Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership Is Unknown

Chang-Ching Lin* Serena Ng†

December 2007

Academia Sinica Columbia University

Abstract

This paper proposes a data-dependent, semi-parametric method for estimating panel data models with grouped specific parameters when group membership is not known. We first create a set of "pseudo" threshold variables based on time series estimation of the individual specific parameters. We then use these variables to stratify individuals. The problem of parameter heterogeneity is turned into estimation of a panel threshold model in which the threshold variables are themselves being estimated. The framework can accommodate fixed effects as well as cross-sectionally heterogeneous variances. We show that individuals can be consistently sorted into groups distinguished by parameter heterogeneity when N and T are large. We also extend the K-means algorithm to panel data regressions with fixed effects. Results with the pseudo threshold approach are compared.

Keywords: Parameter Heterogeneity, Threshold Models, Cluster Analysis.

*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. The author gratefully acknowledges financial support from Taiwan National Science Council grant NSC-96-2415-H-001-025.

†Department of Economics, Columbia University, 420 West 118 St., New York, NY 10027 USA tel: 212-854-5488 Email: serena.ng@columbia.edu.

1 INTRODUCTION

This paper considers estimation of panel data models with fixed effects when some of the units in the panel share common parameters and can be grouped (or clustered), but that group membership is not known to the econometrician. Our approach is to turn the problem into one of estimating threshold regressions, but that the threshold variable is itself being estimated. We refer to this as a 'pseudo' threshold approach.

Our analysis is motivated by the fact that the data that are commonly used in economic analysis are heterogeneous in multiple dimensions. While pooled estimation is convenient and the convergence rate is increasing in the number of cross-section units, pooling is inefficient when there is parameter heterogeneity, even though consistent estimation of the population mean of the parameters can still be possible. Evidence of parameter heterogeneity is not uncommon. For example, Hsiao and Tahmiscioglu (1997) find heterogeneity in the parameters of equations that describe investment dynamics and observed that such differences cannot be explained by commonly considered firm characteristics. Using the Penn World Table (PTW) data, Lee, Pesaran, and Smith (1997) rejected the hypotheses that the rate of technological growth and the rate of convergence of per capita output to the steady state level are the same across countries. On the other hand, assuming complete parameter heterogeneity is also inefficient as the problem reduces to time series estimation on a unit by unit basis which does not take advantage of the panel structure of the data at all. Partitioning of the data into groups is thus an immediate approach that permits pooling and yet still accommodates the heterogeneous nature of the panel. The main obstacle is how to form clusters when group membership is not known.

The most common approach is to form clusters using a priori information from economic theory or simply by intuition. For example, age and occupation are often used to split a sample of household data. In some analysis, households are considered liquidity constrained if their wealth exceeds a certain level. In firm level analysis, units are grouped by their capital intensity. In cross-country analysis, groups are sometimes formed depending on whether a country is a member of the OECD. Spatial information can also provide a meaningful one-dimensional criterion for clustering. However, in regressions with multiple covariates including the examples above, there may be several logical ways to partition the sample, and in such cases, how to actually split the sample is no longer clear-cut. Even if it is feasible to cluster the data by observed variables as guided by intuition or theory, the resulting clusters need not be optimal from a statistical point of view. This is not to mention that sometimes we may not have prior information about the group structure at all.

We handle the problem of identifying unknown group structure by estimating a panel threshold model. Goldfeld and Quandt (1973) were the first to use threshold variables, also referred to as transition variables, to form clusters. They considered a model in which the clusters are determined by a linear function of several transition variables. They proposed a so-called D-method within the maximum likelihood framework to enable estimation of the parameters in the transition function. The D method assumes deterministic switching of regimes, and stands in contrast to the so-called λ-method in which units are assigned to regimes in a random manner. A more popular idea, also due to Goldfeld and Quandt (1973), is to partition a data set based on a known threshold variable taking on an unknown threshold value. Threshold autoregressive models, structure break models of the suddenly changing

autoregressive model of Tyssedal and Tjøstheim (1988), are variations of this idea in a time series context. Hansen (1999) extended time series threshold regressions to non-dynamic panel regressions. The coefficients are allowed to vary not only across individuals, but also over time for a given individual. The regimes or groups in Hansen's analysis are clustered according to some pre-specified observed variable. While the final model restricts the parameters to be the same within groups, group membership can change over time.

Our analysis differs from a conventional threshold or breakpoint problem in that we do not use observed covariates to split the sample. Instead, we exploit the fact that when T is large, the unit-specific parameters can be consistently estimated. From these estimated coefficients, we form threshold variables to partition the sample. Our analysis proceeds in three steps. First, we obtain time series estimates of the response coefficients for every unit. Standardizing the individual parameter estimates yields a set of "pseudo" threshold variables. Second, the threshold value is estimated by fixed-effects estimation, conditional on the pseudo variables. Third, the model is re-estimated by groups. Thus, units within a group have homogeneous parameters but the parameters are heterogeneous across groups.

Our pseudo-threshold analysis has several advantages. First, it is entirely data dependent and does not require selection of a transition variable(s) in advance. It is a model rather than a graphical based approach and formal statistical statements can be made about the estimated groups. Second, the method is valid even in the presence of fixed effects, and it can be modified to accommodate the downward bias that is often found in dynamic panels. Third, the method also allows us to control for cross-sectionally heterogeneous variances. As we will see, we can distinguish cluster-specific slope coefficients

that are different in the order $T^{-1/2+\Delta}$, $0 < \Delta \leq 1/2$.

A widely popular tool in cluster analysis is the K-means method. While in theory, the algorithm may converge to a local optimum and the method can also be computationally demanding, it tends to have good finite sample properties. We extend the K-means method to panel regressions with fixed effects and use simulations to compare its performance with our method. We apply the methods to study economic growth across countries. The results indicate that some countries in the low growth club have high per capita income, or belong to the OECD. Splitting the data by geographical location, OECD membership, the level of income do not seem to produce clusters that are statistically optimal.

2 PRELIMINARIES

Several methods have been proposed to deal with parameter heterogeneity. Of the non-Bayesian options, one approach is to parametrize the coefficients as a function of individual characteristics, see for example, Alvarez, Browning, and Ejrnæs (2002). Although easy to implement, the approach is sensitive to the functional form used. One can also non-parametrically use a random coefficient model to allow for variations in the coefficients. See, for example, Swamy (1970) and Hsiao and Pesaran (2004). This approach provides efficient estimates for the average effect of y on x, but is uninformative about the response at a more disaggregated level, which is sometimes an object of interest. We consider a hybrid model that is less restrictive than assuming homogeneous parameters but puts more structure than a model that allows complete parameter heterogeneity. Specifically, we allow the parameters to be homogeneous within groups but heterogeneous across groups, which is a form

of model based clustering.

Model based clustering has been used in the statistics literature to partition a set of data, $x_i, i = 1, \ldots N$, into K groups. See, for example, Fraley and Raftery (2002), Hall, Muller, and Wang (2006), and Chiou and Li (2007). The approach is motivated by the idea that a set of data with a group structure can be thought of as generated by a mixture of distributions such that an observation drawn from sub-population k has density $f_k(, x_i, |\theta_k)$. If q_i is the identifying label, i.e. $q_i = k$ if unit i belongs to group k, then one can maximize the likelihood $L(x; \theta, q) = \prod_{i=1}^{N} \prod_{t=1}^{T} f_{q_i}(x_{it}; \beta)$ with respect to β. In practice, the EM algorithm is often used for estimation; the identifier q_i is treated as an unknown and set according to the empirical probability of the group to which unit i belongs. The method can be cumbersome if N is large because we need to consider up to 2^N possible combinations to find the maximum.

Sun (2005) estimated panel models with unknown group structure using a modified EM algorithm with the restriction that the units in a cluster share the common parameters. Consistency and asymptotic normality of the maximum likelihood estimator are proved under the assumption that N is large and T is fixed. He used a logit regression to infer which group unit i belongs, and then applied weighted least squares to estimate the group parameters. The analysis heavily depends on the parametric assumptions made. Recently, Juárez and Steel (2007) proposed a Bayesian method to cluster units in panel data. Their analysis, however, is based on two strong assumptions: (1) the errors are cross-sectionally homogeneous and are independently drawn from a t-distribution, and (2) the individual-specific fixed effects are normally distributed.

Likelihood based clustering methods tend to split the sample according to the estimated probability of which group a unit belongs. We now present a

semi-parametric method that is not likelihood based, is computationally less demanding, and can consistently estimate the partitions in a sense to be made precise. We assume that we have a balanced panel of data with observations on N cross-section units over T time periods. There are G regressors and K clusters, and to introduce the main idea, we start with the simple case of $G = 1$ and $K = 2$. Let N_1 and N_2 denote the number of individuals in clusters I_1 and I_2, respectively. The data are assumed to be generated as

$$\tilde{y}_{it} = \alpha_i + \beta_i \tilde{x}_{it} + \tilde{e}_{it} \quad \tilde{e}_{it} \sim (0, \sigma_i^2) \tag{1}$$

where α_i denotes the individual fixed-effect, $\tilde{x}_{it}$ is an exogenous variable. The coefficients for the two clusters are denoted B_1 and B_2. That is, $\beta_i = B_1$ if $i \in I_1$ and $\beta_i = B_2$ if $i \in I_2$. Without loss of generality, assume that $B_1 < B_2$. The case of homogeneous parameters obtains if $B_1 = B_2$.

Suppose that we can find a variable q_i which, along with a set of cut-off parameter values Γ^0, will lead to perfect information about I_1 and I_2 in the sense that $i \in I_1$ if $q_i \leq \gamma^0$ for any $\gamma^0 \in \Gamma^0$ and $i \in I_2$ otherwise. We can write

$$\tilde{y}_{it} = \alpha_i + B_1 \tilde{x}_{it} 1(q_i \leq \gamma^0) + B_2 \tilde{x}_{it} 1(q_i > \gamma^0) + \tilde{e}_{it}. \tag{2}$$

We can also consider a threshold representation of the model

$$\begin{aligned} \tilde{y}_{it} &= \alpha_i + B_1 \tilde{x}_{it} + \tilde{e}_{it} \qquad \text{if} \quad q_i \leq \gamma^0 \\ &= \alpha_i + B_2 \tilde{x}_{it} + \tilde{e}_{it} \qquad \text{if} \quad q_i > \gamma^0. \end{aligned}$$

Hansen (1999) considered threshold panel regression models where the sample is split according to whether q_{it} is less than some γ. In his analysis, q_{it} is an observed variable that is often one of the $\tilde{x}_{it}$, and it is time-varying. Unit i can be in one group in period t if $q_{it} \geq \gamma^0$ but is in another group in period $t + 1$ if $q_{it+1} < \gamma^0$. In contrast, our threshold variable q_i is not observed, and

group structure does not change over time. Because of these differences, we call q_i a 'pseudo threshold variable' and γ the 'pseudo threshold parameter' to distinguish it from the usual definitions used in the threshold literature.

If q_i and Γ^0 were both known, estimates of B_1 and B_2 can be obtained using a threshold, or split-sample, regression, once we control for the fixed effects such as by letting $z_{it} = \tilde{z}_{it} - \overline{\tilde{z}}_i$ where $\overline{\tilde{z}}_i = \frac{1}{T}\sum_{t=1}^{T} \tilde{z}_{it}$, and $\tilde{z}_{it}$ can be $\tilde{y}_{it}, \tilde{x}_{it}$, or $\tilde{e}_{it}$. Defining z_{it} as $\tilde{z}_{it} - \tilde{z}_{it-1}$ is also possible. The model becomes

$$y_{it} = B_1 x_{it} 1(i \in I_1) + B_1 x_{it} 1(i \in I_2) + e_{it}, \tag{3}$$

Split sample regressions can then be performed using (y_{it}, x_{it}) as data. Observations with $q_i \leq \gamma^0$ for any $\gamma^0 \in \Gamma^0$ can be pooled to estimate B_1, while observations with $q_i > \gamma^0$ can be pooled to estimate B_2. The problem, however, is that neither q_i nor Γ_0 is observed. We first discuss how to estimate γ assuming q_i is known. We then consider two possible choices of q_i.

We will be concerned with both individual and pooled estimates of β_i. For each i, let $\hat{\beta}_i$ be the least squares estimate of β_i obtained from a regression of y_{it} on x_{it}. Define $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{it}^2$, where $\hat{e}_{it} = \tilde{y}_{it} - \hat{\alpha}_i - \tilde{x}_{it}'\hat{\beta}_i$.

To obtain a pooled estimate of β_i, we use the fixed effect model:

$$\tilde{y}_{it} = \alpha_i + B\tilde{x}_{it} + \tilde{e}_{it}. \tag{4}$$

The pooled least squares estimator is

$$\hat{B}_\omega = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \delta_i y_{it} x_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} \delta_i x_{it}^2} = \sum_{j=1}^{2} \left(\frac{\sum_{i\in I_j} \delta_i \hat{Q}_i}{\sum_{i=1}^{N} \delta_i \hat{Q}_i} \cdot \frac{\sum_{i\in I_1}\sum_{t=1}^{T} \delta_i x_{it} y_{it}}{\sum_{i\in I_1}\sum_{t=1}^{T} \delta_i x_{it}^2} \right). \tag{5}$$

The unweighted fixed effect estimator obtains when $\delta_i = 1$. The weighted fixed-effects estimator obtains when $\delta_i = 1/\hat{\sigma}_i^2$. $\hat{B}_\omega$ was used in Swamy (1970) to test for parameter homogeneity in large T small N panel data models. It has recently been extended by Pesaran and Yamagata (2007) to allow both N and T to be large. We use the following:

Assumption A: 1. e_{it} is cross-sectionally independent and uncorrelated with β_1 and β_2 for all i's and t's. 2. For each i, let $x_{it} = \tilde{x}_{it} - \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}$. (2a) $\hat{Q}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}^2$ is finite and positive. The second moments of $\hat{Q}_i^{-1}$ are finite with $\hat{Q}_i \xrightarrow{p} Q_i > 0$. (2b) $\hat{Q}^j = N_j^{-1}\sum_{i\in I_j}\hat{Q}_i/\sigma_i^2 > 0$ and $\hat{Q}^j \xrightarrow{p} Q^j > 0$ as $(N,T) \to \infty$ jointly; (2c) the pooled matrix $Q = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2$ converges in probability to a positive and finite Q as (N,T) approaches infinity; 3. $\hat{\sigma}_i^2/\sigma_i^2 = 1 + O_p(T^{-1})$ and $\mathrm{E}(1/\sigma^2)$ exists and bounded. 4. For $j = 1, 2$, $N_j/N > 0$ and $N_1/N \to \pi$ with $\pi \in (0,1)$. 5. $T^{-1/2}\sum_{t=1}^{T} x_{it}e_{it} \xrightarrow{d} N(0, \sigma_i^2 Q_i)$ is cross-sectionally independent.

Lemma 1 *Suppose that Assumptions A.1–A.4 hold. Then* $\hat{\beta}_i = \beta_i + O_p(T^{-1/2})$. *Let* $B_\omega = \omega B_1 + (1-\omega)B_2$, *where* $\omega = \mathrm{plim}\sum_{i\in I_1}\hat{Q}_i/\sum_{i=1}^{N}\hat{Q}_i$ *if* $\delta_i = 1$, *and* $\omega = \mathrm{plim}\sum_{i\in I_1}\hat{\sigma}_i^{-2}\hat{Q}_i/\sum_{i=1}^{N}\hat{\sigma}_i^{-2}\hat{Q}_i$ *if* $\delta_i = 1/\hat{\sigma}_i^2$. *Then* $\hat{B}_\omega = B_\omega + O_p((NT)^{-1/2})$.

The regression model (4) is mis-specified if $B_1 \neq B_2$. However, $\hat{B}_\omega$ is $\sqrt{NT}$ consistent for the population mean, B_ω.

3 ESTIMATION OF γ WHEN q_i IS OBSERVED

When q_i is known but Γ^0 is not observed, a $\gamma^0 \in \Gamma^0$ can be estimated as follows. Order the observations by q_i. For a given $\gamma \in \Gamma = [q_{min}, q_{max}]$, let $\hat{B}_1(\gamma)$ and $\hat{B}_2(\gamma)$ denote the least squares estimator of B_1 and B_2 using observations with $q_i \leq \gamma$ and $q_i > \gamma$ respectively. Then

$$\tilde{\gamma} = \arg\min_{\gamma\in[q_{min},q_{max}]} S_{NT}(\gamma),$$

where the weighted sum of squared residuals with weights δ_i is defined as

$$\begin{aligned} S_{NT}(\gamma) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\delta_i\left(y_{it} - \hat{B}_1(\gamma)x_i 1(q_i \leq \gamma) - \hat{B}_2(\gamma)x_i 1(q_i > \gamma)\right)^2 \\ &= \sum_{i|q_i\leq\gamma}\sum_{t=1}^{T}\delta_i(y_{it} - \hat{B}_1(\gamma)x_{it})^2 + \sum_{i|q_i>\gamma}\sum_{t=1}^{T}\delta_i(y_{it} - \hat{B}_2(\gamma)x_{it})^2. \end{aligned}$$

Since q_i can be used to order the data, using q_i to split the unordered sample at some γ is isomorphic to splitting the ordered sample at some observation i^* that has $q_{i^*} = \gamma$. Therefore, even though there are 2^N possible groupings of the data, we only need to consider at most $N-1$ possible values of γ. If the trial value of γ is too low, $\hat{B}_2$ will be calculated with some observations from group 1 and will not be consistent for B_2. Similarly, at too high a value of γ, $\hat{B}_1$ will be calculated with observations from group 2, and hence not consistent for B_1. Minimizing $S_{NT}(\gamma)$ should then yield a $\tilde{\gamma}$ that also minimizes the number of misclassified units.

Theorem 1 *Let $\tilde{\gamma} = \arg\min_{\gamma\in\Gamma} S_{NT}(\gamma)$. Suppose that the data are generated by (1) and that Assumption A holds. Let $N_{jk}(\gamma)$ be the number of units that belong to group j but are classified into group k when the sample is partitioned at γ and let $N_s = \max_{j\neq k} N_{jk}(\tilde{\gamma})$. Then (i) for fixed $B_2 - B_1 = O(1)$, $N_s = O\left(N^{1/2}T^{-1/2}\right)$, and (ii) for $B_2 - B_1 = cT^{-(1/2-\Delta)}$, $0 \leq \Delta < 1/2$, $0 < c < \infty$, $N_s = O\left(N^{1/2}T^{-\Delta}\right)$.*

In fact, we only need to consider where to position γ in relation to the N ordered observations of q_i.There will be a $\tilde{\gamma}$ that minimizes the size of the misclassified set. For fixed $B_2 - B_1 = O(1)$, Theorem 1 implies that the maximum mis-classification rate is $N_s/N = O(N^{-1/2}T^{-1/2})$. If $B_2 - B_1$ is in the $T^{-(1/2-\Delta)}$ neighborhood of zero, the mis-classification rate is $N_s/N = O(N^{-1/2}T^{-\triangle})$. Thus the misclassification rate tends to zero as $N, T \to \infty$ jointly.

Given $\tilde{\gamma}$, the two groups can be estimated as $I_1(\tilde{\gamma}) = \{i|q_i \leq \tilde{\gamma}\}$ and $I_2(\tilde{\gamma}) = \{i|q_i > \tilde{\gamma}\}$. In our analysis, group membership is known once we have an appropriate q_i. We now propose two $\sqrt{T}$ consistent estimators of q_i with these properties. The first one does not account for sampling uncertainty. The

second estimator does, and can be more appropriate when there is substantial dispersion in the error variances.

4 A TWO-STEP PSEUDO THRESHOLD APPROACH

In practice, q_i is not observed. We propose to replace q_i by some $\hat{q}_i$ that has the same information as q_i in the sense that $\hat{q}_i \leq \gamma$ when $q_i \leq \gamma$ as $T \to \infty$. All variables indexed by i are assumed to be ordered once q_i is estimated. Then

$$\hat{\gamma} = \arg \min_{\gamma \in [\hat{q}_{min}, \hat{q}_{max}]} \sum_{i|\hat{q}_i \leq \gamma} \sum_{t=1}^{T} \delta_i (y_{it} - \hat{B}_1(\gamma) x_{it})^2 + \sum_{i|\hat{q}_i > \gamma} \sum_{t=1}^{T} \delta_i (y_{it} - \hat{B}_2(\gamma) x_{it})^2$$

The two groups are then estimated as $\hat{I}_1 = \{i|\hat{q}_i \leq \hat{\gamma}\}$ and $\hat{I}_2 = \{i|\hat{q}_i > \hat{\gamma}\}$. We now propose two estimate of q_i.

A Non-normalized $\hat{q}_i = \hat{\beta}_i$: Consider defining $q_i = \beta_i - B_\omega$. It is easy to see that

$$q_i = \begin{cases} \beta_i - B_1 - (1-\omega)(B_2 - B_1) & \text{for } i \in I_1 \\ \beta_i - B_2 + \omega(B_2 - B_1) & \text{for } i \in I_2. \end{cases}$$

Now $\beta_i - B_1 = 0$ if $i \in I_1$ and $B_2 \neq B_1$ by assumption. Thus, $q_i = -(1-\omega)(B_2 - B_1) < 0$ if $i \in I_1$. On the other hand, $q_i = \omega(B_2 - B_1) > 0$ if $i \in I_2$. The pseudo variable q_i along with $\gamma^0 = 0$ completely summarizes group membership. In fact, any $\gamma^0 \in [0, B_2 - B_1)$ will also identify group membership, not just $\gamma^0 = 0$. Although β_i is not known, $\hat{\beta}_i$ is $\sqrt{T}$ consistent for β_i, which can be estimated using the time series observation on unit i only. Since $\hat{\beta}_i$ is $\sqrt{T}$ consistent for β_i, it follows that $\hat{q}_i = q_i + O_p(T^{-1/2})$ where $\hat{q}_i = \hat{\beta}_i - \hat{B}_\omega$ and $q_i = \beta_i - B_\omega$. Importantly, $\hat{q}_i = \hat{\beta}_i - \hat{B}_\omega$, like q_i, is negative when $i \in I_1$, and is positive when $i \in I_2$. Note also that B_ω is common across i. Thus $q_i = \beta_i$ along with any $\gamma^0 \in \Gamma^0 = [B_1, B_2)$ will also identify group membership. The classification can also be obtained without estimation of B_ω. We now consider an alternative $\hat{q}_i$ that will make explicit use of $\hat{B}_\omega$.

A Normalized $\hat{q}_i$ Letting $\hat{q}_i = \hat{\beta}_i$ is natural, but this ignores the sampling uncertainty of $\hat{\beta}_i$. When there is heterogeneity in error variances, ie. σ_i^2 varies across i, $\hat{\beta}_i$ can provide rather inaccurate information of clustering. We therefore consider a standardized pseudo threshold variable in view of this concern. This variable is defined as

$$\hat{q}_i = \hat{\tau}_i \quad \text{where} \quad \hat{\tau}_i = (\hat{\beta}_i - \hat{B}_\omega)/\sqrt{\hat{\sigma}_i^2 \hat{Q}_i^{-1}},$$

$\hat{Q}_i$ is defined in Assumption A with $\hat{Q}_i \to Q_i > 0$ as $T \to \infty$. By standardizing the deviation between the individual estimate of β_i and an estimate of B_ω, we account for the sampling variability arising from time series estimation of β_i as well as fixed-effect estimation of B_ω.

The ability of this pseudo threshold in identifying the clusters can be seen as follows. Let τ_i be the analog of $\hat{q}_i = \hat{\tau}_i$ when B_ω, σ_i and Q_i are observed. Using the definition of B_ω,

$$\sqrt{T}\tau_i = \begin{cases} \frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\sigma_i Q_i^{-1/2}} - \frac{\sqrt{T}(1-\omega)(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1), & \text{for } i \in I_1, \\ \frac{\sqrt{T}(\hat{\beta}_i - B_2)}{\sigma_i Q_i^{-1/2}} + \frac{\sqrt{T}\omega(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1), & \text{for } i \in I_2. \end{cases} \tag{6}$$

Let $z \sim N(0,1)$ be a standard normal random variable, It can be seen that as $N, T \to \infty$ jointly, for $i \in I_1$, $\sqrt{T}\tau_i = -\sqrt{T}(1-\omega)(B_2 - B_1)/(\sigma_i^2 Q_i^{-1})^{1/2} + z + o_p(1)$; and for $i \in I_2$, $\sqrt{T}\tau_i = \sqrt{T}\omega(B_2 - B_1)/(\sigma_i^2 Q_i^{-1})^{1/2} + z + o_p(1)$. It is obvious that τ_i will eventually identify the clusters in which $B_1 \neq B_2$, $B_2 - B_1 = O(T^{-1/2+\Delta})$, for $0 < \Delta \leq 1/2$ when T is large. The above arguments assume that $\hat{\beta}_i$ is the only quantity in τ_i that is being estimated, which is not the case in practice. But $\hat{\tau}_i$ is consistent for τ_i whenever $\hat{B}_\omega \to B_\omega$ and $\hat{Q}_i \to Q_i$. Lemma A.1 in the Appendix shows that $\hat{\tau}_i = \tau_i + o_p(T^{-1/2})$.

In Theorem 1, we have shown that the classification error rate is $O_p((NT)^{-1/2})$ when q_i is known. That is, if $B_2 - B_1$ is fixed, $P(N_s/N|q_i) = O_p((NT)^{-1/2})$.

But $\hat{q}_i$ only has a convergence rate of $O_p(T^{-1/2})$. A consequence of the two step procedure is that

$$P(N_s/N|\hat{q}_i) = O_p(\max((NT)^{-1/2}, T^{-1/2})) = O_p(T^{-1/2}).$$

The overall correct classification rate is therefore dominated by how precisely we can estimate q_i. Inevitably, when T is small, the classification error can be high. Simulations confirm that the misclassification rate decreases with T.

Test for Parameter Homogeneity: It is also of interest to test the null hypothesis of parameter homogeneity: $H_0^A : \beta_i = B \ \forall i$, which can also be stated as $H_0 : B_1 = B_2$. In standard threshold models, we can test for H_0 based on the sup-Wald type test. See, for example, Andrews and Ploberger, 1994, Hansen, 1996. However, there are three features in our cases that make the SupW test for parameter homogeneity infeasible. First, $\hat{B}_1$ and $\hat{B}_2$ are estimated from two split samples ordered by $\hat{\beta}_i$ and these will be biased if $B_1 = B_2$. Second, $\hat{B}_1$ and $\hat{B}_2$ are correlated when $B_1 = B_2$ when the cross-section independence assumption fails. Third, as $\hat{q}_i^*$ is ordered, Hansen's bootstrap that is valid for iid data does not apply.

Instead, we use the dispersion test proposed by Pesaran and Yamagata (2007) to detect the parameter heterogeneity. The test is defined as

$$PY = \frac{\sqrt{N}(S/N - G)}{\sqrt{2G}}, \tag{7}$$

where G denotes the number of the regressors, $\tilde{\sigma}_i^2$ is a consistent estimator of σ_i^2 based on restricted fixed-effect estimator of B, and $S = \sum_{i=1}^{N} \tilde{\sigma}_i^{-2}(\hat{\beta}_i - \hat{B}_w)'(x_i x_i')(\hat{\beta}_i - \hat{B}_w)'$. This test allows for cross-sectional fixed effects and heteroskedasticity and has a standard normal distribution as N and T go to infinity in the models with strictly exogenous regressors.

5 MULTIPLE REGRESSORS

We have shown that when there is a single regressor, either $\hat{\beta}_i$ or a t-ratio (divided by $\sqrt{T}$) for the significance of $\beta_i = B_\omega$ can be used to split the sample. When there are G regressors, and only a subset of the parameters are heterogeneous across groups, we can still split the sample based on one $\hat{\beta}_{ig}$, $g = 1, \ldots, G$, at a time, where $\hat{\beta}_i = (\hat{\beta}_{i1}, \ldots, \hat{\beta}_{iG})'$ denotes the OLS estimator of the slope parameters for individual i. The hypothesis of homogeneous parameter can be rejected when the parameter homogeneity test is rejected for some g.

Suppose that it is suspected that all the parameters are different between groups. The natural extension to the multiple regressor case is to use a linear combination of the estimated parameters for unit i, $R\hat{q}_i$, where R is a $1 \times G$ vector. Yet, how to define R is not so straightforward.

To see why, consider the case of two regressors, $x_{1,it}$ and $x_{2,it}$. There are $K = 2$ clusters. Let $B_j = (B_{j,1}, B_{j,2})'$ be slope parameters for Group j, $j = 1, 2$ and $\hat{B}_j = (\hat{B}_{j,1}, \hat{B}_{j,2})'$ be a consistent estimator of B_j. Suppose that for $j_1, j_2, = 1, 2$, $j_1 \neq j_2$, we have $B_{j_1,1} > B_{j_2,1}$ and $B_{j_1,2} > B_{j_2,2}$. We will refer to this as Case 1. Since both parameters are strictly larger in one group than in the other, a natural pseudo transition variable is $\hat{\beta}_i^+ = \hat{\beta}_{i1} + \hat{\beta}_{i2}$. When $B_{11} + B_{12} = B_{21} + B_{22}$, however, the sum of the coefficients is no longer a infomative statistic for group membership even though we can still split the sample using $\hat{\beta}_{ig}$ alone, $g = 1$ or 2.

Notice that $B_{11} + B_{12} = B_{21} + B_{22}$ happens when $B_{j_1,1} > B_{j_2,1}$ but $B_{j_1,2} < B_{j_2,2}$. We will refer to this as Case 2. To resolve this problem, consider the transition variable $\hat{\beta}_i^- = \hat{\beta}_{i1} - \hat{\beta}_{i2}$. Notice that if $B_{j_1,2} < B_{j_2,2}$, then $-B_{j_1,2} > -B_{j_2,2}$. Thus, it can again be seen that $\hat{\beta}_i^-$ will separate those $i \in I_1$ from those $i \in I_2$ when T is large.

Evidently, how to split the data depends on the case under consideration. Fortunately, there are notable differences between the two cases. Under Case 1, $\hat{\beta}_{i1} - \hat{B}_{\omega 1}$ and $\hat{\beta}_{i2} - \hat{B}_{\omega 2}$ should have the same sign; and opposite signs under Case 2. This suggests the following. Let n_{++} n_{--}, n_{+-}, and n_{-+} denote the number of individuals whose $sgn(\hat{\beta}_{i1} - \hat{B}_{\omega 1}, \hat{\beta}_{i2} - \hat{B}_{\omega 2})$ are, respectively, $(+,+)$, $(-,-)$, $(+,-)$, and $(-,-)$. Calculate a suitable non-parametric measure of association between pairs of concordant data (with the same sign) and discordant data (with opposite sign), such as the Goodman-Kruskal's (gamma) statistic, defined as

$$\psi = (\psi_1 - \psi_2)/(\psi_1 + \psi_2). \tag{8}$$

where $\psi_1 = n_{++} \cdot n_{--}$, $\psi_2 = n_{+-} \cdot n_{+-}$. Then define $R = (1,1)$ if $\psi > 0$; otherwise, $R = (1,-1)$. Given R, our non-normalized and normalized pseudo threshold variables are

$$\hat{q}_i(\hat{\beta}_i) = R\hat{\beta}_i, \quad \text{and} \quad q_i(\hat{\tau}_i) = R(\hat{\beta}_i - \hat{B}_\omega)(R\widehat{\text{avar}}(\hat{\beta}_i)R')^{-1/2}, \tag{9}$$

where $\widehat{\text{avar}}(\hat{\beta}_i)$ denotes a consistent estimator of the variance of $\sqrt{T}(\hat{\beta}_i - B_\omega)$. Minimization of the weighted or unweighted sum of squared errors gives estimates of γ from which the sample can be grouped.

The analysis can easily be extended to the case of G regressors. Let $\widehat{\beta}_i = (\hat{\beta}_{i1}, \ldots, \hat{\beta}_{iG})'$ denote the least squares estimator for the i^{th} individual and $\widehat{\mathbf{B}}_\omega = (\hat{B}_{\omega 1}, \ldots, \hat{B}_{\omega G})'$ denote the weighted fixed-effects estimator under the restriction of parameter homogeneity. Then, we can create R by the following steps: 1. For $g = 1, \ldots, G$, $i = 1, \ldots, N$, calculate $\hat{q}_{ig}$, which can be $\hat{\beta}_{iq}$ or $\hat{\tau}_{iq}$. 2. Let $W = \arg\max_{g=\{1,\ldots,G\}} ||\hat{q}_g||_2$, where $\hat{q}_g = (\hat{q}_{1g}, \hat{q}_{2g}, \ldots, \hat{q}_{Ng})'$. 3. Calculate the Goodman-Kruskal's statistic, ψ_g, between $(\hat{\beta}_{ig} - \hat{B}_{\omega g}, \hat{\beta}_{iW} - \hat{B}_{\omega W})$, $i = 1, \ldots N$ as defined in (8). 4. Create a $1 \times G$ vector R, where the g^{th}

position is 1 if $\psi_g \geq 0$; and -1, otherwise. 5. Once R is obtained, the pseudo threshold variable for a model with G regressors can be either non-normalized or normalized pseudo threshold variable in (9) .

6 K-MEANS CLUSTERING

As our work is a form of cluster analysis, it is useful to also review related methods. The aim of cluster analysis is to classify a set of data into groups. The simplest way to estimate group membership is plot $\hat{\beta}_i$ and then 'eyeball' to see when $\hat{\beta}_i$ abruptly shifts from one mean to another. Such a graphical approach is often a useful diagnostic, but does not permit formal statistical statements to be made. A more systematic approach is K-means clustering. See, for example, Hartigan (1975), Abraham, Cornillion, Matzner-Lober, and Molinari (2003). It is well known that the K-means method can be sensitive to the initial choice of the centroids, and the method is not guaranteed to find the global minimizer. In spite of these shortcomings, the algorithm is quite popular in applied statistical work. It is therefore worthwhile considering a modification of the K-means method to allow for covariates and fixed effects in panel data models. Consider again the transformed fixed-effect model in (3). Let $\ddot{B}_1$ and $\ddot{B}_2$ be the (fixed effect) estimator of B_1 and B_2 conditional on the assigned group membership. Define

$$SSR_i^1 = \sum_{t=1}^{T}(y_{it} - \ddot{B}_1 x_{it})^2 \quad \text{and} \quad SSR_i^2 = \sum_{t=1}^{T}(y_{it} - \ddot{B}_2 x_{it})^2.$$

The algorithm then consists of the following steps: 1. remove individual effects for each individual. 2. randomly assign individuals into two groups $\{\ddot{I}_1, \ddot{I}_2\}$. 3. calculate fixed effects estimator $(\ddot{B}_1, \ddot{B}_2)$ based on $\{\ddot{I}_1, \ddot{I}_2\}$. 4. repeat (4.a) and (4.b) until no individual is changed from one group to another: (4.a) For each i, Calculate $SSE_i^j = \sum_{t=1}^{T}(y_{it} - \ddot{B}_j x_{it})^2$. (4.b) If $SSE_i^{j_1} \leq SSE_i^{j_2}$, individual i

is re-assigned to group j_1. Then, (4.c) update $\{\ddot{I}_1, \ddot{I}_2\}$ and recalculate the fixed effects estimator $(\ddot{B}_1, \ddot{B}_2)$. 5. re-shuffle individuals unit by unit to reform new grouping $\{\ddot{I}'_1, \ddot{I}'_2\}$ and calculate $(\ddot{B}'_1, \ddot{B}'_2)$. If $\sum_j \sum_{i \in I'_j} SSR_i^{j'} < \sum_j \sum_{i \in I_j} SSR_i$, then repeat (4.a)–(4.c) with $(\ddot{B}'_1, \ddot{B}'_2)$.

Steps 1 to 5 are repeated several times to reduce the effects of the initial group assignment. The algorithm can be easily adapted to allow for heterogeneous errors by defining SSR_i^1 and SSR_i^2 as follows:

$$SSR_i^{1*} = \sum_{t=1}^{T} (y_{it} - \hat{B}_1 x_{it})^2 / \hat{\sigma}_i^2 \quad \text{and} \quad SSR_i^{2*} = \sum_{t=1}^{T} (y_{it} - \hat{B}_2 x_{it})^2 / \hat{\sigma}_i^2.$$

Obviously, $\hat{\sigma}_i^2$ can introduce extra sampling variations. It is always desirable to use the unscaled sum of squared residuals when there is prior information that the variances are homogeneous across units.

Like the pseudo-threshold approach, the K-means method is also non-parametric. But there are some differences between them. The K-means algorithm considers every unit in the sample for a move to a different group. This makes the K-means method computationally costly when N is large. Our pseudo variable is ordered and can be thought of as using more information to move all those units with $\hat{q}_i$ above and below the threshold value simultaneously. The simultaneous move method is fast, but can be inaccurate when the ordering of $\hat{q}_i$ does not agree with q_i, as may be the case when the sample size is small, or when q_i does not provide complete information about the group structure. Also, the K-means method relies only on the pooled estimates $\hat{B}_g$ which is $\sqrt{NT}$ consistent and does not require the individual estimates $\hat{\beta}_i$ which is $\sqrt{T}$ consistent. Thus the K-means will be more precise even when N or T is small. In contrast, the pseudo threshold approach requires both N and T to be large. We can therefore expect a trade-off between precision and

speed in the different methods.

7 SIMULATIONS AND APPLICATIONS

We now use Monte Carlo simulations to examine the finite sample properties of the methods considered. We generate data as follows. For $t = -49, \ldots, 0, \ldots, T$ and $i = 1, \ldots, N$,

$$\begin{aligned} \tilde{y}_{it} &= \alpha_i + B_1 \tilde{x}_{it} 1(i \in I_1^0) + B_2 \tilde{x}_{it} 1(i \in I_2^0) + \tilde{e}_{it} \\ \tilde{x}_{it} &= \alpha_i (1 - \rho_i) + \rho_i \tilde{x}_{i,t-1} + (1 - \rho_i^2)^{1/2} \tilde{v}_{it}, \end{aligned}$$

where $\alpha_i \sim$ i.i.d. $N(1,1)$, $\rho_i \sim$ i.i.d. $U(0.05, 0.95)$, $\tilde{v}_{it} \sim$ i.i.d. $N(0, \sigma_{ix}^2)$ with $\sigma_{ix}^2 \sim$ i.i.d. $\chi^2(1)$ and independent from $\tilde{e}_{it}$, while $\tilde{e}_{it} \sim N(0, \sigma_i^2)$ is i.i.d. over i and t. The first 50 observations are discarded. In the case with homoskedastic errors, we let $\sigma_i^2 = 1$ for all i. In the case with heterogeneous errors, we assume $\sigma_i^2 \sim$ i.i.d. $U(0.5, 2.5)$. We set $B_1 = 0.7$ and $B_2 = 1$ and randomly assign individuals into two groups $\{I_1^0, I_2^0\}$ with size $N_1 = 2N/3$ and $N_2 = N/3$ respectively. We use $M = 1000$ replications, holding $\{I_1^0, I_2^0\}$, α_i, ρ_i, and σ_{ix}^2 fixed in each repetition. This data generating process is modified from Table 1 of Pesaran and Yamagata (2007).

For each (N, T) with $N = (20, 50, 200)$ and $T = (20, 50, 200)$, we keep track of the computation time, the root-mean-squared error of the estimates (RMSE), and the accuracy of classification (CR). For $k = 1, 2$ and $G = 1$, these are defined as

$$RMSE = \left\{ \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{2} \left[\left(\frac{N_k}{N} \sum_{g=1}^{G} (\hat{B}_{kg}^{(m)} - B_{kg})^2 \right) \right] \right\}^{1/2}.$$

$$CR = \frac{1}{M} \sum_{m=1}^{M} CR^{(m)} \text{ where } CR^{(m)} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{2} (i \in \hat{I}_j^{(m)}) \bigcap (i \in I_j^0),$$

where the superscript (m) denotes the estimate from the m^{th} trial. Results are reported in Table 1 (homoskedastic errors) and 2 (heterogeneous errors). In the tables, PS_1 denotes the pseudo threshold approach using $\hat{q}_i = \hat{\beta}_i$, while PS_2 denotes the method using $\hat{q}_i = \hat{\tau}_i$ with the unweighted objective functions. Analogously, PS_1^H and PS_2^H are based on the weighted objective function with $\delta_i = 1/\hat{\sigma}_i^2$. The panel K-means method with and without controlling for heterogeneity in variances are denoted K_1^H and K_1, respectively. To reduce the effects of the initial group assignment, we randomly draw seven sets of initial groups for K_1^H and K_1 in each repetition and take the results from the set with minimum total sum of squared residuals. It should be mentioned that in results unreported, ignoring the individual-specific fixed effects leads to more inaccurate estimates of the clusters as expected.

We first turn to Table 1, which reports results for homoskedastic errors. Overall, all methods perform quite well. The RMSEs tend to decrease as N or T increases. Moreover, the effect of an increase in T on RMSE is larger than an increase in N. It is also interesting that increasing N rarely affects the CR, but increasing T can improve the CR. For example, the CR improves form around 63% to 85% when T increases from 20 to 200.

Because K_1, PS_1 and PS_2 minimize the same objective function, while K_1^H, P_1^H, and $P^H S_2$ have the same objective function, it is not surprising that their finite sample performances are very similar. However, there are some subtle differences between these methods. The computation time used to calculate K_1 is more than the time used in PS_1 and PS_2. Although all three methods have similar classification rates, the K_1 has much smaller RMSE for both parameters when T is very small, even though when $T = 200$ all three methods have similar RMSEs. Comparing PS_1 with PS_2, the PS_2 has a

slightly lower RMSE for all combinations of (N, T). Using a standardized $\hat{q}_i$ is thus useful, especially when T is small. Implementations of the three methods robust to heterogeneous variances yield slightly higher RMSEs as expected because the errors here are homoskedastic. Overall, $\hat{q}_i = \hat{\beta}_i$ is dominated by $\hat{q}_i = \hat{\tau}_i$ and the K-means method.

Table 2 reports the results when the errors are heterogeneous. The K-means approach is again more time consuming. The groups are as accurately classified whether or not heterogeneity in variances is taken into account, and all three methods have similar classification accuracy. The K_2^H has smaller RMSE when T is small, but the PS_2^H is more precise when T is large, and both dominate PS_1^H. When T is very small, accounting for heterogeneity in variances is not recommended because the data is not informative enough to permit estimation of σ_i^2. However, not taking account of heterogeneity in the error variances is undesirable. The RMSEs of all three methods are lower when heterogeneity is taken into account.

The design of the experiment for $G = 2$ is similar to above. In the case with homoskedastic errors, we let $\sigma_i^2 = 2$ for all i. In the case with heterogeneous error variances, we assume $\sigma_i^2 \sim$ i.i.d. $U(1, 3)$. We set $(B_{11}, B_{12}) = (0.7, 1)$ and $(B_{21}, B_{22}) = (1, 1.1)$. The computation time here is roughly 1.5 times longer than what is in Table 1. Suffice it to mention that the K-means method remains more time consuming than the PS methods. However, as seen from Table 3, the classification error rate is higher than when $G = 1$, but the K-means method is slightly more accurate. Its CRs are higher than those of the PS methods by $2 - 6\%$. However, with large N and T, the PS methods seem to have smaller RMSE than the K-means methods.

Overall, the K-means method is more accurate when T is small. However,

for $G = 1$, the computation cost is not matched with a significant improvement in RMSE over the pseudo threshold approach with $\hat{q}_i = \hat{\tau}_i$. For $G = 2$, the K-means method provides somewhat better classification at the cost of higher computation time.

7.1 Empirical Study

A group of countries with a similar steady state and can be characterized by the same linear model is said to form a convergence club. However, as is well documented, the heterogeneity in country growth rates is not easily explained by observable country-specific characteristics. See, for example, Durlauf, Kourtellos, and Tan (2005). We use our method to estimate the clusters.

To motivate the estimation issue when group membership is not known, consider the regression model used in Lee, Pesaran, and Smith (1997):

$$\tilde{y}_{it} = \mu_i + g_i t + \tilde{\epsilon}_{it}, \quad \tilde{\epsilon}_{it} = \gamma_i \tilde{\epsilon}_{it-1} + \eta_t + \tilde{e}_{it}, \tag{10}$$

where $\tilde{y}_{it}$ is the logarithm of real per capita output (2000 as base year), μ_i absorbs all the invariant parameters, $\gamma_i = e^{-\lambda_i}$, λ_i is rate of convergence to steady state per capita output for country i, η_t are common time dummies, and the parameter of interest is g_i. We apply the above model for intermediate countries considered in Mankiw, Romer, and Weil (1992). Data is taken from the PWT v6.2 for the sample from 1965 to 2000. Due to consolidation and limitation of data we remove Germany, Bangladesh, Bolivia, Haiti, and Myanmar. Eventually, there are 70 countries in our data set. Figure 1(a) shows the same scatter plot for countries in different geographical regions. It is obvious that the growth rates from data set vary by geography, the membership of OECD (mainly in the category of Western Europe and North America), and GDP per capita. But many countries with similar growth rates and yet have

different observed characteristics. A pattern for growth rates cannot be easily established. Figures 1(b)–(c) plot the distribution of the non-standardized and standardized $\hat{g}_i$ estimated by kernel smoothing.

We use our methods to split countries into two groups. Based on (10), Table 4 shows that five out of six methods suggest a grouping of between 28 to 30 countries in the low growth group with a growth rate around 0.5%, and between 40 and 42 countries in the high growth group with a growth rate of about 2.7%. From Table 5, the methods differ primarily over the classification of 11 countries around the threshold value. Clearly, subgroups may be formed within the two blocks of countries. Of note is that some developed countries/OECD like New Zealand and Switzerland are classified in the low growth group. Also, it should be mentioned that in the results not reported here, all methods except the K-means method that allows for heterogeneous error variance give similar clusters when the model is re-estimated using first differences to remove the fixed effect. It would be useful to further understand what are the trade-offs amongst observed characteristics that make rather dissimilar countries to belong to the same growth club.

8 CONCLUSION

We use time series estimates of the coefficients for each unit to form 'pseudo threshold variables'. These are then used to partition the panel into groups. Our model based method is shown to consistently estimate the true number of groups identified by distinct coefficients on the covariates. The analysis is valid for whether or not the regression error variance is equal across units.

References

Abraham, C., P. Cornillion, E. Matzner-Lober, and N. Molinari (2003): "Unsupervised Curve Clustering Using B-Splines," *Scandinavian Journal of Statistics*, 30, 581–595.

Alvarez, J., M. Browning, and M. Ejrnæs (2002): "Modelling Income Processes with Lots of Heterogeneity," 10th International Conferences on Panel Data, Berlin D2-3.

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

Chiou, J.-M., and P.-L. Li (2007): "Functional Clustering and Identifying Substructure of Longitudinal Data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 679–699.

Durlauf, S. N., A. Kourtellos, and C. M. Tan (2005): "Empirics of Growth and Development," Discussion Papers Series: 0520, Dept. of Economics, Tufts University.

Fraley, C., and A. E. Raftery (2002): "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association*, 97, 611–631.

Goldfeld, S., and R. Quandt (1973): "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement*, 2, 475–485.

Hall, P., H. Muller, and J. Wang (2006): "Properties of Principal Component Methods for Functional and Longitudinal Data Analysis," *The Annals of Statistics*, 34, 1493–1517.

Hansen, B. E. (1999): "Threshold Effects in Non-dynamic Panels: Estimation, Testing, and Inference," *Journal of Econometrics*, 93, 345–368.

Hartigan, J. A. (1975): *Clustering Algorithms*. Wiley.

Hsiao, C., and M. H. Pesaran (2004): "Random Coefficient Panel Data Models," *CESifo Working Paper:* 1223.

Hsiao, C., and A. K. Tahmiscioglu (1997): "A Panel Analysis of Liquidity Constraints and Firm Investment," *Journal of the American Statistical Association*, 92, 455–465.

Juárez, M. A., and M. F. J. Steel (2007): "Model-based Clustering of non-Gaussian Panel Data Based on Skew-t Distributions," *Working Paper*.

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, 407–437.

Pesaran, M. H., and T. Yamagata (2007): "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics*, forthcoming.

Sun, Y. (2005): "Estimation and Inference in Panel Structure Models," *Working Paper:* 2005-11, Dept. of Economics, University of California, San Diego.

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

Tyssedal, J., and D. Tjøstheim (1988): "An Autoregressive Model with Suddenly Changing Parameters and an Application to Stock-Market Prices," *Journal of the American Statistical Association*, 93, 1188–1202.

APPENDIX: PROOFS

Proof of Lemma 1: We want to show that $\sqrt{NT}\left[\hat{B}_\omega - (\omega B_1 + (1-\omega)B_2)\right] = O_p(1)$. Let $\hat{B}_j = \sum_{i\in I_j}\sum_{t=1}^T \delta_i x_{it} y_{it} / \sum_{i\in I_j}\sum_{t=1}^T \delta_i x_{it}^2$ with $\delta_i = 1$ or $1/\hat{\sigma}_i^2$. By the definition of $\hat{B}_\omega$ in (5) and that $\omega = \text{plim}\sum_{i\in I_1}\hat{\sigma}_i^{-2}\hat{Q}_i / \sum_{i=1}^N \hat{\sigma}_i^{-2}\hat{Q}_i$, we have

$$\begin{aligned}
&\sqrt{NT}\left[\hat{B}_\omega - (\omega B_1 + (1-\omega)B_2)\right] \\
&= \sqrt{\frac{N}{N_1}}\left[\omega\sqrt{N_1 T}\left(\hat{B}_1 - B_1\right)\right] + \sqrt{\frac{N}{N_2}}\left[(1-\omega)\sqrt{N_2 T}\left(\hat{B}_2 - B_2\right)\right] + o_p(1)
\end{aligned}$$

To prove Lemma 1 with $\delta_i = \hat{\sigma}_i^{-2}$, it suffices to show that $\sqrt{N_j T}(\hat{B}_j - B_j) = O_p(1)$. By Assumptions A.1 and A.2, we have $N_j^{-1}\sum_{i\in I_j}\sigma_i^{-2}\hat{Q}_i = \hat{Q}^j = O_p(1)$,

$$\text{and} \quad \text{Avar}\left(\sqrt{NT}\cdot\sum_{i\in I_j}\sigma_i^{-2}\sum_{t=1}^T x_{it}e_{it}\right) = \lim_{N\to\infty}\left(N^{-1}\sum_{i\in I_j}\sigma_i^{-2}Q_i\right) = Q^j = O(1).$$

Under Assumptions A.1–A.3, we also have

$$\begin{aligned}
(N_j T)^{-1/2}\cdot\sum_{i\in I_j}\hat{\sigma}_i^{-2}\sum_{t=1}^T x_{it}e_{it} &= (N_j T)^{-1/2}\cdot\sum_{i\in I_j}\sigma_i^{-2}\sum_{t=1}^T x_{it}e_{it} + O_p\left(\frac{\sqrt{N_j}}{T}\right) \\
N_j^{-1}\sum_{i\in I_j}\hat{\sigma}_i^{-2}\hat{Q}_i &= N_j^{-1}\sum_{i\in I_j}\sigma_i^{-2}\hat{Q}_i + O_p\left(\frac{1}{T}\right).
\end{aligned}$$

Therefore, if $\sqrt{N}/T \to 0$ and Assumptions A.1–A.4 hold,

$$(N_j T)^{1/2}\left(\hat{B}_j - B_j\right) = \frac{(N_j T)^{-1/2}\sum_{i\in I_j}\sum_{t=1}^T \sigma_i^{-2}x_{it}e_{it}}{N_j^{-1}\sum_{i=1}^N \sigma_i^{-2}\hat{Q}_i} + O_p(\sqrt{N}/T) = O_p(1).$$

The case with $\delta = 1$ is similar. □

Proof of Theorem 1: Let (I_1, I_2) denote group membership other than the true group membership (I_1^0, I_2^0). We will consider the general case where $B_2 - B_1 = cT^{-\alpha}$, c does not depend on T, $\alpha = 1/2 - \Delta$, and $0 < \Delta \leq 1/2$.

For $j, k = 1, 2$, let N_{kj} be the number of individuals assigned to be in group j by (I_1, I_2) when they truly belong to group k and let $\hat{B}_j$ denote the estimator of slope parameter for $i \in I_j$, and $\hat{B}_j^0$ for $i \in I_j^0$ $j = 1, 2$. Without loss of generality, we assume $N_{11} > N_{21}$ and let $N_s = \max\{N_{12}, N_{21}\}$.

We first consider the case in which $N_{12} > 0$, $N_{21} > 0$. Let $z_{it} = \tilde{z}_{it} - \overline{\tilde{z}}_i$, where $\tilde{z}_{it}$ can be $\tilde{y}_{it}, \tilde{x}_{it}, \tilde{e}_{it}, \widehat{\tilde{e}}_{it}$, and $\overline{\tilde{z}}_i = \frac{1}{T}\sum_{t-1}^{T} \tilde{z}_{it}$. Then for $j, k = 1, 2$,

$$y_{it} = B_k x_{it} + e_{it} = \hat{B}_j x_{it} + \hat{e}_{it}^{kj}.$$

where $\hat{e}_{it}^{kj} = e_{it} + (B_k - \hat{B}_j)x_{it}$. We have

$$\begin{aligned}
(\hat{e}_{it}^{kj})^2 &= e_{it}^2 + (B_k - \hat{B}_j)^2 x_{it} + 2(B_k - \hat{B}_j)x_{it}, \\
S_{NT}(I_1, I_2) &= \sum_{j,k=1,2} \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^{T} (\hat{e}_{it}^{kj})^2 \\
&= \sum_{j,k=1,2} \sum_{i \in I_k^0 \cap I_j} \left[\sum_{t=1}^{T} e_{it}^2 + (B_k - \hat{B}_j)^2 \sum_{t=1}^{T} x_{it}^2 + 2(B_k - \hat{B}_j) \sum_{t=1}^{T} x_{it} e_{it} \right].
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{T^{-1+2\alpha}}{N_s} \left[S_{NT}(I_1, I_2) - S_{NT}(I_1^0, I_2^0) \right] \\
&= \sum_{j,k=1,2} \left[\frac{T^{-1+2\alpha}}{N_s} \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^{T} (\hat{e}_{it}^{kj})^2 - (\hat{e}_{it}(I_1^0, I_2^0))^2 \right] = \sum_{\substack{j,k=1,2 \\ j \neq k}} [\phi_{1,kj} + \phi_{2,kj}]
\end{aligned}$$

where $\phi_{1,kj}$ and $\phi_{2,kj}$ are

$$\begin{aligned}
\phi_{1,kj} &= \frac{T^{2\alpha-1}}{N_s} \left\{ (\hat{B}_j^0 - \hat{B}_j)^2 \sum_{i \in I_j^0 \cap I_j} \sum_{t=1}^{T} x_{it}^2 + (\hat{B}_k^0 - \hat{B}_j)^2 \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^{T} x_{it}^2 \right\}, \\
\phi_{2,kj} &= \frac{2T^{2\alpha-1}}{N_s} \left\{ (\hat{B}_j^0 - \hat{B}_j) \sum_{i \in I_j^0 \cap I_j} \sum_{t=1}^{T} x_{it} e_{it} + (\hat{B}_k^0 - \hat{B}_j) \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^{T} x_{it} e_{it} \right\}.
\end{aligned}$$

We want to show that for any $(I_1, I_2) \neq (I_1^0, I_2^0)$, if $N^{1/2}T^{-1/2+\alpha}/N_s = o(1)$,

$$\text{Prob}(T^{-1+2\alpha}/N_s \left[S_{NT}(I_1, I_2) - S_{NT}(I_1^0, I_2^0)\right] > 0) \xrightarrow{p} 1, \tag{11}$$

or equivalently $\text{Prob}\left(\sum_{j,k=1,2,j\neq k} \phi_{1,kj} + \phi_{2,kj} > 0 \right) \xrightarrow{p} 1$. We will first show that given any (I_1, I_2) such that for $j, k = 1, 2$, $\sum_{j\neq k} \phi_{2,kj} = o_p(1)$ and $\sum_{j\neq k} \phi_{1,kj} = M^* + o_p(1)$, where M^* is a strictly positive number.

Define $C_{kj} = \sum_{i\in I_k^0 \cap I_j} \sum_{t=1}^T x_{it}y_{it}$, $D_{kj} = \sum_{i\in I_k^0 \cap I_j} \sum_{t=1}^T x_{it}^2$, $\hat{B}_{jj}^* = C_{jj}/D_{jj}$, $\hat{B}_{kj}^* = C_{kj}/D_{kj}$, and $D_j = D_{jj} + D_{kj}$. Note that for $j, k = 1, 2$, $j \neq k$,

$$\begin{aligned}
\hat{B}_j &= \frac{C_{jj} + C_{kj}}{D_j} = \frac{C_{jj}}{D_{jj}}\left(1 - \frac{D_{kj}}{D_j}\right) + \frac{C_{kj}}{D_{kj}}\frac{D_{kj}}{D_j} = \hat{B}_{jj}^* + (\hat{B}_{kj}^* - \hat{B}_{jj}^*)\frac{D_{kj}}{D_j} \\
&= B_j + \left[(B_k - B_j) + (\hat{B}_{kj}^* - B_k) - (\hat{B}_{jj}^* - B_j)\right]\frac{D_{kj}}{D_j} + (\hat{B}_{jj}^* - B_j).
\end{aligned}$$

Notice that $\phi_{2,kj} = \phi_{21,kj} + \phi_{22,kj}$, where

$$\phi_{21,kj} = T^\alpha(B_k - B_j)\left[\frac{N_{jj}}{N_s}\frac{D_{kj}T^{\alpha-1}}{D_j N_{jj}} \sum_{i\in I_j^0\cap I_j}\sum_{t=1}^T x_{it}e_{it} - \frac{N_{kj}}{N_s}\frac{D_{jj}T^{\alpha-1}}{D_j N_{kj}} \sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}\right],$$

$$\begin{aligned}
\phi_{22,kj} &= \left\{\left[(\hat{B}_{kj}^* - B_k) - (\hat{B}_{jj}^* - B_j)\right]\frac{D_{kj}}{D_j} + (\hat{B}_{jj}^* - B_j)\right\} \\
&\quad \times \left(\frac{N_{jj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{jj}T} + \frac{N_{kj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{kj}T}\right) \\
+(\hat{B}_j^0 - B_j)&\frac{N_{jj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{jj}T} + (\hat{B}_k^0 - B_k)\frac{N_{kj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{kj}T}.
\end{aligned}$$

To show $\phi_{2,kj} = o_p(1)$ is equivalent to showing both $\phi_{21,kj}$ and $\phi_{22,kj}$ are $o_p(1)$. By assumption, $E(x_{it}e_{it}) = 0$ for all i and t and

$$(N_{kj}T)^{-1}\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it} = O_p((N_{kj}T)^{-1/2}).$$

Also, $T^\alpha(B_2 - B_1) = c$ for $0 \leq \alpha < 1/2$, and for $j, k = 1, 2$, $j \neq k$, $N_j = N_{kj} + N_{jj}$, $N_{kj}D_{kj}(N_sD_j)^{-1} = O_p(1)$, and $N_{jj}D_{kj}(N_sD_j)^{-1} = O_p(1)$. Thus,

$$\phi_{21,kj} = O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) = o_p(1).$$

To show that $\phi_{22,kj} = o_p(1)$, note first that $\hat{B}^*_{kj} - B_k = O_p((N_{kj}T)^{-1/2})$, $\hat{B}^*_{jj} - B_j = O_p((N_{jj}T)^{-1/2})$, $\hat{B}^0_j - B_j = O_p(((N_{jk}+N_{jj})T)^{-1/2})$, $\hat{B}^0_k - B_j = O_p(((N_{kj}+N_{kk})T)^{-1/2})$, and $T^\alpha(B_k - B_j) = c$. After some tedious computation, it can be shown that $\phi_{22,kj} = o_p(1)$; and, therefore, $\sum_{j\neq k}\phi_{2,kj} = \sum_{j\neq k}\phi_{21,kj} + \phi_{22,kj} = o_p(1)$.

Next, consider $\sum_{j\neq k}\phi_{1,kj}$. Notice that $\phi_{1,kj} = \phi_{11,kj} + \phi_{12,kj}$, where

$$\phi_{11,kj} = \frac{D_{jj}T^{2\alpha}}{N_s T}\left[(B_k - B_j)\frac{D_{kj}}{D_j}\right]^2 + \frac{D_{kj}T^{2\alpha}}{N_s T}\left[(B_k - B_j)\frac{D_{jj}}{D_j}\right]^2,$$

$$\begin{aligned}
&\phi_{12,kj} = \phi_{1,kj} - \phi_{11,kj} = \phi^A_{12,kj} + \phi^B_{12,kj} + \phi^C_{12,kj},\\
&\phi^A_{12,kj} = \left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j)\right]\left[2(B_k - B_j) + \hat{B}^*_{jj} - B_j + \hat{B}^0_j - B_j\right]\left(\tfrac{T^{2\alpha}}{N_s}\tfrac{D_{jj}}{T}\right)\\
&\phi^B_{12,kj} = \left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j)\right]\left[2(B_k - B_j) + \hat{B}^*_{jj} - B_j + \hat{B}^0_j - B_j\right]\left(\frac{T^{2\alpha}}{N_s}\frac{D_{kj}}{T}\right)\\
&\phi^C_{12,kj} = \left[\left((\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j)\right)\tfrac{D_{kj}}{D_j}\right]\left(\tfrac{T^{2\alpha}}{N_s}\tfrac{D_j}{T}\right)\\
&\qquad\times\left[2(B_k - B_j) + (\hat{B}^*_{jj} - B_j) + \hat{B}^0_j - B_j + \left(\hat{B}^*_{kj} - B_k - \hat{B}^*_{jj} + B_j\right)\tfrac{D_{kj}}{D_j}\right].
\end{aligned}$$

Under the assumption that $(N^{1/2}T^{-1/2+\alpha})/N_s = o(1)$, we have

$$\begin{aligned}
&\tfrac{N_{jj}T^{2\alpha}}{N_s}(B_k - B_j)\left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j)\right]\left(\tfrac{D_{jj}}{TN_{jj}}\right) = O_p\left(\tfrac{N^{1/2}N_s^{-1}}{T^{1/2-\alpha}}\right) = o_p(1).\\
&\left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j)\right]^2\tfrac{T^{2\alpha}D_{jj}}{N_{jj}T}\tfrac{N_{jj}}{N_s} = \tfrac{1}{N_sT^{1/2-\alpha}}\cdot O_p\left(\tfrac{N^{1/2}}{T^{1/2-\alpha}}\right) = o_p(1),
\end{aligned}$$

Thus, $\phi^A_{12,kj} = o_p(1)$. Analogously, $\phi^B_{12,kj} = o_p(1)$. Similarly, because $D_{jj}/D_j = O_p(1)$ and $D_{kj}/D_j = O_p(1)$, it follows that $\phi^C_{12,kj} = o_p(1)$. Together with the results above, we obtain $\phi_{12,kj} = o_p(1)$. Also,

$$\textstyle\sum_{\substack{j,k=1,2,\\ j\neq k}}\phi_{11,kj} = c^2\left[\frac{D_{11}D_{21}}{D_1} + \frac{D_{22}D_{12}}{D_2}\right]\frac{1}{N_sT} + o_p(1) = M^* + o_p(1).$$

where M^* is a strictly positive number. Therefore,

$$\sum_{\substack{j,k=1,2,\\ j\neq k}}\phi_{1,kj} = \sum_{\substack{j,k=1,2,\\ j\neq k}}(\phi_{11,kj} + \phi_{12,kj})M^* + o_p(1),$$

Thus, for any $(I_1, I_2) \neq (I_1^0, I_2^0)$, if $(N^{1/2}T^{-1/2+\alpha})/N_s = o(1)$, $\text{Prob}(\sum_{j\neq k} \phi_{1,kj} + \phi_{2,kj} > 0) \xrightarrow{p} 1$ and thus (11) holds. The result also holds when $N_{12} > N_{21} = 0$ or $N_{21} > N_{12} = 0$.

Recall that $\triangle = 1/2 - \alpha$. By our method, the number of misspecified individuals, N_s, asymptotically is bounded by $N^{1/2}T^{-1/2+\alpha} = N^{1/2}T^{-\triangle}$. Particularly, when $\alpha = 0$, we have $\triangle = 1/2$ and $B_2 - B_1 = c > 0$. In this case, the number of misspecified individuals is asymptotically bounded by $N^{1/2}T^{-1/2}$ and the rate of misclassification, N_s/N, diminishes at rate $N^{-1/2}T^{-1/2}$. □

Lemma A.1: Suppose that the data are generated by (1) and Assumption A holds. Then,

$$\sqrt{T}\hat{\tau}_i = \sqrt{T}\tau_i + o_p(1).$$

Proof of Lemma A.1: By Lemma 1, for $i \in I_1$, $\sqrt{T}\hat{\tau}_i$ can expressed as

$$\sqrt{T}\hat{\tau}_i = \frac{\sqrt{T}(\hat{\beta}_i - \hat{B}_\omega)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} = \frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} - (1-\omega)\frac{\sqrt{T}(B_2 - B_1)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} + O_p(N^{-1/2})$$

Similarly, for $i \in I_2$, we have

$$\sqrt{T}\hat{\tau}_i = \frac{\sqrt{T}(\hat{\beta}_i - B_2)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} + \omega\frac{\sqrt{T}(B_2 - B_1)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} + o_p(1)$$

By Assumptions A.2 and A.3, $\hat{\sigma}^2 \hat{Q}_i{}^{-1} \xrightarrow{p} \sigma^2 Q_i^{-1}$ and the result follows.

□.

Table 1: Homoskedastic Errors when $G = 1$

		RMSE			CR (%)			Time (sec/rep)		
	N\T	20	50	200	20	50	200	20	50	200
PS_1	20	0.436	0.119	0.027	64.38	73.67	86.55	0.005	0.006	0.011
	50	0.246	0.112	0.029	62.76	70.42	83.39	0.017	0.020	0.049
	200	0.080	0.030	0.009	66.29	74.69	86.00	0.099	0.187	0.512
PS_2	20	0.431	0.113	0.027	64.96	74.36	86.85	0.005	0.007	0.010
	50	0.221	0.100	0.030	62.79	70.33	83.75	0.016	0.019	0.047
	200	0.079	0.031	0.009	66.42	74.99	86.41	0.100	0.185	0.495
K_1	20	0.188	0.081	0.027	65.46	73.97	86.54	0.051	0.059	0.101
	50	0.189	0.092	0.031	61.86	69.60	83.19	0.069	0.090	0.194
	200	0.080	0.030	0.009	66.26	74.66	86.00	0.115	0.222	0.574
PS_1^H	20	0.497	0.104	0.027	64.06	73.69	86.53	0.005	0.007	0.012
	50	0.349	0.108	0.029	62.71	70.41	83.37	0.016	0.020	0.049
	200	0.081	0.030	0.009	66.31	74.69	86.00	0.099	0.186	0.498
PS_2^H	20	0.536	0.101	0.027	64.45	74.28	86.86	0.006	0.008	0.013
	50	0.331	0.101	0.031	62.64	70.28	83.73	0.018	0.021	0.052
	200	0.080	0.031	0.009	66.43	75.02	86.43	0.101	0.192	0.504
K_1^H	20	0.281	0.086	0.027	65.23	73.86	86.53	0.053	0.061	0.103
	50	0.197	0.094	0.030	61.81	69.56	83.25	0.074	0.096	0.192
	200	0.081	0.031	0.009	66.22	74.64	86.00	0.128	0.241	0.604

Table 2: Heteroskedastic Errors when $G = 1$

		RMSE			CR (%)			Time (sec/rep)		
	N\T	20	50	200	20	50	200	20	50	200
PS_1	20	0.534	0.149	0.031	62.47	71.10	84.51	0.005	0.007	0.009
	50	0.378	0.153	0.040	61.88	68.47	81.43	0.016	0.017	0.038
	200	0.106	0.042	0.012	64.60	72.17	83.52	0.088	0.171	0.507
PS_2	20	0.519	0.142	0.031	63.13	71.83	84.90	0.005	0.006	0.009
	50	0.317	0.139	0.042	61.43	68.17	81.39	0.014	0.017	0.036
	200	0.103	0.043	0.012	64.68	72.45	84.17	0.087	0.170	0.490
K_1	20	0.265	0.105	0.031	63.46	71.71	84.51	0.039	0.047	0.063
	50	0.240	0.118	0.042	60.94	67.76	81.11	0.057	0.066	0.138
	200	0.104	0.043	0.012	64.54	72.12	83.52	0.113	0.202	0.604
PS_1^H	20	0.602	0.126	0.029	62.94	71.64	84.50	0.005	0.007	0.009
	50	0.432	0.142	0.038	61.00	68.01	81.38	0.015	0.017	0.037
	200	0.090	0.035	0.010	64.76	72.23	83.53	0.089	0.171	0.493
PS_2^H	20	0.600	0.123	0.029	63.33	72.12	84.86	0.006	0.008	0.010
	50	0.331	0.127	0.041	60.55	67.81	81.37	0.016	0.019	0.040
	200	0.088	0.036	0.010	64.99	72.79	84.25	0.090	0.174	0.499
K_1^H	20	0.275	0.095	0.029	63.76	71.99	84.47	0.041	0.048	0.065
	50	0.225	0.115	0.042	60.12	66.93	80.90	0.059	0.070	0.153
	200	0.089	0.035	0.010	64.57	72.17	83.53	0.124	0.215	0.619

Table 3: RMSE's and CR's when $G = 2$

Homoskedastic Errors										
		RMSE			CR (%)			Time (sec/rep)		
	N\T	20	50	200	20	50	200	20	50	200
PS_1	20	2.280	1.512	0.829	56.06	60.57	70.13	0.008	0.009	0.013
	50	1.324	0.291	0.067	55.92	60.45	72.34	0.023	0.028	0.053
	200	0.142	0.079	0.031	57.38	62.80	74.80	0.136	0.196	0.630
PS_2	20	1.769	1.448	0.424	56.13	60.92	72.74	0.009	0.009	0.014
	50	0.477	0.299	0.063	56.62	61.76	75.33	0.023	0.027	0.051
	200	0.165	0.091	0.034	57.28	63.08	76.32	0.132	0.193	0.615
K_1	20	0.490	0.245	0.076	58.77	64.67	80.48	0.067	0.080	0.124
	50	0.301	0.160	0.045	59.14	65.92	81.08	0.083	0.101	0.204
	200	0.194	0.084	0.021	60.65	68.50	82.03	0.145	0.245	0.641
PS_1^H	20	2.298	1.435	0.787	55.79	60.54	70.33	0.009	0.009	0.014
	50	1.342	0.571	0.067	55.67	60.26	72.30	0.023	0.027	0.053
	200	0.148	0.080	0.031	57.35	62.77	74.81	0.133	0.199	0.626
PS_2^H	20	1.884	1.351	0.388	55.73	60.92	72.80	0.010	0.011	0.016
	50	0.683	0.328	0.064	56.46	61.23	75.36	0.025	0.028	0.054
	200	0.180	0.091	0.034	57.17	63.11	76.34	0.137	0.201	0.634
K_1^H	20	0.571	0.260	0.076	58.22	64.67	80.43	0.070	0.083	0.127
	50	0.323	0.162	0.045	59.18	66.03	81.14	0.091	0.110	0.215
	200	0.200	0.086	0.021	60.50	68.33	82.02	0.168	0.266	0.671
Heteroskedastic Errors										
		RMSE			CR (%)			Time (sec/rep)		
	N\T	20	50	200	20	50	200	20	50	200
PS_1	20	2.852	1.985	0.965	54.93	57.74	64.67	0.008	0.010	0.014
	50	1.570	0.853	0.107	56.01	59.89	70.27	0.023	0.027	0.063
	200	0.163	0.092	0.044	57.05	62.37	74.78	0.126	0.229	0.816
PS_2	20	2.092	1.486	0.455	55.35	58.70	70.07	0.008	0.009	0.013
	50	0.561	0.288	0.074	56.17	60.44	72.55	0.025	0.028	0.062
	200	0.179	0.100	0.041	57.22	62.90	76.53	0.125	0.227	0.833
K_1	20	0.378	0.212	0.072	58.64	64.63	78.91	0.048	0.064	0.095
	50	0.314	0.167	0.052	58.83	65.51	80.05	0.087	0.160	0.359
	200	0.201	0.087	0.022	61.11	69.66	83.16	0.171	0.554	2.011
PS_1^H	20	2.712	1.795	0.834	54.99	58.59	66.10	0.008	0.010	0.014
	50	1.597	0.871	0.140	55.60	59.45	69.85	0.024	0.028	0.064
	200	0.163	0.093	0.046	56.90	62.16	74.71	0.127	0.230	0.831
PS_2^H	20	2.286	1.375	0.344	55.72	60.38	72.08	0.010	0.011	0.017
	50	0.594	0.363	0.074	55.82	60.41	73.10	0.026	0.028	0.069
	200	0.186	0.099	0.041	57.15	62.85	76.62	0.128	0.236	0.840
K_1^H	20	0.378	0.194	0.069	58.38	64.66	79.47	0.051	0.066	0.100
	50	0.313	0.163	0.050	58.72	65.61	80.13	0.093	0.165	0.368
	200	0.195	0.081	0.020	61.07	69.64	83.16	0.199	0.582	2.098

Table 4: Estimation of g_j, $j = 1, 2$

Cluster based on level equation (10)							
Whole Sample		PY-statistic = 2.940*					
		PS_1	PS_2	PS_1^H	PS_2^H	K_1	K_1^H
	N_1	30	41	30	28	30	30
Low	$\hat{g}_1(\%)$	0.371	0.411	0.614	0.611	0.371	0.614
	PY_1	1.481	2.523*	1.308	1.526	1.481	1.308
Threshold	$\hat{\gamma}$	0.015	-0.001	0.015	-4.492	—	—
	N_2	40	29	40	42	40	40
High	$\hat{g}_2(\%)$	2.720	2.749	2.507	2.501	2.720	2.507
	PY_2	1.823	1.721	1.913	2.051*	1.823	1.913
Cluster based on first difference, $\overline{\Delta y_{it}}$							
Whole Sample		PY-statistic =4.649*					
		PS_1	PS_2	PS_1^H	PS_2^H	K_1	K_1^H
	N_1	38	43	38	37	38	45
Low	$\hat{g}_1(\%)$	0.929	0.958	1.264	1.259	0.929	1.130
	PY_1	-0.012	0.978	-0.077	0.002	-0.012	0.960
Threshold	$\hat{\gamma}$	0.020	0.002	0.020	-0.449	—	—
	N_2	32	27	32	33	32	25
High	$\hat{g}_2(\%)$	3.120	3.143	2.781	2.777	3.110	3.359
	PY_2	2.040*	1.853	1.793	1.669	2.040*	1.841

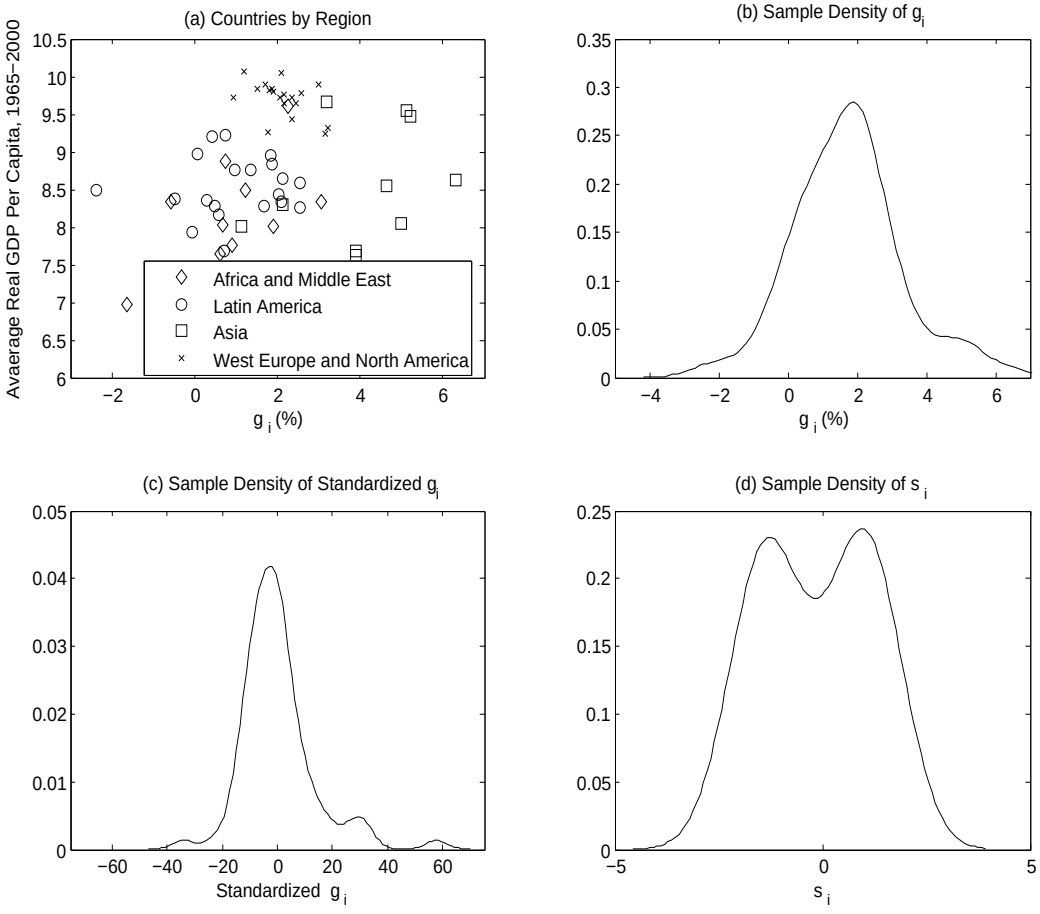
Note: (A) We apply ordinal least squares method to estimate (10). The detailed model please see Lee, Pesaran, and Smith (1997). (B) PY_1 and PY_2 are the PY statistics for the estimated low-growth and high-growth clubs, respectively. * indicates that the null hypothesis of parameter homogeneity is rejected at 5% significance level. (C) Let $\hat{g}_w$ denote the pooled weighted fixed effect estimator of g_0 defined in (5), and let $\tilde{\sigma}(\hat{g}_i)$ and $\hat{\sigma}(\hat{g}_i)$, respectively, denote the restricted and unrestricted long-run variance of $\hat{g}_i$ by the method of Andrews (1991) with Quadratic Spectral kernel. In this case, the PY statistic can be expressed as $PY = \sqrt{N}(\sum_i^N s_i^2/N - 1)/\sqrt{2}$, where $s_i = \sqrt{T}(\hat{g}_i - \hat{g}_w)/\tilde{\sigma}(\hat{g}_i)$, Similarly, for PS_2 and PS_2^H the standardized $\hat{g}_i$ is $\sqrt{T}(\hat{g}_i - \hat{g}_w)/\hat{\sigma}(\hat{g}_i)$.

Table 5: Estimation of Clusters Based on (10)

Group	Country
Low-Growth Club	Algeria, Argentina, Bolivia, Cameroon**, Costa Rica, Cote d'Ivoire, El Salvador, Ethiopia**, Guatemala, Honduras, Jamaica, Jordan, Kenya, Madagascar, Malawi, Mexico, New Zealand, Nicaragua, Nigeria, Peru, Philippines, Senegal, South Africa, Sweden, Switzerland, Tanzania, Trinidad-Tobago, Venezuela, Zambia, Zimbabwe
High-Growth Club	Australia*, Austria, Belgium, Brazil, Canada*, Chile*, Colombia, Denmark*, Dominican Rep, Ecuador*, Finland, France, Greece*, Hong Kong, India, Indonesia, Ireland, Israel, Italy, Japan, Rep of Korea, Malaysia, Mali*, Morocco*, Netherlands*, Norway, Pakistan, Panama, Paraguay, Portugal, Singapore, Spain, Sri Lanka, Syria*, Thailand, Tunisia Turkey, United Kingdom, Uruguay*, USA

Note: (A) Estimated results are based on PS_1, PS_1^H, PS_2^H, K_1, and K_1^H.
(B) * denotes the countries which are assigned to Low-Growth club by PS_2 but to High-Growth club by the other methods.
(C) ** denotes the countries which are assigned to High-Growth club by PS_1^H but to Low-Growth club by the other methods.

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



For the detailed definitions of standardized $\hat{g}_i$ and s_i, please see Table 4.