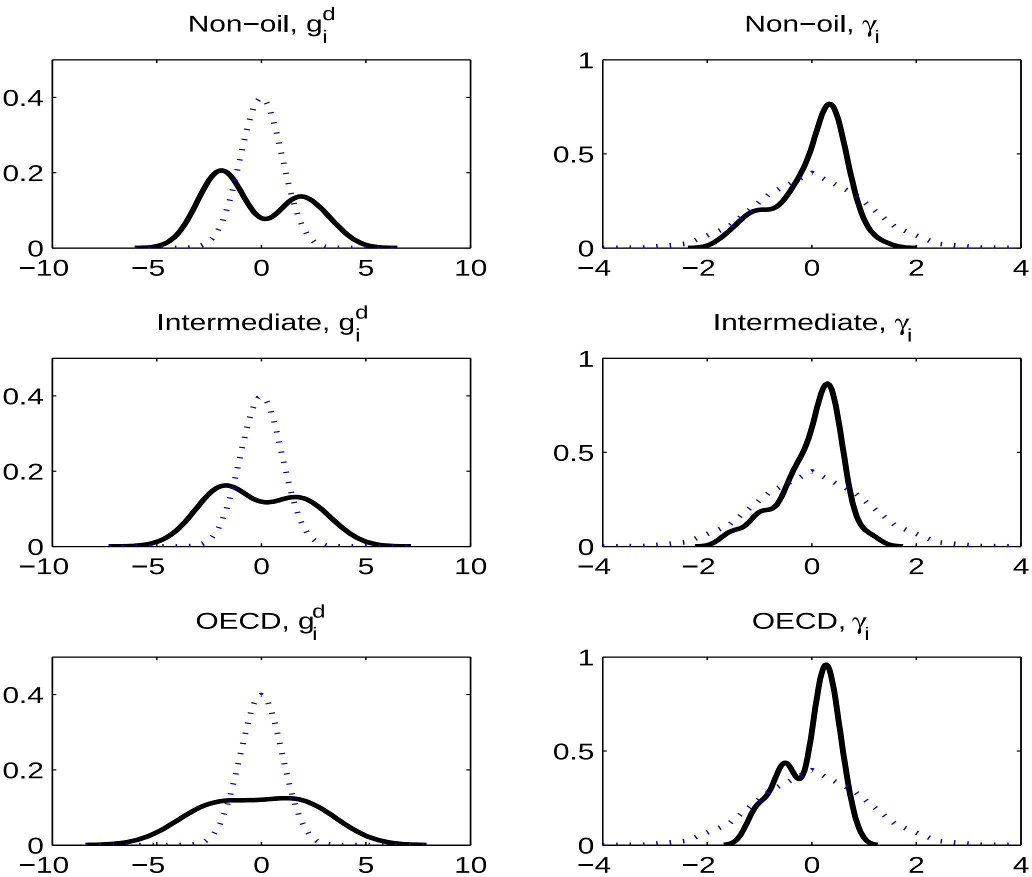
	Non-oil (N =93)		Intermediate (N =73)		OECD (N =21)	
LL	4797.199		4029.733		1429.038	
	$\hat{g}_i$	$\hat{\gamma}_i$	$\hat{g}_i$	$\hat{\gamma}_i$	$\hat{g}_i$	$\hat{\gamma}_i$
Mean	0.0136	0.7139	0.0177	0.7283	0.0276	0.7635
Standard Deviation	0.0202	0.1945	0.0180	0.1928	0.0072	0.1402
Median	0.0161	0.7374	0.0196	0.7457	0.0272	0.7997
Maximum	0.0642	1.0000	0.0642	1.0000	0.0420	1.0000
Minimum	-0.0333	0.0769	-0.0280	0.0246	0.0120	0.5398
	$H_0 : g_i = g$ for all i					
	Non-oil (N =93)		Intermediate (N =73)		OECD (N =21)	
LL	2958.323		2624.148		1169.874	
LR	3677.752*		2811.172*		518.327*	
PY	24.166*		18.700*		10.823*	
L_n	15.288*		12.509*		7.244*	
	$\hat{g}$	$\hat{\gamma}_i$	$\hat{g}$	$\hat{\gamma}_i$	$\hat{g}$	$\hat{\gamma}_i$
Mean	0.0194	0.7923	0.0206	0.7897	0.0274	0.7890
Standard Deviation		0.1209		0.1304		0.1086
Median		0.8039		0.7975		0.7931
Maximum		1.0000		1.0000		1.0000
Minimum		0.2920		0.2413		0.5984
	$H_0 : \gamma_i = \gamma$ for all i					
	Non-oil (N =93)		Intermediate (N =73)		OECD (N =21)	
LL	4687.207		3943.720		1410.506	
LR	219.984*		172.026*		37.065*	
PY	-3.922*		-4.025*		-2.434*	
L_n	-4.786*		-5.160*		-3.090*	
	$\hat{g}_i$	$\hat{\gamma}$	$\hat{g}_i$	$\hat{\gamma}$	$\hat{g}_i$	$\hat{\gamma}$
Mean	0.0136	0.6771	0.0176	0.6989	0.0275	0.7289
Standard Deviation	0.0201		0.0179		0.0072	
Median	0.0162		0.0190		0.0270	
Maximum	0.0643		0.0643		0.0424	
Minimum	-0.0329		-0.0282		0.0122	

(a) LL is the maximized value of the quasi log-likelihood function based on (19). $\hat{g}_i = \hat{g}_i^d + \hat{\bar{g}}$. $\hat{\gamma}_i$ and $\hat{g}_i^d$ are estimated by (19) and $\hat{\bar{g}}$ is estimated by (20) for each group. The estimates of g_i^d and γ_i have been standardized by (21) and (22), respectively. Then, the L_N test and the PY test are calculated by (2) and (12), respectively.

(b) The mean, standard deviation, median, maximum, and minimum of $\hat{\gamma}_i$ and $\hat{g}_i$ are calculated.

(c) * indicates that the hypothesis is rejected at the 5% significace level.

Figure 1: Sample Density of the Standardized Estimates of the Rates of Technology Growth and the Speeds of Convergence.



(a) The estimates of g_i^d and γ_i are standardized by (21) and (22), respectively.

(b) The probability density estimates of the standardized estimates are based on a normal kernel function.

(c) The black solid line denotes the kernel density of standardized estimates and the blue dashed line denotes the standard normal density.

(d) The non-oil producing group consists of 93 countries, the intermediate group consists of 73 countries, and the OECD group consists of 21 countries.

Table 7: Sample Mean and Asymptotic Variance of $l_2(\tau)$ when $\tau \sim \chi^2(K)$.

K	2		3		4		5		10		15	
N	$\bar{\lambda}_K$	$\widehat{\Lambda}_{N,K}$	$\bar{\lambda}_K$	$\widehat{\Lambda}_{N,K}$	$\bar{\lambda}_K$	$\widehat{\Lambda}_{N,K}$	$\bar{\lambda}_K$	$\widehat{\Lambda}_{N,K}$	$\bar{\lambda}_K$	$\widehat{\Lambda}_{N,K}$	$\bar{\lambda}_K$	$\widehat{\Lambda}_{N,K}$
10	1.000	1.409	1.272	1.809	1.499	2.197	1.697	2.577	2.461	4.456	3.040	6.304
15	1.000	1.380	1.274	1.773	1.500	2.142	1.698	2.504	2.461	4.293	3.039	6.083
20	1.000	1.365	1.273	1.750	1.500	2.117	1.698	2.472	2.461	4.223	3.040	5.952
25	1.000	1.363	1.273	1.742	1.500	2.100	1.697	2.452	2.460	4.183	3.039	5.894
30	1.000	1.355	1.273	1.729	1.500	2.087	1.698	2.436	2.462	4.157	3.040	5.858
40	1.000	1.352	1.273	1.723	1.500	2.073	1.698	2.419	2.461	4.117	3.039	5.792
50	1.000	1.350	1.273	1.716	1.500	2.070	1.698	2.417	2.460	4.100	3.038	5.784
100	1.000	1.343	1.273	1.708	1.500	2.055	1.698	2.397	2.461	4.067	3.039	5.717
150	1.000	1.340	1.273	1.702	1.500	2.052	1.698	2.387	2.461	4.049	3.039	5.694
200	1.000	1.337	1.273	1.701	1.500	2.047	1.698	2.384	2.461	4.043	3.039	5.689
1000	1.000	1.333	1.273	1.688	1.500	2.030	1.698	2.366	2.461	4.031	3.039	5.675

(a) one million repetitions for each (N, K) are used to generate this table.