

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

CHANG-CHING LIN* SERENA NG†

January 30, 2009

Academia Sinica Columbia University

Abstract

This paper considers two ways of estimating panel data models with group specific parameters when group membership is not known. The first is based on the K-means algorithm, adapted to panel data regressions with fixed effects and heteroskedastic variances. The second is a new data-dependent, two step approach. 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. We show that individuals can be consistently sorted into groups distinguished by parameter heterogeneity when N and T are large.

KEYWORDS: Parameter Heterogeneity, Threshold Models, Cluster Analysis.

*Institute of Economics, Academia Sinica, 128 Academia Road, Sec 2, Taipei 115, Taiwan. Tel: 886-2-27822791 ext. 301. E-mail: lincc@econ.sinica.edu.tw

†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 work is motivated by the concern that while heterogeneous response of endogenous variables to covariates is often predicted by economic theory, estimation of panel data models often take parameter homogeneity as a maintained assumption, even though evidence for parameter is not hard to find.¹

There are advantages to the assumption parameter homogeneity. Pooled estimation is convenient and the convergence rate of the estimator is increasing in the number of cross-section units, However, pooling is inefficient when there is parameter heterogeneity, even though consistent estimation of the population mean of the parameters can still be possible. 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.

We consider two methods. The first does not require estimation of the individual parameters and is a variation of the K-means algorithm modified to partition the data using the slope coefficients. The second method is new, and is suited for panel data with a large number of time series observations. By using consistent estimates of the unit-specific slope parameters, we turn the problem of identifying group membership into one of estimating a threshold panel regression, where the threshold variable is itself being estimated. We refer to this as a ‘pseudo’ threshold approach.

K-means is a widely popular algorithm used in cluster analysis. While in theory,

¹ 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.

the algorithm may converge to a local optimum and the method can also be computationally demanding, it tends to have good finite sample properties. The algorithm is (to our knowledge) not widely used in regression analysis, and certainly not in panel data regressions. We modify the K-means method to allow for group-specific slope parameters in panel regressions with fixed effects. The method, to be presented in Section 3, is non-parametric, and does not require the estimation of the individual slope parameters. As such, its properties are not sensitive to the sample size in the cross-section dimension.

Our pseudo threshold approach 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.

The two methods considered have several advantages. First, they are entirely data dependent and do not require selection of a transition variable(s) in advance. They are a model rather than a graphical based approach and formal statistical statements can be made about the estimated groups. Second, the methods are valid even in the presence of fixed effects. Third, the methods also allow us to control for cross-sectionally heterogeneous variances. As we will see, the pseudothreshold method allows us to distinguish cluster-specific slope coefficients that are different in the order $T^{-\alpha}$, $0 \leq \alpha < 1/2$.

The remainder of this paper is organized as follows. After a review of related work, the K-means method is presented in Section 3, while the pseudo threshold method is introduced in Section 4. Extension to the case of multiple groups and multiple covariates is considered in Section 5. Simulations are then presented. As an application, we apply our methods to study economic growth across countries.

2 Related Literature and the Econometric Framework

Our work is related to clustering analysis. 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. In addition, sometimes we may not have prior information about the group structure at all.

Regression based methods have also been proposed to deal with parameter heterogeneity. One approach is to parameterize the coefficients as a function of individual characteristics, see, for example, Alvarez, Browning, and Ejrnaes (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.² 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, $x_i, i = 1, \dots, N$, into K groups.³ The approach is motivated by the idea that

²See, for example, Swamy (1970) and Hsiao and Pesaran (2004).

³See, for example, Fraley and Raftery (2002), Hall, Muller, and Wang (2006), and Chiou and

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 $f_k(x_i|\beta_k)$. If q_i is the identifying label, i.e. $q_i = k$ if unit i belongs to group k , then one can maximize the likelihood

$$L(x; \theta, q) = \prod_{i=1}^N \prod_{t=1}^T f_{q_i}(x_{it}; \beta)$$

with respect to β . In practice, the EM algorithm is often used for estimation; the identifier q_i is treated as an unknown and set according to the empirical probability of the group to which unit i belongs. The method can be cumbersome if N is large because we need to consider up to 2^N possible combinations to find the maximum (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 to 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 (2008) 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 will consider two approaches that are not likelihood based, 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 $I^0 = (I_1^0, I_2^0)$ be the true group

Li (2007).

membership and let N_1^0 and N_2^0 denote the number of individuals in clusters I_1^0 and I_2^0 , respectively. The data are assumed to be generated as

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

where α_i denotes the individual fixed-effect, and \tilde{x}_{it} is a vector of exogenous variables. The coefficients for the two clusters are denoted as B_1 and B_2 . That is, $\beta_i = B_1$ if $i \in I_1^0$ and $\beta_i = B_2$ if $i \in I_2^0$. Without loss of generality, assume that $B_1 < B_2$. The case of homogeneous parameters is obtained if $B_1 = B_2$, and the case of complete heterogeneity is obtained when there are N clusters each consisting of only one unit. Moreover, (1) can be rewritten as

$$\tilde{y}_{it} = \alpha_i + \tilde{x}_{it}B_1\mathbf{1}(i \in I_1^0) + \tilde{x}_{it}B_2\mathbf{1}(i \in I_2^0) + \tilde{e}_{it}. \quad (2)$$

We control for the fixed effects by letting $z_{it} = \tilde{z}_{it} - \bar{\tilde{z}}_i$ where $\bar{\tilde{z}}_i = \frac{1}{T} \sum_{t=1}^T \tilde{z}_{it}$, 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

$$y_{it} = x_{it}B_1\mathbf{1}(i \in I_1^0) + x_{it}B_2\mathbf{1}(i \in I_2^0) + e_{it}. \quad (3)$$

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 $\hat{\sigma}_i^2 = \frac{1}{T-G-1} \sum_{t=1}^T \hat{e}_{it}^2$, where $\hat{e}_{it} = y_{it} - x_{it}\hat{\beta}_i$. To obtain a pooled estimate of β_i , we use the fixed effect model:

$$\tilde{y}_{it} = \alpha_i + \tilde{x}_{it}B_\omega + \tilde{e}_{it}. \quad (4)$$

The unweighted ($\delta_i = 1$) and weighted ($\delta_i = 1/\hat{\sigma}_i^2$) pooled fixed effect least squares estimators are then defined as

$$\hat{B}_\omega = \frac{\sum_{i=1}^N \sum_{t=1}^T \delta_i x'_{it} y_{it}}{\sum_{i=1}^N \sum_{t=1}^T \delta_i x'_{it} x_{it}} = \sum_{i=1}^N \left(\frac{\sum_{t=1}^T \delta_i x'_{it} x_{it}}{\sum_{i=1}^N \sum_{t=1}^T \delta_i x'_{it} x_{it}} \right) \hat{\beta}_i.$$

The weighted 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 (2008) to allow both N and T to be large.

Let $I = (I_1, I_2)$ denote a group membership other than $I^0 = (I_1^0, I_2^0)$. For $j, k = 1, 2$, let N_{jk} be the number of individuals assigned to be in group k when

they truly belong to group j . Let $N_1 = N_{11} + N_{21}$ and $N_2 = N_{22} + N_{12}$ and let $N_s = N_s(I, I^0) = N_{12} + N_{21}$ be the number of misclassified units. The following assumptions will be used throughout.

Assumption

1. For all i 's and t 's, (a) $\tilde{e}_{it}|\tilde{x}_i \sim IID(0, \sigma_i^2)$ and uncorrelated with B_1, B_2 . (b) $\max_{1 \leq i \leq N}(\sigma_i^2)$ is finite and $\min_{1 \leq i \leq N}(\sigma_i^2) > 0$. (c) $E(\tilde{e}_{it}^9|\tilde{x}_i)$ is finite.
2. For each i , there exists a finite T_0 such that for $T > T_0$, $E[(\hat{\sigma}_i^2/\sigma_i^2)^{-4-\epsilon}]$ is finite for some small positive constant ϵ .
3. For $j = 1, 2$, $N_j^0/N > 0$ and $N_1^0/N \rightarrow \pi$ with $0 < \pi < 1$.
4. (a) For each i , $\hat{Q}_i = \frac{1}{T} \sum_{t=1}^T x'_{it}x_{it}$ is finite and positive definite and $\max_{1 \leq i \leq N} E\|\hat{Q}_i\|$ is finite with $\hat{Q}_i \xrightarrow{p} Q_i > 0$ as $T \rightarrow \infty$, where Q_i is a non-stochastic positive definite and $\max_{1 \leq i \leq N} E\|Q_i\|$ is finite.
 (b) Let I_l be a subset of the whole sample and let N_l denote the number of units in I_l . If $N_l \geq 1$, $N_l^{-1} \sum_{i \in I_l} \hat{Q}_i$ and $N_l^{-1} \sum_{i \in I_l} \sigma_i^{-2} \hat{Q}_i$ has the minimal eigenvalue bounded away from zero in probability as $(N_l, T) \rightarrow \infty$ jointly.

Assumption 1(a) is commonly imposed in panel data models with fixed effects and strictly exogenous regressors. Following Pesaran and Yamagata (2008), Assumptions 1(b), 1(c), and 2 are imposed when $\delta_i = \hat{\sigma}_i^{-2}$. Assumption 3 is imposed for the situation when q_i is not observed. We can relax this assumption by letting $T \rightarrow \infty$ to increase the information around the breaks when q_i is observed and exogenous. Assumption 4 concerns identification condition. Under the assumptions above, Lemma A7 in the Appendix shows that \hat{B}_ω is consistent for the population mean, B_ω , even though the regression model (4) is misspecified in the model (1) with the cross-sectionally heterogeneous slope parameters.

3 K-means Clustering

The aim of cluster analysis is to classify a set of data into groups. Suppose that we observe y_{it} , $i = 1, \dots, N$, $t = 1, \dots, T$, and there are no covariates. The simplest way to estimate group membership is to 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 Essentially, the algorithm 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. Harmonic and fuzzy means have also been used instead of simple means. The main idea of the K-means method remains that two clusters in y_{it} are identified by comparing the data to the sequentially updated centroids.

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. Suppose that there are two groups in the data and consider again the transformed fixed-effect model in (3). The algorithm then consists of the following steps.

1. Randomly assign individuals into two groups $\{\ddot{I}_1, \ddot{I}_2\}$, and calculate fixed effects estimator (\ddot{B}_1, \ddot{B}_2) based on $\{\ddot{I}_1, \ddot{I}_2\}$
2. Repeat (a) and (b) until no individual is changed from one group to another:
 - (a) Calculate $SSR_i^j = \sum_{t=1}^T (y_{it} - x_{it}\ddot{B}_j)^2$, $i = 1, \dots, N$, and $j = 1, 2$
 - (b) If $SSR_i^1 \leq SSR_i^2$, individual i is re-assigned to group 1; otherwise, i stays with group 2. Then, (c) update $\{\ddot{I}_1, \ddot{I}_2\}$ and recalculate the fixed effects estimator (\ddot{B}_1, \ddot{B}_2) and SSR_i^j .
3. Re-shuffle individuals unit by unit to form new grouping $\{\ddot{I}'_1, \ddot{I}'_2\}$ and calculate $(\ddot{B}'_1, \ddot{B}'_2)$ and $SSR_i^{j'} = \sum_{t=1}^T (y_{it} - x_{it}\ddot{B}'_{j'})^2$. If $\sum_j \sum_{i \in I'_j} SSR_i^{j'} < \sum_j \sum_{i \in I_j} SSR_i^j$, then repeat 2.(a)–(c) with $(\ddot{B}'_1, \ddot{B}'_2)$.

Steps 1 to 3 are repeated several times to reduce the effects of the initial group assignment. The algorithm can be extended to the situations with $K > 2$ groups in

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

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

the data. It can be easily adapted to allow for heterogeneous errors by redefining SSR^j and $SSR^{j'}$ in Steps 2 and 3, respectively, as follows:

$$SSR_i^j = \sum_{t=1}^T (y_{it} - x_{it}\ddot{B}_j)^2 / \hat{\sigma}_i^2 \quad \text{and} \quad SSR_i^{j'} = \sum_{t=1}^T (y_{it} - x_{it}\ddot{B}'_j)^2 / \hat{\sigma}_i^2, \quad j = 1, 2.$$

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

4 A Threshold Approach

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 and structure break models are variations of this idea in a time series context.

Suppose we can find a variable q_i that, along with a set of cut-off parameter values Γ^0 , will lead to perfect information about I_1^0 and I_2^0 in the sense that $i \in I_1^0$ if $q_i \leq \gamma^0$ for any $\gamma^0 \in \Gamma^0$ and $i \in I_2^0$ otherwise. Then (1) can be written as

$$\tilde{y}_{it} = \alpha_i + \tilde{x}_{it}B_11(q_i \leq \gamma^0) + \tilde{x}_{it}B_21(q_i > \gamma^0) + \tilde{e}_{it}. \quad (5)$$

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

‘pseudo threshold parameter’ to distinguish them from the usual definitions used in the threshold literature.

If q_i and Γ^0 are both known, estimates of B_1 and B_2 can be obtained using a threshold, or split-sample, regression. Observations with $q_i \leq \gamma^0 \in \Gamma^0$ 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 .

4.1 Estimation of γ when q_i is Observed and Exogenous

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

$$\tilde{\gamma} = \arg \min_{\gamma \in [q_{\min}, q_{\max}]} S_{NT}(\gamma), \quad (6)$$

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

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

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 too many units are put in Group 1 (and thus $N_{21}(\gamma) > 0$), then it will also be the case that $N_{12}(\gamma) = 0$. Thus, one of the misclassified 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 *Suppose the data are generated by (1) and q_i is exogenous and observed. Suppose that $\tilde{\gamma}$ is obtained from (6) and $B_2 - B_1 = \nu T^{-\alpha}$ with $0 < \|\nu\| < \infty$ and $0 \leq \alpha < 1/2$. When $\delta_i = 1$ for all i 's, $N_s(\tilde{\gamma}) = O_p(T^{-1+2\alpha})$.*

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 will not be 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 misclassified set. In fact, any $\tilde{\gamma} \in [q_{[i^*]}, q_{[i^*+1]})$, where i^* is such that $q_{[i^*]} = \tilde{\gamma}$, will yield the same classification. For fixed $B_2 - B_1 = O(1)$, Theorem 1 implies that the maximum misclassification rate is $N_s/N = O_p(\frac{1}{NT})$. If $B_2 - B_1$ is in the $T^{-\alpha}$ neighborhood of zero, the misclassification rate is $N_s/N = O_p(\frac{1}{NT^{1-2\alpha}})$. Thus the misclassification rate tends to zero as $N, T \rightarrow \infty$ jointly.

The classification error rate also tends to zero when $\delta_i = \hat{\sigma}_i^{-2}$ in (6). However, $\hat{\sigma}_i^2$ can introduce extra sampling variations to the objective function. Because the sum of these variations over units can render the group estimation inconsistent when N is large, we need to control for the growth rate of N related to T to ensure group membership is consistently estimated. This leads to the following.

Corollary 1: *Suppose that the conditions in Theorem 1 are satisfied and $B_2 - B_1 = \nu T^{-\alpha}$ with $0 < \|\nu\| < \infty$ and $0 \leq \alpha < 1/2$. When $\delta_i = \hat{\sigma}_i^{-2}$ and $(N, T) \rightarrow \infty$ jointly such that $\sqrt{N}/T^{3/2-2\alpha} \rightarrow 0$, we have $N_s(\tilde{\gamma}) = O_p(T^{-1+2\alpha})$.*

Given $\tilde{\gamma}$, the two groups can be estimated as $I_1(\tilde{\gamma}) = \{i | q_i \leq \tilde{\gamma}\}$ and $I_2(\tilde{\gamma}) = \{i | q_i > \tilde{\gamma}\}$. 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 by 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 .

4.2 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 $T \rightarrow \infty$. Hereafter, all variables indexed by i are assumed to be ordered once q_i is estimated. Given \hat{q}_i , the problem is to find an estimate of γ . Let

$$\hat{\gamma} = \arg \min_{\gamma \in [\hat{q}_{\min}, \hat{q}_{\max}]} S_{NT}(\gamma, \hat{q}), \quad (7)$$

where

$$S_{NT}(\gamma, \hat{q}) = \sum_{i|\hat{q}_i \leq \gamma} \sum_{t=1}^T \delta_i(y_{it} - x_{it}\hat{B}_1(\gamma))^2 + \sum_{i|\hat{q}_i > \gamma} \sum_{t=1}^T \delta_i(y_{it} - x_{it}\hat{B}_2(\gamma))^2. \quad (8)$$

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 \sqrt{T} consistent estimators of q_i . The first one does not account for sampling uncertainty. The second estimator does, and is more appropriate when there is substantial dispersion in the error variances.

A Non-normalized Estimator: $\hat{q}_i = \hat{\beta}_i - \hat{B}_\omega$ or $\hat{\beta}_i$.

To motivate this choice of \hat{q}_i , first consider defining $q_i = \beta_i - B_\omega$. It is easy to see that

$$q_i = \begin{cases} \beta_i - B_1 - (1 - \omega)(B_2 - B_1) & \text{for } i \in I_1^0, \\ \beta_i - B_2 + \omega(B_2 - B_1) & \text{for } i \in I_2^0, \end{cases}$$

where $\omega = \frac{\sum_{i \in I_1^0} Q_i}{\sum_i Q_i}$ and Q_i is defined in Assumption 4(a). Now $\beta_i - B_1 = 0$ if $i \in I_1^0$ and $B_2 \neq B_1$ by assumption. Thus, $q_i = -(1 - \omega)(B_2 - B_1) < 0$ if $i \in I_1^0$. On the other hand, $q_i = \omega(B_2 - B_1) > 0$ if $i \in I_2^0$. The pseudo variable q_i along with $\gamma^0 = 0$ completely summarizes group membership. In fact, any $\gamma^0 \in [-(1 - \omega)(B_2 - B_1), \omega(B_2 - B_1))$ 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 observations on unit i only. 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, when T becomes large, $\hat{q}_i = \hat{\beta}_i - \hat{B}_\omega$, like q_i , is negative for those $i \in I_1^0$, and is positive for those $i \in I_2^0$. 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}_ω .

A Normalized Estimator 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 on clustering. We therefore consider a standardized pseudo threshold variable in view of this concern. This variable is defined as

$$\hat{q}_i = \hat{\tau}_i \quad \text{where} \quad \hat{\tau}_i = \frac{(\hat{\beta}_i - \hat{B}_\omega)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}}, \quad (9)$$

\hat{Q}_i is defined in Assumption 4(a) with $\hat{Q}_i \rightarrow Q_i > 0$ as $T \rightarrow \infty$. 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 known. Using the definition of B_ω , q_i can be expressed as:

$$q_i = \tau_i = \begin{cases} \frac{(\hat{\beta}_i - B_1)}{\sigma_i Q_i^{-1/2}} - \frac{(1-\omega)(B_2 - B_1)}{\sigma_i Q_i^{-1/2}}, & \text{for } i \in I_1^0, \\ \frac{(\hat{\beta}_i - B_2)}{\sigma_i Q_i^{-1/2}} + \frac{\omega(B_2 - B_1)}{\sigma_i Q_i^{-1/2}}, & \text{for } i \in I_2^0. \end{cases} \quad (10)$$

It can be seen that as $T \rightarrow \infty$,

$$q_i = \tau_i = \begin{cases} -\frac{(1-\omega)(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1), & \text{for } i \in I_1^0 \\ \frac{\omega(B_2 - B_1)}{\sigma_i Q_i^{-1/2}} + o_p(1), & \text{for } i \in I_2^0. \end{cases}$$

This implies that if $B_2 > B_1$, q_i will become negative for those $i \in I_1^0$ and positive for those $i \in I_2^0$ as T increases. On the other hand, when $B_1 = B_2 = B_\omega$,

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

In this case, $\sqrt{T}\tau_i$ will be approximately normal with mean zero as T becomes large. It is obvious that τ_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 $B_2 - B_1 = O(T^{-\alpha})$ for $0 \leq \alpha < 1/2$.

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 $\hat{B}_\omega \rightarrow B_\omega$ and $\hat{Q}_i \rightarrow Q_i$. Lemma A8 in the Appendix shows that $\sqrt{T}(\hat{\tau}_i - \tau_i) = o_p(1)$ under Assumptions 1–4.

In Theorem 1, we have shown that the classification error rate is $O_p(N^{-1}T^{-1+2\alpha})$ when q_i is known and exogenous. That is, if $B_2 - B_1$ is fixed, $P(N_s/N|q_i) = O_p((NT)^{-1})$. However, $\hat{\beta}_i$ and $\hat{\tau}_i$ are no longer exogenous. Nevertheless, the classification error rates of our two-step approach still converge to zero.

Theorem 2 *Let \hat{q}_i be $\hat{\beta}_i$ or $\hat{\tau}_i$. Suppose the data are generated by (1) and $\hat{\gamma}$ is obtained from (8). Then for $B_2 - B_1 = \nu T^{-\alpha}$ with $0 < \|\nu\| < \infty$ and $0 \leq \alpha < 1/2$, $N_s(\hat{\gamma})/N = O_p(T^{-1/2+\alpha})$ if (i) $\delta_i = 1$ or if (ii) $\delta_i = \hat{\sigma}_i^{-2}$ with $\sqrt{N}/T^{3/2-2\alpha} \rightarrow 0$ as $(N, T) \rightarrow \infty$ jointly.*

The consequence of having to estimate q_i is that the classification error rate tends to zero at a slower rate since

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

The rate of $N^{-1}T^{-1+2\alpha}$ arises from having to estimate γ , while the rate of $T^{-1/2+\alpha}$ arises from having to estimate q_i . 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.

5 Multiple Regressors

We have shown that when there is a single regressor, either $\hat{\beta}_i$ or $\hat{\tau}_i$ can be used to split the sample. Suppose it is suspected that all the parameters are different between groups. The natural extension from the two regressor 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 $j, k, = 1, 2, j \neq k$, we have $B_{j1} > B_{k1}$ and $B_{j2} > B_{k2}$. We will refer to this as Case 1. Since both parameters are strictly larger in one group than in another group, a natural pseudo transition variable is

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

Using arguments analogous to those in the previous section, one can expect $\hat{\beta}_i^+$ to completely separate those $i \in I_1^0$ from those $i \in I_2^0$ when T is large. However, this method does not always work. For example if $(B_{11}, B_{12}) = (0.8, 1)$ and $(B_{21}, B_{22}) = (1, 0.8)$, 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 $B_{j1} > B_{k1}$ but $B_{j2} < B_{k2}$. We will refer to this as Case 2. That is, when the ranking of the parameter between groups depends on which parameter is being compared. To resolve this problem, consider the transition variable

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

If $B_{j2} < B_{k2}$, then $-B_{j2} > -B_{k2}$. Thus, it can again be seen that $\hat{\beta}_i^-$ will separate

those $i \in I_1^0$ from those $i \in I_2^0$ when T is large.

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

$$\psi = (n_I \cdot n_{III} - n_{II} \cdot n_{IV}) / (n_I \cdot n_{III} + n_{II} \cdot n_{IV}). \quad (11)$$

Then define $R = (1, 1)$ if $\psi > 0$; otherwise, $R = (1, -1)$. Given R , our non-normalized and normalized pseudo threshold variables are

$$\hat{q}_i(\hat{\beta}_i) = R\hat{\beta}_i, \quad \text{and} \quad q_i(\hat{\tau}_i) = R(\hat{\beta}_i - \hat{B}_\omega)(\widehat{\text{avar}}(\hat{\beta}_i)R')^{-1/2}, \quad (12)$$

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

The analysis can easily be extended to the case of G regressors. Let $\hat{\beta}_i = (\hat{\beta}_{i1}, \dots, \hat{\beta}_{iG})'$ denote the least squares estimator for the i^{th} individual and $\hat{\mathbf{B}}_\omega = (\hat{B}_{\omega 1}, \dots, \hat{B}_{\omega G})'$ denote the weighted fixed-effects estimator under the restriction of parameter homogeneity.

1. For $g = 1, \dots, G$, $i = 1, \dots, N$, calculate \hat{q}_{ig} , which can be $\hat{\beta}_{ig}$ or $\hat{\tau}_{ig}$.
2. Let $W = \arg \max_{g=\{1, \dots, G\}} \|\hat{q}_g\|_2$, where $\hat{q}_g = (\hat{q}_{1g}, \hat{q}_{2g}, \dots, \hat{q}_{Ng})'$.
3. Calculate the Goodman-Kruskal's statistic, ψ_g , between $(\hat{\beta}_{ig} - \hat{B}_{\omega g}, \hat{\beta}_{iW} - \hat{B}_{\omega W})$, $i = 1, \dots, N$ as defined in (11). Create a $1 \times G$ vector R , where the g^{th} position is 1 if $\psi_g \geq 0$; and -1, otherwise.
4. Once R is obtained, the pseudo threshold variable for a model with G regressors can be either $\hat{q}_i(\hat{\beta}_i)$ or $\hat{q}_i(\hat{\tau}_i)$ as defined in (12).

Finally, it remains to consider the case of partial parameter heterogeneity. That is, there are G regressors, and only a subset of the parameters are heterogeneous across groups. We can still split the sample using the two-step approach. For example, if we suspect that the second and third slope coefficients vary between groups, we can let $\hat{q}_i = \hat{\beta}_{i2}$ (or $\hat{\beta}_{i3}$). Upon splitting the sample, we can test if the β_{i2} is the same within group. This requires a test for parameter homogeneity.

5.1 Testing Parameter Homogeneity

Once we have estimated which group each unit belongs, we can test the null hypothesis of parameter homogeneity: $H_0^A : \beta_i = B \forall i$, which can also be stated as $H_0 : B_1 = B_2$. In standard threshold models, we can use the sup-Wald type test. See, for example, Davies (1977), Andrews and Ploberger (1994), Hansen (1996), Bai (1997), and Caner and Hansen (2004). However, there are three features in our cases that make the SupW test for parameter homogeneity infeasible. First, \hat{B}_1 and \hat{B}_2 are estimated from two split samples ordered by $\hat{\beta}_i$ and these will be biased if $B_1 = B_2$. Second, \hat{B}_1 and \hat{B}_2 are correlated when $B_1 = B_2$ because the cross-sectional independence assumption fails. Third, as \hat{q}_i is ordered, the conventional bootstrap that is valid for cross-sectionally independent data does not apply.

Instead, we use the dispersion test proposed by Pesaran and Yamagata (2008) to detect parameter heterogeneity. The test is defined as

$$PY = \sqrt{N}(S/N - G)/\sqrt{2G}, \quad (13)$$

where G denotes the number of the regressors, $S = \sum_{i=1}^N \tilde{\sigma}_i^{-2}(\hat{\beta}_i - \hat{B}_w)'(x_i x_i')(\hat{\beta}_i - \hat{B}_w)'$, and $\tilde{\sigma}_i^2$ is a consistent estimator of σ_i^2 based on a restricted fixed-effect estimator of B . This test allows for cross-sectional fixed effects and heteroskedasticity and non-normally distributed errors in the models with strictly exogenous regressors as N and T go to infinity jointly such that $\sqrt{N}/T^2 \rightarrow 0$.

6 Simulations and Applications

The pseudo threshold approach and the K-means method are both non-parametric procedures, but there are some differences between them. The K-means algorithm

makes random initial guesses of the centroids and then evaluates if a move to a different group is desirable unit by unit. This makes the K-means method computationally costly when N is large. Our pseudo variable is ordered and can be thought of as using more information to 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 with large N can be computationally 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 are \sqrt{NT} consistent, and does not require the individual estimates $\hat{\beta}_i$, which are \sqrt{T} consistent. Thus the K-means method will be more precise even when N or T is small. In contrast, the pseudo threshold approach requires 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. When there are at least two regressors in a panel data model, estimating ψ in (11) can introduce extra errors into the pseudo threshold method. We can therefore expect a trade-off between precision and speed in the different methods.

We now use Monte Carlo simulations to examine the finite sample properties of the methods considered. We generate data as follows. For $K = 2, 3$, $t = -49, \dots, 0, \dots, T$, and $i = 1, \dots, N$,

$$\begin{aligned}\tilde{y}_{it} &= \alpha_i + \sum_{k=1}^K \tilde{x}_{it} B_k 1(i \in I_k^0) + \tilde{\epsilon}_{it} \\ \tilde{x}_{it} &= \alpha_i(1 - \rho_i) + \rho_i \tilde{x}_{i,t-1} + (1 - \rho_i^2)^{1/2} \tilde{v}_{it},\end{aligned}$$

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 of $\tilde{\epsilon}_{it}$, $\tilde{\epsilon}_{it} \sim N(0, \sigma_i^2)$ is i.i.d. over i and t , and $\sigma_i^2 \sim$ i.i.d. $U(1, 3)$. The first 50 observations are discarded. For the experiment with $K = 2$ (and $G = 1$), we set $B_1 = 0.7$ and $B_2 = 1$ and randomly assign individuals into two groups $\{I_1^0, I_2^0\}$ with size $N_1^0 = \lfloor 2N/3 \rfloor$ and $N_2^0 = N - N_1^0$, respectively, where $\lfloor A \rfloor$ denotes the maximum integer that does not exceed real number A . For $K = 3$, we set $(B_1, B_2, B_3) = (0.7, 1, 1.3)$ and randomly assign individuals into three groups $\{I_1^0, I_2^0, I_3^0\}$ with size $N_k^0 = \lfloor N/3 \rfloor$ for $k = 1, 2$ and $N_3^0 = N - N_1^0 - N_2^0$. Throughout all experiments, we use $M = 1000$ replications, holding $\{I_1^0, \dots, I_K^0\}$, α_i , ρ_i , and σ_{ix}^2 fixed over all repetitions in each combination of (N, T) . This data generating process is modified from Table 1 of Pesaran and Yamagata (2008).

For each (N, T) with $N = (20, 100, 500, 1000)$ and $T = (20, 200)$, we keep track of the pooled root mean squared error of the estimates (RMSE) and the accuracy of classification (CR). These are defined as

$$RMSE = \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^K \left[\left(\frac{N_k^0}{N} \sum_{g=1}^G (\hat{B}_{kg}^{(m)} - B_{kg})^2 \right) \right] \right\}^{1/2}$$

$$CR = \frac{1}{M} \sum_{m=1}^M CR^{(m)} \text{ where } CR^{(m)} = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K 1 \left((i \in \hat{I}_k^{(m)}) \cap (i \in I_k^0) \right),$$

where the superscript (m) denotes the estimate from the m^{th} trial. Results are reported in Figures 1 and 2. In the Figures, PS_1^H denotes the pseudo threshold approach using $\hat{q}_i = \hat{\beta}_i$, while PS_2^H denotes the method using $\hat{q}_i = \hat{\tau}_i$. Both PS_1^H and PS_2^H are based on the weighted objective function with $\delta_i = \hat{\sigma}_i^{-2}$. The panel K-means method controlling for heterogeneity in variances is denoted as K_1^H . To reduce the effect of the initial group assignment, we randomly draw seven sets of initial groups for K_1^H in each repetition and take the results from the set with the minimum sum of squared residuals. It should be mentioned that in unreported results, ignoring the individual-specific fixed effects leads to more inaccurate estimates of the clusters, as expected. Also, while the groups are accurately classified regardless of whether heterogeneity in variances is taken into account or not, the three methods robust to heterogeneous variances have slightly smaller RMSEs because the errors are heteroskedastic by design.

We first turn to Figure 1, which reports the results for the experiment with

$(K, G) = (2, 1)$. Overall, all methods perform quite similarly. The RMSEs tend to decrease as N or T increases. Moreover, the impact 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 from around 60% to 80% when T increases from 20 to 200. Because K_1^H , PS_1^H and PS_2^H minimize the same objective function, it is not surprising that they have similar CRs and RMSEs.

Figure 2 reports the CRs and RMSEs for the experiment with $(K, G) = (3, 1)$. Similar to the above, the RMSEs tend to decrease as N or T increases, while the CR improves with T along. Because there are three instead of two groups, the classification error rate is also higher, as expected. Although the K_1^H tends to have better CR than the PS approach, no method outperforms the others in terms of RMSEs over all considered combinations of N and T .

The results for the experiments with $(K, G) = (2, 2)$ and $(2, 3)$ are reported in Figures 3 and 4, respectively. The design of the experiments for multiple regressors are similar to those when $G = 1$. When $K = 2$, we set $B_1 = (B_{11}, B_{12})' = (0.7, 1)'$ and $B_2 = (B_{21}, B_{22})' = (1, 1.1)'$. When $K = 3$, we let B_1 and B_2 be the same as $(G, K) = (2, 2)$, and set $B_3 = (B_{31}, B_{32})' = (1.3, 1.2)'$. In these experiments, we calculate \hat{q}_i based on (12) PS_2^H . Similar to the results for $G = 1$, the RMSEs are affected by N and T and the CRs improve as T increases. The K-means method again has smaller classification error rates than the PS methods over all combinations of N and T . This can be traced to the fact that estimating the second slope parameter and R both introduce sampling errors. However, the K-means method is not always preferred in terms of RMSE. The PS_1^H and PS_2^H have smaller RMSEs for quite a few configurations of N and T .

Table 1 shows the empirical size and power of the PY test for parameter homogeneity. The DGPs are the same as those in Figures 1 and 3, respectively. The size for each case is roughly around the 5% nominal level. Clearly, when T or N increases, the PY test will have more power to detect the alternative. Also, the finite sample power is higher in the case with only one regressor than with two regressors.

6.1 Empirical Study

The existence of “convergence clubs” has generated much research interests in the growth literature. A group of countries with a similar steady state that can be characterized by the same linear model are 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 by Heston, Summers, and Aten (2006) for the sample 1965 to 2000.⁶ Figure 5(a) shows the same scatter plot for countries in different geographical regions. It is obvious that the growth rates vary by geography, membership in the OECD (mainly in the category of Western Europe and North America), and GDP per capita. However, many countries with similar growth rates have different observed characteristics. A pattern for growth rates cannot be easily established.

The regression model is

$$\tilde{y}_{it} = \mu'_i + g_i t + \tilde{\epsilon}_{it} \quad (14)$$

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 5(b)–(c) plot the distribution of the non-standardized and standardized \hat{g}_i estimated by kernel smoothing.

⁶See 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.

⁷Specifically, $\tilde{y}_{it} = \gamma_i \tilde{y}_{i,t-1} + \tilde{x}_{it} \beta_i + \phi_i t + c_i + D_t + \tilde{\epsilon}_{it}$ 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 $\tilde{\epsilon}_{it} = \gamma_i \tilde{\epsilon}_{i,t-1} + \eta_t + \tilde{\epsilon}_{it}$ gives the regression being estimated. See Islam (1995) and Lee, Pesaran, and Smith (1997) for details.

We use our methods to split countries into two groups. Table 2 shows that these methods suggest a very similar grouping of between 28 to 30 countries in the low growth group with a growth rate around 0.5%, and between 40 and 42 countries in the high growth group with a growth rate of about 2.7%. The methods differ primarily over the classification of two countries. Clearly, subgroups may be formed within the high growth group. Of note is that some developed countries/OECD such as New Zealand, Switzerland, and Sweden 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. Table 4 shows that our threshold methods give similar clusters when the model is specified in first difference form. These results suggest 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 belong to the same growth club.

7 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 whether or not the regression error variance is equal across units.

References

- ABRAHAM, C., P. CORNILLION, E. MATZNER-LOBER, AND N. MOLINARI (2003): “Unsupervised Curve Clustering Using B-Splines,” *Scandinavian Journal of Statistics*, 30, 581–595.
- ALVAREZ, J., M. BROWNING, AND M. EJRNÆS (2002): “Modelling Income Processes with Lots of Heterogeneity,” 10th International Conferences on Panel Data, Berlin D2-3.
- ANDREWS, D., AND W. PLOBERGER (1994): “Optimal Tests When a Nuisance Parameter is Present only under the Alternative,” *Econometrica*, 62, 1383–1414.
- 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.
- CANER, M., AND B. E. HANSEN (2004): “Instrumental Variable Estimation of a Threshold Model,” *Econometric Theory*, 20, 813–843.
- CHIOU, J.-M., AND P.-L. LI (2007): “Functional Clustering and Identifying Substructure of Longitudinal Data,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 679–699.
- DAVIES, R. B. (1977): “Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative,” *Biometrika*, 64, 247–254.
- DURLAUF, S. N., A. KOURTELLOS, AND C. M. TAN (2005): “Empirics of Growth and Development,” Discussion Papers Series: 0520, Dept. of Economics, Tufts University.
- FRALEY, C., AND A. E. RAFTERY (2002): “Model-Based Clustering, Discriminant Analysis, and Density Estimation,” *Journal of the American Statistical Association*, 97, 611–631.
- GOLDFELD, S., AND R. QUANDT (1973): “The Estimation of Structural Shifts by Switching Regressions,” *Annals of Economic and Social Measurement*, 2, 475–485.
- HALL, P., H. MULLER, AND J. WANG (2006): “Properties of Principal Component Methods for Functional and Longitudinal Data Analysis,” *The Annals of Statistics*, 34, 1493–1517.
- HANSEN, B. E. (1996): “Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis,” *Econometrica*, 64, 413–430.
- (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.
- HESTON, A., R. SUMMERS, AND B. ATEN (2006): *Penn World Table Version 6.2*. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- HSIAO, C., AND M. H. PESARAN (2004): “Random Coefficient Panel Data Models,” *CEifo Working Paper*: 1223.
- HSIAO, C., AND A. K. TAHMISCIOGLU (1997): “A Panel Analysis of Liquidity Constraints and Firm Investment,” *Journal of the American Statistical Association*, 92, 455–465.
- ISLAM, N. (1995): “Growth Empirics: A Panel Approach,” *The Quarterly Journal of Economics*, 110(4), 1127–1170.
- JUÁREZ, M. A., AND M. F. J. STEEL (2008): “Model-based Clustering of non-Gaussian Panel Data Based on Skew- t Distributions,” *Journal of Business and Economic Statistics*, Forthcoming.
- LEE, K., M. H. PESARAN, AND R. SMITH (1997): “Growth and Convergence in a Multi-country Empirical Stochastic Solow Model,” *Journal of Applied Econometrics*, 12, 357–392.
- MANKIW, N. G., D. ROMER, AND D. N. WEIL (1992): “A Contribution to the Empirics of Economic Growth,” *The Quarterly Journal of Economics*, 107, 407–437.
- PESARAN, M. H., AND T. YAMAGATA (2008): “Testing Slope Homogeneity in Large Panels,” *Journal of Econometrics*, 142, 50–93.
- SUN, Y. (2005): “Estimation and Inference in Panel Structure Models,” *Working Paper*: 2005-11, Dept. of Economics, University of California, San Diego.
- SWAMY, P. A. V. B. (1970): “Efficient Inference in a Random Coefficient Regression Model,” *Econometrica*, 38, 311–323.

APPENDIX: PROOFS

To prove Theorem 1, we let $I^0 = (I_1^0, I_2^0)$ be the true group membership and let $I = (I_1, I_2)$ denote group membership other than (I_1^0, I_2^0) . Suppose that the DGP is

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

We will consider the general case where $B_2 - B_1 = \nu T^{-\alpha}$, $0 \leq \alpha < 1/2$ and ν does not depend on T and $|\nu| > 0$. Then $\alpha = 0$ corresponds to the case when $B_2 - B_1 = \nu \neq 0$.

For $j, k = 1, 2$, let N_{kj} be the number of individuals assigned to be in group j by $I = (I_1, I_2)$ when they truly belong to group k and let \hat{B}_{kj} denote the estimator of slope parameter for $i \in I_k^0 \cap I_j$, \hat{B}_j for $i \in I_j$, and \hat{B}_j^0 for $i \in I_j^0$. Let N_j^0 denote the number of individuals truly belonging to group j ; let $N_1 = N_{11} + N_{21}$ and $N_2 = N_{22} + N_{12}$. Without loss of generality, we assume $N_{11} > N_{21}$ and let $N_s = N_s(I, I^0) = N_{12} + N_{21}$ be the number of misclassified units. We then define $\mathcal{I}(C) = \{I : N_s(I, I^0) \geq C/T^{1-2\alpha}\}$.

Let $z_{it} = \tilde{z}_{it} - \bar{z}_i$, where \tilde{z}_{it} can be $\tilde{y}_{it}, \tilde{x}_{it}, \tilde{e}_{it}, \hat{e}_{it}$, and $\bar{z}_i = \frac{1}{T} \sum_{t=1}^T \tilde{z}_{it}$. Then for $j, k = 1, 2$,

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

where $\hat{e}_{it}^{kj} = e_{it} + x_{it}(B_k - \hat{B}_j)$. For $i \in I_k^0 \cap I_j$, $j, k = 1, 2$, we have

$$(\hat{e}_{it}^{kj})^2 = e_{it}^2 + (\hat{B}_j - B_k)' x_{it}' x_{it} (\hat{B}_j - B_k) + 2x_{it}e_{it}(B_k - \hat{B}_j).$$

It is convenient to define

$$H_{kj} = \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^T x_{it}' x_{it}$$

with $H_1 = H_{21} + H_{11}$, $H_2 = H_{12} + H_{22}$, $H_1^0 = H_{12} + H_{11}$, and $H_2^0 = H_{21} + H_{22}$. Also let

$$\Delta_{kj} = \hat{B}_{kj} - B_k = H_{kj}^{-1} \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^T x_{it}e_{it}$$

with $\Delta_k^0 = \hat{B}_k^0 - B_k = (H_k^0)^{-1} \sum_{i \in I_k^0} \sum_{t=1}^T x_{it}e_{it}$. Finally, let

$$\xi_1 = H_1^{-1}[H_{11}\Delta_{11} + H_{21}\Delta_{21}] \quad \text{and} \quad \xi_2 = H_2^{-1}[H_{22}\Delta_{22} + H_{12}\Delta_{12}].$$

Let $S_{NT}(I)$ and $S_{NT}(I^0)$ denote the unweighted sum of squared residuals under I and I^0 , respectively.

$$S_{NT}(I) - S_{NT}(I^0) = \phi_1 + \phi_2 + \phi_3,$$

where, for $j, k = 1, 2$

$$\begin{aligned}\phi_1 &= \sum_{j,k} (\hat{B}_j - B_k)' H_{kj} (\hat{B}_j - B_k), \\ \phi_2 &= 2 \sum_{j,k} (\hat{B}_k^0 - \hat{B}_j)' \sum_{i \in I_k^0 \cap I_j} \sum_{t=1}^T x_{it} e_{it} = 2 \sum_{j,k} (\hat{B}_k^0 - \hat{B}_j)' H_{kj} (\hat{B}_{kj} - B_k), \\ \phi_3 &= - \sum_{j,k} (\hat{B}_k^0 - B_k) H_{kj} (\hat{B}_k^0 - B_k).\end{aligned}$$

Notice that $\hat{B}_1 = H_1^{-1} H_{11} \hat{B}_{11} + H_1^{-1} H_{21} \hat{B}_{21}$ and $\hat{B}_2 = H_2^{-1} H_{22} \hat{B}_{22} + H_2^{-1} H_{12} \hat{B}_{12}$. Recall that $B_2 - B_1 = T^{-\alpha} \nu$. We have

$$\begin{aligned}\hat{B}_1 - B_2 &= H_1^{-1} H_{11} \Delta_{11} - H_1^{-1} H_{21} B_1 + H_1^{-1} H_{21} \Delta_{21} + H_1^{-1} H_{21} B_2, \\ &= H_1^{-1} [H_{11} \Delta_{11} + H_{21} \Delta_{21}] + H_1^{-1} H_{21} (B_2 - B_1) \\ &= \xi_1 + H_1^{-1} H_{21} \nu T^{-\alpha}.\end{aligned}$$

Similarly,

$$\hat{B}_1 - B_2 = \xi_1 - H_1^{-1} H_{11} \nu T^{-\alpha}, \hat{B}_2 - B_2 = \xi_2 - H_2^{-1} H_{12} \nu T^{-\alpha}, \text{ and } \hat{B}_2 - B_1 = \xi_2 + H_2^{-1} H_{22} \nu T^{-\alpha}.$$

We will rewrite $\phi_1 + \phi_2 + \phi_3$ into three terms: one without $\nu T^{-\alpha}$, one that is a linear term in $\nu T^{-\alpha}$, and one that involves a quadratic term of $\nu T^{-\alpha}$. First consider ϕ_1 .

$$\begin{aligned}\phi_1 &= \sum_{j,k=1,2} (\hat{B}_j - B_k)' H_{kj} (\hat{B}_j - B_k) \\ &= T^{-\alpha} \nu' (H_{21} H_1^{-1} H_{11} H_1^{-1} H_{21} + H_{11} H_1^{-1} H_{21} H_1^{-1} H_{11}) \nu T^{-\alpha} \\ &\quad + T^{-\alpha} \nu' (H_{12} H_2^{-1} H_{22} H_2^{-1} H_{21} + H_{22} H_2^{-1} H_{12} H_2^{-1} H_{22}) \nu T^{-\alpha} \\ &\quad + 2\xi_1' H_{11} H_1^{-1} H_{21} \nu T^{-\alpha} - 2\xi_1' H_{21} H_1^{-1} H_{11} \nu T^{-\alpha} + 2\xi_2' H_{22} H_2^{-1} H_{12} \nu T^{-\alpha} \\ &\quad - 2\xi_2' H_{12} H_2^{-1} H_{22} \nu T^{-\alpha} \\ &\quad + \xi_1' (H_{11} + H_{21}) \xi_1 + \xi_2' (H_{22} + H_{12}) \xi_2.\end{aligned}$$

Now $H_1 = H_{11} + H_{21}$ and $H_2 = H_{22} + H_{12}$. Thus, $H_{21} H_1^{-1} H_{11} = H_{11} H_1^{-1} H_{21}$ and $H_{22} H_2^{-1} H_{12} = H_{11} H_2^{-1} H_{12}$. We have

$$\begin{aligned}&T^{-\alpha} \nu' (H_{21} H_1^{-1} H_{11} H_1^{-1} H_{21} + H_{11} H_1^{-1} H_{21} H_1^{-1} H_{11}) \nu T^{-\alpha} \\ &\quad + T^{-\alpha} \nu' (H_{12} H_2^{-1} H_{22} H_2^{-1} H_{21} + H_{22} H_2^{-1} H_{12} H_2^{-1} H_{22}) \nu T^{-\alpha} \\ &= T^{-\alpha} \nu' (H_{21} H_1^{-1} H_{11} + H_{12} H_2^{-1} H_{22}) \nu T^{-\alpha}\end{aligned}$$

and

$$2\xi_1' H_{11} H_1^{-1} H_{21} \nu T^{-\alpha} - 2\xi_1' H_{21} H_1^{-1} H_{11} \nu T^{-\alpha} + 2\xi_2' H_{22} H_2^{-1} H_{12} \nu T^{-\alpha} - 2\xi_2' H_{12} H_2^{-1} H_{22} \nu T^{-\alpha} = 0.$$

We also have

$$\begin{aligned}\xi_1' H_{11} \xi_1 + \xi_1' H_{21} \xi_1 + \xi_2' H_{22} \xi_2 + \xi_2' H_{12} \xi_2 &= [H_{11} \Delta_{11} + H_{21} \Delta_{21}]' H_1^{-1} [H_{11} \Delta_{11} + H_{21} \Delta_{21}] \\ &\quad + [H_{22} \Delta_{22} + H_{12} \Delta_{12}]' H_2^{-1} [H_{22} \Delta_{22} + H_{12} \Delta_{12}].\end{aligned}$$

Together with these results, we have

$$\begin{aligned}\phi_1 &= T^{-\alpha}\nu'(H_{21}H_1^{-1}H_{11} + H_{12}H_2^{-1}H_{22})\nu T^{-\alpha} + [H_{11}\Delta_{11} + H_{21}\Delta_{21}]'H_1^{-1}[H_{11}\Delta_{11} + H_{21}\Delta_{21}] + \\ &\quad [H_{22}\Delta_{22} + H_{12}\Delta_{12}]'H_2^{-1}[H_{22}\Delta_{22} + H_{12}\Delta_{12}].\end{aligned}$$

Next, consider ϕ_2 . Notice that

$$\begin{aligned}(\hat{B}_1^0 - \hat{B}_1)H_{11}\Delta_{11} &= -[(\hat{B}_1 - B_1) - (\hat{B}_1^0 - B_1)]'H_{11}\Delta_{11} \\ &= -(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}H_{11}\Delta_{11} - T^{-\alpha}\nu'H_{21}H_1^{-1}H_{11}\Delta_{11} + \Delta_1^{0'}H_{11}\Delta_{11}, \\ (\hat{B}_2^0 - \hat{B}_1)H_{21}\Delta_{21} &= -(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}H_{21}\Delta_{21} + T^{-\alpha}\nu'H_{11}H_1^{-1}H_{21}\Delta_{21} + \Delta_2^{0'}H_{21}\Delta_{21}, \\ (\hat{B}_2^0 - \hat{B}_2)H_{22}\Delta_{22} &= -(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}H_{22}\Delta_{22} + T^{-\alpha}\nu'H_{12}H_2^{-1}H_{22}\Delta_{22} + \Delta_2^{0'}H_{22}\Delta_{22}, \\ (\hat{B}_1^0 - \hat{B}_2)H_{12}\Delta_{12} &= -(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}H_{12}\Delta_{12} - T^{-\alpha}\nu'H_{22}H_2^{-1}H_{12}\Delta_{12} + \Delta_1^{0'}H_{12}\Delta_{12}.\end{aligned}$$

Adding up these terms, we obtain

$$\begin{aligned}\phi_2 &= -2(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}(H_{11}\Delta_{11} + H_{21}\Delta_{21}) \\ &\quad -2(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}(H_{22}\Delta_{22} + H_{12}\Delta_{12}) \\ &\quad -2T^{-\alpha}\nu'(H_{21}H_1^{-1}H_{11}\Delta_{11}) + 2T^{-\alpha}\nu'(H_{11}H_1^{-1}H_{21}\Delta_{21}) \\ &\quad + 2T^{-\alpha}\nu'(H_{12}H_2^{-1}H_{22}\Delta_{22}) - 2T^{-\alpha}\nu'(H_{22}H_2^{-1}H_{12}\Delta_{12}) \\ &\quad + 2\Delta_1^{0'}H_1^0\Delta_1^0 + 2\Delta_2^{0'}H_2^0\Delta_2^0.\end{aligned}$$

Finally, consider ϕ_3 . Since $H_1^0 = H_{11} + H_{12}$ and $H_2^0 = H_{22} + H_{21}$,

$$\phi_3 = -\Delta_1^{0'}H_1^0\Delta_1^0 - \Delta_2^{0'}H_2^0\Delta_2^0.$$

Collecting the results, $S_{NT}(I) - S_{NT}(I^0) = \phi_1 + \phi_2 + \phi_3 = \psi_1 + \psi_2 + \psi_3$, where

$$\begin{aligned}\psi_1 &= T^{-\alpha}\nu'(H_{21}H_1^{-1}H_{11} + H_{12}H_2^{-1}H_{22})\nu T^{-\alpha} \\ \psi_2 &= -2T^{-\alpha}\nu'(H_{21}H_1^{-1}H_{11}\Delta_{11}) + 2T^{-\alpha}\nu'(H_{11}H_1^{-1}H_{21}\Delta_{21}) \\ &\quad + 2T^{-\alpha}\nu'(H_{12}H_2^{-1}H_{22}\Delta_{22}) - 2T^{-\alpha}\nu'(H_{22}H_2^{-1}H_{12}\Delta_{12}) \\ \psi_3 &= -(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}(H_{11}\Delta_{11} + H_{21}\Delta_{21}) + \Delta_1^{0'}H_1^0\Delta_1^0 \\ &\quad - (H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}(H_{22}\Delta_{22} + H_{12}\Delta_{12}) + \Delta_2^{0'}H_2^0\Delta_2^0.\end{aligned}$$

Lemma A1 Under Assumptions 1 and 4,

(a) For each i , $T^{-1/2} \sum_{t=1}^T x_{it}e_{it} \xrightarrow{d} N(0, \sigma_i^2 Q_i)$ as $T \rightarrow \infty$.

(b) Let I_l denote a subset of the whole sample and let N_l be the number of units in I_l . If $N_l \geq 1$, $(N_l T)^{-1/2} \sum_{i \in I_l} \sum_{t=1}^T x_{it}e_{it} \xrightarrow{d} N(0, Q_l)$ as $(N_l, T) \rightarrow \infty$ jointly, where $Q_l = \lim_{N_l \rightarrow \infty} N_l^{-1} \sum_{i \in I_l} \sigma_i^2 Q_i$.

Proof of Lemma A1: Lemma A1(a) follows from Assumption 1 by central limit theorems used in classical linear regression models. Since e_{it} and is cross-sectionally independent and $E\|\hat{Q}_i\| < K$ for all i , Lemma A1(b) directly follows from Assumptions 1 and 4. See Lemma 2 in Pesaran and Yamagata (2008). \square

Lemma A2 Let I_l denote a subset of the whole sample and let N_l be the number of units in I_l . Under Assumptions 1,2 and 4,

(a) $\hat{\sigma}_i^2 - \sigma_i^2 = O_p(T^{-1/2})$ for each i .

(b) $N_l^{-1/2} \sum_{i \in I_l} \hat{\sigma}_i^{-2} \left(T^{-1/2} \sum_{t=1}^T x_{it} e_{it} \right) = O_p(1) + O_p(\sqrt{N_l/T})$ if $N_l \geq 1$.

(c) $\left(N_l^{-1} \sum_{i \in I_l} \hat{\sigma}_i^{-2} \hat{Q}_i \right)^{-1} = \left(N_l^{-1} \sum_{i \in I_l} \sigma_i^{-2} \hat{Q}_i \right)^{-1} + O_p(T^{-1/2})$ if $N_l \geq 1$.

Proof of Lemma A2: (a) directly follows from the fact that $\hat{\beta}_i$ is \sqrt{T} consistent. For (b) and (c), see Equations (A.10) and (A.11) in Pesaran and Yamagata (2008). \square

Lemma A3 Under Assumptions 1–4,

(a) For $j, k = 1, 2$ and $j \neq k$, if $N_{jk} \geq 1$, $H_{jk}/(N_{jk}T) = O_p(1)$ and $H_{jk}\Delta_{jk}/\sqrt{N_{jk}T} = O_p(1)$. Otherwise, $H_{jk}/(N_sT) = 0$ and $H_{jk}\Delta_{jk}/\sqrt{N_sT} = 0$

(b) For $j = 1, 2$, $H_j^0/(N_j^0T) = O_p(1)$ and $H_j^0\Delta_j^0/\sqrt{N_j^0T} = O_p(1)$.

(c) For $j = 1, 2$, $H_j/(N_jT) = O_p(1)$ and $(H_{jk}\Delta_{jk} + H_{jj}\Delta_{jj})/\sqrt{N_jT} = O_p(1)$.

Proof of Lemma A3: First, consider (a). When $N_{jk} > 0$, $H_{jk}/(N_{jk}T) = O_p(1)$ directly follows from Assumption 4. Since $\Delta_{jk} = H_{jk}^{-1} \sum_{i \in I_j^0 \cap I_k} \sum_{t=1}^T x_{it} e_{it}$ and by Lemma A1,

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

When $I_k^0 \cap I_j = \emptyset$, $N_{jk} = 0$. It follows that $H_{jk} = 0$ and $\sum_{i \in I_j^0 \cap I_k} \sum_{t=1}^T x_{it} e_{it} = 0$. Analogously, (b) holds by Assumptions 1–4, Lemma A1, and the fact that $H_j^0 \Delta_j^0 = \sum_{i \in I_j^0} \sum_{t=1}^T x_{it} e_{it}$. For (c), the first claim holds because $H_1 = H_{11} + H_{21}$, $H_2 = H_{22} + H_{12}$, $N_1 = N_{11} + N_{21}$, and $N_2 = N_{22} + N_{12}$. Because $(H_{jk}\Delta_{jk} + H_{jj}\Delta_{jj}) = \sum_{i \in I_j} \sum_{t=1}^T x_{it} e_{it}$, the second claim follows by Lemma A1. \square

Lemma A4 Under Assumptions 1–4, $\inf_{I \in \mathcal{I}(C)} N_s^{-1} T^{-1+2\alpha} \psi_1 > 0$.

Proof of Lemma A4: Notice that

$$(N_s^{-1}T^{-1+2\alpha})\psi_1 = \nu'T^{-1}N_s^{-1}(H_{21}H_1^{-1}H_{11} + H_{12}H_2^{-1}H_{22})\nu. \quad (15)$$

Lemma A4 follows if we can show that the minimal eigenvalue of $T^{-1}N_s^{-1}((H_{21}H_1^{-1}H_{11} + H_{12}H_2^{-1}H_{22}))$ is bounded away from zero uniformly in $\mathcal{I}(C)$. Notice that

$$\begin{aligned} & T^{-1}N_s^{-1}(H_{21}H_1^{-1}H_{11} + H_{12}H_2^{-1}H_{22}) \\ &= \frac{H_{21}}{N_sT} \left(\frac{H_1}{N_1T} \right)^{-1} \frac{H_{11}}{N_1T} + \frac{H_{12}}{N_sT} \left(\frac{H_2}{N_2T} \right)^{-1} \frac{H_{22}}{N_2T}. \end{aligned}$$

By $N_s = N_{21} + N_{12}$, $N_1 = N_{11} + N_{21}$, $N_2 = N_{22} + N_{12}$ and Assumption 4, to show (15) is strictly positive is then equivalent to show at least the minimal eigenvalue of $\frac{H_{21}}{N_sT}$ or $\frac{H_{12}}{N_sT}$ is bounded away from zero uniformly in $\mathcal{I}(C)$.

Consider three possible cases for $I \in \mathcal{I}(C)$: (a) $N_{21}/N_s \rightarrow 0$ (and $N_{12}/N_s \rightarrow 1$), (b) $N_{21}/N_s \rightarrow 1$ (and $N_{12}/N_s \rightarrow 0$), and (c) neither N_{21}/N_s nor N_{12}/N_s converges to zero. In case (a), while $\frac{H_{12}}{N_sT} \rightarrow 0$, $\frac{H_{21}}{N_sT} \rightarrow \frac{H_{21}}{N_{21}T}$ and its minimal eigenvalue bounded away from zero by Assumption 4. Analogously, in case (b) $\frac{H_{21}}{N_sT} \rightarrow 0$ but $\frac{H_{12}}{N_sT} \rightarrow \frac{H_{12}}{N_{12}T}$ has the minimal eigenvalue bounded away from zero. In case (c), both N_{21}/N_s and N_{12}/N_s are positive numbers between 0 and 1. By Assumption 4, the minimal eigenvalue of $T^{-1}N_s^{-1}(H_{21}H_1^{-1}H_{11} + H_{12}H_2^{-1}H_{22})$ is then positive and bounded away from zero. \square

Lemma A5 Under Assumptions 1-4, $N_s^{-1}T^{-1+2\alpha}\psi_2 = o_p(1)$ uniformly in $\mathcal{I}(C)$.

Proof of Lemma A5: Notice that

$$\begin{aligned} (N_s^{-1}T^{-1+2\alpha})\psi_2 &= -2\nu'T^{-1/2+\alpha} \left[N_s^{-1}T^{-1/2} (H_{21}H_1^{-1}H_{11}\Delta_{11} - H_{11}H_1^{-1}H_{21}\Delta_{21}) \right] \\ &\quad + 2\nu'T^{-1/2+\alpha} \left[N_s^{-1}T^{-1/2} (H_{12}H_2^{-1}H_{22}\Delta_{22} - H_{22}H_2^{-1}H_{12}\Delta_{12}) \right]. \end{aligned}$$

We first show that the first term is $o_p(1)$ regardless of whether $N_{21} = 0$ or not. When $N_{21} = 0$, $H_{21}H_1^{-1}H_{11}\Delta_{11} - H_{11}H_1^{-1}H_{21}\Delta_{21} = 0$. When $N_{21} > 0$, $(N_s^{-1}T^{-1/2}H_{21}H_1^{-1}H_{11}\Delta_{11}) = N^{-1/2}(N_{21}N_s^{-1})O_p(1)$ and $(N_s^{-1}T^{-1/2}H_{11}H_1^{-1}H_{21}\Delta_{21}) = N_{21}^{1/2}N_s^{-1}O_p(1)$ by Lemma A3. Because $-1/2 + \alpha < 0$,

$$-2\nu'T^{-1/2+\alpha} \left[N_s^{-1}T^{-1/2} (H_{21}H_1^{-1}H_{11}\Delta_{11} - H_{11}H_1^{-1}H_{21}\Delta_{21}) \right] = o_p(1).$$

Similarly,

$$-2\nu'T^{-1/2+\alpha} \left[N_s^{-1}T^{-1/2} (H_{12}H_2^{-1}H_{22}\Delta_{22} - H_{22}H_2^{-1}H_{12}\Delta_{12}) \right] = o_p(1).$$

Therefore, $N_s^{-1}T^{-1+2\alpha}\psi_2 = o_p(1)$ uniformly in $\mathcal{I}(C)$. \square

Lemma A6 Under Assumptions 1-4, $(N_s^{-1}T^{-1+2\alpha})\psi_3 = o_p(1)$ uniformly in $\mathcal{I}(C)$.

Proof of Lemma A6: By Lemma A3(c), $(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}(H_{11}\Delta_{11} + H_{21}\Delta_{21}) = O_p(1)$ and $(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}(H_{22}\Delta_{22} + H_{12}\Delta_{12}) = O_p(1)$ uniformly in $\mathcal{I}(C)$. Also, by Lemma A3(b), $\Delta_1^{0'}H_1^0\Delta_1^0 + \Delta_2^{0'}H_2^0\Delta_2^0 = O_p(1)$ uniformly in $\mathcal{I}(C)$. Thus,

$$\begin{aligned} (N_s^{-1}T^{-1+2\alpha})\psi_3 &= (N_s^{-1}T^{-1+2\alpha})(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}(H_{11}\Delta_{11} + H_{21}\Delta_{21}) \\ &\quad + (N_s^{-1}T^{-1+2\alpha})(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}(H_{22}\Delta_{22} + H_{12}\Delta_{12}) \\ &\quad + (N_s^{-1}T^{-1+2\alpha}) \left[\Delta_1^{0'}H_1^0\Delta_1^0 + \Delta_2^{0'}H_2^0\Delta_2^0 \right] \\ &= o_p(1). \quad \square \end{aligned}$$

Proof of Theorem 1: Define $\mathcal{I}(C) = \{I : N_s(I, I^0) \geq C/T^{1-2\alpha}\}$. We want to show that for any ε and $C > 0$, $P(N_s > \frac{C}{T^{1-2\alpha}}) < \varepsilon$. Since $P(N_s > \frac{C}{T^{1-2\alpha}}) < P(\hat{I} \in \mathcal{I}(C))$, by the definition of \hat{I} , it suffices to show that

$$P\left(\sup_{I \in \mathcal{I}(C)} SSR(I^0) \geq SSR(I)\right) < \varepsilon.$$

Notice that

$$P\left(\sup_{I \in \mathcal{I}(C)} SSR(I^0) \geq SSR(I)\right) \leq P\left(\sup_{I \in \mathcal{I}(C)} \frac{-(\psi_2 + \psi_3)}{N_s T^{1-2\alpha}} \geq \inf_{I \in \mathcal{I}(C)} \frac{\psi_1}{N_s T^{1-2\alpha}}\right).$$

By Lemma A4: $\inf_{I \in \mathcal{I}(C)} \psi_1/(N_s T^{1-2\alpha})$ is strictly positive. By Lemmas A5 and A6: $\sup_{I \in \mathcal{I}(C)} -(\psi_2 + \psi_3)/(N_s T^{1-2\alpha}) = o_p(1)$. Together with these lemmas,

$$P\left(\sup_{I \in \mathcal{I}(C)} \frac{-(\psi_2 + \psi_3)}{N_s T^{1-2\alpha}} \geq \inf_{I \in \mathcal{I}(C)} \frac{\psi_1}{N_s T^{1-2\alpha}}\right) < \varepsilon,$$

and, therefore, Theorem 1 follows. \square

Proof of Corollary 1 Let $H_{kj}(\hat{\sigma}^2) = \sum_{i \in I_k^0 \cap I_j} \hat{\sigma}_i^{-2} \sum_{t=1}^T x'_{it} x_{it}$ and $\Xi_{kj} = \sum_{i \in I_k^0 \cap I_j} \hat{\sigma}_i^{-2} \sum_{t=1}^T x'_{it} e_{it}$, for $j, k = 1, 2$. Also, let $\Xi_1 = \Xi_{11} + \Xi_{21}$, $\Xi_2 = \Xi_{22} + \Xi_{12}$, $\Xi_1^0 = \Xi_{11} + \Xi_{12}$, and $\Xi_2^0 = \Xi_{22} + \Xi_{21}$. When $\delta_i = \hat{\sigma}_i^{-2}$, we have

$$(N_s^{-1}T^{-1+2\alpha}) [S_{NT}(I) - S_{NT}(I^0)] = \psi'_1 + \psi'_2 + \psi'_3,$$

where ψ'_1 , ψ'_2 , and ψ'_3 are the counterparts of ψ_1 , ψ_2 , and ψ_3 , respectively, and

$$\psi'_1 = (N_s^{-1}T^{-1}) [\nu'H_{11}H_1^{-1}(H_{21}(\hat{\sigma}^2) + H_{21}H_{11}^{-1}H_{11}(\hat{\sigma}^2)H_{11}^{-1}H_{21})H_1^{-1}H_{11}\nu] \quad (16)$$

$$+ (N_s^{-1}T^{-1}) [\nu'H_{22}H_2^{-1}(H_{12}(\hat{\sigma}^2) + H_{12}H_{22}^{-1}H_{22}(\hat{\sigma}^2)H_{22}^{-1}H_{12})H_2^{-1}H_{22}\nu] \quad (17)$$

$$\psi'_2 = (N_s^{-1}T^{-1+\alpha}) [2(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}H_{11}H_1^{-1} \cdot (H_{21}(\hat{\sigma}^2) - H_{21}H_{11}^{-1}H_{11}(\hat{\sigma}^2))\nu] \quad (18)$$

$$+ (N_s^{-1}T^{-1+\alpha}) [2(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}H_{22}H_2^{-1} \cdot (H_{12}(\hat{\sigma}^2) - H_{12}H_{22}^{-1}H_{22}(\hat{\sigma}^2))\nu] \quad (19)$$

$$+ 2(N_s^{-1}T^{-1+\alpha}) [\nu'H_{11}H_1^{-1}(\Xi_{21} - H_{21}H_{11}^{-1}\Xi_{11})] \quad (20)$$

$$- 2(N_s^{-1}T^{-1+\alpha}) [\nu'H_{22}H_2^{-1}(\Xi_{12} - H_{12}H_{22}^{-1}\Xi_{22})], \quad (21)$$

$$\psi'_3 = (N_s^{-1}T^{-1+2\alpha}) [(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}H_1(\hat{\sigma}^2)H_1^{-1}(H_{11}\Delta_{11} + H_{21}\Delta_{21})] \quad (22)$$

$$+ (N_s^{-1}T^{-1+2\alpha}) [(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}H_2(\hat{\sigma}^2)H_2^{-1}(H_{22}\Delta_{22} + H_{12}\Delta_{12})] \quad (23)$$

$$+ (N_s^{-1}T^{-1+2\alpha}) [\Delta_1^{0'}H_1(\hat{\sigma}^2)\Delta_1^0 + \Delta_2^{0'}H_2(\hat{\sigma}^2)\Delta_2^0] \quad (24)$$

$$- 2(N_s^{-1}T^{-1+2\alpha}) [(H_{11}\Delta_{11} + H_{21}\Delta_{21})'H_1^{-1}\Xi_1] \quad (25)$$

$$- 2(N_s^{-1}T^{-1+2\alpha}) [(H_{22}\Delta_{22} + H_{12}\Delta_{12})'H_2^{-1}\Xi_2] \quad (26)$$

$$+ 2(N_s^{-1}T^{-1+2\alpha}) [\Delta_1^{0'}\Xi_1^0 + \Delta_2^{0'}\Xi_2^0]. \quad (27)$$

By Lemmas A2(c), if $N_{kj} \geq 1$, $H_{kj}(\hat{\sigma}^2)/(N_{kj}T)$ converges to $H_{kj}(\sigma^2)/(N_{kj}T)$ in probability. Together with the results in Lemma A3, (16) and (17) are of $O_p(1)$ and positive and bounded away from zero uniformly in $I \in \mathcal{I}(C)$. Similarly, (18), (19), (22), (23), and (24) are of $o_p(1)$. Also, by Lemmas A2(b), if $N_{kj} \geq 1$, $\Xi_{kj}/\sqrt{N_{kj}T} = O_p(1) + O_p(\sqrt{N_{kj}/T})$ and $\Xi_{kj}/(N_{kj}T) = o_p(1)$. Together with Assumption 4 and the results in Lemma A3, (20) and (21) are of $o_p(1)$ uniformly in $I \in \mathcal{I}(C)$. Finally, (25), (26), and (27) are of $o_p(1)$ when $N^{1/2}T^{-3/2+2\alpha} \rightarrow 0$ uniformly in $I \in \mathcal{I}(C)$. Thus, we obtain the counterparts of Lemmas A4–A6 when $\delta_i = \hat{\sigma}_i^{-2}$. Analogous to the proof of Theorem 1, the desired result follows when $N^{1/2}T^{-3/2+2\alpha} \rightarrow 0$. \square

Lemma A7: Let $B_\omega = \omega B_1 + (1 - \omega)B_2$, where $\omega = \text{plim}_{N \rightarrow \infty} \frac{\sum_{i \in I_1} \delta_i \hat{Q}_i}{\sum_{i=1}^N \delta_i \hat{Q}_i}$ and $\delta_i = 1$ or $1/\hat{\sigma}_i^2$. Suppose that Assumptions 1–4 hold. Then $\sqrt{NT}(\hat{B}_\omega - B_\omega) = O_p(1) + O_p(\sqrt{N/T})$.

Proof of Lemma A7: Let $\hat{B}_j = \frac{\sum_{i \in I_j^0} \sum_{t=1}^T \delta_i x'_{it} y_{it}}{\sum_{i \in I_j^0} \sum_{t=1}^T \delta_i x'_{it} x_{it}}$ with $\delta_i = 1$ or $1/\hat{\sigma}_i^2$. By direct

calculations,

$$\begin{aligned}
& \sqrt{NT} \left[\hat{B}_\omega - (\omega B_1 + (1 - \omega)B_2) \right] \\
&= \sqrt{NT} \left[\frac{\sum_{i \in I_1^0} \sum_{t=1}^T \delta_i x'_{it} y_{it} + \sum_{i \in I_2^0} \sum_{t=1}^T \delta_i x'_{it} y_{it}}{\sum_{i=1}^N \sum_{t=1}^T \delta_i x'_{it} x_{it}} - (\omega B_1 + (1 - \omega)B_2) \right] \\
&= \sqrt{NT} \left(\frac{\sum_{i \in I_1^0} \delta_i \hat{Q}_i}{\sum_{i=1}^N \delta_i \hat{Q}_i} \cdot \frac{\sum_{i \in I_1^0} \sum_{t=1}^T \delta_i x'_{it} y_{it}}{\sum_{i \in I_1^0} \sum_{t=1}^T \delta_i x'_{it} x_{it}} - \omega B_1 \right) \\
&\quad + \sqrt{NT} \left(\frac{\sum_{i \in I_2^0} \delta_i \hat{Q}_i}{\sum_{i=1}^N \delta_i \hat{Q}_i} \cdot \frac{\sum_{i \in I_2^0} \sum_{t=1}^T \delta_i x'_{it} y_{it}}{\sum_{i \in I_2^0} \sum_{t=1}^T \delta_i x'_{it} x_{it}} - (1 - \omega)B_2 \right) \\
&= \sqrt{\frac{N}{N_1^0}} \left[\omega \sqrt{N_1^0 T} (\hat{B}_1 - B_1) \right] + \sqrt{\frac{N}{N_2^0}} \left[(1 - \omega) \sqrt{N_2^0 T} (\hat{B}_2 - B_2) \right] + o_p(1).
\end{aligned}$$

To prove Lemma A7 with $\delta_i = 1/\hat{\sigma}_i^2$, it suffices to show that $(N_j^0 T)^{1/2}(\hat{B}_j - B_j) = O_p(1) + O_p\left(\sqrt{N_j^0/T}\right)$. By Assumption 4, and Lemma A1(c), we have $N_j^{-1} \sum_{i \in I_j^0} \sigma_i^{-2} \hat{Q}_i = \hat{Q}^j \xrightarrow{p} Q^j$, and

$$\text{avar} \left(\sqrt{N_j T} \cdot \sum_{i \in I_j^0} \sigma_i^{-2} \sum_{t=1}^T x_{it} e_{it} \right) = \lim_{N_j \rightarrow \infty} \left(N_j^{-1} \sum_{i \in I_j^0} \hat{Q}_i \right) = O(1).$$

By Results (b) and (c) in Lemma A2, we obtain

$$\begin{aligned}
(N_j^0 T)^{1/2} (\hat{B}_j - B_j) &= \frac{(N_j^0 T)^{-1/2} \sum_{i \in I_j} \sum_{t=1}^T \hat{\sigma}_i^{-2} x'_{it} e_{it}}{(N_j^0)^{-1} \sum_{i=1}^N \hat{\sigma}_i^{-2} \hat{Q}_i} \\
&= \frac{(N_j^0 T)^{-1/2} \sum_{i \in I_j} \sum_{t=1}^T \sigma_i^{-2} x'_{it} e_{it}}{(N_j^0)^{-1} \sum_{i=1}^N \sigma_i^{-2} \hat{Q}_i} + O_p\left(\sqrt{N_j^0/T}\right) \\
&= O_p(1) + O_p\left(\sqrt{N_j^0/T}\right).
\end{aligned}$$

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

Lemma A8: *Suppose that the data are generated by (1) and Assumptions 1–4 hold. Then,*

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

Proof of Lemma A8: By Lemma A7, for $i \in I_1$, $\sqrt{T}\hat{\tau}_i$ can be expressed as

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

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

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

By Assumptions 2 and 4(a), $\hat{\sigma}^2 \hat{Q}_i^{-1} \xrightarrow{p} \sigma^2 Q_i^{-1}$ and the result follows. \square

Proof of Theorem 2: Let Γ^0 be a set of threshold values that will achieve correct clustering. Let $\gamma_{\min}^0 = \min_{\gamma} \{\gamma : \gamma \in \Gamma^0\}$ and $\gamma_{\max}^0 = \max_{\gamma} \{\gamma : \gamma \in \Gamma^0\}$. Then for any $\gamma^0 \in [\gamma_{\min}^0, \gamma_{\max}^0]$,

$$F(\gamma^0) = P(q_i^0 < \gamma^0) = \frac{\sum_{i=1}^N 1(q_i^0 < \gamma_{\max}^0)}{N} = \frac{N_1^0}{N}.$$

First, we prove Theorem 2 with $\hat{q}_i = \hat{\beta}_i$. The case with $\hat{q}_i = \hat{\tau}_i$ is similar. Consider the following cases:

I. Γ^0 known, q_i estimated Suppose we know Γ^0 but not q_i^0 . Let $\frac{\hat{c}_i}{\sqrt{T}} = \hat{q}_i - q_i^0$

$$\begin{aligned} \frac{N_s}{N} &= \frac{1}{N} \sum_{i \in I_1^0} 1(\hat{q}_i > \gamma_{\max}^0) + \frac{1}{N} \sum_{i \in I_2^0} 1(\hat{q}_i < \gamma_{\min}^0) \\ &= \frac{1}{N} \sum_{i \in I_1^0} 1(\gamma_{\max}^0 < q_i^0 + \frac{\hat{c}_i}{\sqrt{T}}) + \frac{1}{N} \sum_{i \in I_2^0} 1(\gamma_{\min}^0 > q_i^0 + \frac{\hat{c}_i}{\sqrt{T}}) \\ &= \frac{1}{N} \sum_{i \in I_1^0} 1(\frac{\hat{c}_i}{\sqrt{T}} > \gamma_{\max}^0 - q_i^0) + \frac{1}{N} \sum_{i \in I_2^0} 1(\frac{\hat{c}_i}{\sqrt{T}} < \gamma_{\min}^0 - q_i^0). \end{aligned}$$

Note that $\hat{c}_i > 0$ if $\hat{q}_i > q_i^0$ and $i \in I_1^0$ and $\hat{c}_i < 0$ if $\hat{q}_i < q_i^0$ and $i \in I_2^0$. Now $\gamma_{\max}^0 = B_2^0$ and $\gamma_{\min}^0 = B_1^0$ for Method 1 with $q_i^0 = B_1^0$ for those $i \in I_1^0$ and $q_i^0 = B_2^0$ for those $i \in I_2^0$,

$$\begin{aligned} \frac{N_s}{N} &= \frac{1}{N} \sum_{i \in I_1^0} 1(\frac{\hat{c}_i}{\sqrt{T}} > B_2^0 - q_i^0) + \frac{1}{N} \sum_{i \in I_2^0} 1(\frac{\hat{c}_i}{\sqrt{T}} < B_1^0 - q_i^0) \\ &= \frac{1}{N} \sum_{i \in I_1^0} 1(\frac{\hat{c}_i}{\sqrt{T}} > B_2^0 - B_1^0) + \frac{1}{N} \sum_{i \in I_2^0} 1(\frac{\hat{c}_i}{\sqrt{T}} < B_1^0 - B_2^0) \\ &= \frac{1}{N} \sum_{i \in I_1^0} 1(\frac{\hat{c}_i}{\sqrt{T}} > \nu T^{-\alpha}) + \frac{1}{N} \sum_{i \in I_2^0} 1(\frac{\hat{c}_i}{\sqrt{T}} < -\nu T^{-\alpha}) \\ &= \frac{1}{N} \sum_{i \in I_1^0} 1(T^{-1/2+\alpha} \hat{c}_i > \nu) + \frac{1}{N} \sum_{i \in I_2^0} 1(T^{-1/2+\alpha} \hat{c}_i < -\nu). \end{aligned}$$

We have $N_s/N = O_p(T^{-1/2+\alpha})$.

Before we turn to the case when Γ^0 and q_i^0 are both unknown, it is useful to consider two intermediate cases. First, suppose we use $\hat{\gamma}$ instead of $\gamma^0 \in \Gamma^0$. Let $\hat{\gamma}_{\min}$ and $\hat{\gamma}_{\max}$ be defined such that for any $\tilde{\gamma} \in [\hat{\gamma}_{\min}, \hat{\gamma}_{\max}]$, if the ordered observation $[i^*] \in I_1(\hat{\gamma})$ but $[i^* + 1] \in I_2(\hat{\gamma})$, then $[i^*] \in I_1(\tilde{\gamma})$ and $[i^* + 1] \in I_2(\tilde{\gamma})$.

Notice that because q_i^0 's are observed, we can treat it as a special case of Theorem 1 and directly assign individuals into groups without uncertainty. Thus, $\hat{\gamma}_{\min} = \min_{\hat{q}_i} \{\hat{q}_i | \hat{q}_i \geq B_1\}$ and $\hat{\gamma}_{\max} = \max_{\hat{q}_i} \{\hat{q}_i | \hat{q}_i < B_2\}$. Under the assumption that $\hat{\gamma}_{\min} - \gamma_{\min}^0$ and $\hat{\gamma}_{\max} - \gamma_{\max}^0$ are of order $O_p(N^{-1}T^{-1/2})$, we can let $\frac{\hat{d}_{\max}}{N\sqrt{T}} = \gamma_{\max}^0 - \hat{\gamma}_{\max}$ and $\frac{\hat{d}_{\min}}{N\sqrt{T}} = \gamma_{\min}^0 - \hat{\gamma}_{\min}$. Note that \hat{d}_{\max} and \hat{d}_{\min} do not depend on i . Also, misclassification arises only if $\hat{\gamma}_{\min} > \gamma_{\min}^0$ or $\hat{\gamma}_{\max} < \gamma_{\max}^0$.

Now suppose Γ^0 is unknown, and we use a ‘‘perturbed’’ q_i as indicator variable. Let $\tilde{q}_i = q_i^0 + c_i T^{-1/2}$ where $c_i \neq 0$ is observed, and is not necessarily the same across i .

$$\begin{aligned}
\frac{\tilde{N}_s}{N} &\leq \frac{1}{N} \sum_{i \in I_1^0; \hat{\gamma}_{\max} < \gamma_{\max}^0} 1(\tilde{q}_i > \hat{\gamma}_{\max}) + \frac{1}{N} \sum_{i \in I_2^0; \hat{\gamma}_{\min} > \gamma_{\min}^0} 1(\tilde{q}_i < \hat{\gamma}_{\min}) \\
&= \frac{1}{N} \sum_{i \in I_1^0; \hat{d}_{\max} > 0} 1\left(\tilde{q}_i > \gamma_{\max}^0 - \frac{\hat{d}_{\max}}{N\sqrt{T}}\right) + \frac{1}{N} \sum_{i \in I_2^0; \hat{d}_{\min} < 0} 1\left(\tilde{q}_i < \gamma_{\min}^0 - \frac{\hat{d}_{\min}}{N\sqrt{T}}\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0; \hat{d}_{\max} > 0} 1\left(\frac{\hat{d}_{\max}}{N\sqrt{T}} + \frac{c_i}{\sqrt{T}} > \gamma_{\max}^0 - q_i^0\right) + \frac{1}{N} \sum_{i \in I_2^0; \hat{d}_{\min} < 0} 1\left(\frac{\hat{d}_{\min}}{N\sqrt{T}} + \frac{c_i}{\sqrt{T}} < \gamma_{\min}^0 - q_i^0\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0; \hat{d}_{\max} > 0} 1\left(\frac{\hat{d}_{\max}}{N\sqrt{T}} + \frac{c_i}{\sqrt{T}} > B_2^0 - B_1^0\right) + \frac{1}{N} \sum_{i \in I_2^0; \hat{d}_{\min} < 0} 1\left(\frac{\hat{d}_{\min}}{N\sqrt{T}} + \frac{c_i}{\sqrt{T}} < B_1^0 - B_2^0\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0; \hat{\gamma}_{\max} < \gamma_{\max}^0} 1\left(\frac{\hat{d}_{\max}}{N\sqrt{T}} + \frac{c_i}{\sqrt{T}} > \nu T^{-\alpha}\right) + \frac{1}{N} \sum_{i \in I_2^0; \hat{d}_{\min} < 0} 1\left(\frac{\hat{d}_{\min}}{N\sqrt{T}} + \frac{c_i}{\sqrt{T}} < -\nu T^{-\alpha}\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0; \hat{d}_{\max} > 0} 1\left(\frac{\hat{d}_{\max}}{NT^{1/2-\alpha}} + \frac{c_i}{T^{1/2-\alpha}} > \nu\right) + \frac{1}{N} \sum_{i \in I_2^0; \hat{d}_{\min} < 0} 1\left(\frac{\hat{d}_{\min}}{NT^{1/2-\alpha}} + \frac{c_i}{T^{1/2-\alpha}} < -\nu\right).
\end{aligned}$$

Since $0 \leq \alpha < 1/2$, we have $1/2 - \alpha > 0$ and therefore \tilde{N}_s/N converges to zero.

We are now in a position to consider the case when \hat{q}_i as a transition variable instead of q_i^0 or \tilde{q}_i .

II. Γ^0 and q_i^0 both unknown Because $\hat{q}_i = q_i^0 + \frac{\hat{c}_i}{\sqrt{T}}$, we have

$$\begin{aligned}
\frac{N_s}{N} &\leq \frac{1}{N} \sum_{i \in I_1^0} 1(\hat{q}_i > \hat{\gamma}_{\max}) + \frac{1}{N} \sum_{i \in I_2^0} 1(\hat{q}_i < \hat{\gamma}_{\min}) \\
&= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{\max}}{N\sqrt{T}} > \gamma_{\max}^0 - q_i^0\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{\min}}{N\sqrt{T}} < \gamma_{\min}^0 - q_i^0\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{\max}}{N\sqrt{T}} > B_2^0 - B_1^0\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{\min}}{N\sqrt{T}} < B_1^0 - B_2^0\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{\max}}{N\sqrt{T}} > \nu T^{-\alpha}\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{\min}}{N\sqrt{T}} < -\nu T^{-\alpha}\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{T^{1/2-\alpha}} + \frac{\hat{d}_{\max}}{NT^{1/2-\alpha}} > \nu\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{T^{1/2-\alpha}} + \frac{\hat{d}_{\min}}{NT^{1/2-\alpha}} < -\nu\right) \\
&= \frac{1}{N} \sum_{i \in I_1^0} 1(T^{\alpha-1/2}\hat{c}_i > \nu) + \frac{1}{N} \sum_{i \in I_2^0} 1(T^{\alpha-1/2}\hat{c}_i < -\nu)
\end{aligned}$$

The above terms tend to zero because $0 \leq \alpha < 1/2$. Also, when $\delta_i = \hat{\sigma}_i^{-2}$ in (8), the result holds when $\sqrt{N}/T^{3/2-2\alpha} \rightarrow \infty$. \square

Table 1: Testing Parameter Homogeneity

		Size:				Power:			
		$(K, G) = (2, 1)$							
		$B_1 = B_2 = 0.8$				$B_1 = 0.7, B_2 = 1$			
N \ T		20	50	100	200	20	50	100	200
100		3.4	4.3	3.2	5.5	13.6	67.0	99.5	100.0
200		3.3	5.0	4.7	4.1	22.2	88.3	100.0	100.0
500		3.9	3.4	4.1	5.7	46.3	100.0	100.0	100.0
1000		3.9	4.2	4.8	4.4	81.7	100.0	100.0	100.0
		$(K, G) = (2, 2)$							
		$B_1 = B_2 = (0.8, 3.2/3)$				$B_1 = (0.7, 1), B_2 = (1, 1.1)$			
N \ T		20	50	100	200	20	50	100	200
100		3.4	3.7	4.9	5.7	7.3	36.2	86.6	100.0
200		2.6	4.1	4.4	3.6	13.1	74.4	99.9	100.0
500		3.4	4.0	4.5	5.9	29.5	98.4	100.0	100.0
1000		3.4	3.9	4.8	6.5	52.2	100.0	100.0	100.0

Table 2: Estimation of the Rate of Technology Growth, g_j , $j = 1, 2$

Cluster based on		level equation (14)			first difference, $\overline{\Delta y_{it}}$		
Whole Sample		PY-statistic = 2.940			PY-statistic = 4.649		
		K_1^H	PS_1^H	PS_2^H	K_1^H	PS_1^H	PS_2^H
Low	N_1	30	30	28	45	38	37
	$\hat{g}_1(\%)$	0.614	0.614	0.611	1.13	1.264	1.259
	PY_1	1.308	1.308	1.526	0.96	-0.077	0.002
Threshold	$\hat{\gamma}$	—	0.015	-4.492	—	0.020	-0.449
High	N_2	40	40	42	25	32	33
	$\hat{g}_2(\%)$	2.507	2.507	2.501	3.359	2.781	2.777
	PY_2	1.913	1.913	2.051	1.841	1.793	1.669

Note: (A) Model:

$$y_{it} = \mu'_i + g_i t + \varepsilon_{it},$$

$$\varepsilon_{it} = \gamma_i \varepsilon_{i,t-1} + \eta_t + e_{it},$$

where $\mu'_i = A_i(0) + g_i + [\alpha_i/(1 - \alpha_i)] \cdot [\log(s_i) + \log(n_i + g_i + \delta)]$. (B) Data: Penn World Tables (PWT v6.2). (C) Estimation: ordinal least squares estimation. For the detailed estimation method please see Lee, Pesaran, and Smith (1997). (D) PY denotes the PY statistic in (13) for parameter homogeneity in the 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. In this case, the PY statistic can be expressed as $PY = \sqrt{N}(\sum_i^N s_i^2/N - 1)/\sqrt{2}$, where

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

\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 . Similarly, for PS_2^H the standardized \hat{g}_i is $\hat{\tau}_i = \sqrt{T}(\hat{g}_i - \hat{g}_w)/\hat{\sigma}(\hat{g}_i)$, where $\hat{\sigma}(\hat{g}_i)$ is the unrestricted long-run variance of \hat{g}_i . Both $\hat{\sigma}(\hat{g}_i)$ and $\tilde{\sigma}(\hat{g}_i)$ are calculated by the automatic method developed by Andrews (1991) with the Quadratic Spectral kernel.

Table 3: Clusters Based on the Level Equation

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

Note: $\ln(y_t)$ denotes the average real GDP per capita over 1965–2000. *Region* denotes geographical code: 1–Africa and Middle East, 2–Latin America, 3–Asia, and 4–Western Europe, Australia, Canada, New Zealand, and USA. For the definitions of $\hat{\tau}_i$ and s_i , please see Table 2. K_1^H , PS_1^H , and PS_2^H denote the group estimates.

Table 4: Clusters Based On the Growth Rate Equation

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

Note: $\ln(y_t)$ denotes the average real GDP per capita over 1965–2000. *Region* denotes geographical code: 1–Africa and Middle East, 2–Latin America, 3–Asia, and 4–Western Europe, Australia, Canada, New Zealand, and USA. For the definitions of $\hat{\tau}_i$ and s_i , please see Table 2. K_1^H , PS_1^H , and PS_2^H denote the group estimates.

Figure 1: Simulation Results when $(K, G) = (2, 1)$

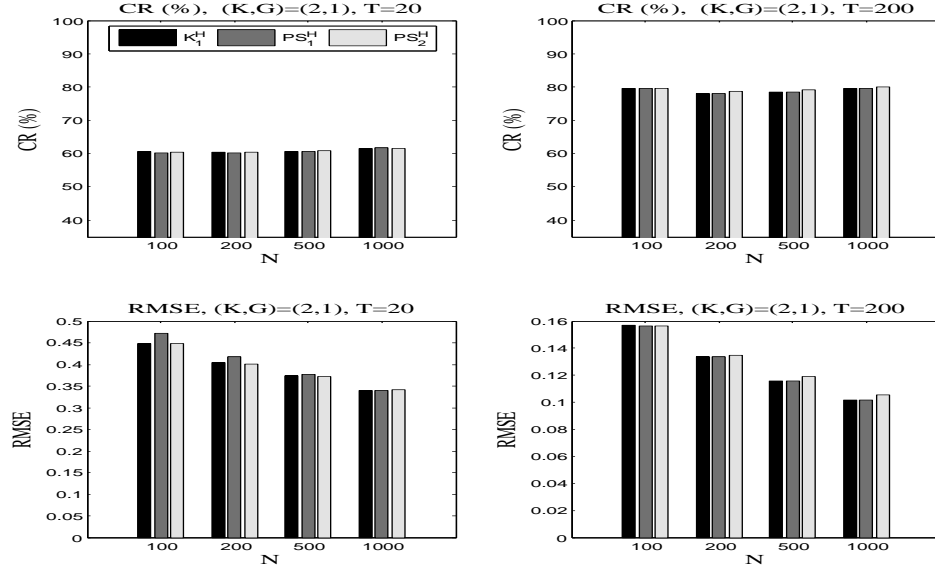


Figure 2: Simulation Results when $(K, G) = (3, 1)$

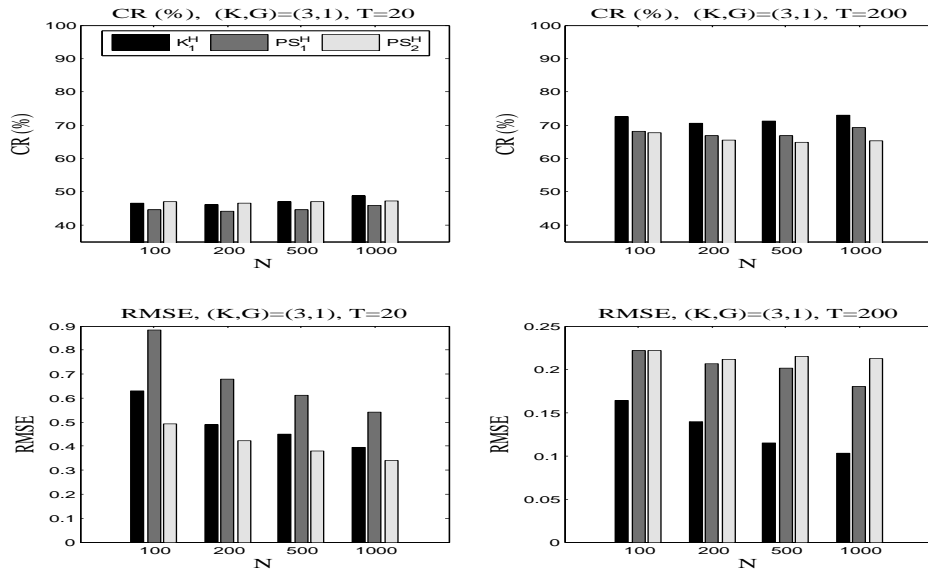


Figure 3: Simulation Results when $(K, G) = (2, 2)$

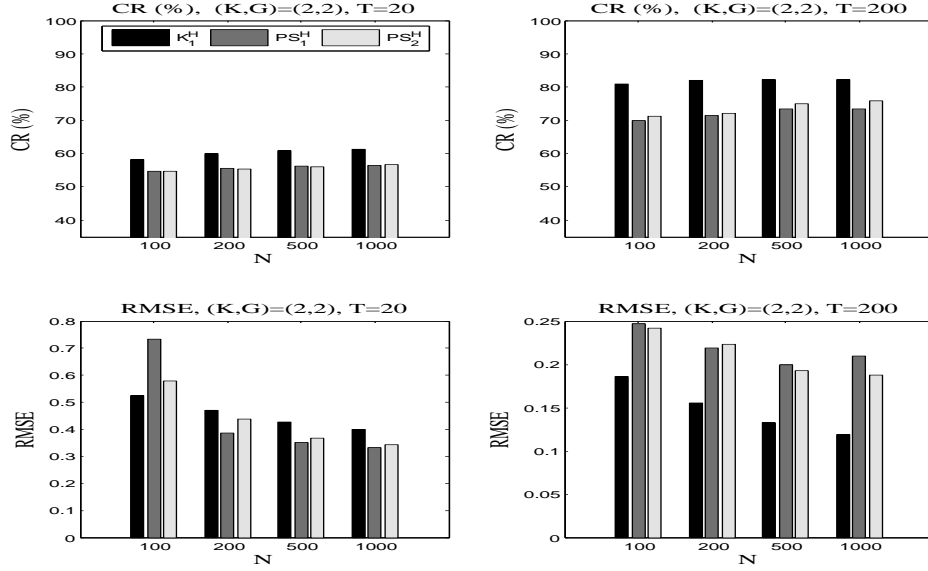


Figure 4: Simulation Results when $(K, G) = (3, 2)$

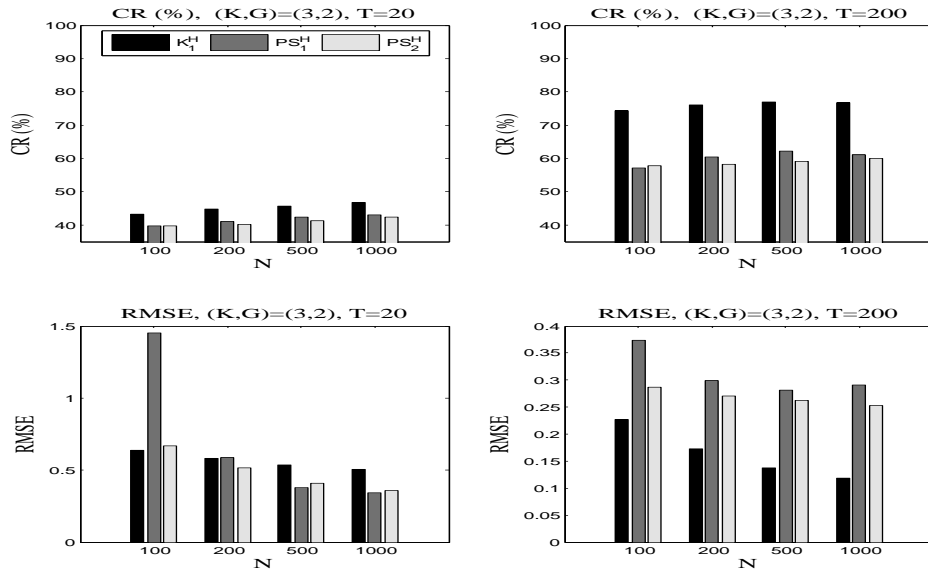
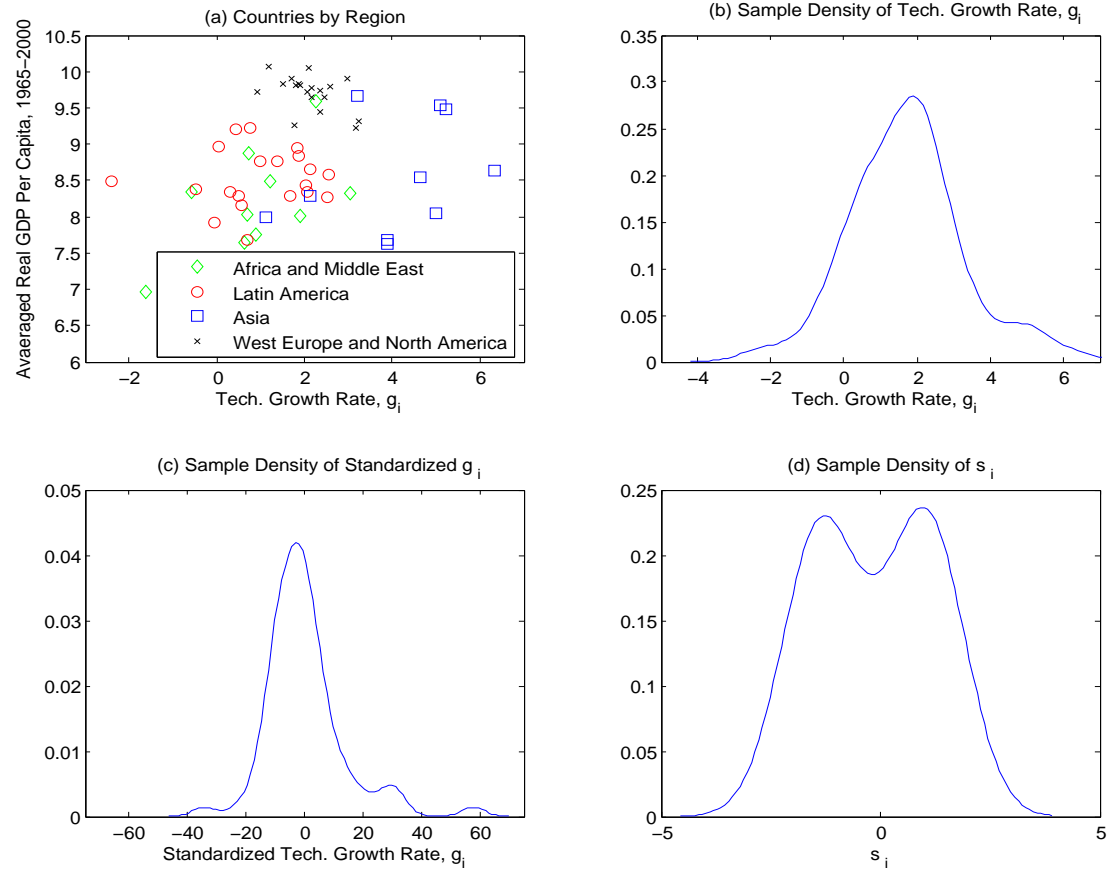


Figure 5: Sample Density of the Standardized Estimates of the Rates of Technology Growth.



For the definitions of standardized \hat{g}_i and s_i , please see Table 2.