Research Article

Chang-Ching Lin and Serena Ng

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

Abstract: This paper proposes two methods for estimating panel data models with group specific parameters when group membership is not known. The first method uses the individual level time series estimates of the parameters to form threshold variables. The problem of parameter heterogeneity is turned into estimation of a panel threshold model with an unknown threshold value. The second method modifies the K-means algorithm to perform conditional clustering. Units are clustered based on the deviations between the individual and the group conditional means. The two approaches are used to analyze growth across countries and housing market dynamics across the states in the U.S.

Chang-Ching Lin: 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
Serena Ng: Department of Economics, Columbia University, 420 West 118, St., New York, NY 10027, USA. Tel: +212-854-5488, E-mail: serena.ng@columbia.edu

1 Introduction

This paper considers estimation of panel data models when the slope parameters are heterogeneous across groups, but that group membership is not known to the econometrician. We consider two methods that let the data determine the grouping. The first method uses the time series estimates of the individual slope coefficients to form threshold variables. The problem of identi-

fying group membership is turned into one of estimating a threshold panel regression with an unknown threshold value. The second method is a modification of the K-means algorithm. The units are clustered according to the deviations between the individual and the group conditional means. Both approaches are agnostic about the sources of parameter heterogeneity.

Panel data models often take parameter homogeneity as a maintained assumption even though evidence against it is not difficult to find. Using data on US manufacturing firms, Burnside (1996) rejects homogeneity of the parameters in the production function. Lee, Pesaran, and Smith (1997) find that the convergence rates of per capita output to the steady state level are heterogeneous across countries. Hsiao and Tahmiscioglu (1997) find heterogeneity in the parameters that describe investment dynamics and observe that such differences cannot be explained by commonly considered firm characteristics. As Browning and Carro (2007) point out, there is usually much more heterogeneity than empirical researchers allow. Robertson and Symons (1992) find using simulations that the Anderson and Hsiao (1982) estimator can be severely biased when parameter heterogeneity is omitted.[1]

While fixed effects estimation allows for time invariant type heterogeneity through the intercept, few methods are available to allow for heterogeneity in the slope parameters. A random coefficient model[2] can estimate the mean of the coefficients but is uninformative about the response at a more disaggregated level, which is sometimes an object of interest. Maddala, Trost, Li, and Joutz (1997) suggest a Bayesian method that shrinks the individual estimates toward the estimator of the overall mean. Alvarez, Browning, and Ejrnæs (2010) parameterize the individual

---

[1] Song (2004) studied the convergence rate of estimating group parameters under cross-section dependence from the individual coefficient estimates and develops a measure of parameter heterogeneity. However, the ideas have not been illustrated using simulations or applications.

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

coefficients as a function of observed characteristics, but the results necessarily depend on the specifications used. While one can assume complete parameter heterogeneity, this would reduce the problem to time series estimation on a unit by unit basis which does not take advantage of the panel structure of the data.

We are interested in precise modeling of the conditional mean function and not group membership per se. We consider two ways of clustering units with the objective of pooling observations to estimate the group specific parameters. In this way, units within a group have the same parameters, but the parameters are heterogeneous across groups. This provides a compromise between complete parameter heterogeneity and parameter homogeneity.

The remainder of this paper is organized as follows. After a review of related work, Section 3 presents the pseudo threshold method. Extensions to multiple groups and multiple covariates are considered in Section 4. The conditional K-means approach is presented in Section 5. Simulations are then presented in Section 6. We apply the methods to study economic growth across countries and regional housing dynamics in the U.S.

2 Related Literature and the Econometric Framework

The simplest way to form clusters from a set of heterogeneous observations on a scalar variable y_{it} is to plot the unconditional mean of the ordered data $\hat{\beta}_{(i)} = \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ for $i = 1,...,N$, and then 'eyeball' to see when $\hat{\beta}_{(i)}$ abruptly shifts from one mean to another.[3] Such a graphical approach is often a useful diagnostic, but does not permit formal statistical statements to be made. One can also use a priori information to organize units into groups, but the approach is not objective. A more systematic approach is model based clustering which assumes that the data are generated by a mixture of distributions. That is, an observation drawn from group g is assumed to have density $f_g(y_{it}, x_{it}; \beta_g)$, where β_g are the group-specific parameters of the conditional mean function, and x_{it} are the covariates. The likelihood is then $L(x; \theta) = \prod_{i=1}^{N}\prod_{t=1}^{T} f_g(y_{it}, x_{it}; \beta_g)$. The parameters are typically estimated using Bayesian methods, though Sun (2005) considers the problem from a frequentist perspective. These likelihood based analy-

ses yield an estimate of the probability of which group a unit belongs, but can be computationally cumbersome if N is large because we need to consider up to 2^N possible combinations of the data.

While the focus of cluster analysis is usually group membership, our interest is precise estimation of the group-specific parameters β_g. We consider a balanced panel of data with observations $(\tilde{y}_{it}, \tilde{x}_{it})$, $i = 1,...,N$, $t = 1,...,T$, and that N and T are large. Let K be the number of regressors and G be the number of clusters. Let $I^0 = (I_1^0,...,I_G^0)$ be indicator variables for true group membership. Then for $g = 1,...,G$, N_g^0 is the number of individuals in cluster I_g^0. The panel data model is

$$\tilde{y}_{it} = \alpha_i + \tilde{x}_{it}\beta_{(i)} + \tilde{e}_{it}$$
$$= \alpha_i + \tilde{x}_{it}\beta_g + \tilde{e}_{it} \qquad i \in I_g^0,$$

where $\tilde{x}_{it} = (\tilde{x}_{1it},...,\tilde{x}_{Kit})$ is a $1 \times K$ vector of regressors, α_i is unobserved heterogeneity, $\beta_{(i)} = (\beta_{(i)1},...,\beta_{(i)K})'$ is a $K \times 1$ vector of slope coefficients for unit i. Group effects are modeled by letting $\beta_g = (\beta_{g1},...,\beta_{gK})'$ be a $K \times 1$ vector of group-specific slope coefficients such that $\beta_{(i)}$ equals or is well approximated by β_g for all i's in I_g^0. Let $y_{it} = \tilde{y}_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $x_{kit} = \tilde{x}_{kit} - \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{kit}$ for $k = 1...,K$. The model in terms of demeaned data becomes

$$y_{it} = x_{it}\beta_g + e_{it} \qquad i \in I_g^0.$$

The econometric exercise is to estimate β_g without knowing I_g^0.[4] This requires a way to pool 'similar' observations for estimation. Two methods are considered. The first defines similarity in terms of the slope coefficients, and the second defines similarity in terms of the conditional mean.

Let N_{gj} be the number of units assigned to group j when they belong to group g. Then $N_s = \Sigma_{g \neq j} N_{gj}$ is the number of misclassified units. Let $I = (I_1,...,I_G)$ denote arbitrary group membership. The size of group j based on an arbitrary classification is $N_j = N_{jj} + \Sigma_{g \neq j} N_{gj}$. The following assumptions will be used.

Assumption 1: For all i and t, (a) $\tilde{e}_{it} \sim (0, \sigma_i^2)$ has finite fourth moments and is cross-sectionally and serially independent. (b) $0 < \sigma_i^2 < \infty$ and (c) $N^{-1}\sum_{i=1}^{N} \alpha_i = O(1)$. Furthermore, $\tilde{e}_{it}$ is independent of $\tilde{y}_{i0}$ and uncorrelated with β_g for all $g = 1,...,G$.

---

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

[4] A group specific intercept can be recovered once group membership is determined by regressing $\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_g \bar{x}_i$ on a set of group dummies.

Assumption 2: (a) $N_g^0/N > 0$ and $N_g^0/N \to \pi$ with $0 < \pi < 1$. (b) $0 \leq \lim N/T < \infty$ as N and T diverge jointly.

In the case of a dynamic panel model with $\tilde{x}_{1it} = \tilde{y}_{i,t-1}$, we also need the following.

Assumption 3: For $g = 1, \ldots, G$ and $i \in I_g^0$, $|\beta_{g1}| < 1$, $\tilde{y}_{i0} = \alpha_i/(1 - \beta_{g1}) + u_{i0}$, where $u_{i0} \sim (0, \sigma_{u,i}^2)$ has finite fourth moments and $0 < \sigma_{u,i}^2 < \infty$. Furthermore, u_{i0} is cross-sectionally independent, independent of $\tilde{e}_{it}$, and uncorrelated with β_{g1}.

We assume cross-section independence of $\tilde{e}_{it}$ to simplify the presentation. Cross-section dependence can be entertained by explicitly controlling for the presence of common factors. Assumptions 2 and 3 are similar to the ones used in Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), and Pesaran and Yamagata (2008) for dynamic panel model with fixed effects. The assumptions can be relaxed if the regressors are strictly exogenous.

3 A Two-Step Pseudo 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 and proposed a *D*-method to estimate the parameters in the transition function by maximum likelihood. The *D*-method assumes deterministic switching of regimes, and stands in contrast to the *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 and structure break models are variations of this approach.

To fix ideas, consider the case of $G = 2$ groups. The model expressed in demeaned data is

$$y_{it} = \begin{cases} x_{it}\beta_1 + e_{it} & i \in I_1^0 \\ x_{it}\beta_2 + e_{it} & i \in I_2^0 \end{cases}. \qquad (1)$$

Suppose there exists a variable q_i^0 and a set of cut-off parameter values Γ^0 such that $i \in I_1^0$ if $q_i^0 \leq \gamma^0$ for any $\gamma^0 \in \Gamma^0$, and $i \in I_2^0$ otherwise. Then the pair of variables (q_i^0, γ^0) provide perfect information about group membership, and (1) can be written as a panel threshold model:

$$y_{it} = \begin{cases} x_{it}\beta_1 + e_{it} & q_i^0 \leq \gamma^0 \\ x_{it}\beta_2 + e_{it} & q_i^0 > \gamma^0 \end{cases}. \qquad (2)$$

Hansen (1999) considers threshold panel regressions where the sample is split according to whether an observed and possibly time varying threshold variable q_{it}^0 is less than some cut-off value γ. Unit i can be in one group in period t if $q_{it}^0 \geq \gamma^0$, but might be in another group in period $t+1$ if $q_{it+1}^0 < \gamma^0$. See also Seo and Shin (2011) for dynamic panel threshold models. Our threshold variable q_i^0 in (1) is not observed, and group structure does not change over time. Because of these differences, we call q_i the 'pseudo threshold variable' and γ the 'pseudo threshold parameter' to distinguish them from the definitions used in the literature.

If q_i^0 and Γ^0 were known, estimates of β_1 could be obtained by pooling observations with $q_i^0 \leq \gamma^0$ for any $\gamma^0 \in \Gamma^0$, while observations with $q_i^0 \leq \gamma^0$ would be pooled to estimate β_2. The problem, however, is that neither q_i^0 nor Γ_0 is observed. We propose to replace q_i^0 by some $\hat{q}_i$ that has the same information as q_i^0 in the sense that $\hat{q}_i \leq \gamma$ when $q_i^0 \leq \gamma$ as $T \to \infty$ for a given γ. Unit i is then classified into group 1 if $\hat{q}_i$ is less than γ; otherwise unit i is in group 2. More precisely, we propose to use $\hat{q}_i$ to order the data, which has the computational advantage that any unit i' with $\hat{q}_{i'} < \hat{q}_i$ will also be classified in the same group as unit i. Thus, even though there are 2^N possible groupings of the data, we only need to consider at most $N-1$ possible values of γ.

Given $\hat{q}_i$ the problem of how to group units with similar coefficients is formulated as finding the value of the threshold parameter γ that minimizes $S_{NT}(\gamma, \hat{q})$:

$$\hat{\gamma} = \arg \min_{\gamma \in [\hat{\underline{q}}, \hat{\bar{q}}]} S_{NT}(\gamma, \hat{q}), \qquad (3)$$

where $S_{NT}(\gamma, \hat{q})$ is the total squared residuals, obtained by summing over groups 1 and 2:

$$S_{NT}(\gamma, \hat{q}) = \sum_{i|\hat{q}_i \leq \gamma} \sum_{t=1}^{T} \left(y_{it} - x_{it}\hat{\beta}_1(\gamma) \right)^2 + \sum_{i|\hat{q}_i > \gamma} \sum_{t=1}^{T} \left(y_{it} - x_{it}\hat{\beta}_2(\gamma) \right)^2. \qquad (4)$$

As in the breakpoint literature, the threshold parameter cannot be too large or too small, as defined by $\hat{\bar{q}}$ and $\hat{\underline{q}}$. If the trial value of γ is too low, $\hat{\beta}_2(\gamma)$ will be estimated with some observations from group 1 and will not be consistent for β_2. Similarly, at too high a value of γ, $\hat{\beta}_1(\gamma)$ will be estimated with observations from group 2, and hence will not precisely estimate β_1. The optimization problem yields $\hat{I}_1 = \{i | \hat{q}_i \leq \hat{\gamma}\}$ and $\hat{I}_2 = \{i | \hat{q}_i > \hat{\gamma}\}$. Since the data are ordered according to $\hat{q}_i$, $\hat{I}_1$ and $\hat{I}_2$ are the same if $\hat{q}_{i^*} = \tilde{\gamma}$ for any $\tilde{\gamma} \in (q_{i^*}, q_{i^*+1})$. We will refer to this procedure as PSEUDO(G,K).

It remains to be precise about the choice of $\hat{q}_i$. Now $\beta_{(i)} = \beta_1$ if $i \in I_1^0$ and $\beta_2 \neq \beta_1$ by assumption. Thus, $q_i = \beta_{(i)}$ along with any $\gamma^0 \in [\beta_1, \beta_2]$ completely summarizes group membership. For example, $\beta_\omega = \omega\beta_1 + (1-\omega)\beta_2$ for $\omega \in (0,1)$ is a valid value of γ^0. This suggests letting $\hat{q}_i = \hat{\beta}_{(i)}$ where for

each i, $\hat{\beta}_{(i)}$ is the least squares estimate from a time series regression of y_{it} on x_{it}.[5] Since T is large by assumption, $\hat{\beta}_{(i)}$ is a consistent estimate of $\beta_{(i)}$.

To see that $\hat{q}_i = \hat{\beta}_{(i)}$ will separate the sample, suppose $\hat{\gamma} = \hat{\beta}_\omega$, where $\hat{\beta}_\omega$ is a consistent estimate of $\beta_\omega = \omega\beta_1 + (1-\omega)\beta_2$. Under Assumptions 1–3:

$$P\left(\hat{\beta}_{(i)} > \hat{\beta}_\omega \mid \beta_{(i)} = \beta_1\right)$$
$$= P\left(\sqrt{T}\left(\hat{\beta}_{(i)} - \beta_1\right) > \sqrt{T}\left(\hat{\beta}_\omega - \beta_1\right) \middle| \beta_{(i)} = \beta_1\right)$$
$$= P\left(\sqrt{T}\left(\hat{\beta}_{(i)} - \beta_1\right) > \sqrt{T}\left[(1-\omega)\left(\beta_2 - \beta_1\right)\right.\right.$$
$$\left.\left. + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)\right] \middle| \beta_{(i)} = \beta_1\right).$$

Since $\beta_2 - \beta_1 > 0$ by assumption, the expected misclassification rate is $E\left(N_s(\hat{\gamma})/N\right) = P\left(\hat{\beta}_{(i)} > \hat{\beta}_\omega \mid \beta_{(i)} = \beta_1\right) + P\left(\hat{\beta}_{(i)} < \hat{\beta}_\omega \mid \beta_{(i)} = \beta_2\right) = O(T^{-1})$. Instead of imposing $\hat{\gamma}$ to $\hat{\beta}_\omega$, we let $\hat{\gamma}$ be determined by (3), but the argument is similar.

Procedure PSEUDO(G,K) for $G=2$, $K=1$ can be summarized as follows.

Algorithm 1: PSEUDO(2,1)
1. For each i, regress y_{it} on x_{it} to obtain $\hat{\beta}_{(i)}$ and let $\hat{q}_i = \hat{\beta}_{(i)}$.
2. Use (3) to obtain $\hat{\gamma}$.
3. Assign unit i into $\hat{I}_1$ if $\hat{q}_i \le \hat{\gamma}$; and into $\hat{I}_2$ if $\hat{q}_i > \hat{\gamma}$.
4. For $g=1, 2$, estimate β_g using all units in $\hat{I}_g$.

It is shown in the Appendix that PSUEDO(2,1) yields a negligible misclassification rate in the sense that $N_s(\hat{\gamma})/N \to 0$ as $(N,T) \to \infty$ jointly. In practice, we restrict the number of units in each group to be at least $\underline{N} = \max(10, 0.1N)$ and $\bar{N} = N - \underline{N}$. This avoids having groups with too few units. As $\hat{q}_i - q_i = O_p\left(T^{-1/2}\right)$, the two-step procedure may be imprecise when T is small.

3.1 Extension to multiple regressors

We now turn to the case when there are $K > 1$ regressors. There are two cases to consider. The first occurs if a subset but not all K parameters are suspicious of being different across groups. In such a case of partial parameter homogeneity, procedure PSEUDO(2,1) is still valid

---

5 We also consider using $\hat{q}_i = \dfrac{\left(\hat{\beta}_{(i)} - \hat{\beta}_\omega\right)}{\hat{\sigma}_i \hat{Q}_i^{-1/2}}$ where $\hat{Q}_i = T^{-1}\sum_{t=1}^{T} x'_{it} x_{it}$. By standardizing $\hat{\beta}_{(i)}$, we control for heterogeneity due to σ_i.

even when $K > 1$. For example, if the second slope coefficient varies between groups, we define $\hat{q}_i = \hat{\beta}_{(i)2}$ and apply Algorithm 1.

More difficult to handle is the second case of complete parameter heterogeneity which arises when all K coefficients are group specific. To see why this is more involved, suppose there are two regressors, x_{1it} and x_{2it} and there are $G=2$ clusters. Let $\beta_g = (\beta_{g1}, \beta_{g2})'$, $g=1,2$ be the group specific slope parameters. Suppose first that $\beta_{11} > \beta_{21}$ and $\beta_{12} > \beta_{22}$. Since both parameters are strictly larger in one group, an obvious pseudo threshold variable is $\hat{\beta}_{(i)}^+ = \hat{\beta}_{(i)1} + \hat{\beta}_{(i)2}$. But this does not always work! In particular, if $\beta_{11} > \beta_{21}$ but $\beta_{12} < \beta_{22}$, the sum of the coefficients may not be a sufficient statistic for group membership. For example, suppose that $(\beta_{11}, \beta_{12}) = (0.8, 1)$ and $(\beta_{21}, \beta_{22}) = (1, 0.8)$. Since $\beta_{11} + \beta_{12} = \beta_{21} + \beta_{22}$, the appropriate pseudo threshold variable is no longer $\hat{\beta}_i^+$, but $\hat{\beta}_{(i)}^- = \hat{\beta}_{(i)1} - \hat{\beta}_{(i)2}$. Although $\hat{\beta}_{(i)}^+$ or $\hat{\beta}_{(i)}^-$ can be used as threshold variable, we would first need to determine the sign of the coefficients before we can classify the units. In an earlier version of this paper, we used the Goodman-Kruskal's gamma statistic to measure the association between pairs of concordant (same sign) and discordant data (opposite sign) data. Although the method works reasonably well in simulations, it is somewhat cumbersome.

A simpler and more effective approach is to recognize that even in the case of complete parameter heterogeneity, we can still split the sample using one of the $\hat{\beta}_{(i)k}$ parameters since each component of $\hat{\beta}_{(i)} = \left(\hat{\beta}_{(i)1}, ..., \hat{\beta}_{(i)K}\right)'$ is informative about group membership. The only issue that remains is which $\hat{\beta}_{(i)k}$ to use. We let the data speak by considering each component as a possible candidate and choose the one that minimizes the sum of squared residuals. More precisely, for given k with $\hat{q}_{ik} = \hat{\beta}_{(i)k}$, let $\hat{\gamma}_k$ be estimated from (3). Define

$$S_{NT,k}\left(\hat{\gamma}_k\right) = \sum_{i|\hat{q}_{ik} \le \hat{\gamma}_k} \sum_{t=1}^{T} \left(y_{it} - x_{it}\hat{\beta}_1\left(\hat{\gamma}_k\right)\right)^2 + \sum_{i|\hat{q}_{ik} > \hat{\gamma}_k} \sum_{t=1}^{T} \left(y_{it} - x_{it}\hat{\beta}_2\left(\hat{\gamma}_k\right)\right)^2,$$

where $x'_{it}, \hat{\beta}_1$, and $\hat{\beta}_2$ are $K \times 1$ vectors. The best threshold variable is $\hat{\beta}_{(i)k^\star}$ where

$$k^\star = \min_k S_{NT,k}\left(\hat{\gamma}_k\right). \tag{5}$$

The procedure for complete parameter heterogeneity when $k > 1$ is summarized as follows:

Algorithm 2: PSEUDO(2,K)
1. For each i, regress y_{it} on x_{it} to obtain $\hat{\beta}_{(i)}$.
2. For $k=1,...,K$, let $\hat{q}_{ik} = \hat{\beta}_{(i)k}$.
   i. Use (3) to obtain $\hat{\gamma}_k$.
   ii. Assign unit i into $\hat{I}_1$ if $\hat{q}_{ik} \le \hat{\gamma}_k$; and into $\hat{I}_2$ if $\hat{q}_{ik} > \hat{\gamma}_k$.

iii   For $g = 1, 2$, estimate β_g using all units in $\hat{I}_g$ and record $S_{NT,k}(\hat{\gamma}_k)$.

3   Let $k^\star = \arg\min_k S_{NT,k}(\hat{\gamma}_k)$, $\hat{q}_i = \hat{\beta}_{(i)k^\star}$ and apply Algorithm 1.

The appeal of this approach is generality, since the procedure is the same for any K. Partial parameter heterogeneity is just a special case when Steps 1 and 2 can be skipped, and $\hat{q}_i$ is simply the parameter estimate associated with the variable whose effect on y_{it} is group specific.

Procedure PSEUDO(2,K) extends naturally to $G > 1$, which would then have $G - 1$ threshold values. In the multiple structural breaks literature, Bai (1997) showed that a sequential approach can consistently estimate the break fractions without the need to search for all break dates simultaneously. This idea is adapted to our threshold problem. We estimate two subgroups from each of the groups identified in the preceding step, subject to the constraint that the size of each subgroup is not too small. For example, given $\hat{I}_1$ and $\hat{I}_2$, we first partition $\hat{I}_1$ into $\hat{I}_{11}$ and $\hat{I}_{12}$. If $\hat{I}_{11}$ and $\hat{I}_{12}$ both have a minimum number of observations, then these together with $\hat{I}_2$ form one possible split of the sample into three groups. Similarly, $\hat{I}_2$ is partitioned into $\hat{I}_{21}$ and $\hat{I}_{22}$, which together with $\hat{I}_1$ form another possible sample split of three groups. We then decide which of the two possibilities for $G = 3$ to keep by comparing the sum of squared residuals.

4 Conditional K-means Clustering

The K-means algorithm is a popular way of forming clusters from a single series with N observations. The algorithm moves unit i to an appropriate group to minimize the sum of squared deviations between the units and the centroids.[6] Except for the gene-array analysis in Qin and Self (2006), the K-means method tends to be used to form clusters from observations on a scalar variable with no reference to covariates. We modify the K-means algorithm for use in regression analysis, which can be thought of as a form conditional clustering. We refer to the procedure as CK-means(G,K).

Algorithm 3: CK-means(G,K)
1   Initialize $\{\ddot{I}_1,...,\ddot{I}_G\}$ randomly and let $(\ddot{\beta}_1,...,\ddot{\beta}_G)$ be the pooled estimates of $\beta_1,...,\beta_G$.

2   Repeat for $i = 1,...,N$ until no individual changes group:
   (a)   Calculate $SSR_i^g = \sum_{t=1}^T (y_{it} - x_{it}\ddot{\beta}_g)^2$, $g = 1,...,G$.
   (b)   For $g = 2,...,G$, $g' = 1,...,g-1$, if $SSR_j^g \leq SSR_i^{g'}$ individual i is reassigned to group g'; otherwise, i stays with group g.
   (c)   Update $\{\ddot{I}_g,...,\ddot{I}_G\}$ and estimate $(\ddot{\beta}_1,...,\ddot{\beta}_G)$.

The unconditional K-means method is known to be sensitive to the initial choice of the centroids and is not guaranteed to find the global minimizer. Thus Steps 1 to 2 are repeated several times with initial group assignment.[7]

Assuming i.i.d. data, Pollard (1981) uses empirical process arguments to obtain a strong consistency result, while Pollard (1982) shows that the centroids estimated by the algorithm are asymptotically normal. However, Pollard (1981) notes that his consistency result does not necessarily apply to algorithms used in practice which involve multiple starting values. The asymptotic properties of K-means algorithm used in practice is not known even in the absence of covariates.

While our pseudo threshold procedure minimizes the same objective function as CK-means, some differences are noteworthy. First, we only estimate the ordered regression once. The CK-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 CK-means method computationally costly when N is large. Furthermore, when there are multiple alternatives and N is large, convergence of the CK-means can be slow. Second, because we follow the structural break literature and search for the optimal threshold value in the subsample $[\underline{N}, \overline{N}]$, our approach is less sensitive to outliers. Simulations bear this out. Third, we locate the threshold values one at a time, starting with the largest. In contrast, the CK-means is a global procedure. Units found to be in Group 1 by the CK-means when $G = 2$ may well be in Group 2 when $G = 3$.

The CK-means method also has some advantages. First, the algorithm relies only on the pooled estimator $\hat{\beta}_g$ which is $\sqrt{N_g T}$ consistent, and does not require the individual estimates $\hat{\beta}_{(i)}$'s, which are $\sqrt{T}$ consistent. Thus the CK-means method should be more precise than the PSEUDO when N or T is small. Second, the CK-means method considers moving every unit to a different group. Our pseudo threshold method 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

---

6 There are many variations to the basic algorithm. Harmonic and fuzzy means have also been used instead of simple means. See, for example, Hartigan (1975), Abraham, Cornillion, Matzner-Lober, and Molinari (2003).

7 Garcia-Escudero and Gordaliza (1999) pointed out the algorithm can be sensitive to outliers.

ordering of $\hat{q}_i$ does not agree with q_i, as may be the case when the sample size is small, or when q_i does not provide complete information about the group structure. We can therefore expect a trade-off between precision and speed in the two methods. Third, the CK-means algorithm is easier to implement when $G > 3$ because it is a global procedure and considers moving every unit to a different group. In contrast, PSEUDO is a sequential procedure; the membership of the subgroups identified by PSEUDO always depends on the outcome of the preceding step.

5 Determining *G*

Both the pseudo threshold and the conditional K-means algorithm require knowledge of the number of groups, G. An informal approach is to graph the value of the objective function S_{NT} for a given G against G and then locate the 'knee point' at which the objective function starts to flatten. More formal procedures have been proposed for unconditional clustering of y_{it}. Milligan and Cooper (1985) consider 30 procedures and find that the global procedure of Calinski and Harabasz (1974) works best, while the local procedure of Duda and Hart (1973) is second. But as Sugar and James (2003) point out, most methods are aimed at clustering data with specific properties and no method works uniformly well. We experimented with many of these methods and found them to be accurate only when the parameters in different groups are very far apart.

The problem of determining the number of clusters is similar to determining the number of break points or thresholds in many ways. In breakpoint problems, we can use a sup-Wald type test for the null hypothesis of no break.[8] However, there are three features that make the SupW test for parameter homogeneity infeasible here. First, $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated from two split samples ordered by $\hat{\beta}_{(i)}$. By construction, one sample will have smaller values of $\hat{\beta}_{(i)}$ and the other will have the larger values. Thus, the pooled estimate will be biased if $\beta_1 = \beta_2$. Second, $\hat{\beta}_1$ and $\hat{\beta}_2$ are correlated when $\beta_1 = \beta_2$, making inference non-standard. Third, as $\hat{q}_i$ is ordered, bootstrap procedures valid for cross-sectionally independent data are now invalid.

We found two ways that determine G quite accurately in our setup. The first uses a sequential test of parameter homogeneity to provide information about the number of

---

8 See, for example, Davies (1977), Andrews and Ploberger (1994), Hansen (1996), Bai (1997), and Caner and Hansen (2004).

groups. Specifically, if we reject parameter homogeneity in the pooled data, we partition the sample into two groups and then test if parameter homogeneity holds in each of the subgroups. If subsample homogeneity is rejected, the sample is split again until the null hypothesis of parameter homogeneity cannot be rejected for the subsamples. We use the dispersion t_g test proposed by Pesaran and Yamagata (2008) to test parameter homogeneity:

$$t_g = \frac{\sqrt{N}\left(\xi_N / N - K\right)}{\sqrt{2KG}}, \tag{6}$$

where K denotes the number of the regressors, $\xi_N = \sum_{i=1}^{N} \tilde{\sigma}_i^{-2}\left(\hat{\beta}_i - \tilde{\beta}_w\right)'\left(\sum_{t=1}^{T} x_{it}'x_{it}\right)\left(\hat{\beta}_i - \tilde{B}_w\right)'$, $\tilde{B}_w$ is the weighted pooled fixed effects estimator of Swamy (1970), and $\tilde{\sigma}_i^2$ is obtained under the null hypothesis of homogeneity. The t_g test allows for heteroskedasticity and non-normally distributed errors and is consistent as N and T go to infinity jointly such that $\sqrt{N} / T^2 \to 0$.

The second approach is a modified BIC criterion:

$$\text{BIC}(\tilde{G}) = \log\left(\sum_{NT}\left(\tilde{G}, \hat{\gamma}, \hat{q}\right)\right)$$
$$+ \tilde{G}K \cdot \frac{c_{NT}\log(NT)}{NT} + \left(\tilde{G} - 1\right)\frac{\log(N^2)}{N^2}, \tag{7}$$

where

$$\sum_{NT}\left(\tilde{G}, \hat{\gamma}, \hat{q}\right) = \frac{1}{NT}\sum_{g=1}^{\tilde{G}}\sum_{i \in \hat{I}_g}\sum_{t=1}^{T}\left(y_{it} - x_{it}\hat{\beta}_g(\hat{\gamma})\right)^2.$$

The goodness of fit component of the BIC is computed as the total sum of squared residuals divided by NT. The penalty term $\log(NT)/NT$ is guided by the fact that $\hat{\beta}_g$ is $\sqrt{NT}$ consistent under the null and thus the BIC should consistently select G if (i) $\frac{c_{NT}^{\star}}{NT} = \frac{c_{NT}\log(NT)}{NT} \to 0$, and (ii) $c_{NT}^{\star} \to \infty$ as $N,T \to \infty$. When all regressors and γ are observed, BIC obtains with $c_{NT}^{\star} = \log(NT)$ and $c_{NT} = 1$. We consider a heavier penalty because $\hat{q}_i$ and $\hat{\gamma}$ are themselves estimated. Based on extensive simulations, we let $c_{NT} = \sqrt{\min[N,T]}$. The required conditions for consistent model selection are satisfied because $c_{NT}^{\star} = c_{NT}\log(NT) \to \infty$ and $\frac{c_{NT}^{\star}}{NT} \to 0$. Furthermore, the breakpoint literature suggests $\hat{\gamma}$ is super-consistent with a convergence rate of N^2. A penalty of $\log(N^2)$ is put on each threshold variable, giving an overall penalty on $\hat{\gamma}$ of $\left(\tilde{G} - 1\right)\log(N^2) / N^2$. The idea of using the BIC or t_g test to determine G is simple, but the exposition of the complete algorithm is notationally involved. Details are given in an appendix available on request.

6 Simulations and Applications

We now use Monte Carlo simulations to examine the finite sample properties of the methods considered. For $G = 2, 3$, $K = 1, 2$, data are generated as

$$\tilde{y}_{it} = \alpha_i + \sum_{g=1}^{G} \sum_{i \in I_g^0} \tilde{x}_{it}' \beta_g + \tilde{e}_{it},$$

where β_g is $K \times 1$, $\alpha_i \sim$ i.i.d. $N(1,1)$, $\tilde{x}_{kit} \sim$ i.i.d. $N(1, 3)$, and independent of $\tilde{e}_{it}$, $\tilde{e}_{it} \sim N(0,1)$, is i.i.d. over i and t. When $G = 2$, we randomly assign individuals into two groups $\{I_1^0, I_2^0\}$ with size $N_1^0 = \lfloor 2N/3 \rfloor$ $N_2^0 = N - N_1^0$, where $\lfloor A \rfloor$ denotes the maximum integer that does not exceed real number A. When $G = 3$, individuals are randomly assigned into three groups $\{I_1^0, I_2^0, I_3^0\}$ of equal size $\lfloor N/3 \rfloor$.

We consider the following configurations:
i $(G, K) = (2, 1)$: $(\beta_1, \beta_2) = (0.3, 0.9)$.
ii $(G, K) = (3, 1)$: $(\beta_1, \beta_2, \beta_3) = (0.3, 0.5, 0.8)$.
iii $(G, K) = (2, 2)$: $\beta_1 = (\beta_{11}, \beta_{12})' = (0.1, 0.3)'$ and $\beta_2 = (\beta_{21}, \beta_{22})' = (2/3, 0.6)'$.
iv $(G, K) = (3, 2)$: $\beta_1 = (\beta_{11}, \beta_{12})' = (0.3, -0.3)'$, $\beta_2 = (\beta_{21}, \beta_{22})' = (0.5, 0)'$ and $\beta_3 = (\beta_{31}, \beta_{32})' = (0.7, 0.3)'$.

These parameterizations give an R^2 of around 0.5. For given N and T with $N = (50, 100, 200, 500)$ and $T = (20, 50, 100, 200, 500)$, group membership is held fixed in the $M = 1000$ replications.

We determine G by the BIC defined in (7) and impose the restriction that each subgroup must contain at least $\underline{N} = \max\{10, 0.1N_p\}$ units, where N_p denotes the number of units in the parent group. The estimates are evaluated using the root mean squared error (RMSE) defined as

$$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{N} \frac{1}{NK} \left\| \hat{\beta}_{(i)}^{(m)}(\hat{G}) - \beta_{(i)}(G) \right\|^2},$$

where $\hat{\beta}_{(i)}^{(m)}(\hat{G})$ is the pooled slope parameter in the m-th replication estimated for the ith unit based on $\hat{G}$, and $\beta_{(i)}(G)$ is the true slope coefficient for the unit. To assess the error in the estimates due to estimation of G, we also consider the RMSE (a) when G and group membership are known, and (b) when G is known but group membership is not. Group membership is determined by PSEUDO or the CK-means.

As can be seen from Table 1, the RMSEs of both PSEUDO and CK-means decrease as N or T increases. An increase in T has a larger impact on RMSE than an increase in N. When $G = 2$, PSEUDO and CK-means yield similar RMSEs when T is large, but the CK-means has smaller errors when T is small. This is to be expected since the pseudo threshold method requires $\sqrt{T}$

consistent estimation of the individual slope parameters. When $G = 3$ and assumed known, the CK-means tends to outperform PSEUDO, perhaps because the former can cluster units more flexibly. When G is estimated, The RMSEs are similar to those when both G and group membership are known, provided T is large. This suggests that estimation of group membership has little impact on the estimated slope parameters when T is large.

To check robustness, we consider smaller differences in the parameters between groups.
i $(G, K) = (2, 1)$: $(\beta_1, \beta_2) = (0.55, 0.65)$.
ii $(G, K) = (3, 1)$: $(\beta_1, \beta_2, \beta_3) = (0.4, 0.5, 0.6)$.
iii $(G, K) = (2, 2)$: $\beta_1 = (\beta_{11}, \beta_{12})' = (0.3, 0.4)'$ and $\beta_2 = (\beta_{21}, \beta_{22})' = (0.4, 0.5)'$.
iv $(G, K) = (3, 2)$: $\beta_1 = (\beta_{11}, \beta_{12})' = (0.4, 0.2)'$, $\beta_2 = (\beta_{21}, \beta_{22})' = (0.5, 0.3)'$ and $\beta_3 = (\beta_{31}, \beta_{32})' = (0.6, 0.4)'$.

The RMSE results are reported in Table 2. Not surprisingly, PSEUDO and CK-means need a larger T to be precise. However, other features of the results are similar to those in Table 1.

Data are also generated from a dynamic panel model with group specific parameters:

$$\tilde{y}_{it} = \alpha_i + \rho_g \tilde{y}_{i,t-1} + \phi_g t + \tilde{e}_{it}, \qquad \text{if } i \in I_g^0. \qquad (8)$$

We set $\alpha_i = 0$ for all i's, $(\rho_1, \rho_2) = (0.3, 0.8)$, $(\phi_1, \phi_2) = (0, 0.03)$, $\tilde{e}_{it} \sim N(0, 1)$ is i.i.d. over i and t. The results are presented in Table 3. As in the static DGP, the RMSE tends to decrease as T increases. Furthermore, the RMSEs for PSEUDO and CK-means are similar to those with known membership when T is large.

Overall, the results find that PSEUDO and CK-means have good properties especially when T is large. For small sample sizes, the CK-means is preferred. When the sample size is large, PSEUDO tends to be as effective as CK-means. The PSEUDO has the distinct computational advantage that the number of regressions is of order N, much smaller than the G^N regressions under CK-means.

6.1 Example 1: growth regressions

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. Lee, Pesaran, and Smith (1997) took data for 69 countries over the sample 1965 to 2003 from

Parameters	T	PSEUDO N=50	100	200	500	CK-means 50	100	200	500	PSEUDO when G is known 50	100	200	500	l⁰ is known 50	100	200	500
	20	8.60	10.86	12.12	12.16	8.95	11.20	12.18	12.37	8.44	8.04	8.03	7.86	2.70	1.88	1.32	0.83
$(G,K)=(2,1)$	50	1.79	1.64	5.51	6.64	1.81	1.63	5.50	6.57	1.79	1.49	1.38	1.10	1.56	1.14	0.83	0.52
$\beta_1=0.3$	100	1.14	0.82	1.25	4.20	1.14	0.82	1.03	2.65	1.14	0.82	0.56	0.37	1.14	0.82	0.56	0.37
$\beta_2=0.9$	200	0.82	0.57	0.40	2.68	0.82	0.57	0.40	0.28	0.82	0.57	0.40	0.27	0.82	0.57	0.40	0.27
	500	0.53	0.37	0.26	0.16	0.53	0.37	0.26	0.16	0.53	0.37	0.26	0.16	0.53	0.37	0.26	0.16
	20	6.62	6.51	8.76	9.82	6.06	5.69	8.24	9.56	6.62	6.47	6.46	6.75	2.60	1.87	1.34	0.84
$(G,K)=(2,2)$	50	1.72	1.34	1.02	4.00	1.72	1.32	1.00	4.52	1.72	1.34	1.01	0.84	1.65	1.17	0.83	0.52
$\beta_1=(0.1,0.3)'$	100	1.17	0.82	0.59	2.27	1.17	0.82	0.59	2.37	1.17	0.82	0.59	0.36	1.16	0.82	0.59	0.36
$\beta_2=(2/3,0.6)'$	200	0.81	0.59	0.41	0.26	0.81	0.59	0.41	0.26	0.81	0.59	0.41	0.26	0.81	0.59	0.41	0.26
	500	0.52	0.36	0.27	0.16	0.52	0.36	0.27	0.16	0.52	0.36	0.27	0.16	0.52	0.36	0.27	0.16
	20	14.02	13.78	13.60	13.57	14.01	13.33	13.52	13.56	13.75	13.74	13.77	13.87	3.30	2.29	1.62	1.03
$(G,K)=(3,1)$	50	10.73	9.14	8.51	8.49	10.44	7.61	7.67	8.36	9.05	9.11	9.09	9.06	1.99	1.40	1.02	0.64
$\beta_1=0.3$	100	6.39	5.57	5.30	5.46	4.65	4.08	3.97	4.83	5.74	5.61	5.38	5.35	1.40	1.01	0.70	0.46
$\beta_2=0.5$	200	2.60	2.28	2.21	2.16	1.81	1.65	1.57	1.50	2.60	2.28	2.21	2.16	1.01	0.70	0.50	0.32
$\beta_3=0.8$	500	0.67	0.53	0.36	0.27	0.66	0.48	0.33	0.23	0.67	0.53	0.36	0.27	0.64	0.46	0.32	0.20
	20	13.03	12.17	11.98	12.25	12.55	9.82	9.70	10.78	11.96	12.00	12.16	12.38	3.23	2.31	1.65	1.03
$(G,K)=(3,2)$	50	9.87	7.07	6.22	5.96	7.24	4.16	3.98	4.21	7.26	7.44	7.74	8.02	2.04	1.44	1.01	0.64
$\beta_1=(0.3,-0.3)'$	100	4.54	3.82	3.41	3.23	1.76	1.43	1.24	1.10	4.39	4.19	3.97	3.82	1.43	1.00	0.72	0.44
$\beta_2=(0.5,0)'$	200	2.02	1.79	1.58	1.36	0.99	0.73	0.50	0.32	2.02	1.79	1.58	1.36	0.99	0.72	0.50	0.32
$\beta_3=(0.7,0.3)'$	500	0.63	0.52	0.41	0.30	0.63	0.44	0.32	0.20	0.63	0.52	0.41	0.30	0.63	0.44	0.32	0.20

Table 1: RMSE: Static Panel.

Data are generated from the following static panel model. For $G=2, 3$, $K=1, 2$, $t=1,\dots,T$, and $i=1,\dots,N$,

$$\tilde{y}_{it}=\alpha_i+\sum_{g=1}^{G}\left(\sum_{k=1}^{K}\tilde{x}_{kit}\,\beta_{gk}\right)1(i\in l_g^0)+\tilde{e}_{it}.$$

The numbers reported above are the RMSEs $\times 100$, where *RMSE* is defined in the text.

The number of units in each group is restricted to be at least $\max(10, 0.1N)$. Columns 1 and 2 assume that G and group membership are both unknown. Column 3 assumes that G is known.

Parameters	T	PSEUDO				CK-means				PSEUDO when G is known				P⁰ is known			
		N=50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
	20	10.02	10.97	11.27	12.87	9.97	10.96	12.09	12.76	11.11	10.97	10.94	10.90	2.70	1.88	1.32	0.83
$(G,K)=(2,1)$	50	4.91	7.05	7.12	7.84	4.91	7.05	7.12	7.68	7.16	7.09	7.12	7.06	1.56	1.14	0.83	0.52
$\beta_1=0.55$	100	4.78	5.10	5.16	5.27	4.78	5.12	5.17	5.46	5.20	5.17	5.16	5.14	1.14	0.82	0.56	0.37
$\beta_2=0.65$	200	4.57	3.64	3.60	3.59	4.58	3.67	3.63	3.80	3.63	3.62	3.60	3.58	0.82	0.57	0.40	0.27
	500	2.74	1.73	1.70	1.68	2.78	1.76	1.71	1.69	1.77	1.73	1.70	1.68	0.53	0.37	0.26	0.16
	20	5.11	8.46	9.84	10.97	5.25	8.55	10.25	11.05	8.97	8.66	8.45	8.35	2.60	1.87	1.34	0.84
$(G,K)=(2,2)$	50	4.81	4.88	5.83	6.96	4.81	4.98	5.30	6.59	6.00	5.90	5.83	5.83	1.65	1.17	0.83	0.52
$\beta_1=(0.3, 0.4)'$	100	4.74	4.73	4.39	4.77	4.74	4.67	3.68	4.07	4.31	4.34	4.39	4.43	1.16	0.82	0.59	0.36
$\beta_2=(0.4, 0.5)'$	200	4.54	3.12	3.05	3.12	4.44	2.32	2.22	2.16	2.82	2.95	3.05	3.11	0.81	0.59	0.41	0.26
	500	1.01	1.13	1.25	1.38	0.85	0.69	0.64	0.60	0.97	1.13	1.25	1.38	0.52	0.36	0.27	0.16
	20	11.55	11.56	12.26	13.02	11.49	11.60	12.55	13.01	12.62	12.44	12.40	12.33	3.30	2.29	1.62	1.03
$(G,K)=(3,1)$	50	8.04	7.73	7.74	8.10	8.04	7.71	7.88	8.00	7.95	7.93	7.95	7.90	1.99	1.40	1.02	0.64
$\beta_1=0.4$	100	6.05	5.94	5.92	5.79	6.05	5.94	5.89	5.75	5.80	5.81	5.82	5.81	1.40	1.01	0.70	0.46
$\beta_2=0.5$	200	5.04	5.01	5.00	4.20	5.05	5.02	4.76	4.00	4.43	4.47	4.51	4.54	1.01	0.70	0.50	0.32
$\beta_3=0.6$	500	4.64	4.13	3.06	2.45	4.71	2.09	1.93	1.92	3.15	3.30	3.38	3.53	0.64	0.46	0.32	0.20
	20	8.85	9.76	11.06	11.43	8.85	9.03	10.72	11.35	11.05	10.80	10.67	10.64	3.23	2.31	1.65	1.03
$(G,K)=(3,2)$	50	8.05	6.86	6.92	7.31	7.82	6.08	6.01	7.00	7.10	7.06	7.05	7.11	2.04	1.44	1.01	0.64
$\beta_1=(0.4, 0.2)'$	100	5.60	5.44	5.50	5.18	5.25	5.07	5.03	4.06	5.06	5.15	5.21	5.30	1.43	1.00	0.72	0.44
$\beta_2=(0.5, 0.3)'$	200	4.76	4.79	4.78	3.55	4.78	4.78	3.95	2.43	3.74	3.86	3.97	4.08	0.99	0.72	0.50	0.32
$\beta_3=(0.6, 0.4)'$	500	3.87	2.21	2.01	2.05	3.32	0.81	0.75	0.68	2.07	2.12	2.07	2.14	0.63	0.44	0.32	0.20

Table 2: RMSE: Static Panel – Robustness Check.
See footnote below Table 1.

T	PSEUDO				CK-means				PSEUDO when G is known				I^0 is known			
	N=50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
20	15.41	15.98	16.69	16.64	13.06	13.35	15.46	15.90	15.38	15.45	15.47	15.48	6.55	6.12	5.86	5.70
50	5.84	6.72	6.67	7.08	3.03	2.67	2.53	5.28	5.84	7.03	7.93	8.49	2.77	2.51	2.33	2.24
100	1.60	1.38	1.27	2.27	1.60	1.38	1.24	2.84	1.60	1.38	1.27	1.23	1.60	1.38	1.24	1.14
200	0.99	0.80	0.67	0.60	0.99	0.80	0.67	0.60	0.99	0.80	0.67	0.60	0.99	0.80	0.67	0.60
500	0.56	0.42	0.34	0.27	0.56	0.42	0.34	0.27	0.56	0.42	0.34	0.27	0.56	0.42	0.34	0.27

Table 3: RMSE: Dynamic Panel Model.
Data are generated from the following dynamic panel model: $\tilde{y}_{it} = \alpha_i + \rho_g \tilde{y}_{i,t-1} + \phi_g t + \tilde{e}_{it}$, if $i \in I_g^0$. See footnote below Table 1.

PWT v6.2 provided by Heston, Summers, and Aten (2006).[9] The regression model is

$$\tilde{y}_{it} = \alpha_i + \rho_g \tilde{y}_{it-1} + \phi_g t + \tilde{e}_{it}, \quad i \in I_g, \quad g=1,2,...,G, \qquad (9)$$

where $\tilde{y}_{it}$ is the log per-capita output, α_i denotes country-specific fixed effects, ρ_g and ϕ_g, $g=1,...,G$, $G=\{1,2,..,5\}$, are group specific. While previous studies allow for differences in α_i, we allow for the possibility that ϕ and ρ are potentially heterogeneous. The BIC suggests $G=1$, but t_g rejects parameter homogeneity, and the CK-means suggests $G=4$. Results for $G=4$ are reported in the left panel (Model A) of Table 4.

Evidently, $\hat{\phi}_g$ is negative in groups 1 and 2, but positive in groups 3 and 4. Furthermore, $\hat{\rho}_g$ is much smaller in groups 1 and 3 than in groups 2 and 4. Thus, both ρ_g and ϕ_g appear to be heterogeneous. Interestingly, the 21 OECD countries do not all belong to the same group while the fast growing countries including Indonesia, Korea, Malaysia, and Thailand are in the same (non-OECD) group. A priori information would unlikely arrive at such a grouping.

Equation (9) assumes cross-section independence in $\tilde{e}_{it}$. Pesaran (2006) suggests to control for cross-correlated errors by adding the cross-section average of appropriate variables to the pooled regression. To avoid simultaneity bias, we add $\Delta \bar{y}_{t-1}$ and $\Delta \bar{y}_{t-2}$ to both the pooled and the individual regressions, where $\bar{y}_s = N^{-1} \sum_i^N \tilde{y}_{is}$, $\Delta \bar{y}_s = \bar{y}_s - \bar{y}_{s-1}$, $s=1,...,T$. The result to highlight in the right panel of Table 4 (labeled Model B) is that countries differ in both the growth rate and in the speed of adjustment to equilibrium.

6.2 Example 2: housing dynamics

Housing wealth is a large component of household wealth and housing market activities are always closely watched

by policy makers and business cycle analysts. Stock and Watson (2009) provide a new data set on state-level monthly seasonally adjusted building permits (H_{it}) from 1969:1 to 2007:4. They use the K-means algorithm to cluster the four quarter change in the idiosyncratic component of $\log(H_{it})$, where the common component is estimated by a dynamic factor model. The K-means algorithm finds $G=5$ groups that roughly define contiguous geographical regions, even though no spatial structure was imposed in the estimation. Their results suggest that variations in housing permits consist of a national, a regional, and an idiosyncratic component.

We analyze the data to study if long run effects of the federal funds rate R_t on H_{it} are possibly group specific. For each $i=1,...,50$, an autoregressive distributed lag model in H_{it} and R_t can be reparameterized as:

$$\Delta^4 \log(H_{it}) = \alpha_i + \beta_{(i)1} \log(H_{it-1}) + \beta_{(i)2} R_{t-1}$$
$$+ \sum_{j=1}^3 \gamma_{(i)1,j} \Delta \log(H_{it-j}) + \sum_{j=1}^3 \gamma_{(i)2,j} \Delta R_{t-j} + e_{it}.$$

The parameter $\beta_{(i)2}$ is the sum of the coefficients on R_{t-1} and its lags, while $\beta_{(i)1}$ is the sum of the coefficients on the autoregressive terms. The long run effect $\frac{\beta_{(i)2}}{1-\beta_{(i)1}}$ can differ across states if $\beta_{(i)1}$ and/or $\beta_{(i)2}$ vary across states. We allow the short-run dynamics ($\gamma_{(i)1,j}$, $\gamma_{(i)2,j}$) to be state specific. As there is no evidence to support the need to consider more than two groups, we report only results of $G=1$ and $G=2$. We consider four models and the results are reported in Table 5.

(A) $\beta_{(i)1} = \beta_1$, $\beta_{(i)2} = \beta_2$ for all i;
(B) $\beta_{(i)1} = \beta_{g1}$ and $\beta_{(i)2} = \beta_{g2}$ for $i \in I_g$;
(C) $\beta_{(i)1} = \beta_1$ for all i, $\beta_{(i)2} = \beta_{g2}$ for $i \in I_g$;
(D) $\beta_{(i)1} = \beta_{g1}$ for $i \in I_g$, $\beta_{(i)2} = \beta_2$ for all i.

Model (A) is a pooled regression that imposes parameter homogeneity. Model (B) allows both $\beta_{(i)1}$ and $\beta_{(i)2}$ to be group specific. Model (C) allows $\beta_{(i)2}$ to be group specific but $\beta_{(i)1}$ is the same across i. Model (D) allows $\beta_{(i)1}$ to be group specific but $\beta_{(i)2}$ is the same across i. Not surprisingly,

---

9 We started with 75 countries, as in Mankiw, Romer, and Weil (1992). From this, Germany is removed from data set due to reunification. Due to limitation of data, we also remove Bangladesh, Bolivia, Botswana, Haiti, and Myanmar.

	Model A				Model B			
$\hat{\rho}_p$	0.9683 (218.0212)				0.9669 (208.6939)			
$\hat{\phi}_p$	0.0001 (0.9487)				0.0005 (4.4844)			
	PSEUDO				PSEUDO			
Group	1	2	3	4	1	2	3	4
N_g	9	13	23	24	20	9	19	21
$\hat{\rho}_g$	0.8663 (35.8575)	0.9693 (125.5729)	0.9178 (79.3623)	0.9592 (123.8460)	0.9606 (114.8522)	0.8736 (41.1522)	0.8866 (59.5900)	0.9681 (131.6144)
$\hat{\phi}_g$	-0.0009 (-3.4227)	-0.0005 (-2.3176)	0.0011 (4.4124)	0.0009 (3.8578)	-0.0002 (-1.0806)	0.0011 (3.1301)	0.0026 (7.3747)	0.0011 (4.6951)
t_g test	2.9625*	-0.2241	0.9209	6.3712*	2.8038*	-0.3125	5.1246*	3.2935*
	CK-means				CK-means			
Group	1	2	3	4	1	2	3	4
N_g	9	20	27	13	11	8	16	34
$\hat{\rho}_g$	0.8394 (31.0516)	0.9727 (166.9385)	0.8764 (73.1236)	0.9519 (93.9476)	0.9279 (57.2315)	0.8771 (42.6843)	0.916 (73.9328)	0.9817 (206.5877)
$\hat{\phi}_g$	-0.0009 (-3.0740)	-0.0004 (-2.5774)	0.0018 (7.3693)	0.0018 (4.8926)	-0.0004 (-1.4769)	0.0009 (2.1326)	0.0023 (7.0390)	0.0004 (2.6644)
t_g test	0.5454	1.1865	0.0757	1.2584	1.5780	0.3537	1.3110	5.2131*

Table 4: Application: Growth Regression.
Note: Regression models for $g = 1,2,...,4$:

Model A : $\tilde{y}_{it} = \alpha_i + \rho_g \tilde{y}_{it-1} + \phi_g t + \tilde{e}_{it}$, $i \in I_g$,
Model B : $\tilde{y}_{it} = \alpha_i + \rho_g \tilde{y}_{it-1} + \phi_g t + \beta_g \Delta \overline{y_{t-1}} + \gamma_g \Delta \overline{y_{t-2}} + \tilde{e}_{it}$, $i \in I_g$,

where α_i denotes country-specific fixed effects, ρ_g, ϕ_g, β_g and γ_g, are group specific, $\overline{y}_s = N^{-1} \sum_i^N \tilde{y}_{is}$, $\Delta \overline{y}_s = \overline{y}_s - \overline{y}_{s-1}$, $s = 1,...,T$. t values are in parentheses. t_g tests the null hypothesis of parameter homogeneity and * denotes the significance at the 5% nominal level.

the pooled estimates of β_1 and β_2 are roughly the average of the group estimates, and the sum of squared residuals of model (A) is largest. The group specific long run effects are always larger in absolute value than the one implied by the pooled regression.

Given the group-specific estimates $\hat{\beta}_1$ and/or $\hat{\beta}_2$, the t_g statistic is used to test for parameter heterogeneity in the full panel, as well as within the identified groups. While parameter homogeneity in the full panel is always rejected, the parameters in the subgroups appear to be homogeneous. The t_g statistic does not take into account that allowing for parameter heterogeneity increases the complexity of the model. When this is taken into account by the BIC, group effect is found for β_1 and only when group membership is determined by the CK-means. Overall, the data find some evidence for parameter heterogeneity but do not strongly reject a pooled regression for the 50 states. This shows that the proposed method favors group effects only if the evidence is strong.

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. A conditional K-means algorithm is also considered. Both methods can be used to estimate groupspecific parameters in panel data models when group membership is not known.

Appendix

In this appendix, we show that PSUEDO(2, 1) yields a negligible misclassification rate in the sense that $N_s(\hat{\gamma})/N \to 0$

Regression	(A)	(B)		(C)		(D)	
Homogenous	$\beta_{\cdot_1}, \beta_{\cdot_2}$	—		$\beta_{\cdot_1}$		$\beta_{\cdot_2}$	
Group-specific	—	$\beta_{\cdot_1}, \beta_{\cdot_2}$		$\beta_{\cdot_2}$		$\beta_{\cdot_1}$	
		PSEUDO	CK-means	PSEUDO	CK-means	PSEUDO	CK-means
$\hat{\beta}_1$	-0.117			-0.120	-0.120		
	(-15.887)			(-18.498)	(-18.498)		
$\hat{\beta}_2$	-0.017					-0.019	-0.019
	(-14.761)					(-18.403)	(-18.218)
$\hat{\beta}_{11}$		-0.261	-0.277			-0.231	-0.271
		(-12.404)	(-12.024)			(-14.059)	(-15.487)
$\hat{\beta}_{12}$		-0.024	-0.028	-0.024	-0.024		
		(-10.378)	(-10.271)	(-17.719)	(-17.719)		
$\hat{\beta}_{21}$		-0.095	-0.095			-0.102	-0.105
		(-12.461)	(-12.728)			(-14.502)	(-13.056)
$\hat{\beta}_{22}$		-0.017	-0.016	-0.010	-0.010		
		(-12.294)	(-12.520)	(-5.921)	(-5.921)		
$\dfrac{\hat{\beta}_2}{1-\hat{\beta}_1}$	-0.0193	-0.0325	-0.0387	-0.0273	-0.0273	-0.0247	-0.0261
	-0.0177	-0.0177	-0.0242	-0.0242	-0.0242	-0.0212	-0.0212
N_g	50	(16, 34)	(14, 36)	(31, 19)	(31, 19)	(14, 36)	(13, 37)
$\hat{G}$ (by BIC)	—	1	1	1	1	1	2
t_g : accept H_0	no	no	no	no	no	no	no
t_g : accept H_0'	—	no	yes	yes	yes	yes	yes
SSR	60.656	59.679	59.627	60.009	60.009	59.874	59.785

Table 5: Housing Dynamics in 50 States in the U.S.

$$\Delta^4 \log(H_{it}) = \alpha_i + \beta_{(i)1} \log(H_{it-1}) + \beta_{(i)2} R_{t-1} + \sum_{j=1}^{3} \gamma_{(i)1,j} \Delta \log(H_{it-j}) + \sum_{j=1}^{3} \gamma_{(i)2,j} \Delta R_{t-j} + e_{it}.$$

Note: α_i, $\gamma_{(i)1,j}$ and $\gamma_{(i)2,j}$ are state-specific. N_g denotes the size of the groups. t values are in parentheses. t_g is used to test for parameter homogeneity for the full sample (H_0), and for the subsamples (H_0') with group membership estimated by PSEUDO or CK-means.

as $(N, T) \to \infty$ jointly. Let $I^0 = (I_1^0, I_2^0)$ be the true group membership, let N_g^0 denote the number of individuals in cluster I_g^0 for $g = 1, 2$, and let $I = (I_1, I_2)$ denote group membership other than (I_1^0, I_2^0). Suppose that the DGP is

$$\tilde{y}_{it} = \alpha_i + \tilde{x}_{it} \beta_1 + \tilde{e}_{it}, \quad \text{for} \quad i \in I_1^0,$$

$$\tilde{y}_{it} = \alpha_i + \tilde{x}_{it} \beta_2 + \tilde{e}_{it}, \quad \text{for} \quad i \in I_2^0.$$

We will consider the general case where $\beta_2 - \beta_1 = \nu T^{-\alpha}$, $0 \leq \alpha < 1/2$, ν does not depend on T, and $\|\nu\| > 0$. Then $\alpha = 0$ corresponds to the case when $\beta_2 - \beta_1 = \nu \neq 0$. For $g, j = 1, 2$, let N_{gj} be the number of individuals assigned to be in group j by $I = (I_1, I_2)$ when individuals truly belong to group g. We also let $N_s = N_s(I, I^0) = N_{21} + N_{12}$ be the number of misclassified units.

If $N_s/N \to 0$, then we have $\hat{\beta}_g(\hat{q}) \to \beta_g$, $g = 1, 2$. In the following, we will show $N_s/N \to 0$ under PSEUDO. 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 $\gamma^0 \in [\gamma_{\min}^0, \gamma_{\max}^0]$,

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

Consider the following cases:

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

$$\frac{N_s}{N} = \frac{1}{N} \sum_{i \in I_1^0}^{N} 1(\hat{q}_i > \gamma_{\max}^0) + \frac{1}{N} \sum_{i \in I_2^0}^{N} 1(\hat{q}_i < \gamma_{\min}^0)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\gamma_{\max}^0 < q_i + \frac{\hat{c}_i}{\sqrt{T}}\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\gamma_{\min}^0 > q_i + \frac{\hat{c}_i}{\sqrt{T}}\right)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} > \gamma_{\max}^0 - q_i\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} < \gamma_{\min}^0 - q_i\right).$$

Note that $\hat{c}_i > 0$ if $\hat{q}_i > q_i$ and $i \in I_1^0$, and $\hat{c}_i < 0$ if $\hat{q}_i < q_i$ and $i \in I_2^0$. Now $\gamma_{max}^0 = \beta_2$ and $\gamma_{min}^0 = \beta_1$ with $q_i = \beta_1$ for those $i \in I_2^0$, and $q_i = \beta_2$ for those $i \in I_2^0$.

$$\frac{N_s}{N} = \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} > \beta_2 - q_i\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} < \beta_1 - q_i\right)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} > \beta_2 - \beta_1\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} < \beta_1 - \beta_2\right)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\hat{c}_i > \nu T^{1/2-\alpha}\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\hat{c}_i < -\nu T^{1/2-\alpha}\right).$$

Since $0 \leq \alpha < 1/2$, N_s/N tends to zero as $T \to \infty$.

II. Γ^0 and q_i both unknown: Now turn to the case when Γ^0 and q_i are both unknown. Under the assumption that $\hat{\gamma}_{min} - \gamma_{min}^0$ and $\hat{\gamma}_{max} - \hat{\gamma}_{max}^0$ 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 $\frac{\hat{d}_{min}}{N\sqrt{T}} = \gamma_{min}^0 - \hat{\gamma}_{max}$. Note that $\hat{d}_{max}$ and $\hat{d}_{min}$ do not depend on i. Because $\hat{q}_i = q_i + \frac{\hat{c}_i}{\sqrt{T}}$, we have

$$\frac{N_s}{N} \leq \frac{1}{N} \sum_{i \in I_1^0}^{N} 1(\hat{q}_i > \hat{\gamma}_{max}) + \frac{1}{N} \sum_{i \in I_2^0}^{N} 1(\hat{q}_i < \hat{\gamma}_{min})$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{max}}{N\sqrt{T}} > \gamma_{max}^0 - q_i\right)$$

$$+ \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{min}}{N\sqrt{T}} < \gamma_{min}^0 - q_i\right)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{max}}{N\sqrt{T}} > \beta_2 - \beta_1\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{min}}{N\sqrt{T}} < \beta_1 - \beta_2\right)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{max}}{N\sqrt{T}} > \nu T^{-\alpha}\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\frac{\hat{c}_i}{\sqrt{T}} + \frac{\hat{d}_{min}}{N\sqrt{T}} < -\nu T^{-\alpha}\right)$$

$$= \frac{1}{N} \sum_{i \in I_1^0} 1\left(\hat{c}_i > \nu T^{-\alpha+1/2} - \frac{\hat{d}_{max}}{N}\right) + \frac{1}{N} \sum_{i \in I_2^0} 1\left(\hat{c}_i < -\nu T^{-\alpha+1/2} - \frac{\hat{d}_{min}}{N}\right).$$

Notice that

$$P\left(\hat{q}_i > \hat{\gamma}_{max} | i \in I_1^0\right) = P\left(\hat{c}_i > \nu T^{-\alpha+1/2} - \frac{\hat{d}_{max}}{N} \Big| i \in I_1^0\right)$$

$$\leq P\left(|\hat{c}_i| > \nu T^{-\alpha+1/2} + O_p\left(N^{-1}\right) \Big| i \in I_1^0\right) = O\left(T^{2\alpha-1}\right),$$

where the last equality comes from the Chebyshev's inequality. Similarly,

$$P\left(\hat{q}_i < \hat{\gamma}_{min} | i \in I_2^0\right) = O\left(T^{2\alpha-1}\right).$$

Thus, $E(N_s/N) = O(T^{2\alpha-1})$. Since $0 \leq \alpha < 1/2$, N_s/N tends to zero as $T \to \infty$.

References

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

Alvarez, J., and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121–1159.

Alvarez, J., M. Browning, and M. Ejrnæs (2010): "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77, 1353–1381.

Anderson, T. W., and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Andrews, D., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter is Present only under the Alternative," *Econometrica*, 62, 1383–1414.

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

Browning, M., and J. Carro (2007): "Heterogeneity and Microeconometric Modeling," *Advances in Economics and Econometrics*, 3, edited by Richard Blundell, Whitney Newey and Torsten Persson, Cambridge University.

Burnside, C. (1996): "Production Function Regressions, Returns to Scale and Externalities," *Journal of Monetary Economics*, 37, 177–201.

Calinski, T., and J. Harabasz (1974): "A Dendrite Method for Cluster Analysis," *Communication in Statistics*, 3, 1–27.

Caner, M., and B. E. Hansen (2004): "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843.

Davies, R. B. (1977): "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 64, 247–254.

Duda, R., and D P. Hart (1973): *Pattern Classification and Scene Analysis*. Wiley.

Garcia-Escudero, L. A., and A. Gordaliza (1999): "Robustness Properties of K Means and Trimmed K Means," *Journal of the American Statistical Association*, 94, 956–969.

Goldfeld, S., and R. Quandt (1973): "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement*, 2, 473–483.

Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for A Dynamic Panel Model with Fixed Effects when Both N and T are Large," *Econometrica*, 70, 1639–1657.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

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

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

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

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," edited by L. Matyas and P. Sevestre, Third Edition, Springer Publishers, Ch. 6.

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.

Lee, K. C., M. H. Pesaran, and R. P. Smith (1997): "Growth and Convergence in a Multi-country Empirical Stochastic Solow Model," *Journal of Applied Econometrics*, 12, 357–392.

Maddala, G. S., R. P. Trost, H. LI, and F. Joutz (1997): "Estimationn of Short Run and Long Run Elasticities of Energy Demand from Panel Data using Shrinkage Estimators," *Journal of Business and Economic Statistics*, 15, 90–100.

Mankiw, N. G., D. Romer, and D. N. Weil (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407–437.

Milligan, G., and W. Cooper (1985): "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, 50, 159–179.

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

Pesaran, M. H., and T. Yamagata (2008): "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics*, 142, 50–93.

Pollard, D. (1981): "Strong Consistency of K-means Clustering," *Annals of Statistics*, 9, 135–140.

Pollard, D. (1982): "A Central Limit Theorem for K-Means Clustering," *Annals of Probability*, 10, 919–926.

Qin, L. X., and S. G. Self (2006): "The Clustering of Regression Model Methods with Applications in Gene Expresssion Data," *Biometrics*, 62, 526–533.

Robertson, D., and J. Symons (1992): "Some Strange Properties of Pooled Data Estimators," *Journal of Applied Econometrics*, 7, 175–189.

Seo, M. K. and Y. Shin (2011): "Inference for Dynamic Panels with Threshold Effects and Endogeneity," mimeo, London School of Economics.

Song, K. K. (2004): "Large Panel Models with Limited Heterogeneity," mimeo, University of Pennsylvannia.

Stock, J. H. and M. W. Watson (2009): "The Evolution of National and Regional Factors in U.S. Housing Construction," in *Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle*, edited by T. Bollerslev, J. Russell and M. W. Watson, Oxford University Press.

Sugar, C., and G. James (2003): "Finding the Number of Clsuters in a Dataset: An Information Theoretic Approach," *Journal of the American Statistical Association*, 98, 750–763.

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.