ESTIMATION OF PANEL DATA MODELS WITH PARAMETER HETEROGENEITY WHEN GROUP MEMBERSHIP IS UNKNOWN

Chang-Ching Lin* Serena Ng†

November 10, 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.

JEL classification: C1, C2, C3, C4

*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

†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 that can come 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 Bai (1997), 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. However, 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. We also provide a test for the null hypothesis that the panel is homogeneous. A non-rejection of the hypothesis then supports pooling the entire panel to further improve efficiency of the estimates.

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. The method, which is also non-parametric, is compared to the pseudo threshold approach. Simulations show that the two methods

have rather similar performances. As an application, we apply our methods to study economic growth across counters. 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 the proposed method to estimate the clusters. The results indicate that some countries in the low growth club have high per capita income, or belong to the OECD. On the other hand, some countries in the high growth club do not belong to the OECD and are not located in what are generally thought to be high growth regions. Splitting the data by geographical location, OECD membership, the level of income do not seem to produce clusters that are statistically optimal.

The remainder of this paper is organized as follows. After some preliminaries, Section 3 considers how to form clusters when a perfect indicator of group membership is observed. In Section 4, a feasible approach that replaces the latent indicator by two estimated indicators will be considered. A test for homogeneity is introduced. The analysis is extended to the case of multiple regressors in Section 5. The relation of our method to the K-means algorithm will be discussed in Section 6. Simulations and an empirical application are presented in Section 7.

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.[1] 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.[2] 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

[1] See, for example, Swamy (1970) and Hsiao and Pesaran (2004).

[2] See, for example, Fraley and Raftery (2002), Hall, Muller, and Wang (2006), and Chiou and Li (2006).

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 (or minimum).

The idea of model-based clustering is further extended by Sun (2005) to estimation of panel models with unknown group structure. The likelihood is maximized 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. Sun (2005) 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 (2006) 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 coefficient 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 $B_1=B_2$, and the case of complete heterogeneity obtains when there are N clusters each consisting of only one unit.

Suppose 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 $q_i\le\gamma^0$ for any $\gamma^0\in\Gamma^0$ and $i\in I_2$ otherwise. Then (1) can be written as

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

We can also consider a threshold representation of the model

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

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 $q_{it} \ge \gamma^0$ 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. We can rewrite the model as

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

Split sample regressions can then be performed using (y_{it}, x_{it}) as data. Observations with $q_i \le \gamma^0$ 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

$$\hat{B}_\omega = \frac{\sum_{i=1}^N \sum_{t=1}^T \delta_i(\tilde{y}_{it} - \overline{\tilde{y}_i})(\tilde{x}_{it} - \overline{\tilde{x}_i})}{\sum_{i=1}^N \sum_{t=1}^T \delta_i(\tilde{x}_{it} - \overline{\tilde{x}_i})^2} = \frac{\sum_{i=1}^N \sum_{t=1}^T \delta_i y_{it} x_{it}}{\sum_{i=1}^N \sum_{t=1}^T \delta_i x_{it}^2}.$$

The unweighted fixed effect estimator obtains when $\delta_i = 1$. The weighted fixed-effects estimator obtains when $\delta_i = 1/\hat{\sigma}_i^2$. It can be rewritten as

$$\hat{B}_\omega = \sum_{i=1}^N \kappa_i \hat{\beta}_i, \qquad \kappa_i = \frac{\sum_{t=1}^T \delta_i x_{it}^2}{\sum_{i=1}^N \sum_{t=1}^T \delta_i x_{it}^2}.$$

This estimator 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) and Lin (2006) to allow both N and T to be large.

The following assumptions will be used throughout.

Assumption A

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

Lemma 1 *Suppose Assumption A holds. Then* $\hat{\beta}_i = \beta_i + O_p(T^{-1/2})$. *Let* $B_\omega = \omega B_1 + (1-\omega)B_2$, *where* $\omega = \pi \in (0,1)$ *if* $\delta_i = 1$, *and* $\omega = \text{plim}\frac{\sum_{i\in I_1}\hat{\sigma}_i^{-2}\hat{Q}_i}{\sum_{i=1}^{N}\hat{\sigma}_i^{-2}\hat{Q}_i}$ *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 the slope parameters are cross-sectionally heterogeneous. 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 $q_i \le \gamma$ 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

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

Since q_i can be used to order the data, this means that if q_i is less than some trial value of γ and the i-th unit is classified into group 1, any unit j with $q_j < q_i$ will also be classified in the group. 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.

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. Notice that when $K = 2$ too many units are put in group 1, say, ie $N_{21}(\gamma) > 0$, it will also be the case that $N_{12}(\gamma) = 0$. Thus, one of the missclassified set is always empty. A unit misclassified into Group 1 will contribute a larger squared error than if the unit was put into group 2 since B_2 is closer to $\hat{B}_2(\gamma)$ than $\hat{B}_1(\gamma)$. 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 the data are generated by (1). Let N_{jk} be number of units in group j being mis-classified into group k and let $N_s = \max_{j\neq k} N_{jk}$. Then (i) for fixed $B_2 - B_1 = O(1)$, $N_s = O(\sqrt{\frac{N}{T}})$, and (ii) for $B_2 - B_1 = cT^{-(1/2-\Delta)}$, $0 \leq \Delta < 1/2$, $0 < c < \infty$, $N_s = O(\frac{\sqrt{N}}{T^{\Delta}})$.*

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. We only need to consider where to position γ in relation to the N ordered observations of q_i, denoted $q_{[i]}$. There will be a $\tilde{\gamma}$ that minimizes the size of the mis-classified set. In fact, any $\tilde{\gamma} \in [q_{[i^*]}, q_{[i^*]+1})$, where i^* is such that $\tilde{q}_{[i^*]} = \tilde{\gamma}$, will yield the same classification. For fixed $B_2 - B_1 = O(1)$, Theorem 1 implies that the maximum mis-classification rate is $N_s/N = O(\frac{1}{\sqrt{NT}})$. If $B_2 - B_1$ is in the $T^{-(1/2-\Delta)}$ neighborhood of zero, the mis-classification rate is $N_s/N = O(\frac{1}{\sqrt{N}T^{\Delta}})$. 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 $I_2(\tilde{\gamma}) = \{i|q_i > \tilde{\gamma}\}$. The miss-classified units are then in the set $I_{jk}(\tilde{\gamma}) = \{i : i \notin (I_j(\tilde{\gamma}) \cap I_j(\gamma^0))\}$ for $j, k = 1, 2$ and $j \neq k$. Once group membership is consistently estimated, units within a group can be pooled to yield more efficient cluster-specific parameters. Consistency and asymptotic normality of $\hat{B}_1$ and $\hat{B}_2$ can be established treating $\tilde{\gamma}$ as though it was known. Note that likelihood based cluster analysis yields a probability that unit i belongs to a group. Markov switching and threshold models also share the same conceptual difference. In our analysis, group membership is known once we can find 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

$$\hat{\gamma} = \arg \min_{\gamma \in [\hat{q}_{min}, \hat{q}_{max}]} S_{NT}(\gamma, \hat{q}), \tag{5}$$

where

$$S_{NT}(\gamma, \hat{q}) = \sum_{i|\hat{q}_i \leq \gamma} \sum_{t=1}^{T} \delta_i (y_{it} - \hat{B}_1(\gamma) x_{it})^2 + \sum_{i|\hat{q}_i > \gamma} \sum_{t=1}^{T} \delta_i (y_{it} - \hat{B}_2(\gamma) x_{it})^2. \tag{6}$$

The two groups are then estimated as

$$\hat{I}_1 = \{i|\hat{q}_i \leq \hat{\gamma}\} \quad \text{and} \quad \hat{I}_2 = \{i|\hat{q}_i > \hat{\gamma}\}.$$

We now propose two estimate of q_i.

A Non-normalized $\hat{q}_i$ 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

$$\hat{q}_i = \hat{\tau}_i \quad \text{where} \quad \hat{\tau}_i = \frac{\sqrt{T}(\hat{\beta}_i - \hat{B}_\omega)}{\sqrt{\hat{\sigma}_i^2 \hat{Q}_i^{-1}}}, \tag{7}$$

$\hat{Q}_i$ 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 $q_i = \tau_i$ 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, q_i can be expressed as:

$$\tau_i = \begin{cases} \frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\sigma_i Q_i^{-1/2}} - \frac{\sqrt{T}(1-\omega)(B_2-B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1), & \text{for } i \in I_1, \\ \frac{\sqrt{T}(\hat{\beta}_i - B_2)}{\sigma_i Q_i^{-1/2}} + \frac{\sqrt{T}\omega(B_2-B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1), & \text{for } i \in I_2. \end{cases} \tag{8}$$

It can be seen that as $N, T \to \infty$ jointly,

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

where $z \sim N(0,1)$ is a standard normal random variable. This implies that if $B_2 > B_1$, q_i will become more negative for those $i \in I_1$ and more positive for those $i \in I_2$ as T increases. On the other hand, when $B_1 = B_2 = B_\omega$,

$$\tau_i = \frac{\sqrt{T}\left(\hat{\beta}_i - B_\omega\right)}{\sigma_i Q_i^{-1/2}} + o_p(1) \quad \forall i.$$

In this case, τ_i will be approximately normal with mean zero as N and T become large. It is obvious that q_i will eventually separate the sample into two groups when T is large and $B_1 \neq B_2$. It can also be seen that τ_i will identify the clusters in which the parameters are such that

$$B_2 - B_1 = O(T^{-1/2+\Delta})$$

for $0 < \Delta \leq 1/2$.

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. That the misclassification rate decreases with T will be confirmed in the simulations.

4.1 Testing for Homogeneity

The estimation procedure above proceeds assuming that $B_1 \neq B_2$. We are interested in testing two formulations of the null hypothesis of parameter homogeneity: $H_0^A : \beta_i = B \quad \forall i$, 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, Davies, 1977, Andrews and Ploberger, 1994, Hansen, 1996, Bai, 1997, and Caner and Hansen, 2004. Briefly, for a fixed $\pi \in [\pi_{min}, \pi_{max}]$, where $q_{[\pi N]} = \gamma$, and under the assumption of homoskedasticity, the Wald statistic for testing parameter homogeneity in the two samples split at $[\pi N]$ is

$$W(\pi) = \left(\hat{B}_1(\gamma) - \hat{B}_2(\gamma)\right)' \left[\widehat{\text{var}}(\hat{B}_1(\gamma)) + \widehat{\text{var}}(\hat{B}_2(\gamma))\right]^{-1} \left(\hat{B}_1(\gamma) - \hat{B}_2(\gamma)\right),$$

where $\hat{B}_j$ are unbiased estimates of B_j based on the split samples. The sup-Wald-type statistic is defined as

$$\text{SupW} = \sup_{\pi \in [\pi_{min}, \pi_{max}]} W(\gamma)$$

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 $\hat{\beta}_i$. By the nature that the sample is truncated, $\hat{B}_1$ will always be biased downward while $\hat{B}_2$ will be upward biased even if $B_1 = B_2 = B$. For this reason, we need to obtain estimates that are unbiased for B under the null hypothesis.

Second, in conventional threshold models, the parameter estimators in different groups are uncorrelated because individuals are assumed to be cross-sectionally independent. However, our grouping method is based on $\hat{\beta}_i$ and, therefore, under the null that $B_1 = B_2 = B$, $\hat{B}_1$ and $\hat{B}_2$ are correlated. To the best of our knowledge, there exists no satisfactory method to estimate covariance between two estimators from split samples with different sizes.

Third, in the conventional threshold models with fixed threshold variable q_i, we can randomly reshuffle the errors to find asymptotic critical values (see, for example, Caner and Hansen 2004, p.823–824). But our $\hat{q}_i$'s are estimated and when the pseudo- dependent variable y_i^* is randomly generated[3] the ordering of $\hat{q}_i^*$ will be changed. Therefore, we cannot simply apply Hansen's method (which assumes $\hat{q}_i^*$ does not change) to obtain asymptotic critical values.

[3]The pseudo-dependent variable can be generated by $y_i^* = \hat{e}_i(\pi)\eta_i$, where $\hat{e}_i(\pi)$ is the estimated error under the unrestricted model for each π, and $\eta_i \sim iidN(0,1)$.

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

$$PY = \frac{\sqrt{N}(S/N - K)}{\sqrt{2K}}, \tag{9}$$

where K 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

$$S = \sum_{i=1}^{N} (\hat{\beta}_i - \hat{B}_w)' \frac{(x_i x_i')}{\tilde{\sigma}_i^2} (\hat{\beta}_i - \hat{B}_w)'.$$

This test allows for cross-sectional fixed effects and heteroskedasticity and has a standard normal distribution as N and T go to infinity jointly in the models with strictly exogenous regressors and normally distributed errors. It is worth noting that the SupW test in conventional threshold model has more power against the alternative than this dispersion test, because the asymptotic distribution of the PY test is established based on a large N asymptotics. In our case, however, the SupW test is infeasible; but the simulation results show that the PY test still can have satisfactory size and power.

5 Multiple Regressors and Dynamic Panels

We have shown that when there is a single regressor, either $\hat{\beta}_i$ or a t-type test 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, a case we refer to as partial parameter heterogeneity, we can still split the sample based on one $\hat{\beta}_{ig}$, $g = 1, \ldots, G$, at a time. The hypothesis of homogeneous parameter can be rejected when the parameter homogeneity test is rejected for some g.

Suppose it is suspected that all the parameters are different between groups. The natural extension to the multiple regressor case is to use the sum of the estimated parameters for unit i, or some form of joint test for the hypothesis that the parameters of each unit equal the corresponding group mean in the population. Complete parameter homogeneity would then imply restrictions of the form $R\hat{q}_i$, where R is a $1 \times G$ vector. This intuition is generally correct, but 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 $B_1 = (B_{11}, B_{12})'$ and $B_2 = (B_{21}, B_{22})'$ be slope parameters for Group 1 and Group 2, respectively. Let $\hat{\beta}_i = (\hat{\beta}_{i1}, \hat{\beta}_{i2})'$ denote the OLS estimate of the slope parameters for individual i and let $\hat{B}_\omega = (\hat{B}_{\omega 1}, \hat{B}_{\omega 2})'$ be a consistent estimate of B_ω. By assumption, the slope parameters of the units in Group 1 are the same, but are different from those in Group 2. In principle, we can take $\hat{\beta}_{i1} - \hat{B}_{\omega 1}$ or $\hat{\beta}_{i2} - \hat{B}_{\omega 2}$ as the pseudo transition variable to split the sample. However, intuition suggests that $\hat{\beta}_{i1}$ and $\hat{\beta}_{i2}$ together should provide

more information about group structure than using one variable alone. But how should we combine the information? There are actually two cases to consider.

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 Group 1 than in Group 2, a natural pseudo transition variable is

$$\hat{\beta}_i^+ = \hat{\beta}_{i1} + \hat{\beta}_{i2}.$$

Using arguments analogous to the previous section, one can expect $\hat{\beta}_i^+$ to completely separate those $i \in I_1$ from those $i \in I_2$ when T is large. However, this method does not always work. For example if $(B_{11}, B_{12}) = (0.8, 1)$ and $(B_{11}, B_{12}) = (1, 0.8)$, we have $B_{11} + B_{12} = B_{21} + B_{22}$. The sum of the coefficients is no longer a sufficient statistic for group membership even though we can still split the sample using $\hat{\beta}_{ig} - \hat{B}_{\omega g}$ 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 $B_{j_1,_2} < B_{j_2,2}$, $j_1, j_2, = 1, 2$, $j_1 \neq j_2$. We will refer to this as Case 2. That is, when the ranking of the parameter between groups depend on which parameter is being compared. 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. But can the data tell us which case will being analyzed? Fortunately, the answer is yes because there are notable differences between the two cases. Under Case 1, $\hat{\beta}_{i1} - \hat{B}_{w1}$ and $\hat{\beta}_{i2} - \hat{B}_{\omega 2}$ should have the same sign. Under Case 2, these quantities should have the opposite signs. The signs of the paired coefficients are thus informative about which case should be considered. This suggests the following. Calculate the Goodman-Kruskal's statistic, defined as

$$\psi = N^{-1} \sum_i^N \left[\frac{\hat{\beta}_{i1} - \hat{B}_{\omega 1}}{|\hat{\beta}_{i1} - \hat{B}_{\omega 1}|} \right] \left[\frac{\hat{\beta}_{i2} - \hat{B}_{\omega 2}}{|\hat{\beta}_{i2} - \hat{B}_{\omega 2}|} \right].$$

It is a symmetric and non-parametric measure of association between pairs of concordant data (with the same sign) and discordant data (with opposite sign). Then define

$$R = \begin{cases} (1,1) & \text{if } \psi > 0 \\ (1,-1) & \text{if } \psi \leq 0. \end{cases}$$

Given R, our proposed non-normalized pseudo threshold variable is

$$\hat{q}_i = \hat{q}_i(\hat{\beta}_i) = R\hat{\beta}_i. \tag{10}$$

The normalized threshold variable is then

$$\hat{q}_i = q_i(\hat{\tau}_i) = \frac{R(\hat{\beta}_i - \hat{B}_\omega)}{\sqrt{(R\widehat{\text{var}(\hat{\beta}_i)}R')}}. \tag{11}$$

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..

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. Let

$$\psi_g = N^{-1} \sum_i^N \left[\frac{\hat{\beta}_{ig} - \hat{B}_{\omega g}}{|\hat{\beta}_{ig} - \hat{B}_{\omega g}|}\right] \left[\frac{\hat{\beta}_{iW} - \hat{B}_{\omega W}}{|\hat{\beta}_{iW} - \hat{B}_{\omega W}|}\right], \quad g = 1, \ldots, G$$

   Create a $1 \times G$ vector R, where the g^{th} position is 1 if $\psi_g \geq 0$ and -1, otherwise;

4. The pseudo threshold variable for a model with G regressors is

$$\hat{q}_i = R\hat{\beta}_i \qquad \text{or} \qquad \hat{q}_i = \frac{R(\widehat{\beta}_i - \widehat{\mathbf{B}}_\omega)}{\sqrt{(R\widehat{\text{var}(\hat{\beta}_i)}R')}}. \tag{12}$$

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. To fix ideas, suppose we observe y_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, and there are no covariates. Suppose there are two groups in the data. Let I_1 and I_2 be two clusters of size N_1 and N_2, with $N_1 + N_2 = N$.

The simplest way to estimate group membership is plot $\hat{\beta}_i (= \bar{y}_i)$ and then 'eyeball' to see when $\hat{\beta}_i$ abruptly shifts from one mean to another.[4] 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.[5] The basic algorithm consists of the following steps: (i) initialize K points to represent the initial group mean (often referred to as centroid in this literature); (ii) assign y_{it} to the group that has the closet centroid; (ii) recalculate the positions of the K centroids when all objects have been assigned. Steps (ii) and (iii) are then repeated until the centroids no longer change. Essentially, the algorithm

[4] See, for example, Henderson and Russell (2005).

[5] See, for example, Hartigan (1975), Abraham, Cornillion, Matzner-Lober, and Molinari (2003).

produces K clusters by moving unit i to a different group to minimize the sum of squared deviations between the units and the centroids. There are many variations to the basic algorithm. Instead of "batch" updating, which recalculates the centroids after all units are assigned, "online" updating which recalculates the centroids as each unit is re-assigned has also been considered. Harmonic and fuzzy means have also been used instead of simple means. The main idea of K-means method remains that two clusters in y_{it} are identified by comparing the data to the sequentially updated centroids. The clusters I_1 and I_2 are then estimated as

$$(\hat{I}_1, \hat{I}_2) = \text{argmin}_{I_1,I_2} \sum_{i \in I_1} \sum_{t=1}^{T} (y_{it} - \hat{B}_1)^2 + \sum_{i \in I_2} \sum_{t=1}^{T} (y_{it} - \hat{B}_2)^2.$$

It is well known that the K-means method can be sensitive to the initial choice of the centroids. A different choice of the initial partition can result in different final clusters. 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

$$y_{it} = B_1 x_i 1(i \in I_1) + B_2 x_i 1(i \in I_2) + e_{it}$$

where for a generic variable $\tilde{z}_{it}$, z_{it} is the demeaned data. 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

$$SSR_i^1 = \sum_{t=1}^{T} (y_{it} - \ddot{B}_1 x_{it})^2 \quad \text{and} \quad SSR_i^2 = \sum_{t=1}^{T} (y_{it} - \ddot{B}_2 x_{it})^2. \tag{13}$$

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 $(\ddot{B}_1, \ddot{B}_1)$ based on $\{\ddot{I}_1, \ddot{I}_2\}$.
4. repeat (a) and (b) until no individual is changed from one group to another.
   a calculate $SSE_i^j = \sum_{t=1}^{T} (y_{it} - \ddot{B}_j x_{it})^2, \quad i = 1, \ldots, N$
   b if $SSE_i^1 \leq SSE_i^2$, individual i is re-assigned to group 1; otherwise, i stays with group 2.
   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)–(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^1 and SSR^2 as follows:

$$SSR_i^{1*} = \sum_{t=1}^{T}(y_{it} - \hat{B}_1 x_{it})^2/\hat{\sigma}_i^2 \quad \text{and} \quad SSR_i^{2*} = \sum_{t=1}^{T}(y_{it} - \hat{B}_2 x_{it})^2/\hat{\sigma}_i^2. \tag{14}$$

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 makes random initial guesses of the centroids and then evaluate if a move to a different group is desirable unit by unit. This also 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 decide which units to move. To the extent that q_i is a sufficient statistic about β_i and β_i is indeed the true classifier, a global minimizer of the sum of squared residuals of the panel threshold model exists and is unique. The pseudo threshold approach can be easily extended from two to more groups by first splitting the sample into two groups, and then looking for two groups within each of the subgroups. With the K-means method, the reassignment step when there are multiple alternatives can be computationally very costly, and convergence can also be slow.

The K-means method also has two advantages. First, the algorithm 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. Second, the method considers every unit in the sample for a move to a different group. This is unlike the pseudo threshold method which moves 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. 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 $B_1 = 1$ and $B_2 = 0.7$ 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 $N = (20, 50, 100, 200)$ and $T = (20, 50, 100, 200)$, we keep track of the computation time, the root-mean-squared error of the estimates (RMSE), and the accuracy of classification (CR). For $j = 1, 2$, these are defined as

$$\begin{aligned} RMSE(\hat{B}_j) &= \left[\frac{1}{M}\sum_{m=1}^{M}(\hat{B}_j^{(m)} - B_j)^2\right]^{1/2} \\ CR &= \frac{1}{M}\sum_{m=1}^{M} CR^{(m)} \quad \text{where} \quad CR^{(m)} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{2}(i \in \hat{I}_j^{(m)}) \bigcap (i \in I_j^0), \end{aligned}$$

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 denote the pseudo threshold approach using $\hat{\beta}_i$ as a transition variables, while PS_2 denote the method using the standardized transition variable 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. 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 60.5% to 82% when T increases from 20 to 200.

Because K_1 (K-means), PS_1 (using $\hat{q}_i = \hat{\beta}_i$) and PS_2 (using $\hat{q}_i = \hat{\tau}_i$) 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. Notably, the computation time used to calculate K_1 is much 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 $T = 100$ 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, with $T \geq 20$, 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 $(B_{21}, B_{22}) = (1, 1.1)$ 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 hold $\{I_1^0, I_2^0\}$, α_i, ρ_i, and σ_{ix}^2 fixed in the replications. Because there are two parameters, the RMSE is now computed as

$$RMSE = \left[\frac{1}{M}\sum_{m=1}^{M}\left(\frac{N_1}{N}\left[(\hat{B}_{11}^{(m)} - B_{11})^2 + (\hat{B}_{12}^{(m)} - B_{12})^2\right] + \frac{N_2}{N}\left[(\hat{B}_{21}^{(m)} - B_{21})^2 + (\hat{B}_{22}^{(m)} - B_{22})^2\right]\right)\right]^{1/2}.$$

The computation time here is roughly 1.5 times longer than what is in Table 1 and will not be reported to conserve space. 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.

Table 8 shows the empirical size and power of the PY test. The DGPs are the same with those in Tables 1, 2, and 3, respectively. The size for each case is roughly around the 5% nominal level except two cases in which $(K, T, N) = (2, 20, 100)$ and $(K, T, N) = (2, 20, 200)$. Clearly, when T or N increases, the PY test will have more power to detect the alternative. In the case with only one regressor, its finite sample power is usually

higher when errors are homoskedastic than when errors are heteroskedastic. Interestingly, this result is reverse in the case with two regressors.

7.1 Empirical Study

The existence of "convergence clubs" has generated much research interests in the growth literature. A groups of countries with a similar steady state and can be characterized by the same linear model is said to form a convergence club. Most studies use observed variables to group the countries and then estimate the group specific parameters. See Durlauf, Kourtellos, and Tan (2005) for a survey. Some find that the quality of institutions and ethnic fractionalization are the most important determinants of economic growth. Others argue that the savings rate is more important, as are education-related variables. See Barro and Sala-i-Martin (2003) for a discussion on issues relating to empirical growth regressions.

To motivate the estimation issue when group membership is not known, consider the model used in Lee, Pesaran, and Smith (1997) for 70 countries, taken from the PWT v6.2 for the sample 1965 to 2000.[6] Figure 1(a) shows the same scatterplot for countries for Liberia, Myanmar, Somalia, and Sudan. Also, Germany is removedin different geographical regions. It is obvious that the growth ratesfrom data set due to consolidation 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.

The regression model is

$$\tilde{y}_{it} = \mu_i' + g_i t + \tilde{\epsilon}_{it} \tag{15}$$

and the parameter of interest is g_i. The regression can be motivated by a neoclassical growth model as in Mankiw, Romer, and Weil (1992).[7] Figures 1(b)–(c) plot the distribution of the non-standardized and standardized $\hat{g}_i$ estimated by kernel smoothing.

We use the proposed methods to split countries into two (or more) groups. Table 5 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 percent, and between 40 and

[6] See Mankiw, Romer, and Weil (1992) on how to select 75 intermediate countries. However, Germany is removed from data set due to consolidation. Due to limitation of data, we also remove Bangladesh, Bolivia, Haiti, and Myanmar.

[7] Specifically, $\tilde{y}_{it} = \gamma_i \tilde{y}_{i,t-1} + \tilde{x}_{it}\beta_i + \phi_i t + c_i + D_t + \tilde{e}_{it}$ where c_i captures the country-specific effects, D_t are common time dummies, $\tilde{y}_{it}$ is the logarithm of real per capita output (2000 as base year); $\tilde{x}_{it} = [\ln(\tilde{s}_{it}), \ln(\tilde{n}_{it} + g_i + \delta)]$, $\tilde{s}_{it}$ is the savings rate at time t, $\tilde{n}_{it}$ is the rate of growth of population, g_i is the rate of technology growth, and δ is the rate of depreciation. If λ_i is rate of convergence to steady state per capita output for country i, then $\gamma_i = e^{-\lambda_i}$. $\phi_i = g_i(1-\gamma_i)$ and $c_i = \gamma_i A_i(0) + \phi_i$, where $A_i(0)$ denotes the initial endowment for country i. Letting μ_i' absorbs all the invariant parameters, with $\tilde{\epsilon}_{it} = \gamma_i \tilde{\epsilon}_{it-1} + \eta_t + \tilde{e}_{it}$ gives the regression being estimated. See Lee, Pesaran, and Smith (1997) for details.

42 countries in the high growth group with a growth rate of about 2.7 percent. From Table 6, the methods differ primarily over the classification of eleven 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. Meanwhile, countries such as Ecuador and Syria are geographically distinct from the major OECD countries and have relatively low average GDP per capita. Yet, they are in the high growth group. 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. As seen from Table 7, a qualitatively similar picture emerges, namely, that splitting the data by geographical location, OECD membership, or the level of income can produce clusters of the data that cannot be statistically optimal. 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," (Berlin D2-3), 10th International Conferences on Panel Data in its series.

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

Bai, J. (1997): "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics*, 79, 551–563.

Barro, R. J., and X. Sala-i-Martin (2003): *Economic Growth*. The MIT Press.

Chiou, J.-M., and P.-L. Li (2006): "Functional Clustering of Longitudinal Data," Talk at 2006 Taipei International Statistic Workshop at National Taiwan University, Taiwan.

Durlauf, S. N., A. Kourtellos, and C. M. Tan (2005): "Empirics of Growth and Development," .

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.

Henderson, D. J., and R. R. Russell (2005): "Human Capital and Convergence: A Production-Frontier Approach," *International Economic Review*, 46, 1167–1205.

Hsiao, C., and M. H. Pesaran (2004): "Random Coefficient Panel Data Models," *CESifo Working Paper*, (1233).

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 (2006): "Model-based Clustering of non-Gaussian Panel Data," *Unpublished*.

Lee, K., M. H. Pesaran, and R. Smith (1997): "Growth and Convergence in a Multicountry 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," .

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

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 $\hat{B}_j = \frac{\sum_{i\in I_j}\sum_{t=1}^T \delta_i x_{it} y_{it}}{\sum_{i\in I_j}\sum_{t=1}^T \delta_i x_{it}^2}$ with $\delta_i = 1$ or $1/\hat{\sigma}_i^2$.

By direct calculations,

$$\begin{aligned}
& \sqrt{NT}\left[\hat{B}_\omega - (\omega B_1 + (1-\omega)B_2)\right] \\
= \; & \sqrt{NT}\left[\frac{\sum_{i\in I_1}\sum_{t=1}^T \delta_i x_{it} y_{it} + \sum_{i\in I_2}\sum_{t=1}^T \delta_i x_{it} y_{it}}{\sum_{i=1}^N \sum_{t=1}^T \delta_i x_{it}^2} - (\omega B_1 + (1-\omega)B_2)\right] \\
= \; & \sqrt{NT}\left(\frac{\sum_{i\in I_1}\delta_i \hat{Q}_i}{\sum_{i=1}^N \delta_i \hat{Q}_i} \cdot \frac{\sum_{i\in I_1}\sum_{t=1}^T \delta_i x_{it} y_{it}}{\sum_{i\in I_1}\sum_{t=1}^T \delta_i x_{it}^2} - \omega B_1\right) \\
& \quad + \sqrt{NT}\left(\frac{\sum_{i\in I_2}\delta_i \hat{Q}_i}{\sum_{i=1}^N \delta_i \hat{Q}_i} \cdot \frac{\sum_{i\in I_2}\sum_{t=1}^T \delta_i x_{it} y_{it}}{\sum_{i\in I_2}\sum_{t=1}^T \delta_i x_{it}^2} - (1-\omega)B_2\right) \\
= \; & \sqrt{\frac{N}{N_1}}\left[\omega\sqrt{N_1 T}\left(\hat{B}_1 - B_1\right)\right] + \sqrt{\frac{N}{N_2}}\left[(1-\omega)\sqrt{N_2 T}\left(\hat{B}_2 - B_2\right)\right] + o_p(1) \; = O_p(1).
\end{aligned}$$

To prove Lemma 1 with $\delta_i = 1/\hat{\sigma}_i^2$, it suffices to show that $(N_j T)^{1/2}(\hat{B}_j - B_j) = O_p(1)$. By Assumption A,

$$N_j^{-1}\sum_{i\in I_j}\sigma_i^{-2}\hat{Q}_i = Q^j = O_p(1).$$

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

We also have

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

$$N_j^{-1}\sum_{i\in I_j}\hat{\sigma}_i^{-2}\hat{Q}_i = N_j^{-1}\sum_{i\in I_j}\sigma_i^{-2}\hat{Q}_i + O_p\left(\frac{1}{T}\right).$$

Therefore, if $\sqrt{N}/T \to 0$,

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

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

Proof of Theorem 1: Let (I_1^0, I_2^0) be the true group membership and let (I_1, I_2) denote group membership other than (I_1^0, I_2^0). Suppose that the DGP is

$$\begin{aligned}\tilde{y}_{it} &= \alpha_i + B_1\tilde{x}_{it} + \tilde{e}_{it}, \text{ as } i \in I_1^0,\\ \tilde{y}_{it} &= \alpha_i + B_2\tilde{x}_{it} + \tilde{e}_{it}, \text{ as } i \in I_2^0.\end{aligned}$$

We will consider the general case where $B_2 - B_1 = cT^{-(1/2-\Delta)}$, $0 < \Delta \leq 1/2$. Let $\alpha = 1/2 - \Delta$. Then $\alpha = 0$ corresponds to the case when $B_2 - B_1 = c \neq 0$, c does not depend on T.

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 $\tilde{y}_{it}, \tilde{x}_{it}, \tilde{e}_{it}, \widetilde{e}_{it}$, and $\overline{\tilde{z}}_i = \frac{1}{T}\sum_{t-1}^{T}\tilde{z}_{it}$. Then for $j, k = 1, 2$,

$$y_{it} = B_k x_{it} + e_{it} = \hat{B}_j(I_1, I_2)x_{it} + \hat{e}_{it}^{kj}.$$

where $\hat{e}_{it}^{kj} = e_{it} + (B_k - \hat{B}_j(I_1, I_2))x_{it}$. We have

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

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 the scaled sample moments of the misclassified units and are defined as

$$\begin{aligned}\phi_{1,kj} &= \frac{T^{2\alpha}}{N_s}\left\{N_{jj}(\hat{B}_j^0 - \hat{B}_j(I_1, I_2))^2\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^{T}x_{it}^2}{N_{jj}T} + N_{kj}(\hat{B}_k^0 - \hat{B}_j(I_1, I_2))^2\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^{T}x_{it}^2}{N_{kj}T}\right\},\\ \phi_{2,kj} &= \frac{2T^{2\alpha}}{N_s}\left\{N_{jj}(\hat{B}_j^0 - \hat{B}_j(I_1, I_2))\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^{T}x_{it}e_{it}}{N_{jj}T} + N_{kj}(\hat{B}_k^0 - \hat{B}_j(I_1, I_2))\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^{T}x_{it}e_{it}}{N_{kj}T}\right\}.\end{aligned}$$

We want to show that for any $(I_1, I_2) \neq (I_1^0, I_2^0)$, the sum of squared residuals with mis-classified units must be strictly greater than the sum of squared residuals when all units are correctly classified. That is,

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

if $N^{1/2}T^{-1/2+\alpha}/N_s = o(1)$, or equivalently

$$\text{Prob}\Big(\sum_{j,k=1,2,j\neq k} \phi_{1,kj} + \phi_{2,kj} > 0\Big) \xrightarrow{p} 1.$$

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 $\sum_{j\neq k}\phi_{1,kj} = O_p(1)$.

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

$$\begin{aligned}
\hat{B}_j(I_1,I_2) &= \frac{C_{jj}+C_{kj}}{D_{jj}+D_{kj}} = \frac{C_{jj}}{D_{jj}}\frac{D_{jj}}{D_{jj}+D_{kj}} + \frac{C_{kj}}{D_{kj}}\frac{D_{kj}}{D_{jj}+D_{kj}} \\
&= \hat{B}_{jj}^*\frac{D_{jj}}{D_{jj}+D_{kj}} + \hat{B}_{kj}^*\frac{D_{kj}}{D_{kj}+D_{kj}} \\
&= \hat{B}_{jj}^* + (\hat{B}_{kj}^* - \hat{B}_{jj}^*)\frac{D_{kj}}{D_{jj}+D_{kj}} \\
&= B_j + \Big[(B_k - B_j) + (\hat{B}_{kj}^* - B_k) - (\hat{B}_{jj}^* - B_j)\Big]\frac{D_{kj}}{D_{jj}+D_{kj}} + (\hat{B}_{jj}^* - B_j).
\end{aligned}$$

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

$$\begin{aligned}
\phi_{21,kj} &= \frac{N_{jj}T^{2\alpha}}{N_s}\left[(B_k - B_j)\frac{D_{kj}}{D_{jj}+D_{kj}}\right]\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{jj}T} \\
&\quad + \frac{N_{kj}T^{2\alpha}}{N_s}\left[(B_j - B_k)\frac{D_{jj}}{D_{jj}+D_{kj}}\right]\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{kj}T}, \\
\phi_{22,kj} &= \left\{\Big[(\hat{B}_{kj}^* - B_k) - (\hat{B}_{jj}^* - B_j)\Big]\frac{D_{kj}}{D_{jj}+D_{kj}} + (\hat{B}_{jj}^* - B_j)\right\}\times\frac{N_{jj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{jj}T} \\
&\quad + \left\{\Big[(\hat{B}_{kj}^* - B_k) - (\hat{B}_{jj}^* - B_j)\Big]\frac{D_{kj}}{D_{jj}+D_{kj}} + (\hat{B}_{jj}^* - B_j)\right\}\times\frac{N_{kj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{kj}T} \\
&\quad + (\hat{B}_j^0 - B_j)\frac{N_{jj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_j^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{jj}T} + (\hat{B}_k^0 - B_k)\frac{N_{kj}T^{2\alpha}}{N_s}\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^T x_{it}e_{it}}{N_{kj}T}.
\end{aligned}$$

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. We have

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

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

$$\frac{N_{kj}}{N_s}\frac{D_{kj}}{D_{jj}+D_{kj}} = O_p(1), \quad \frac{N_{jj}}{N_s}\frac{D_{kj}}{D_{jj}+D_{kj}} = O_p(1).$$

Thus,

$$\begin{aligned}
\phi_{21,kj} &= T^{\alpha}(B_k - B_j)\left(\frac{N_{jj}}{N_s}\frac{D_{kj}}{D_{jj}+D_{kj}}\right)\left(T^{\alpha}\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^{T}x_{it}e_{it}}{N_{jj}T}\right) \\
&\quad + T^{\alpha}(B_k - B_j)\left(\frac{N_{kj}}{N_s}\frac{D_{jj}}{D_{jj}+D_{kj}}\right)\left(T^{\alpha}\frac{\sum_{i\in I_k^0\cap I_j}\sum_{t=1}^{T}x_{it}e_{it}}{N_{kj}T}\right) \\
&= c\cdot\left(\frac{N_{jj}}{N_s}\frac{D_{kj}}{D_{jj}+D_{kj}}\right)O_p((N_{jj}T)^{-1/2}T^{\alpha}) + c\cdot\left(\frac{N_{kj}}{N_s}\frac{D_{jj}}{D_{jj}+D_{kj}}\right)O_p((N_{kj}T)^{-1/2}T^{\alpha}) \\
&= O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) = o_p(1).
\end{aligned}$$

To show that $\phi_{22,kj} = o_p(1)$, note first that

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

But $\hat{B}^*_{kj} - B_k = O_p((N_{kj}T)^{-1/2})$, $\hat{B}^*_{jj} - B_j = O_p((N_{jj}T)^{-1/2})$, $\hat{B}^0_j - B_j = O_p(((N_{jk} + N_{jj})T)^{-1/2})$, and $\hat{B}^0_k - B_j = O_p(((N_{kj} + N_{kk})T)^{-1/2})$. We have

$$\begin{aligned}
\phi_{22,kj} &= \left[O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) + O_p(N_{jj}^{-1/2}T^{-1/2+\alpha})\right]\left(\frac{N_{jj}}{N_s}\frac{D_{kj}}{D_{jj}+D_{kj}}\right)\times O_p(N_{jj}^{-1/2}T^{-1/2+\alpha}) \\
&\quad + O_p(N_{jj}^{-1/2}T^{-1/2+\alpha})\frac{N_{jj}}{N_s}O_p(N_{jj}^{-1/2}T^{-1/2+\alpha}) \\
&\quad + \left[O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) + O_p(N_{jj}^{-1/2}T^{-1/2+\alpha})\right]\left(\frac{N_{kj}}{N_s}\frac{D_{kj}}{D_{jj}+D_{kj}}\right)\times O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) \\
&\quad + O_p(N_{jj}^{-1/2}T^{-1/2+\alpha})\frac{N_{kj}}{N_s}O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) \\
&\quad + O_p((N_{jj}+N_{jk})^{-1/2}T^{-1/2+\alpha})\times O_p(N_{jj}^{-1/2}T^{-1/2+\alpha}) \\
&\quad + O_p((N_{kj}+N_{kk})^{-1/2}T^{-1/2+\alpha})\times O_p(N_{kj}^{-1/2}T^{-1/2+\alpha}) \\
&= o_p(1).
\end{aligned}$$

Thus, $\sum_{j\neq k}\phi_{2,kj} = \sum_{j\neq k}\phi_{21,kj} + \phi_{22,kj} = o_p(1)$.

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

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

$$
\begin{aligned}
\phi_{12,kj} &= \phi_{1,kj} - \phi_{11,kj} \\
&= \left\{ 2(B_k - B_j) \left\{ \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj}}{D_{jj} + D_{kj}} + (\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j) \right\} \right. \\
&\quad \left. + \left\{ \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj}}{D_{jj} + D_{kj}} + (\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j) \right\}^2 \right\} \frac{N_{jj} T^{2\alpha}}{N_s} \frac{D_{jj}}{N_{jj} T} \\
&\quad + \left\{ 2(B_k - B_j) \left\{ \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj}}{D_{jj} + D_{kj}} + (\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j) \right\} \right. \\
&\quad \left. + \left\{ \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj}}{D_{jj} + D_{kj}} + (\hat{B}^*_{jj} - B_j) + (\hat{B}^0_k - B_k) \right\}^2 \right\} \frac{N_{kj} T^{2\alpha}}{N_s} \frac{D_{kj}}{N_{kj} T}.
\end{aligned}
$$

If $(N^{1/2} T^{-1/2+\alpha})/N_s = o(1)$,

$$
\begin{aligned}
&\frac{N_{jj} T^{2\alpha}}{N_s} (B_k - B_j) \left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j) \right] \left(\frac{D_{jj}}{T N_{jj}} \right) \\
&\quad = \frac{N_{jj}}{N_s} \left(T^\alpha (B_k - B_j) \right) T^\alpha \left[\frac{\sqrt{T N_{jj}}}{\sqrt{T N_{jj}}} (\hat{B}^*_{jj} - B_j) + \frac{\sqrt{T(N_{jj} + N_{jk})}}{\sqrt{T(N_{jj} + N_{jk})}} (\hat{B}^0_j - B_j) \right] \left(\frac{D_{jj}}{T N_{jj}} \right) \\
&\quad = \frac{c N_{jj}}{N_s} T^\alpha \left[\frac{1}{\sqrt{T N_{jj}}} O_p(1) + \frac{1}{\sqrt{T(N_{jj} + N_{jk})}} O_p(1) \right] \left(\frac{D_{jj}}{T N_{jj}} \right) \\
&\quad = \frac{c N_{jj}}{N_s} \left[O_p(N_{jj}^{-1/2}) + O_p(N_{jj} + N_{jk})^{-1/2} \right] O_p(T^{-1/2+\alpha}) = o_p(1)
\end{aligned}
$$

Also, we have

$$
\begin{aligned}
&T^\alpha (B_k - B_j) T^\alpha \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj} D_{jj}}{D_{jj} + D_{kj}} \frac{1}{T N_s} \\
&\quad = \frac{c}{N_s T} \frac{D_{kj} D_{jj}}{D_{jj} + D_{kj}} \left(O_p(N_{kj}^{-1/2} T^{-1/2+\alpha}) + O_p(N_{jj}^{-1/2} T^{-1/2+\alpha}) \right) = o_p(1), \\
&T^{2\alpha} \left\{ (\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j) \right\}^2 \frac{D_{jj}}{N_{jj} T} \frac{N_{jj}}{N_s} = O_p(N^{-\tau} T^{-1+2\alpha}) = o_p(1), \\
&T^{2\alpha} \left\{ \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj}}{D_{jj} + D_{kj}} \right\}^2 \frac{D_{jj}}{N_s T} = [O_p(N_{jk}^{-1}) + O_p(N_{jj}^{-1})] O_p(T^{-1+2\alpha}) = o_p(1), \\
&2T^{2\alpha} \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_j - B_j) \right] \frac{D_{kj} D_{jj}}{D_{jj} + D_{kj}} \frac{1}{T N_s} \\
&\quad = (O_p(N_{kj}^{-1/2}) + O_p(N_{jj}^{-1/2})) O_p(N_{jj}^{-1/2} T^{-1+2\alpha}) = o_p(1)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&T^\alpha (B_k - B_j) T^\alpha \left[(\hat{B}^*_{jj} - B_j) + (\hat{B}^0_k - B_k) \right] \frac{D_{kj}}{T N_s} \\
&\qquad = c \left[O_p(N_{jj}^{-1/2}) + O_p((N_{kk}, N_{kj})^{-1/2}) \right] O_p(T^{-1/2+\alpha}) = o_p(1), \\
&T^\alpha (B_k - B_j) T^\alpha \left[(\hat{B}^*_{kj} - B_k) - (\hat{B}^*_{jj} - B_j) \right] \frac{D_{kj}^2}{D_{jj} + D_{kj}} \frac{1}{T N_s} \\
&\quad = c \left(O_p(N_{kj}^{-1/2} T^{-1/2+\alpha}) + O_p(N_{jj}^{-1/2} T^{-1/2+\alpha}) \right) \frac{D_{kj} D_{kj}}{D_{jj} + D_{kj}} \frac{1}{N_s T} = o_p(1),
\end{aligned}
$$

$$T^{2\alpha}\left\{(\hat{B}^*_{jj}-B_j)+(\hat{B}^0_j-B_j)\right\}^2\frac{D_{kj}}{N_sT}=O_p(N^{-\tau}T^{-1+2\alpha})=o_p(1),$$

$$T^{2\alpha}\left\{\left[(\hat{B}^*_{kj}-B_k)-(\hat{B}^*_{jj}-B_j)\right]\frac{D_{kj}}{D_{jj}+D_{kj}}\right\}^2\frac{D_{kj}}{N_sT}=[O_p(N_{jk}^{-1})+O_p(N_{jj}^{-1})]O_p(T^{-1+2\alpha})=o_p(1),$$

$$\begin{aligned}&2T^{2\alpha}\left[(\hat{B}^*_{kj}-B_k)-(\hat{B}^*_{jj}-B_j)\right]\left[(\hat{B}^*_{jj}-B_j)+(\hat{B}^0_k-B_k)\right]\frac{D_{kj}D_{jj}}{D_{jj}+D_{kj}}\frac{1}{TN_s}\\&\quad=\left[O_p(N_{kk}^{-1/2})+O_p(N_{jk}^{-1/2})\right]\left[O_p(N_{jj}^{-1/2})+O_p((N_{kk}+N_{jk})^{-1/2})\right]O_p(T^{-1+2\alpha})=o_p(1)\end{aligned}$$

Together with the results above, we obtain $\phi_{12,kj}=o_p(1)$. Therefore,

$$\begin{aligned}\sum_{\substack{j,k=1,2,\\ j\neq k}}\phi_{1,kj}&=\sum_{\substack{j,k=1,2,\\ j\neq k}}\phi_{11,kj}+o_p(1)\\&=T^{2\alpha}\left\{\left[(B_2-B_1)\frac{D_{21}}{D_{11}+D_{21}}\right]^2\frac{D_{11}}{N_sT}+\left[(B_2-B_1)\frac{D_{11}}{D_{11}+D_{21}}\right]^2\frac{D_{21}}{N_sT}\right.\\&\qquad\left.+\left[(B_2-B_1)\frac{D_{22}}{D_{22}+D_{12}}\right]^2\frac{D_{12}}{N_sT}+\left[(B_2-B_1)\frac{D_{12}}{D_{22}+D_{12}}\right]^2\frac{D_{22}}{N_sT}\right\}+o_p(1)\\&=c^2\left[\frac{D_{11}D_{21}}{D_{11}+D_{21}}+\frac{D_{22}D_{12}}{D_{22}+D_{12}}\right]\frac{1}{N_sT}+o_p(1)\\&=M^*+o_p(1),\end{aligned}$$

where M^* is a strictly positive number. Thus, for any $(I_1,I_2)\neq(I_1^0,I_2^0)$, $\text{Prob}(\sum_{j\neq k}\phi_{1,kj}+$ $+\phi_{2,kj}>0)\xrightarrow{p}1$ and thus

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

if $(N^{1/2}T^{-1/2+\alpha})/N_s=o(1)$. The result also holds when $N_{12}>N_{21}=0$ or $N_{21}>N_{12}=0$.

Recall that $\triangle=1/2-\alpha$. By the proposed method, the number of misspecified individuals, N_s, asymptotically is bounded by $N^{1/2}T^{-1/2+\alpha}=\frac{\sqrt{N}}{T^{\triangle}}$. 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 the data are generated by (2) and Assumption A holds. Then,

$$\hat{\tau}_i=\tau_i+O_p(T^{-1/2}).$$

Proof of Lemma A.1: Under the assumed DGP and and Assumption A, we can write

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

For $i \in I_1$, $\hat{q}_i = \hat{\tau}_i$ can expressed as

$$\begin{aligned}
\hat{\tau}_i &= \frac{\sqrt{T}(\hat{\beta}_i - \hat{B}_\omega)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} = \frac{\sqrt{T}[\hat{\beta}_i - \omega B_1 - (1-\omega)B_2]}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} + O_p(N^{-1/2}) \\
&= \frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} - (1-\omega)\frac{\sqrt{T}(B_2 - B_1)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}} + O_p(N^{-1/2}) \\
&= \frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\sigma_i \hat{Q}_i^{-1/2}} - (1-\omega)\frac{\sqrt{T}(B_2 - B_1)}{\sigma_i \hat{Q}_i^{-1/2}} + O_p(N^{-1/2}) + O_p(T^{-1}).
\end{aligned}$$

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

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

By assumption, $\hat{Q}_i \xrightarrow{p} Q_i$ and the result follows.

□.

Lemma A.2 Suppose the data are generated by (2) and Assumption A holds. Also, suppose that $B_2 - B_1 = cT^{-\alpha}$, where $\alpha = 1/2 - \triangle$, and c does not depend on T. Then $\alpha = 0$ corresponds to the case when $B_2 - B_1 = c \neq 0$. Let $\gamma^* \in (B_1, B_2)$ be such that $\gamma^* - B_1 = c_1 T^{-\alpha}$, and let $B_2 - \gamma^* = c_2 T^{-\alpha}$, $c_1, c_2 > 0$, $c_1 + c_2 = c$. As $T \to \infty$

$$\text{(i)} \quad \begin{cases} \text{Prob}\left(T^\alpha(\hat{\beta}_i) \leq T^\alpha(\gamma^*)\right) \xrightarrow{p} 1, \forall i \in I_1, \\ \text{Prob}\left(T^\alpha(\hat{\beta}_i) > T^\alpha(\gamma^*)\right) \xrightarrow{p} 1, \forall i \in I_2. \end{cases}$$

$$\text{(ii)} \quad \begin{cases} \text{Prob}\left(T^{-1/2+\alpha}\hat{\tau}_i \leq T^\alpha \gamma^{*\prime}\right) \xrightarrow{p} 1, \forall i \in I_1, \\ \text{Prob}\left(T^{-1/2+\alpha}\hat{\tau}_i > T^\alpha \gamma^{*\prime}\right) \xrightarrow{p} 1, \forall i \in I_2. \end{cases}$$

where $\gamma^{*\prime} = \frac{[(\omega-1)(B_2-B_1)+(\gamma^*-B_1)]}{\sigma_i^2 Q_i^{-1/2}}$.

Proof of Lemma A.2:
Notice that

$$\text{Prob}\left(T^\alpha(\hat{\beta}_i) \leq T^\alpha(\gamma^*)\right) = \text{Prob}\left(T^\alpha(\hat{\beta}_i - B_1) \leq T^\alpha(\gamma^* - B_1)\right).$$

Thus, the first part of (i) follows from the fact that $T^\alpha(\hat{\beta}_i - B_1) \xrightarrow{p} 0$, and $T^\alpha(\gamma^* - B_1) \xrightarrow{p} c_1 > 0$, as $T \to \infty$. The second part (i) follows analogously.
To prove (ii), note that from (7), for $i \in I_1$,

$$T^{-1/2+\alpha}\hat{\tau}_i = \frac{T^\alpha(1-\omega)(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} + \frac{T^\alpha(\hat{\beta}_i - B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1) = \frac{T^\alpha(1-\omega)(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} + o_P(1),$$

and the first part of (ii) follows from

$$\frac{T^\alpha[(\omega-1)(B_2 - B_1) + (\gamma^* - B_1)]}{\sigma_i^2 Q_i^{-1/2}} - \frac{T^\alpha(\omega-1)(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} = \frac{c_1}{\sigma_i^2 Q_i^{-1/2}} > 0.$$

The second part of (ii) follows analogously. $\square$.

Remark: Lemma $A.2$ indicates that there exists a $\gamma^* \in (B_1, B_2)$ such that the probability of misclassification based on the proposed method can be asymptotically zero as $T \to \infty$. We will further discuss the rate of misclassification in the following.

By Assumption A, it is well-known that $\sqrt{T}(\hat{\beta}_i - B_1)$ is $O_p(1)$ for $i \in I_1$. Also, suppose that $B_2 - B_1 = cT^{-\alpha}$, where $\alpha = 1/2 - \Delta$, and c does not depend on T. Then, we can find a $\gamma^* \in (B_1, B_2)$ such that $\gamma^* - B_1 = c_1 T^{-\alpha}$, and $B_2 - \gamma^* = c_2 T^{-\alpha}$, $c_1, c_2 > 0$, $c_1 + c_2 = c$.

Given that $B_2 > B_1$, the probability of misclassification for the the individuals in Group 1 is equivalent to the probability that $\hat{\beta}_i > \gamma^*$. Also, Prob$(\hat{\beta}_i > \gamma^*)$ for $i \in I_1$ is equivalent to

$$\text{Prob}\left(\frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\sigma_{\hat{\beta}_i}} > \frac{\sqrt{T}(\gamma^* - B_1)}{\sigma_{\hat{\beta}_i}}\right) \quad = \quad \text{Prob}\left(\frac{\sqrt{T}(\hat{\beta}_i - B_1)}{\sigma_{\hat{\beta}_i}} > c_{i,1} T^{1/2-\alpha}\right)$$

where $\sigma_{\hat{\beta}_i} = \sigma_i Q_i^{-1/2}$, and $c_{i,1} = c_1/\sigma_{\hat{\beta}_i}$, $0 < c_{i,1} < \infty$ by assumption. By a central limiting theorem, we have $(\hat{\beta}_i - B_1) = O_p(1/\sqrt{T})$ and Prob$(\hat{\beta}_i > \gamma^*) = O_p(T^{-(1/2-\alpha)}) = O_p(T^{-\Delta})$. Therefore, for the individuals truly in Group 1, the rate of misclassification is $N_s/N = O_p(T^{-(1/2-\alpha)}) = O_p(T^{-\Delta})$. In particular, when $\alpha = 0$, then $B_2 - B_1 = c \neq 0$ and $N_s/N = O_p(T^{-1/2})$. The analogous logic holds for the individuals truly in Group 2.

Table 1: Homoskedastic Errors

		RMSE(B_1)				RMSE (B_2)			
	N\T	20	50	100	200	20	50	100	200
PS_1	20	0.505	0.123	0.044	0.023	0.244	0.110	0.058	0.034
	50	0.221	0.097	0.032	0.016	0.290	0.137	0.088	0.044
	100	0.134	0.051	0.025	0.015	0.061	0.028	0.018	0.011
	200	0.089	0.033	0.015	0.008	0.060	0.025	0.015	0.009
PS_2	20	0.573	0.101	0.118	0.023	0.290	0.109	0.082	0.034
	50	0.368	0.088	0.032	0.016	0.310	0.140	0.088	0.045
	100	0.136	0.052	0.026	0.015	0.068	0.029	0.018	0.011
	200	0.089	0.033	0.015	0.008	0.062	0.025	0.015	0.009
K_1	20	0.163	0.067	0.036	0.023	0.167	0.083	0.055	0.034
	50	0.169	0.076	0.041	0.019	0.151	0.105	0.088	0.054
	100	0.113	0.050	0.025	0.015	0.059	0.028	0.018	0.011
	200	0.090	0.034	0.015	0.008	0.052	0.024	0.015	0.009
PS_1^H	20	0.502	0.117	0.043	0.023	0.229	0.105	0.057	0.034
	50	0.209	0.084	0.034	0.017	0.244	0.126	0.087	0.047
	100	0.116	0.050	0.025	0.015	0.060	0.028	0.018	0.011
	200	0.090	0.034	0.016	0.008	0.053	0.023	0.015	0.010
PS_2^H	20	0.623	0.098	0.119	0.023	0.292	0.106	0.081	0.034
	50	0.360	0.085	0.035	0.017	0.264	0.127	0.087	0.047
	100	0.119	0.051	0.025	0.015	0.063	0.029	0.018	0.011
	200	0.090	0.034	0.016	0.008	0.055	0.023	0.015	0.010
K_1^H	20	0.164	0.068	0.036	0.023	0.180	0.086	0.054	0.034
	50	0.172	0.078	0.041	0.018	0.169	0.106	0.088	0.054
	100	0.115	0.050	0.025	0.015	0.062	0.028	0.018	0.011
	200	0.090	0.035	0.015	0.008	0.054	0.024	0.015	0.009
		CR (%)				Computation Time (sec/rep)			
	N\T	20	50	100	200	20	50	100	200
PS_1	20	64.380	73.670	80.500	86.550	0.005	0.006	0.007	0.009
	50	62.760	70.420	77.158	83.388	0.014	0.017	0.023	0.042
	100	63.731	71.150	77.088	82.532	0.032	0.046	0.082	0.202
	200	66.287	74.688	80.805	86.003	0.082	0.179	0.586	1.370
PS_2	20	64.055	73.690	80.405	86.525	0.005	0.006	0.007	0.009
	50	62.706	70.410	77.122	83.368	0.014	0.017	0.023	0.041
	100	63.708	71.086	77.062	82.529	0.032	0.045	0.081	0.204
	200	66.305	74.686	80.807	85.997	0.082	0.178	0.601	1.369
K_1	20	65.555	74.050	80.440	86.550	0.041	0.048	0.055	0.068
	50	61.054	68.434	75.054	82.516	0.078	0.124	0.170	0.268
	100	64.186	71.254	77.129	82.537	0.114	0.201	0.321	0.755
	200	66.133	74.496	80.748	85.984	0.165	0.362	1.192	3.402
PS_1^H	20	64.955	74.360	80.980	86.845	0.005	0.006	0.007	0.010
	50	62.792	70.334	77.142	83.746	0.014	0.017	0.023	0.042
	100	63.514	70.884	76.984	82.493	0.032	0.045	0.081	0.202
	200	66.415	74.988	81.329	86.413	0.082	0.179	0.587	1.370
PS_2^H	20	64.445	74.280	80.865	86.855	0.006	0.007	0.008	0.010
	50	62.642	70.276	77.168	83.726	0.015	0.018	0.024	0.044
	100	63.471	70.796	76.959	82.495	0.033	0.047	0.083	0.211
	200	66.431	75.019	81.314	86.427	0.083	0.181	0.611	1.393
K_1^H	20	65.575	73.965	80.545	86.490	0.042	0.049	0.056	0.069
	50	61.250	68.356	75.010	82.510	0.079	0.126	0.173	0.271
	100	64.197	71.219	77.091	82.541	0.118	0.206	0.326	0.763
	200	66.135	74.483	80.755	85.974	0.174	0.372	1.206	3.426

Table 2: Heteroskedastic Errors

		RMSE(B_1)				RMSE(B_2)			
	N\T	20	50	100	200	20	50	100	200
PS_1	20	0.636	0.257	0.107	0.020	0.476	0.146	0.059	0.030
	50	0.616	0.165	0.029	0.014	0.459	0.203	0.105	0.048
	100	0.245	0.062	0.033	0.019	0.162	0.039	0.025	0.016
	200	0.074	0.031	0.017	0.009	0.097	0.037	0.020	0.012
PS_2	20	0.161	0.014	0.010	0.007	0.088	0.032	0.022	0.015
	50	0.185	0.013	0.007	0.005	0.246	0.080	0.045	0.028
	100	0.032	0.016	0.010	0.007	0.016	0.009	0.006	0.004
	200	0.007	0.004	0.003	0.002	0.028	0.013	0.008	0.005
K_1	20	0.185	0.065	0.029	0.020	0.157	0.066	0.044	0.031
	50	0.180	0.063	0.028	0.015	0.254	0.126	0.085	0.052
	100	0.127	0.057	0.033	0.019	0.084	0.038	0.025	0.016
	200	0.071	0.033	0.018	0.009	0.090	0.036	0.020	0.012
PS_1^H	20	0.450	0.114	0.050	0.020	0.178	0.067	0.044	0.030
	50	0.430	0.133	0.030	0.016	0.284	0.130	0.088	0.056
	100	0.105	0.053	0.031	0.018	0.089	0.039	0.025	0.016
	200	0.083	0.037	0.019	0.010	0.064	0.028	0.019	0.012
PS_2^H	20	0.162	0.015	0.010	0.007	0.088	0.033	0.022	0.015
	50	0.147	0.014	0.007	0.005	0.216	0.088	0.054	0.030
	100	0.030	0.016	0.010	0.007	0.017	0.009	0.006	0.004
	200	0.008	0.004	0.003	0.002	0.025	0.013	0.008	0.005
K_1^H	20	0.039	0.014	0.010	0.007	0.061	0.032	0.022	0.015
	50	0.062	0.020	0.008	0.005	0.179	0.109	0.067	0.030
	100	0.031	0.016	0.010	0.007	0.016	0.009	0.006	0.004
	200	0.007	0.004	0.003	0.002	0.028	0.013	0.008	0.005
		CR (%)				Computation Time (sec/rep)			
	N\T	20	50	100	200	20	50	100	200
PS_1	20	66.180	77.255	84.375	89.875	0.006	0.006	0.007	0.009
	50	65.874	75.080	81.358	86.564	0.014	0.017	0.023	0.041
	100	66.311	74.016	78.734	83.290	0.032	0.045	0.082	0.259
	200	70.222	76.960	81.927	86.355	0.082	0.167	0.522	1.220
PS_2	20	71.035	79.110	84.880	89.960	0.006	0.006	0.007	0.009
	50	69.174	76.840	81.800	86.530	0.014	0.017	0.023	0.041
	100	68.809	74.560	79.068	83.488	0.032	0.045	0.081	0.260
	200	70.352	77.076	81.987	86.387	0.082	0.167	0.522	1.223
K_1	20	69.115	78.360	84.580	89.850	0.041	0.046	0.052	0.065
	50	66.148	74.404	80.698	86.176	0.071	0.107	0.143	0.220
	100	67.924	74.175	78.773	83.304	0.109	0.183	0.284	0.946
	200	70.181	76.899	81.884	86.329	0.158	0.338	1.214	3.000
PS_1^H	20	68.985	78.970	85.200	90.455	0.006	0.006	0.007	0.009
	50	65.402	73.234	79.724	84.990	0.014	0.017	0.023	0.041
	100	67.820	73.425	77.912	82.467	0.032	0.045	0.082	0.259
	200	69.670	77.116	82.680	87.310	0.082	0.167	0.522	1.221
PS_2^H	20	71.930	80.230	85.955	91.210	0.006	0.007	0.008	0.010
	50	68.148	75.574	80.620	85.270	0.015	0.018	0.024	0.044
	100	67.259	73.282	77.903	82.411	0.033	0.047	0.084	0.268
	200	72.665	79.322	83.960	87.978	0.083	0.169	0.531	1.243
K_1^H	20	70.965	79.045	84.850	89.955	0.040	0.045	0.051	0.063
	50	66.032	74.568	80.936	86.438	0.090	0.115	0.143	0.210
	100	68.861	74.568	79.070	83.492	0.128	0.193	0.288	0.932
	200	70.331	77.069	81.987	86.386	0.198	0.370	1.246	2.994

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

Homoskedastic Errors									
		RMSE				CR(%)			
	N\T	20	50	100	200	20	50	100	200
PS_1	20	2.352	1.501	1.241	0.826	55.795	60.540	64.620	69.960
	50	2.624	0.290	0.365	0.065	56.074	60.648	66.216	72.670
	100	0.315	0.133	0.078	0.054	55.848	60.571	66.014	72.783
	200	0.140	0.077	0.048	0.031	57.438	62.945	68.867	74.823
PS_2	20	1.947	1.427	0.956	0.418	55.925	61.045	65.695	72.835
	50	0.684	0.293	0.127	0.062	56.730	61.970	68.298	75.716
	100	0.247	0.145	0.079	0.050	56.414	61.728	67.739	75.218
	200	0.162	0.089	0.053	0.034	57.392	63.241	69.968	76.324
K_1	20	0.476	0.232	0.129	0.076	58.835	64.655	71.950	80.285
	50	0.299	0.157	0.088	0.046	59.290	65.868	73.432	80.964
	100	0.256	0.124	0.066	0.033	60.187	67.773	75.086	82.591
	200	0.197	0.086	0.042	0.021	60.336	68.146	75.534	82.024
PS_1^H	20	2.362	1.423	1.248	0.784	55.540	60.540	64.680	70.190
	50	2.832	0.352	0.145	0.066	55.782	60.498	66.134	72.624
	100	0.371	0.141	0.079	0.061	55.816	60.553	66.002	72.754
	200	0.146	0.078	0.048	0.031	57.424	62.928	68.846	74.827
PS_2^H	20	1.979	1.355	0.954	0.407	55.665	61.000	65.380	72.770
	50	0.720	0.324	0.131	0.062	56.646	61.486	68.292	75.736
	100	0.274	0.148	0.080	0.050	56.116	61.520	67.646	75.258
	200	0.178	0.089	0.053	0.034	57.324	63.312	69.994	76.345
K_1^H	20	0.528	0.250	0.138	0.075	58.270	64.550	71.250	80.240
	50	0.317	0.160	0.088	0.046	59.292	66.004	73.430	80.922
	100	0.264	0.126	0.065	0.034	60.044	67.561	75.201	82.585
	200	0.202	0.087	0.042	0.021	60.233	68.116	75.515	82.020
Heteroskedastic Errors									
		RMSE				CR(%)			
	N\T	20	50	100	200	20	50	100	200
PS_1	20	2.864	2.010	1.416	0.987	55.065	57.495	60.925	64.290
	50	1.598	0.836	0.226	0.106	56.304	60.052	64.758	70.412
	100	0.249	0.118	0.065	0.047	56.660	62.124	67.649	73.916
	200	0.163	0.091	0.063	0.044	57.096	62.489	68.066	74.850
PS_2	20	2.080	1.507	0.930	0.448	55.725	58.650	64.110	70.105
	50	0.588	0.288	0.132	0.073	56.346	60.638	66.012	72.740
	100	0.219	0.106	0.062	0.042	57.514	63.424	69.985	76.934
	200	0.177	0.100	0.063	0.040	57.328	63.003	69.460	76.587
K_1	20	0.378	0.212	0.116	0.072	58.640	64.630	71.010	78.910
	50	0.314	0.167	0.096	0.052	58.826	65.514	72.928	80.052
	100	0.212	0.097	0.048	0.026	61.616	69.813	77.144	83.501
	200	0.201	0.087	0.041	0.022	61.107	69.664	76.832	83.156
PS_1^H	20	2.753	1.802	1.218	0.858	55.280	58.365	61.945	66.005
	50	1.524	0.868	0.246	0.140	55.752	59.618	64.634	69.982
	100	0.323	0.148	0.073	0.047	56.632	62.117	67.499	73.782
	200	0.163	0.092	0.065	0.046	56.923	62.289	68.040	74.785
PS_2^H	20	2.324	1.386	0.864	0.335	55.990	60.390	65.500	72.290
	50	0.626	0.363	0.130	0.072	56.006	60.598	66.170	73.270
	100	0.237	0.100	0.060	0.041	57.756	64.045	70.127	76.929
	200	0.185	0.098	0.063	0.041	57.187	62.944	69.559	76.692
K_1^H	20	0.378	0.194	0.114	0.069	58.380	64.660	71.180	79.465
	50	0.313	0.163	0.093	0.050	58.722	65.612	72.964	80.130
	100	0.192	0.085	0.042	0.024	62.029	70.325	77.314	83.571
	200	0.195	0.081	0.038	0.020	61.069	69.639	76.746	83.156

Table 4: Testing Parameter Homogeneity

	Homoskedastic Errors, $K = 1$							
	Size:				Power:			
	$B_1 = B_2 = 0.8$				$B_1 = 0.7,\ B_2 = 1$			
N\T	20	50	100	200	20	50	100	200
20	2.2	2.6	4.1	4.1	5.7	29.7	76.2	99.9
50	3.0	3.3	3.4	4.9	7.8	42.5	86.7	99.9
100	3.6	4.7	4.2	5.7	35.4	99.2	100.0	100.0
200	4.7	3.7	4.8	5.3	48.9	100.0	100.0	100.0
	Heteroskedastic Errors, $K = 1$							
	Size:				Power:			
	$B_1 = B_2 = 0.8$				$B_1 = 0.7,\ B_2 = 1$			
N\T	20	50	100	200	20	50	100	200
20	2.2	2.7	4.0	4.4	4.5	17.8	54.8	96.5
50	2.9	3.2	3.3	4.8	7.8	36.7	81.1	99.9
100	3.8	4.6	4.3	5.8	17.2	90.9	100.0	100.0
200	4.7	3.8	4.9	5.5	26.7	99.4	100.0	100.0
	Homoskedastic Errors, $K = 2$							
	Size:				Power:			
	$B_1 = B_2 = (0.8, 3.2/3)$				$B_1 = (0.7, 1),\ B_2 = (1, 1.1)$			
N\T	20	50	100	200	20	50	100	200
20	4.2	3.2	4.0	3.9	6.9	3.6	7.5	23.8
50	7.9	4.7	5.3	3.5	3.3	3.3	9.8	32.8
100	15.8	4.0	5.4	4.9	8.7	8.5	47.1	98.0
200	30.3	7.3	5.0	4.0	13.6	11.9	78.8	100.0
	Heteroskedastic Errors, $K = 2$							
	Size:				Power:			
	$B_1 = B_2 = (0.8, 3.2/3)$				$B_1 = (0.7, 1),\ B_2 = (1, 1.1)$			
N\T	20	50	100	200	20	50	100	200
20	4.2	3.3	4.5	4.1	6.9	3.9	10.5	35.2
50	8.6	4.6	4.8	3.9	2.7	4.8	20.0	62.4
100	16.0	4.1	5.2	4.9	6.8	18.7	81.0	100.0
200	30.3	7.4	4.9	3.9	10.9	25.3	96.5	100.0

Table 5: Heterogeneity in $\hat{g}_i$ estimated from (15)

Cluster based on level equation (15)							
All	PY	2.940					
		PS_1	PS_2	PS_1^H	PS_2^H	K_1	K_1^H
Low	N_1	30	41	30	28	30	30
	$\hat{g}_1(\%)$	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	—	—
High	N_2	40	29	40	42	40	40
	$\hat{g}_2(\%)$	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, $\overline{\Delta y_{it}}$							
All	PY	4.649					
		PS_1	PS_2	PS_1^H	PS_2^H	K_1	K_1^H
Low	N_1	38	43	38	37	38	45
	$\hat{g}_1(\%)$	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	—	—
High	N_2	32	27	32	33	32	25
	$\hat{g}_2(\%)$	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

1. Model:

$$y_{it} = \mu_i' + g_i t + \varepsilon_{it},$$
$$\varepsilon_{it} = \gamma_i \varepsilon_{i,t-1} + \eta_t + e_{it},$$

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

2. Data: Penn World Tables (PWT v6.2).

3. Estimation: see Lee, Pesaran, and Smith (1997) for details.

4. PY denotes the PY statistic in (9) for parameter homogeneity in whole sample. PY_1 and PY_2 are, respectively, for the estimated low-growth and high-growth groups. The asymptotic 5% critical value is 1.96.

5. In this case, the PY statistic can be expressed as $PY = \frac{\sqrt{N}(\sum_i^N s_i^2/N-1)}{\sqrt{2}}$, where

$$s_i = \sqrt{T}(\hat{g}_i - \hat{g}_w)/\tilde{\sigma}(\hat{g}_i),$$

$\hat{g}_w$ denotes the weighted fixed effect estimator of g_0 under the null that $g_i = g_0$ for all individuals, and $\tilde{\sigma}(\hat{g}_i)$ is the restricted long-run variance of $\hat{g}_i$ based on the automatic method developed by Andrews (1991) with Quadratic Spectral kernel.

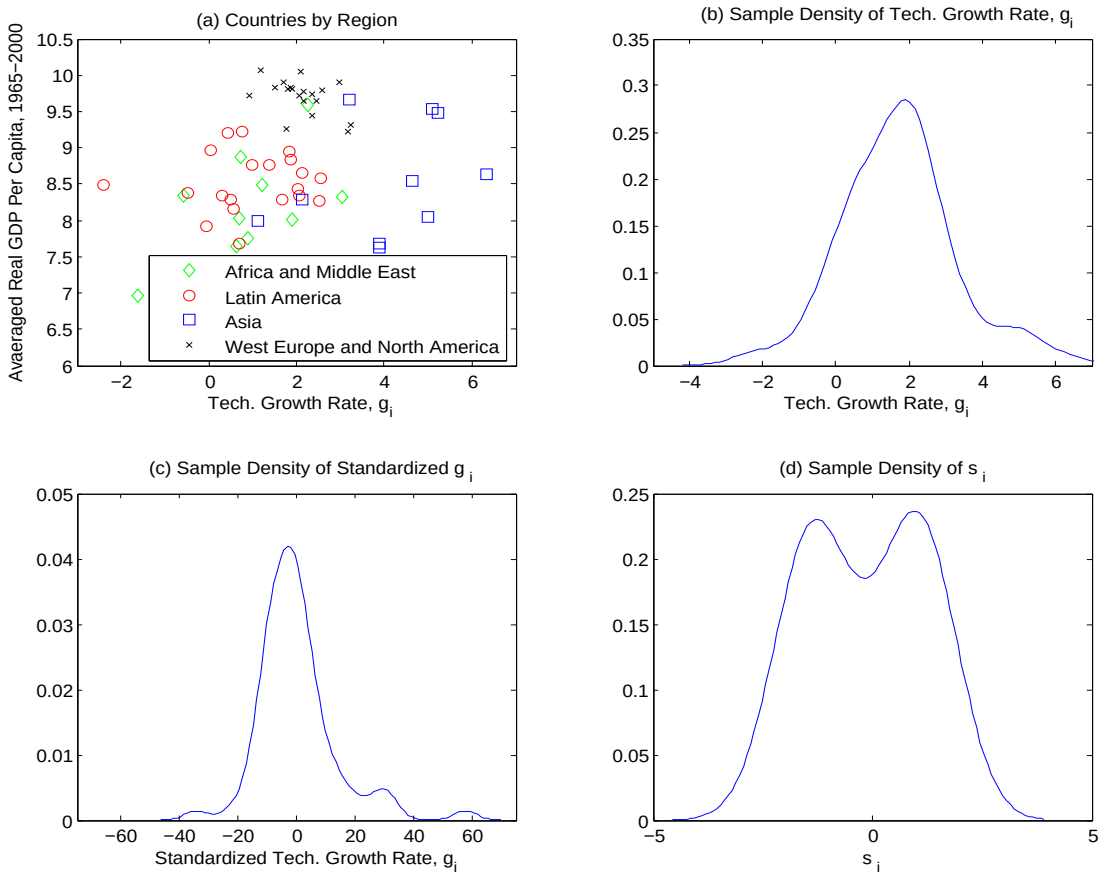
Table 6: Clusters Based on the Level Equation

COUNTRY	Ave. $\ln(y_t)$	Region	$\hat{g}_i$	Std. $\hat{g}_i$	s_i	PS_1	PS_2	PS_1^H	PS_2^H	K_1	K_1^H
Nicaragua	8.486	2	-2.383	-17.251	-1.228	1	1	1	1	1	1
Madagascar	6.964	1	-1.620	-34.241	-1.242	1	1	1	1	1	1
Zambia	7.024	1	-0.789	-6.961	-1.245	1	1	1	1	1	1
Jordan	8.335	1	-0.561	-7.994	-1.479	1	1	1	1	1	1
Peru	8.382	2	-0.477	-13.230	-1.238	1	1	1	1	1	1
Senegal	7.284	1	-0.259	-10.825	-1.531	1	1	1	1	1	1
Bolivia	7.926	2	-0.067	-13.169	-1.413	1	1	1	1	1	1
Venezuela	8.969	2	0.059	-13.045	-1.268	1	1	1	1	1	1
Nigeria	6.940	1	0.097	-9.679	-1.606	1	1	1	1	1	1
Tanzania	6.335	1	0.132	-4.887	-1.502	1	1	1	1	1	1
Kenya	7.128	1	0.163	-8.602	-1.375	1	1	1	1	1	1
Jamaica	8.351	2	0.290	-6.452	-1.311	1	1	1	1	1	1
Argentina	9.205	2	0.422	-8.034	-1.367	1	1	1	1	1	1
El Salvador	8.287	2	0.500	-8.617	-1.361	1	1	1	1	1	1
Guatemala	8.156	2	0.571	-12.041	-1.276	1	1	1	1	1	1
Cote d'Ivoire	7.640	1	0.617	-6.742	-1.291	1	1	1	1	1	1
Zimbabwe	8.022	1	0.689	-5.031	-1.344	1	1	1	1	1	1
Honduras	7.684	2	0.706	-9.239	-1.331	1	1	1	1	1	1
South Africa	8.872	1	0.742	-22.029	-1.323	1	1	1	1	1	1
Trinidad-Tobago	9.226	2	0.754	-4.915	-1.466	1	1	1	1	1	1
Cameroon	7.759	1	0.895	-3.582	-1.301	1	1	1	2	1	1
Ethiopia	6.235	1	0.904	-3.254	-1.395	1	1	1	2	1	1
New Zealand	9.713	4	0.941	-8.099	-1.635	1	1	1	1	1	1
Costa Rica	8.763	2	0.985	-11.049	-1.455	1	1	1	1	1	1
Philippines	8.001	3	1.125	-9.910	-1.701	1	1	1	1	1	1
Switzerland	10.076	4	1.195	-11.404	-1.431	1	1	1	1	1	1
Algeria	8.496	1	1.222	-5.199	-1.496	1	1	1	1	1	1
Mexico	8.758	2	1.366	-4.752	-1.372	1	1	1	1	1	1
Malawi	6.495	1	1.392	-4.492	-2.111	1	1	1	1	1	1
Sweden	9.829	4	1.522	-7.972	-1.819	1	1	1	1	1	1
Mali	6.669	1	1.592	-2.570	-1.181	2	1	2	2	2	2
Ecuador	8.288	2	1.679	-0.782	-0.284	2	1	2	2	2	2
Denmark	9.899	4	1.703	-3.006	-0.881	2	1	2	2	2	2
Syria	7.375	1	1.750	-0.800	-0.141	2	1	2	2	2	2
Greece	9.268	4	1.778	-0.802	-0.056	2	1	2	2	2	2
Netherlands	9.813	4	1.809	-2.356	0.151	2	1	2	2	2	2
Uruguay	8.949	2	1.826	-0.788	0.155	2	1	2	2	2	2
Chile	8.846	2	1.869	-0.260	0.154	2	1	2	2	2	2
Canada	9.834	4	1.872	-3.609	0.587	2	1	2	2	2	2
Australia	9.805	4	1.890	-1.021	0.573	2	1	2	2	2	2
Morocco	8.009	1	1.901	-0.597	0.656	2	1	2	2	2	2
Colombia	8.438	2	2.021	0.759	0.982	2	2	2	2	2	2
United Kingdom	9.717	4	2.053	0.443	0.837	2	2	2	2	2	2
Paraguay	8.344	2	2.081	0.402	0.731	2	2	2	2	2	2
USA	10.043	4	2.107	0.925	0.902	2	2	2	2	2	2
Turkey	8.289	3	2.138	1.741	1.202	2	2	2	2	2	2
Brazil	8.653	2	2.141	0.491	0.788	2	2	2	2	2	2
France	9.769	4	2.148	2.386	1.267	2	2	2	2	2	2
Finland	9.645	4	2.158	1.328	1.153	2	2	2	2	2	2
Israel	9.597	1	2.250	1.957	1.257	2	2	2	2	2	2
Belgium	9.729	4	2.348	7.504	1.213	2	2	2	2	2	2
Spain	9.436	4	2.368	4.478	1.211	2	2	2	2	2	2
Italy	9.652	4	2.464	5.350	1.388	2	2	2	2	2	2
Dominican, Rep	8.263	2	2.536	5.942	1.115	2	2	2	2	2	2
Panama	8.589	2	2.550	4.391	1.279	2	2	2	2	2	2
Austria	9.787	4	2.583	9.123	1.245	2	2	2	2	2	2
India	7.320	3	2.689	3.001	1.069	2	2	2	2	2	2
Pakistan	7.431	3	2.828	15.057	1.140	2	2	2	2	2	2
Norway	9.897	4	2.992	16.481	1.134	2	2	2	2	2	2
Tunisia	8.332	1	3.047	11.535	1.149	2	2	2	2	2	2
Portugal	9.231	4	3.170	9.549	1.351	2	2	2	2	2	2
Japan	9.670	3	3.190	5.614	1.452	2	2	2	2	2	2
Ireland	9.324	4	3.232	4.815	1.152	2	2	2	2	2	2
Indonesia	7.679	3	3.891	57.693	1.170	2	2	2	2	2	2
Sri Lanka	7.619	3	3.903	12.353	1.161	2	2	2	2	2	2
Malaysia	8.548	3	4.631	31.532	1.186	2	2	2	2	2	2
Thailand	8.056	3	4.994	25.725	1.173	2	2	2	2	2	2
Hong Kong	9.540	3	5.108	19.826	1.170	2	2	2	2	2	2
Singapore	9.479	3	5.209	31.366	1.190	2	2	2	2	2	2
Korea, Rep of	8.633	3	6.320	28.621	1.189	2	2	2	2	2	2

Table 7: Clusters Based On the Growth Rate Equation

COUNTRY	Ave. $\ln(y_t)$	Region	$\hat{g}_i$	Std. $\hat{g}_i$	s_i	PS_1	PS_2	PS_1^H	PS_2^H	K_1	K_1^H
Nicaragua	8.486	2	-1.291	-3.727	-2.209	1	1	1	1	1	1
Madagascar	6.964	1	-1.200	-6.573	-2.704	1	1	1	1	1	1
Jordan	8.335	1	-0.619	-1.720	-1.432	1	1	1	1	1	1
Zambia	7.024	1	-0.168	-2.682	-2.185	1	1	1	1	1	1
Senegal	7.284	1	-0.110	-2.884	-2.466	1	1	1	1	1	1
Nigeria	6.940	1	-0.066	-1.605	-1.473	1	1	1	1	1	1
Kenya	7.128	1	0.100	-3.237	-2.677	1	1	1	1	1	1
Venezuela	8.969	2	0.165	-2.505	-2.098	1	1	1	1	1	1
Bolivia	7.926	2	0.314	-3.002	-2.233	1	1	1	1	1	1
Peru	8.382	2	0.338	-1.851	-1.622	1	1	1	1	1	1
Jamaica	8.351	2	0.421	-2.082	-1.735	1	1	1	1	1	1
Cameroon	7.759	1	0.611	-0.894	-0.817	1	1	1	1	1	1
Honduras	7.684	2	0.774	-2.514	-2.228	1	1	1	1	1	1
Cote d'Ivoire	7.640	1	0.784	-1.262	-1.177	1	1	1	1	1	1
Argentina	9.205	2	0.957	-1.710	-1.574	1	1	1	1	1	1
El Salvador	8.287	2	0.960	-2.007	-1.497	1	1	1	1	1	1
Tanzania	6.335	1	1.021	-0.771	-0.720	1	1	1	1	1	1
Guatemala	8.156	2	1.032	-3.478	-2.378	1	1	1	1	1	1
Zimbabwe	8.022	1	1.048	-0.826	-0.784	1	1	1	1	1	1
South Africa	8.872	1	1.113	-3.338	-2.428	1	1	1	1	1	1
Algeria	8.496	1	1.168	-1.513	-1.382	1	1	1	1	1	1
New Zealand	9.713	4	1.175	-2.166	-1.852	1	1	1	1	1	1
Switzerland	10.076	4	1.335	-1.760	-1.524	1	1	1	1	1	1
Ethiopia	6.235	1	1.469	-0.819	-0.754	1	1	1	1	1	1
Philippines	8.001	3	1.504	-1.355	-1.239	1	1	1	1	1	1
Ecuador	8.288	2	1.550	-0.575	-0.518	1	1	1	1	1	1
Mali	6.669	1	1.581	-0.802	-0.730	1	1	1	1	1	1
Costa Rica	8.763	2	1.637	-1.204	-1.081	1	1	1	1	1	1
Syria	7.375	1	1.648	-0.383	-0.337	1	1	1	1	1	1
Morocco	8.009	1	1.655	-0.919	-0.794	1	1	1	1	1	1
Uruguay	8.949	2	1.671	-0.477	-0.410	1	1	1	1	1	1
Paraguay	8.344	2	1.676	-0.582	-0.513	1	1	1	1	1	1
Malawi	6.495	1	1.683	-0.687	-0.624	1	1	1	1	1	1
Mexico	8.758	2	1.734	-0.797	-0.694	1	1	1	1	1	1
Sweden	9.829	4	1.820	-0.956	-0.813	1	1	1	1	1	1
Colombia	8.438	2	1.897	-0.734	-0.519	1	1	1	1	1	1
Trinidad-Tobago	9.226	2	1.931	-0.113	-0.072	1	1	1	2	1	1
Denmark	9.899	4	1.973	-0.746	-0.581	1	1	1	1	1	1
Australia	9.805	4	2.083	-0.227	-0.018	2	1	2	2	2	1
Chile	8.846	2	2.109	0.012	0.058	2	1	2	2	2	1
United Kingdom	9.717	4	2.151	-0.070	0.073	2	1	2	2	2	1
Netherlands	9.813	4	2.187	-0.115	0.062	2	1	2	2	2	2
Canada	9.834	4	2.198	-0.122	0.024	2	1	2	2	2	1
Turkey	8.289	3	2.252	0.201	0.293	2	2	2	2	2	1
Brazil	8.653	2	2.323	0.184	0.257	2	2	2	2	2	1
USA	10.043	4	2.345	0.203	0.336	2	2	2	2	2	2
Dominican, Rep	8.263	2	2.422	0.940	0.950	2	2	2	2	2	1
France	9.769	4	2.452	0.965	0.978	2	2	2	2	2	2
Belgium	9.729	4	2.569	1.614	1.518	2	2	2	2	2	2
Finland	9.645	4	2.573	0.434	0.495	2	2	2	2	2	2
Greece	9.268	4	2.595	0.434	0.514	2	2	2	2	2	2
India	7.320	3	2.637	1.146	1.128	2	2	2	2	2	2
Italy	9.652	4	2.666	1.471	1.303	2	2	2	2	2	2
Panama	8.589	2	2.686	0.582	0.645	2	2	2	2	2	2
Pakistan	7.431	3	2.700	0.964	1.031	2	2	2	2	2	2
Israel	9.597	1	2.724	0.745	0.777	2	2	2	2	2	2
Austria	9.787	4	2.796	2.608	2.024	2	2	2	2	2	2
Spain	9.436	4	2.908	1.236	1.095	2	2	2	2	2	2
Norway	9.897	4	3.099	2.450	1.641	2	2	2	2	2	2
Tunisia	8.332	1	3.183	2.532	1.968	2	2	2	2	2	2
Indonesia	7.679	3	3.505	1.976	1.526	2	2	2	2	2	2
Japan	9.670	3	3.583	1.464	1.163	2	2	2	2	2	2
Portugal	9.231	4	3.692	2.265	1.642	2	2	2	2	2	2
Sri Lanka	7.619	3	3.784	3.155	2.304	2	2	2	2	2	2
Ireland	9.324	4	3.831	2.124	1.335	2	2	2	2	2	2
Thailand	8.056	3	4.572	2.439	1.529	2	2	2	2	2	2
Malaysia	8.548	3	4.687	4.605	2.159	2	2	2	2	2	2
Hong Kong	9.540	3	4.822	2.833	1.658	2	2	2	2	2	2
Singapore	9.479	3	5.284	4.933	1.951	2	2	2	2	2	2
Korea, Rep of	8.633	3	6.089	5.474	1.977	2	2	2	2	2	2

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



See the footnote of Table 5 for details. Particularly,

$$s_i = \sqrt{T}(\hat{g}_i - \hat{g}_w)/\tilde{\sigma}(\hat{g}_i),$$

$\hat{g}_w$ denotes the weighted fixed effect estimator of g_0 under the null that $g_i = g_0$ for all individuals, and $\tilde{\sigma}(\hat{g}_i)$ is the restricted long-run variance of $\hat{g}_i$ based on the automatic method developed by Andrews (1991) with Quadratic Spectral kernel. Under the null that parameters are the same, the shape of s_i should not be too different from the standard normal distribution.