Journal of Econometrics 213 (2019) 54–67

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Quantile-regression-based clustering for panel data

Yingying Zhang [a], Huixia Judy Wang [b,*], Zhongyi Zhu [a]

[a] Department of Statistics, Fudan University, China
[b] Department of Statistics, The George Washington University, USA

ARTICLE INFO

Article history:
Available online 11 April 2019

JEL classification:
C13
C21
C23

Keywords:
Fixed effects
Heterogeneity
Panel data
Quantile regression
Subgroup identification

ABSTRACT

In panel data analysis, it is important to identify subgroups of units with heterogeneous parameters. This can not only increase the model flexibility but also produce more efficient estimation by pooling information across units within the same group. In this paper, we propose a new quantile-regression-based clustering method for panel data. We develop an iterative algorithm using a similar idea of k-means clustering to identify subgroups with heterogeneous slopes at a single quantile level or across multiple quantiles. The asymptotic properties of the group membership estimator and corresponding group-specific slope estimator are established. The finite sample performance of the proposed method is assessed through simulation and the analysis of an economic growth data.

1. Introduction

For panel data analysis, one common practice is to assume heterogeneous unit-specific intercepts but homogeneous slopes across units. While this allows researchers to pool information across units to improve the estimation efficiency of slopes, empirical evidences have suggested that the homogeneous slope assumption may be too restrictive in applications; see Burnside (1996), Lee et al. (1997), Hsiao and Tahmiscioglu (1997), and Browning and Carro (2007). On the other hand, assuming heterogeneity for both intercept and slope parameters would reduce the problem to separate analysis for each unit, which does not utilize the panel data feature and often leads to estimation with large variation. One natural alternative approach is to consider grouped panel data models where the slope parameters are group-specific. Such models allow information sharing across units while accommodating heterogeneity in the regression function, and thus balance better between model flexibility and parsimony.

In practice, the group membership of units is often unknown and has to be estimated from the observed data. Identifying groups with homogeneous slope parameters can be viewed as a model-based clustering problem. In this paper, we propose a new quantile-regression-based clustering method for panel data to identify subgroups of units with the same covariate effect. Our proposed method provides a natural way to capture the heteroscedasticity both across and within units, and allows us to explore subgroup structure at different quantile levels.

We first develop an iterative two-step algorithm using a similar idea of k-means clustering to identify subgroups at a single quantile level. Analysis at a given quantile level separately can help us identify quantile-specific subgroups. To obtain a complete picture of the heterogeneity in the conditional distribution, we extend the clustering method to multiple quantiles to identify subgroups whose slopes are heterogeneous across groups for at least one quantile level. For both single and multiple quantile clustering methods, we establish the consistency of the proposed group-specific slope

* Corresponding author.
E-mail address: judywang@gwu.edu (H.J. Wang).

https://doi.org/10.1016/j.jeconom.2019.04.005
0304-4076/

estimator with respect to Hausdorff distance, as well as the consistency of the group membership estimator. In addition, we provide the asymptotic distribution of the group-specific slope estimator by establishing its asymptotic equivalency with the infeasible estimator with known group membership when both dimensions of the panel diverge. Finally, for cases where the group membership is believed to be common across quantiles, we consider a stability measurement to choose an empirically "optimal" quantile that gives the most stable clustering results. The proposed measurement leads to a data-adaptive criterion that can improve the accuracy of subgroup identification.

In the literature there exist several classes of work for identifying group membership in regression setup. One class of work is based on mixture modeling and EM algorithms; see for instance Rubin and Wu (1997), Sun et al. (2007), and Tang and Qu (2016). Another class of work is based on some penalization of slope differences between pairs of units to identify the sparsity, that is, the homogeneity of coefficients; see for instance Ma and Huang (2017) and Su et al. (2016). The penalization methods share some similar spirit as those used in regression homogeneity pursuit to group predictors with the same coefficients (Bondell and Reich, 2008; Shen and Huang, 2010; Zhu et al., 2013; Ke et al., 2015). One common limitation of the mixture modeling and penalization methods is the computational challenge when the number of units is large, for which a huge number of possible combinations need to be involved in the estimation algorithms. Another class of work is similar to ours, which aims to solve the minimization of an objective function over possible groupings by using some iterative algorithms that are in the similar spirit of k-means clustering for univariate data. Research in this direction include Lin and Ng (2012), Bonhomme and Manresa (2015), Ando and Bai (2017, 2016). Recently, Ke et al. (2016) considered a panel data model that allows homogeneity among all regression coefficients, and they proposed an estimation procedure by converting the subgroup identification problem to the detection of multiple change points based on the sorted initial estimation of unit-specific coefficients. All the existing works focus on mean regression, and most assume homoscedastic regression errors. However, heteroscedasticity is often seen in applications, which may cause the covariates to have different impact at different location of the response distribution. Our proposed method can automatically capture such heteroscedasticity and provide a comprehensive picture of the relationship between the response and explanatory variables.

The rest of the article is organized as follows. In Section 2, we first describe the proposed iterative algorithm at a single quantile level, and then extend the algorithm to multiple quantiles. In Section 3, we establish the theoretical properties of the proposed subgroup estimator and the resulting group-specific quantile coefficient estimator. In Section 4, we discuss some computational issues, including the choice of the number of groups, and the selection of an empirically optimal quantile under a stronger assumption of common group membership across quantiles. The numerical performance of the proposed methods is evaluated through simulation and the analysis of an economic growth data in Sections 5 and 6, respectively. Section 7 concludes the paper with some discussion. The technical proofs are provided in the online supplementary material.

2. Proposed method

2.1. Single quantile clustering

Let y_{it} be the response of unit i at time t, and $\boldsymbol{x}_{it}$ be the p-dimensional vector of covariates, where $t = 1, \ldots, T_i$ and $i = 1, \ldots, N$. For simplicity we assume balanced designs with $T_i = T$ for $i = 1, \ldots, N$. At a given quantile level $\tau \in (0, 1)$, we consider the following quantile regression grouped panel data model:

$$Q_{y_{it}}(\tau|\boldsymbol{x}_{it}) = \alpha_i(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}_{g_i(\tau)}(\tau), \tag{1}$$

where $Q_{y_{it}}(\tau|\boldsymbol{x}_{it})$ denotes the τth conditional quantile of y_{it} given $\boldsymbol{x}_{it}$, $\alpha_i(\tau)$ are the unit-specific fixed effect parameters, and $g_i(\tau)$ are the group memberships that take values in set $\{1, \ldots, G\}$ with G being the number of subgroups. In Model (1), both the parameters $(\alpha_i(\tau), \boldsymbol{\beta}_g(\tau))$ and the group membership $g_i(\tau)$ can depend on the quantile level τ. We assume that the number of subgroups G is known, and defer the discussion of how to determine G in practice to Section 4.

Quantile regression panel models have been also considered in Koenker (2004), Harding and Lamarche (2009), Lamarche (2010), Galvao and Montes-Rojas (2010), Canay (2011), Galvao (2011), Rosen (2012), Kato et al. (2012), Galvao and Wang (2015) and Galvao and Montes-Rojas (2015). A more comprehensive review of recent developments can be found in Galvao and Kato (2017). All these works assume that the covariate effects $\boldsymbol{\beta}(\tau)$ are homogeneous across units. In contrast, Model (1) assumes that the unobserved heterogeneity lies not only in the intercept but also in the slope coefficients in the sense that $\boldsymbol{\beta}(\tau)$ is constant within a subgroup but varies across subgroups. To test the slope homogeneity across units, Galvao et al. (2017) proposed a testing procedure, which can be used as a first step to determine whether a grouped panel data model is suitable prior to applying our proposed method. Chetverikov et al. (2016) considered another setup which models the unit-specific slopes of covariates as a linear combination of another set of unit-level covariates. If the unit-level covariates indicate the group memberships of units, the model in Chetverikov et al. (2016) will reduce to the grouped panel Model (1). However, Chetverikov et al. (2016) requires the unit-level covariates to be observable, while we assume that the group membership is unknown and we aim to identify the membership through clustering.

We first propose a quantile-regression-based clustering method for Model (1) at a given quantile level τ. Our goal is twofold: one is to identify the subgroup membership $g_i(\tau)$ and the other is to obtain accurate estimation of the group-specific parameters $\boldsymbol{\beta}_g(\tau)$ for $g = 1, \ldots, G$. Let $B(\tau) = \{\boldsymbol{\beta}_g(\tau) : g = 1, \ldots, G\}$ be the set comprising of all group-specific

slopes and $\boldsymbol{\gamma}(\tau) = \{g_i(\tau), i = 1, \ldots, N\}$ be the set of group memberships for N units. Thus, $\boldsymbol{\gamma}(\tau) \in \mathcal{F}_G$ denotes a particular partition of the N units, where $\mathcal{F}_G$ is the set of all partitions of $\{1, \ldots, N\}$ into G groups.

To remove the unit-specific fixed effects, one common practice used in the mean-regression-based methods is to use demeaned data obtained by subtracting the unit-specific means from y_{it} and $\boldsymbol{x}_{it}$; see for instance Lin and Ng (2012). Unfortunately, the simple demeaning approach cannot be applied in quantile regression as quantiles are not linearly additive. Instead, we consider a two-stage estimation procedure. Let $\mathcal{A}(\tau)$ be a compact subset of $\mathbb{R}$ and $\Theta(\tau)$ be a compact subset of $\mathbb{R}^p$. In the first stage, we fit the quantile regression for each unit and estimate the fixed effect $\alpha_i(\tau)$ by $\tilde{\alpha}_i(\tau)$, where

$$(\tilde{\alpha}_i(\tau), \tilde{\boldsymbol{\beta}}_i(\tau)) = \underset{a \in \mathcal{A}(\tau), \mathbf{b} \in \Theta(\tau)}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \rho_\tau(y_{it} - a - \boldsymbol{x}_{it}'\mathbf{b}), \tag{2}$$

where $\rho_\tau(u) = \{\tau - I(u < 0)\}u$ is the quantile loss function. In the second stage, we estimate the group memberships and the group-specific parameters by

$$(\hat{B}(\tau), \hat{\boldsymbol{\gamma}}(\tau)) = \underset{B(\tau) \subseteq \Theta(\tau), \boldsymbol{\gamma}(\tau) \in \mathcal{F}_G}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_g(\tau) - \tilde{\alpha}_i(\tau)\}, \tag{3}$$

where the minimum is taken over all possible partitions of the N units into G groups and the group-specific parameters from a compact subset of $\mathbb{R}^p$. Intuitively, the two-stage estimation works for large T since $\tilde{\alpha}_i(\tau)$ is a $\sqrt{T}$-consistent estimator of $\alpha_i(\tau)$ so that $y_{it} - \tilde{\alpha}_i(\tau)$ can be viewed as an approximation of $y_{it} - \alpha_i(\tau)$, whose τth conditional quantile is $\boldsymbol{x}_{it}'\boldsymbol{\beta}_{g_i(\tau)}(\tau)$. However, the error involved in $\tilde{\alpha}_i(\tau)$ will lead to dependence among $y_{it} - \tilde{\alpha}_i(\tau)$ and thus complicate the theoretical development.

For the optimization in the second stage, exhaustive search is virtually impossible since the number of combinations for partitioning N units into G groups increases steeply with N. To bypass the computational challenge, we propose an iterative procedure by a similar idea as k-means clustering. The idea of the iteration is as follows. Note that for any given group-specific coefficients $\hat{B}(\tau) = \{\hat{\boldsymbol{\beta}}_g(\tau), g = 1, \ldots, G\}$, the optimal subgroup assignment for unit i is

$$\hat{g}_i\{\hat{B}(\tau), \tau\} = \underset{g \in \{1, \ldots, G\}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \rho_\tau\{y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}_g(\tau) - \tilde{\alpha}_i(\tau)\}, \tag{4}$$

where in the presence of non-unique solutions an arbitrary solution can be assigned to $\hat{g}_i$. On the other hand, given the group membership $\hat{\boldsymbol{\gamma}}(\tau) = \{\hat{g}_i(\tau), i = 1, \ldots, N\}$, the estimator of $\boldsymbol{\beta}_g(\tau)$ can be updated by

$$\hat{\boldsymbol{\beta}}_g\{\tau, \hat{\boldsymbol{\gamma}}(\tau)\} = \underset{\mathbf{b} \in \Theta(\tau)}{\operatorname{argmin}} \frac{1}{NT} \sum_{i: \hat{g}_i(\tau) = g} \sum_{t=1}^{T} \rho_\tau\{y_{it} - \boldsymbol{x}_{it}'\mathbf{b} - \tilde{\alpha}_i(\tau)\}. \tag{5}$$

Below we describe the details of the proposed two-stage estimation procedure.

Stage 1. Obtain the unit-specific fixed-effect estimator $\tilde{\alpha}_i(\tau)$ by solving (2) for each unit $i = 1, \ldots, N$.
Stage 2. Initialize the unknown group-specific parameters $\boldsymbol{\beta}_g(\tau)$. Then iteratively update $\hat{g}_i\{\hat{B}(\tau), \tau\}$ for each unit by (4) with the given values of $\hat{\boldsymbol{\beta}}_g(\tau)$, and update $\hat{\boldsymbol{\beta}}_g\{\tau, \hat{\boldsymbol{\gamma}}(\tau)\}$ by (5) with the given values of $\hat{\boldsymbol{\gamma}}(\tau)$ until convergence is met, that is, when no unit changes the group membership.

For the initialization of $\boldsymbol{\beta}_g(\tau)$ in the second stage, we can randomly assign N units into G groups and then estimate the quantile coefficient $\boldsymbol{\beta}_g(\tau)$ based on units assigned to the gth group. Since the solution depends on the starting values, we repeat this procedure a number of times and select the solution that yields the lowest overall quantile objective function value across all units. Throughout the iteration, the algorithm may lead to empty groups with no members. When this happens, we make a simple modification by randomly reassigning members in the biggest group to the empty groups, and continue the iteration until the algorithm converges and no empty groups are present.

Remark 1. The proposed procedure shares the similar spirit with the classic k-means clustering for univariate data in the sense that in each iteration step a unit is assigned to the cluster with the closest conditional quantile, where the distance is measured by the quantile objective function. Bottou and Bengio (1995) showed that the classic k-mean clustering algorithm converges to a local minimum of the quantization error; see more recent discussions on convergence issues of k-means algorithms in Li et al. (2015), Tang and Monteleoni (2016). To facilitate convergence and to mitigate the dependence of the algorithm on the initial partition, we draw multiple starting values randomly and choose the one yielding the smallest objective function value. Our numerical results show that this practice works well for data with modest N and G. For problems at larger scales, we may consider applying the k-mean clustering on the unit-specific estimators $\{\tilde{\boldsymbol{\beta}}_i(\tau), i = 1, \ldots, N\}$ to obtain starting values for the grouping, and adopt the local searching and neighborhood jumping idea in Algorithm 2 of Bonhomme and Manresa (2015) to achieve better convergence.

2.2. Clustering across multiple quantiles

The single-quantile clustering algorithm can help identify subgroups at a specific quantile level τ. In applications, it may happen that $\boldsymbol{\beta}_g(\tau) = \boldsymbol{\beta}_{g'}(\tau)$ at the quantile level τ, but $\boldsymbol{\beta}_g(\tau') \neq \boldsymbol{\beta}_{g'}(\tau')$ at another quantile level τ' or vice versa. To capture a more complete picture of the heterogeneity, we extend the algorithm in Section 2.1 to clustering across multiple quantiles. This shares some similar idea as in Qu (2008), and Oka and Qu (2011), which studied structural changes in multiple quantiles.

Suppose that $\boldsymbol{\beta}_g(\tau)$ has heterogeneity for $\tau \in \mathcal{T} = [\omega_1, \omega_2]$ with $0 < \omega_1 < \omega_2 < 1$. One natural approach is to identify subgroups by clustering $\boldsymbol{\beta}_g(\tau)$ across a grid of quantile levels within the interval, referred to as $\omega_1 = \tau_1 < \cdots < \tau_K = \omega_2$. Let $B(\boldsymbol{\tau}) = \{\boldsymbol{\beta}_g(\boldsymbol{\tau}), g = 1, \ldots, G\}$ be the collection of group-specific slopes at these K quantiles, where $\boldsymbol{\beta}_g(\boldsymbol{\tau}) = (\boldsymbol{\beta}_g(\tau_1)', \ldots, \boldsymbol{\beta}_g(\tau_K)')'$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$. Two units are considered to be in different groups g and g' if their slopes differ at some quantile, that is, $\boldsymbol{\beta}_g(\boldsymbol{\tau}) \neq \boldsymbol{\beta}_{g'}(\boldsymbol{\tau})$. In addition, denote $\boldsymbol{\gamma} = \{g_i(\boldsymbol{\tau}) : i = 1, \ldots, N\}$ as the group memberships of each unit, which is determined by the heterogeneity of slopes at K quantiles.

Similar as in Section 2.1, we define the estimator of $(B(\boldsymbol{\tau}), \boldsymbol{\gamma})$ as

$$(\hat{B}(\boldsymbol{\tau}), \hat{\boldsymbol{\gamma}}) = \underset{B(\boldsymbol{\tau})\in\Theta(\boldsymbol{\tau}), \boldsymbol{\gamma}\in\Gamma_G}{\operatorname{argmin}} \frac{1}{NT} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau_k}\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_g(\tau_k) - \tilde{\alpha}_i(\tau_k)\}, \tag{6}$$

where the optimization is over all possible groupings $\boldsymbol{\gamma}$ and the parameter space for slopes at K quantiles, $\Theta(\boldsymbol{\tau}) \doteq \Theta(\tau_1) \times \Theta(\tau_2) \times \cdots \times \Theta(\tau_K)$. To solve this minimization problem, we can modify the iteration algorithm in Section 2.1 by updating $\boldsymbol{\gamma}$ with

$$\hat{g}_i\{B(\boldsymbol{\tau})\} = \underset{g\in\{1,\ldots,G\}}{\operatorname{argmin}} \frac{1}{T} \sum_{k=1}^{K} \sum_{t=1}^{T} \rho_{\tau_k}\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_g(\tau_k) - \tilde{\alpha}_i(\tau_k)\} \text{ for } i = 1, \ldots, N,$$

and updating the quantile slope coefficients by

$$\hat{\boldsymbol{\beta}}_g(\tau_k, \hat{\boldsymbol{\gamma}}) = \underset{\mathbf{b}\in\Theta(\tau_k)}{\operatorname{argmin}} \frac{1}{NT} \sum_{i:\hat{g}_i=g} \sum_{t=1}^{T} \rho_{\tau_k}\{y_{it} - \boldsymbol{x}_{it}'\mathbf{b} - \tilde{\alpha}_i(\tau_k)\}$$

for $g = 1, \ldots, G$ and $k = 1, \ldots, K$. The choice of $\mathcal{T}$ and the quantile grid is an empirical issue. The interval $[\omega_1, \omega_2]$ can often be determined by the research interest. As for the quantile grid, evidence from empirical studies (Oka and Qu, 2011; Chamberlain, 1994; Angrist et al., 2006) suggested that a coarse grid of quantiles with space between 5% and 15% often suffices to obtain an overall picture of the conditional distribution. Throughout our numerical studies, we consider a grid of five quantiles, 0.3, 0.4, ..., 0.7.

3. Asymptotic properties

In Section 3.1, we establish the consistency and asymptotic normality of the proposed single-quantile-based estimator for identifying quantile-specific group membership. In Section 3.2, we present the asymptotic properties of the multiple-quantile clustering method. The statistical properties are established in an asymptotic sense as both N and T tend to infinity.

3.1. Single quantile

In this subsection, group memberships $g_i(\tau)$ are allowed to be quantile-specific. We first establish the consistency of the group-specific coefficient estimator and the group membership estimator. To establish the asymptotic distribution of the group-specific coefficient estimator, we introduce an oracle-type estimator of $B(\tau)$ obtained by assuming known group membership. Let $\boldsymbol{\gamma}^0(\tau) = \{g_1^0(\tau), \ldots, g_N^0(\tau)\}$ denote the true group membership. The oracle estimator is defined as $B^*(\tau) = \{\boldsymbol{\beta}_g^*(\tau) : g = 1, \ldots, G\}$, where

$$\boldsymbol{\beta}_g^*(\tau) = \underset{\mathbf{b}\in\Theta(\tau)}{\operatorname{argmin}} \frac{1}{NT} \sum_{i:g_i^0(\tau)=g} \sum_{t=1}^{T} \rho_\tau\{y_{it} - \boldsymbol{x}_{it}'\mathbf{b} - \tilde{\alpha}_i(\tau)\}. \tag{7}$$

The oracle estimator $B^*(\tau)$ can be regarded as the infeasible counterpart of $\hat{B}(\tau)$. Note that $B^*(\tau)$ is a two-step estimator with known group membership based on the first-step unit-specific estimator $\tilde{\alpha}_i(\tau)$, and thus differs from the one-step estimator in Kato et al. (2012). For large N and T, we establish conditions under which $\hat{\boldsymbol{\beta}}_g(\tau)$ and $\boldsymbol{\beta}_g^*(\tau)$ are asymptotically equivalent so that the asymptotic distribution can be obtained through the Bahadur representation of the latter.

We fix some notation and introduce the assumptions required to establish the consistency. Let $B^0(\tau) = \{\boldsymbol{\beta}_g^0(\tau) : g = 1, \ldots, G\}$ and $\{\alpha_i^0(\tau), i = 1, \ldots, N\}$ denote the true parameters. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, define

$v_{it}(\tau) = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_{g_i^0}^0(\tau) - \alpha_i^0(\tau)$, and let $F_{\tau i}(\cdot|\boldsymbol{x}_{it})$ and $f_{\tau i}(\cdot|\boldsymbol{x}_{it})$ be the conditional and density functions of $v_{it}(\tau)$ given $\boldsymbol{x}_{it}$, respectively. In addition, define $\Delta_G(\tau) := \{B(\tau) \subseteq \Theta(\tau) : B(\tau)$ contains G or fewer points$\}$. For any $B(\tau) = \{\boldsymbol{\beta}_g(\tau) : g = 1, \ldots, G\}$, let $\varpi_g\{B(\tau)\}$ denote the minimum eigenvalue of the matrix

$$N^{-1} \sum_{i:g_i^0(\tau)=g} \mathrm{E}\left\{f_{\tau i}[\boldsymbol{x}_{i1}'\{\boldsymbol{\beta}_{g_i^0}(\tau) - \boldsymbol{\beta}_{g_i^0}^0(\tau)\}|\boldsymbol{x}_{i1}]\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}'\right\}. \tag{8}$$

Assumption 1.

(a) For all $g \in \{1, \ldots, G\}$, $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N I\{g_i^0(\tau) = g\} = \pi_g(\tau) > 0$.

(b) For each $i = 1, \ldots, N$, $\alpha_i^0(\tau) \in \mathcal{A}(\tau)$, and $B^0(\tau) \subseteq \Theta(\tau)$. Furthermore, $\boldsymbol{\beta}_g^0(\tau)$ are distinct across $g = 1, \ldots, G$.

(c) There exists an M such that $\sup_{i\geq 1}\|\boldsymbol{x}_{i1}\| \leq M$ a.s.

(d) $\{(y_{it}, \boldsymbol{x}_{it}), t \geq 1\}$ is stationary and β-mixing for each i, and independent across i. Let $\beta_i(j)$ denote the β-mixing coefficients of $\{(y_{it}, \boldsymbol{x}_{it}), t \geq 1\}$. Then there exist constants $a \in (0, 1)$ and $C_1 > 0$ such that $\sup_{i\geq 1}\beta_i(j) \leq C_1 a^j$ for all $j \geq 1$.

(e) For each $\kappa > 0$, $\epsilon_\kappa = \inf_{1\leq i\leq n}\inf_{\|(a,\mathbf{b})\|=\kappa} \mathrm{E}\left[\int_0^{a+\boldsymbol{x}_{it}'\mathbf{b}}\{F_{\tau i}(s|\boldsymbol{x}_{it}) - \tau\}ds\right] > 0$.

(f) For all g, there exist an $\varepsilon > 0$ such that $\inf_{B(\tau)\in\mathcal{N}^*(\varepsilon)}\varpi_g\{B(\tau)\} \to \varpi_g > 0$, where $\mathcal{N}^*(\varepsilon)$ is the set of parameters $B(\tau) \in \Delta_G$ that satisfy $\|\boldsymbol{\beta}_g(\tau) - \boldsymbol{\beta}_g^0(\tau)\| < \varepsilon$ for all $g \in \{1, \ldots, G\}$.

Assumptions 1(a)–(b) ensure that the G subgroups are well separated so that the parameters $B^0(\tau)$ and $\boldsymbol{\gamma}^0(\tau)$ are identifiable (subject to permutations of group labels). Assumptions 1(c)–(e) are standard conditions in the literature of quantile regression for panel data; see for instance Kato et al. (2012), Galvao and Wang (2015). Specifically, 1(c) assumes the uniform boundedness of the covariates; 1(d) is considered to accommodate β-mixing data; 1(e) is an identification condition for $(\alpha_i^0(\tau), \boldsymbol{\beta}_i^0(\tau)) = \mathrm{argmin}_{(a,\mathbf{b})\in\mathbb{R}^{p+1}} E\left[\rho_\tau(y_{it} - a - \mathbf{x}_{it}'\mathbf{b})\right]$ and is needed to establish the consistency of both $\tilde{\alpha}_i(\tau)$ and $\hat{B}(\tau)$. Assumption 1(f) is a condition that ensures that the matrix in (8) behaves well in the neighborhood of the true parameters, which is needed to establish the asymptotic equivalency of $\hat{B}(\tau)$ with its infeasible counterpart $B^*(\tau)$.

Since the objective function in (3) is invariant to permutations of group labels, we establish the consistency of $\hat{B}(\tau)$ with respect to the Hausdorff distance d_H in $\Delta_G(\tau)$, defined by

$$d_H\{C(\tau), D(\tau)\} = \max\Big\{\max_{g\in\{1,\ldots,G_2\}}\min_{\tilde{g}\in\{1,\ldots,G_1\}}\|\boldsymbol{\theta}_{\tilde{g}}(\tau) - \boldsymbol{\xi}_g(\tau)\|, \max_{\tilde{g}\in\{1,\ldots,G_1\}}\min_{g\in\{1,\ldots,G_2\}}\|\boldsymbol{\theta}_{\tilde{g}}(\tau) - \boldsymbol{\xi}_g(\tau)\|\Big\},$$

where $C(\tau) = \{\boldsymbol{\theta}_g(\tau) \in \mathbb{R}^p : g = 1, \ldots, G_1\}$ and $D(\tau) = \{\boldsymbol{\xi}_g(\tau) \in \mathbb{R}^p : g = 1, \ldots, G_2\}$ with $G_1 \leq G$ and $G_2 \leq G$.

Theorem 1. *Suppose that Model (1) and Assumption 1 hold at $\tau \in (0, 1)$, as $N, T \to \infty$, we have*

(i) if $(\log N)/T \to 0$, then $d_H\{\hat{B}(\tau), B^0(\tau)\} \xrightarrow{p} 0$;
(ii) if $(\log N)^2/T \to 0$, then $Pr\left\{\sup_{i\in\{1,\ldots,N\}}|\hat{g}_i(\tau) - g_i^0(\tau)| > 0\right\} = o(1)$;
(iii) if $(\log N)^2/T \to 0$, and $T = o(N^\upsilon)$ with some $\upsilon > 0$ for $0 < \delta \leq 2$ and $0 < \upsilon < 2/(\delta - 2)$ for $\delta > 2$, then $\hat{\boldsymbol{\beta}}_g(\tau) = \boldsymbol{\beta}_g^*(\tau) + o_p(T^{-\delta/2})$ for all $g \in \{1, \ldots, G\}$.

As Pollard (1981) has pointed out, the consistency with respect to Hausdorff distance in Theorem 1(i) indicates that there exists a permutation $\sigma : \{1, \ldots, G\} \to \{1, \ldots, G\}$ such that $\|\hat{\boldsymbol{\beta}}_{\sigma(g)}(\tau) - \boldsymbol{\beta}_g^0(\tau)\| \xrightarrow{p} 0$. By a simple relabeling, we may take $\sigma(g) = g$, so that we have $\|\hat{\boldsymbol{\beta}}_g(\tau) - \boldsymbol{\beta}_g^0(\tau)\| \xrightarrow{p} 0$ for all g. We adopt this convention in Theorem 1. The condition $(\log N)/T \to 0$ in Theorem 1(i) is to ensure the uniform consistency of the first-step estimator $\tilde{\alpha}_i(\tau)$ across i. Theorem 1(ii) implies that the estimated group membership converges to the truth when $T \gg (\log N)^2$. Theorem 1(iii) states conditions under which the proposed group-specific slope estimator $\hat{\boldsymbol{\beta}}_g(\tau)$ is asymptotically equivalent to the infeasible oracle estimator with known group membership, as both N and T tend to infinity.

We next establish the Bahadur representation for the two-step oracle estimator $\boldsymbol{\beta}_g^*(\tau)$ with known group membership. We introduce some notation and a new set of assumptions. For $i = 1, \ldots, N$, let $f_{\tau i,j}(\cdot, \cdot|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i,1+j})$ be the conditional density of $(v_{i1}(\tau), v_{i,1+j}(\tau))$ given $(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i,1+j})$ for $j = 1, \ldots, T-1$ and $f_{\tau i}(\cdot)$ be the marginal density of $v_{it}(\tau)$. In addition, define $\boldsymbol{\Gamma}_{Ng}(\tau) = N^{-1}\sum_{i:g_i^0(\tau)=g}\mathrm{E}[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}']$, $J_{i0}(\tau) = f_{\tau i}(0) - \mathrm{E}[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]'\{\mathrm{E}[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}']\}^{-1}\mathrm{E}[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]$, $\boldsymbol{J}_{i1}(\tau) = \mathrm{E}[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]$, $\boldsymbol{J}_{i2}(\tau) = \mathrm{E}[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}']$,

$$\boldsymbol{x}_{it}^*(\tau) = \left\{I_p - \frac{\boldsymbol{J}_{i1}(\tau)\boldsymbol{J}_{i1}(\tau)'\boldsymbol{J}_{i2}(\tau)^{-1}}{J_{i0}(\tau)}\right\}\boldsymbol{x}_{it} - \frac{\boldsymbol{J}_{i1}(\tau)}{J_{i0}(\tau)}, \tag{9}$$

and $\boldsymbol{H}_i(\tau) = T^{-1/2}\sum_{t=1}^T \psi_\tau\{v_{it}(\tau)\}\boldsymbol{x}_{it}^*(\tau)$, where I_p denotes a $p \times p$ diagonal matrix and $\psi_\tau(u) = \tau - I(u < 0)$.

Assumption 2.

(a) For each i, $f_{\tau i}(s|\boldsymbol{x})$ is continuously differentiable with respect to s for each $\boldsymbol{x}$; $f_{\tau i}(s|\boldsymbol{x})$ and its first derivative are uniformly bounded over $(s, \boldsymbol{x})$ and i; $f_{\tau i}(0)$ is bounded from below by a positive constant uniformly over i.

(b) For all $g \in \{1, \ldots, G\}$, $\boldsymbol{\Gamma}_{Ng}(\tau)$ is nonsingular for each N and it has a nonsingular limit $\boldsymbol{\Gamma}_g(\tau)$.

(c) For all $i, j \geq 1$, $f_{\tau i,j}(v_1, v_{1+j}|\boldsymbol{x}_1, \boldsymbol{x}_{1+j})$ is uniformly bounded over $(v_1, v_{1+j}, \boldsymbol{x}_1, \boldsymbol{x}_{1+j})$.

(d) For all $g \in \{1, \ldots, G\}$, $\boldsymbol{V}_g(\tau) = \lim_{N,T\to\infty} N^{-1} \sum_{i:g_i^0(\tau)=g} \mathrm{cov}\big[\boldsymbol{H}_i(\tau)\big]$ exists and is nonsingular.

Theorem 2. *Under Model (1) and Assumptions 1 and 2, when* $(\log N)^2/T \to 0$ *and T grows at most polynomially in N, we have the expansion*

$$\begin{aligned}&\boldsymbol{\beta}_g^*(\tau) - \boldsymbol{\beta}_g^0(\tau) + o_p(\|\boldsymbol{\beta}_g^*(\tau) - \boldsymbol{\beta}_g^0(\tau)\|) \\ &= \boldsymbol{\Gamma}_{Ng}^{-1}(\tau)\Big[\frac{1}{N\sqrt{T}} \sum_{i:g_i^0(\tau)=g} \mathbf{H}_i(\tau)\Big] + O_p\{(T/\log N)^{-3/4}\};\end{aligned} \tag{10}$$

if moreover $N^2(\log N)^3/T \to 0$*, we have*

$$\sqrt{NT}\{\boldsymbol{\beta}_g^*(\tau) - \boldsymbol{\beta}_g^0(\tau)\} \xrightarrow{d} N\big(\mathbf{0}, \boldsymbol{\Gamma}_g^{-1}(\tau)\boldsymbol{V}_g(\tau)\boldsymbol{\Gamma}_g^{-1}(\tau)\big)$$

for all $g \in \{1, \ldots, G\}$.

The oracle estimator $\boldsymbol{\beta}_g^*(\tau)$ is a two-step estimator of the quantile slope coefficients based on the first-step fixed effect estimator with known membership. It differs from the one-step fixed-effect estimator in Kato et al. (2012) (with $\{\alpha_i(\tau), i = 1, \ldots, N\}$ and $\{\boldsymbol{\beta}_g(\tau), g = 1, \ldots, G\}$ estimated jointly). Denote the one-step estimator of the quantile slope coefficient by $\boldsymbol{\beta}_g^\star(\tau)$. Following Kato et al. (2012), $\boldsymbol{\beta}_g^\star(\tau)$ has the following Bahadur representation:

$$\begin{aligned}&\boldsymbol{\beta}_g^\star(\tau) - \boldsymbol{\beta}_g^0(\tau) + o_p(\|\boldsymbol{\beta}_g^\star(\tau) - \boldsymbol{\beta}_g^0(\tau)\|) \\ &= \boldsymbol{\Gamma}_{Ng}^{\star-1}(\tau)\Big[\frac{1}{NT} \sum_{i:g_i^0(\tau)=g} \sum_{t=1}^{T} \psi_\tau\{v_{it}(\tau)\}\Big\{\boldsymbol{x}_{it} - \frac{\boldsymbol{J}_{i1}(\tau)}{f_{\tau i}(0)}\Big\}\Big] + O_p\{(T/\log N)^{-3/4}\}.\end{aligned} \tag{11}$$

where $\boldsymbol{\Gamma}_{Ng}^\star(\tau) = \boldsymbol{\Gamma}_{Ng}(\tau) - N^{-1} \sum_{i:g_i^0(\tau)=g} \mathrm{E}\big[f_{\tau i}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{J}_{\mathbf{i1}}(\tau)'/f_{\tau i}(0)\big]$. The projection term $\boldsymbol{x}_{it}^*(\tau)$ in the Bahadur representation of $\boldsymbol{\beta}_g^*(\tau)$ is due to the estimation of $\alpha_i(\tau)$ by $\tilde{\alpha}_i(\tau)$ in the first step, while the influence of fixed effects on the representation of the one-step estimator $\boldsymbol{\beta}_g^\star(\tau)$ is mainly captured by subtracting the interaction term between the intercept and slopes in the matrix $\boldsymbol{\Gamma}_{Ng}^\star(\tau)$.

Combining the results in Theorem 1(iii) and Theorem 2, we obtain the asymptotic representation and distribution of $\hat{\boldsymbol{\beta}}_g(\tau)$ as follows.

Corollary 1. *Assuming the conditions of Theorem 2, then*

$$\begin{aligned}&\hat{\boldsymbol{\beta}}_g(\tau) - \boldsymbol{\beta}_g^0(\tau) + o_p(\|\hat{\boldsymbol{\beta}}_g(\tau) - \boldsymbol{\beta}_g^0(\tau)\|) \\ &= \boldsymbol{\Gamma}_{Ng}^{-1}(\tau)\Big[\frac{1}{N\sqrt{T}} \sum_{i:g_i^0(\tau)=g} \mathbf{H}_i(\tau)\Big] + O_p\{(T/\log N)^{-3/4}\} + o_p(T^{-\delta/2}),\end{aligned} \tag{12}$$

if $(\log N)^2/T \to 0$*, and* $T = o(N^\upsilon)$ *with some* $\upsilon > 0$ *for* $0 < \delta \leq 2$ *and* $0 < \upsilon < 2/(\delta - 2)$ *for* $\delta > 2$*; if moreover* $N^2(\log N)^3/T \to 0$ *and* $NT^{1-\delta} \to 0$*, we have* $\sqrt{NT}\{\hat{\boldsymbol{\beta}}_g(\tau) - \boldsymbol{\beta}_g^0(\tau)\} \xrightarrow{d} N\big(\mathbf{0}, \boldsymbol{\Gamma}_g^{-1}(\tau)\boldsymbol{V}_g(\tau)\boldsymbol{\Gamma}_g^{-1}(\tau)\big)$ *for all* $g \in \{1, \ldots, G\}$.

Remark 2. Corollary 1 suggests that the proposed group-specific quantile slope estimator is asymptotically equivalent to its infeasible counterpart with known group membership, when both N and T go to infinity with appropriate rates. For the asymptotic normality result to hold, one sufficient (but not necessary) condition is that $T \gg N^2(\log N)^3$ but T grows at most polynomially in N. The asymptotic equivalency implies that the group membership estimation will not affect the inference on the slope coefficients, so standard inference methods for quantile regression with an interaction term of the subgroup and covariates can be used. The asymptotics and inference for fixed T are more complex and deserve further investigation.

3.2. Multiple quantiles

For clustering across multiple quantiles, we assume the following model holds:

$$Q_{y_{it}}(\tau_k|\boldsymbol{x}_{it}) = \alpha_i(\tau_k) + \boldsymbol{x}_{it}'\boldsymbol{\beta}_{g_i}(\tau_k), k = 1, \ldots, K. \tag{13}$$

Any two units i and j are said to belong to the same group if $\boldsymbol{\beta}_{g_i}(\tau_k) = \boldsymbol{\beta}_{g_j}(\tau_k)$ for all $k = 1, \ldots, K$, and to different groups if the equality is violated for at least one quantile level.

Let $\boldsymbol{\gamma}^0 = \{g_1^0, \ldots, g_N^0\}$ denote the true group membership across the K quantiles. We discuss the asymptotic properties of the multiple-quantile estimator $(\hat{B}(\boldsymbol{\tau}), \hat{\boldsymbol{\gamma}})$, the minimizer of the combined quantile loss function across K quantiles defined in (6). Define $B^*(\boldsymbol{\tau}) = \{\boldsymbol{\beta}_g^*(\boldsymbol{\tau}) : g = 1, \ldots, G\}$, where $\boldsymbol{\beta}_g^*(\boldsymbol{\tau}) = (\boldsymbol{\beta}_g^*(\tau_1)', \ldots, \boldsymbol{\beta}_g^*(\tau_K)')'$ with $\boldsymbol{\beta}_g^*(\tau_k)$ being the two-step slope estimator at the quantile level τ_k obtained with known group membership. That is, $B^*(\boldsymbol{\tau})$ is the infeasible counterpart of $\hat{B}(\boldsymbol{\tau})$, and can be regarded as the oracle estimator of $B(\boldsymbol{\tau})$. Let $B^0(\boldsymbol{\tau}) = \{\boldsymbol{\beta}_g^0(\boldsymbol{\tau}) : g = 1, \ldots, G\}$ denote the truth of $B(\boldsymbol{\tau})$, and $\boldsymbol{H}_i(\boldsymbol{\tau}) = (\boldsymbol{H}_i(\tau_1)', \ldots, \boldsymbol{H}_i(\tau_K)')'$ with $\boldsymbol{H}_i(\tau_k)$ defined in Section 2.1. We make the following additional assumptions.

Assumption 3.

(a) For all $g \in \{1, \ldots, G\}$, $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N I\{g_i^0 = g\} = \pi_g > 0$.

(b) For each $i = 1, \ldots, N$, $\alpha_i^0(\boldsymbol{\tau}) = (\alpha_i(\tau_1), \ldots, \alpha_i(\tau_K))' \in \mathcal{A}(\boldsymbol{\tau})$, $B^0(\boldsymbol{\tau}) \subseteq \Theta(\boldsymbol{\tau})$ where $\mathcal{A}(\boldsymbol{\tau})$ is a compact subset of $\mathbb{R}^K$ and $\Theta(\boldsymbol{\tau})$ is a compact subset of $\mathbb{R}^{Kp}$; $\boldsymbol{\beta}_g^0(\boldsymbol{\tau}) = (\boldsymbol{\beta}_g^0(\tau_1)', \ldots, \boldsymbol{\beta}_g^0(\tau_K)')'$ are distinct across $g = 1, \ldots, G$.

(c) The limit $\boldsymbol{V}_g(\boldsymbol{\tau}) = \lim_{N,T\to\infty} N^{-1} \sum_{i:g_i^0=g} \mathrm{cov}\big[\boldsymbol{H}_i(\boldsymbol{\tau})\big]$ exists and is nonsingular.

Theorem 3. *Suppose that Model (13), Assumptions 1(c)–(d) and 3 hold, and Assumptions 1(e)–(f) and 2 hold for all τ_k, $k = 1, \ldots, K$, we have*

(i) if $(\log N)/T \to 0$, $d_H\{\hat{B}(\boldsymbol{\tau}), B^0(\boldsymbol{\tau})\} \xrightarrow{p} 0$;
(ii) if $(\log N)^2/T \to 0$, $Pr\{\sup_{i\in\{1,\ldots,N\}} |\hat{g}_i - g_i^0| > 0\} = o(1)$;
(iii) if $(\log N)^2/T \to 0$, and $T = o(N^\upsilon)$ with some $\upsilon > 0$ for $0 < \delta \le 2$ and $0 < \upsilon < 2/(\delta - 2)$ for $\delta > 2$, then $\hat{\boldsymbol{\beta}}_g(\boldsymbol{\tau}) = \boldsymbol{\beta}_g^*(\boldsymbol{\tau}) + o_p(T^{-\delta/2})$; and if moreover $N^2(\log N)^3/T \to 0$ and $NT^{1-\delta} \to 0$, we have $\sqrt{NT}\{\hat{\boldsymbol{\beta}}_g(\boldsymbol{\tau}) - \boldsymbol{\beta}_g^0(\boldsymbol{\tau})\} \xrightarrow{d} N\big(\boldsymbol{0}, \boldsymbol{\Gamma}_g^{-1}(\boldsymbol{\tau})\boldsymbol{V}_g(\boldsymbol{\tau})\boldsymbol{\Gamma}_g^{-1}(\boldsymbol{\tau})\big)$, where $\boldsymbol{\Gamma}_g^{-1}(\boldsymbol{\tau}) = Diag\big(\boldsymbol{\Gamma}_g^{-1}(\tau_k), k = 1, \ldots, K\big)$ for all $g = 1, \ldots, G$.

4. Some computational issues

4.1. Common group membership

In Models (1) and (13), we allow both quantile coefficients and the group memberships to depend on the quantile level(s). In this section, we discuss a special case where the group membership is common across quantiles. Under this assumption, technically one can use the single-quantile method at any quantile level or the multiple-quantile method to identify subgroups, and it is unclear which one is preferable. Note that in this case the multiple-quantile method can be regarded as a composite-quantile approach, as it is based on minimizing the combined quantile loss function, thus information across quantiles is pooled to identify the common group membership. In threshold quantile regression, the composite approach was shown to often give more efficient estimation of the quantile-invariant threshold parameter (Yu, 2013; Zhang et al., 2017). However, we argue that the multiple-quantile approach is not always the best option, especially when the signals differentiating subgroups vary with quantiles; see the example in Model 2 of Section 5, where the difference in the quantile slope $\beta(\tau)$ between two subgroups is stronger at the left tail and weaker around $\tau = 0.8$. Our numerical investigation suggests that a single quantile associated with stronger signal is often preferred when the interaction effect varies across quantiles, otherwise one would be better off with the multiple-quantile clustering

For practical usage, we propose a data-adaptive approach to select the "empirically optimal" quantile level for clustering. Intuitively we would expect a quantile level with stronger discriminative power to produce more stable clustering results with respect to sampling variability. This motivates us to adapt the clustering consensus method in Monti et al. (2003) to assess the stability of putative clusterings based on different quantiles and choose the most stable one. Specifically, we first obtain H perturbed subsets of data by randomly sampling b among T time points without replacement for each unit. By implementing the clustering method for each perturbed dataset, we obtain H clustering results.[1] Throughout our numerical studies, we follow Monti et al. (2003) and let $b = \lfloor 0.8T \rfloor$. The number of possible subsets is $\binom{T}{b}$, and this adds a constraint on the choice of H for small T. In practice, we can let $H = \binom{T}{b}$ for $T \le 11$, and for larger T we found that a finite H such as $H = 200$ often suffices.

Secondly, we construct a consensus matrix $\mathcal{M} = (\mathcal{M}(i,j))$, an $N \times N$ matrix that stores, for each pair of units, the proportion of times among H clustering results in which the two units are clustered together. To assess the clustering stability, we consider a quantitative consensus measurement using the group assignments obtained from the original data as a benchmark. The clustering consensus (CC) statistic is defined as the average of consensus among G groups:

$$CC = \frac{1}{G} \sum_{g=1}^{G} \frac{2}{N_g(N_g - 1)} \sum_{i,j\in\mathcal{G}_g, i<j} \mathcal{M}(i,j), \tag{14}$$

[1] We suggest sampling from $(y_{it} - \tilde{\alpha}_i(\tau), \boldsymbol{x}_{it})$ and only perform the second stage of the algorithm in Section 2.1 with the perturbed subsets, since $\tilde{\alpha}_i(\tau)$ is better estimated using the full dataset.

where $\mathcal{G}_g$ denotes the set of units assigned to group g based on the observed data, and N_g is the cardinality of $\mathcal{G}_g$. We can then calculate the CC statistic for clustering based on each quantile level τ_k, $k = 1, \ldots, K$, and the multiple-quantile method, and choose the method with the largest CC. Our study shows that this method can help identify the quantile level with more discriminant power and thus improve the accuracy of subgroup identification; see numerical evidences in Section 5.1.

4.2. Choice of G

Our proposed algorithm and the theoretical properties are based on the assumption that the number of groups, G, is prespecified. In practice, however, G is unknown and has to be determined. Like in conventional clustering, determining the number of clusters is challenging. In this article, we propose to choose G by adopting the cross validation with averaging method (CVa) in Wang (2010), which was shown to have asymptotic selection consistency when the data are properly split into subgroups. The main idea of CVa is to estimate the clustering instability using a modified cross validation scheme, and then select G as the number of clusters yielding the smallest estimated instability.

The specific procedure of CVa is as follows. First, we randomly divide N units into three parts with sizes M, M and $N - 2M$, respectively: $z_1^c = (w_1^c, \ldots, w_M^c)$, $z_2^c = (w_{M+1}^c, \ldots, w_{2M}^c)$ and $z_3^c = (w_{2M+1}^c, \ldots, w_N^c)$, where w_i^c denotes the unit assigned to the corresponding subset. Let ψ_h^c, $h = 1, 2$ denote the clustering results obtained by applying the proposed algorithm to z_h^c, $h = 1, 2$ for a specific number of groups G. The clustering instability is estimated by $\hat{s}^c(G) = \sum_{2M+1 \le i \le j \le N} I[I\{\psi_1^c(w_i^c) = \psi_1^c(w_j^c)\} + I\{\psi_2^c(w_i^c) = \psi_2^c(w_j^c)\} = 1]$, where $\psi_1^c(w_i^c)$ denotes the assignment of w_i^c based on the clustering result of ψ_1^c. Second, repeat the first step for $c = 1, \ldots, C$ and define $\hat{s}(G) = C^{-1} \sum_{c=1}^{C} \hat{s}^c(G)$ as the estimated instability measure. Then we choose the number of groups as the one that gives the smallest instability measure.

For panel data analysis, the underestimation of G would have more negative impact than overestimation. The reason is as follows. When G is underestimated, some groups with heterogeneous slopes will be merged incorrectly, giving rise to underfitted models and resulting in biased slope estimation $\hat{\boldsymbol{\beta}}_g(\tau)$. On the other hand, if G is overestimated, units from the same group with homogeneous slopes may be split into multiple groups, leading to an overfitted model. From the clustering perspective, with larger G, the false positive rate (chance of assigning units from different groups to the same cluster) will decrease, but the false negative rate (chance of assigning units from the same group to different clusters) will increase. From the parameter estimation aspect, even though the estimation efficiency may be compromised, the resulting estimator $\hat{\boldsymbol{\beta}}_g(\tau)$ is still consistent. Such phenomena are validated by our numerical studies; see Table S1 in the supplement. Therefore, in practice, we recommend to choose a larger G to ensure the validity of parameter estimation and the interpretability of clustering results.

5. Simulation study

We consider three models for generating the simulation data,

Model 1: $y_{it} = \alpha_i + x_{it}(1 + 0.8d_i) + (1 + 0.5x_{it})e_{it}$;
Model 2: $y_{it} = \alpha_i + x_{it}(1 + d_i) + [1.5 + x_{it}\{1 - d_i/\Phi^{-1}(0.8)\}]e_{it}$;
Model 3: $y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}_{g_i^0} + 0.5x_{2it}e_{it}$,

where $i = 1, \ldots, N = 100$, $t = 1, \ldots, T = 50$ and $\alpha_i \overset{iid}{\sim}$ Uniform(0,1). In Models 1–2, $x_{it} = 0.3\alpha_i + z_{it}$, $z_{it} \overset{iid}{\sim}$ Uniform(-1,1) and $d_i \overset{iid}{\sim}$ Bernoulli(0.5) is the group indicator taking value zero for $g = 1$ and one for $g = 2$. In Model 3, $\boldsymbol{x}_{it} = (x_{1it}, x_{2it})'$ with $x_{1it} = 0.3\alpha_i + z_{it}$, $z_{it} \overset{iid}{\sim} N(0, 1)$ and $x_{2it} \overset{iid}{\sim}$ Uniform(0,1), g_i^0 are uniform from $\{1, 2, 3\}$, $\boldsymbol{\beta}_1 = (0.1, 0.1)$, $\boldsymbol{\beta}_2 = (0.2, 0.2)$ and $\boldsymbol{\beta}_3 = (0.3, 0.3)$. For all models, we consider two distributions for e_{it}: $N(0, 1)$ and $t(3)$.

Models 1–2 contain two groups. Let F_e denote the CDF of e_{it}. The τth quantile slope coefficients are $\beta_1(\tau) = 1 + 0.5F_e^{-1}(\tau)$ for group $g = 1$ and $\beta_2(\tau) = 1.8 + 0.5F_e^{-1}(\tau)$ for $g = 2$ in Model 1, while $\beta_1(\tau) = 1 + F_e^{-1}(\tau)$ and $\beta_2(\tau) = 2 + \{1 - 1/\Phi^{-1}(0.8)\}F_e^{-1}(\tau)$ in Model 2. Model 3 contains three subgroups with $\boldsymbol{\beta}_1(\tau) = (0.1, 0.1 + F_e^{-1}(\tau))'$, $\boldsymbol{\beta}_2(\tau) = (0.2, 0.2 + F_e^{-1}(\tau))'$, and $\boldsymbol{\beta}_3(\tau) = (0.3, 0.3 + F_e^{-1}(\tau))'$. All three models contain heteroscedastic errors so that the covariate effects are quantile-dependent. However, the interaction effect of the covariate and group g_i is constant in Models 1 and 3 but it varies across τ in Model 2. More specifically, in Model 2, the signal differentiating the two groups decays when the quantile level approaches 0.8 from both directions.

5.1. Group identification of different clustering methods

We apply the proposed single-quantile-based clustering method at five quantile levels $\tau = 0.3, 0.4, \ldots, 0.7$ and the multiple-quantile-based (MQ) method across the five quantiles to identify the group membership. In addition, we apply the clustering consensus measurement CC to choose among the five single-quantile-based and the MQ methods, and refer the chosen clustering as the empirically optimal quantile (EOQ) clustering. The CC statistics are calculated based on $H = 200$ resamplings. For comparison, we also include the mean-regression-based method in Lin and Ng (2012),

Table 1
The average misclassification rates (MR) in percentages of different methods in Models 1–3 with $N = 100$ and $T = 50$. Values in the parentheses are standard errors.

Model	Error	Single-quantile-based					MQ	EOQ	Mean
		$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$			
1	Normal	11.1	10.3	10.1	10.3	11.1	8.4	9.3	8.2
		(3.4)	(3.2)	(3.1)	(3.1)	(3.1)	(2.9)	(3.3)	(2.7)
	$t(3)$	15.4	13.1	12.5	13.2	15.4	11.5	12.6	18.9
		(3.6)	(3.6)	(3.4)	(3.6)	(3.7)	(3.3)	(3.7)	(3.9)
2	Normal	4.6	8.1	13.9	22.3	34.2	9.6	4.7	11.8
		(2.3)	(2.9)	(3.6)	(4.4)	(4.7)	(3.2)	(2.5)	(3.4)
	$t(3)$	8.6	11.5	16.6	24.6	35.6	13.4	9.1	22.3
		(2.9)	(3.2)	(3.7)	(4.2)	(4.7)	(3.5)	(3.4)	(4.2)
3	Normal	12.9	11.7	11.4	12.0	13.2	8.7	9.8	16.5
		(4.8)	(4.0)	(4.2)	(4.7)	(5.0)	(3.4)	(4.2)	(5.1)
	$t(3)$	21.0	16.8	16.0	16.8	20.9	13.0	14.7	35.8
		(6.8)	(6.0)	(6.2)	(6.1)	(7.2)	(4.8)	(5.7)	(6.5)

MQ: The multiple-quantile-based clustering; EOQ: The empirically optimal clustering; Mean: The mean regression method in Lin and Ng (2012).

Table 2
The average clustering consensus (CC) statistics of different methods in Models 1–3 with $N = 100$ and $T = 50$. Values in the parentheses are standard errors.

Model	Error	Single-quantile-based					MQ	Mean
		$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$		
1	Normal	0.87	0.88	0.88	0.88	0.87	0.89	0.88
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	$t(3)$	0.85	0.86	0.87	0.86	0.85	0.87	0.83
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
2	Normal	0.93	0.90	0.87	0.83	0.80	0.89	0.86
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	$t(3)$	0.89	0.87	0.85	0.82	0.80	0.86	0.81
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
3	Normal	0.88	0.89	0.89	0.89	0.88	0.91	0.82
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	$t(3)$	0.84	0.86	0.86	0.86	0.84	0.88	0.76
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

MQ: The multiple-quantile-based clustering; EOQ: The empirically optimal clustering; Mean: The mean regression method in Lin and Ng (2012).

i.e., the conditional k-means clustering method, to identify groups. For all methods, we use 20 different starting values and choose the one giving the smallest objective function to mitigate the dependency of the algorithm on initial values. The estimations for the original data are then used as starting values for the resampled datasets to calculate the CC measurement. The simulation is repeated 500 times for each scenario.

For each method, the misclassification rate can be calculated as $N^{-1}\sum_{i=1}^{N} I\{\hat{g}_i \neq g_i^0\}$. Note that there are $G!$ different ways to label the identified groups. To avoid ambiguity, we use the permutation that gives the lowest misclassification rate to label $\hat{g}_i$. Table 1 summarizes the average misclassification rates of different methods in various scenarios. When the signal differentiating groups is constant across quantiles, median performs slightly better than the other four quantiles, but through combining information across quantiles the MQ method gives the best performance. In addition, the MQ is as efficient as the mean-based method for Model 1 with *i.i.d.* normal errors, but the former clearly outperforms for all the other scenarios considered. For Models 1 and 3 with uniform signals across quantiles, the CC statistics from MQ are slightly larger than those from the single quantiles (see Table 2); the CC criterion chooses the MQ method about half of the times. The clustering based on the chosen EOQ is in general more accurate than those based on single quantiles, though slightly worse than the MQ method in Models 1 and 3.

Model 2 is more complicated as the signal differentiating groups decreases in $\tau \in (0, 0.8)$. The single quantile $\tau = 0.3$ leads to the most accurate clustering, and it even outperforms the MQ method, whose performance is somewhere between the best and the worst quantile levels but is overall better than the mean-based method. The CC criterion provides an effective way to identify the empirically optimal quantile level; it chooses $\tau = 0.3$ most of the times (see Table 3), making the EOQ method the second best performer (see Table 1).

5.2. Estimation of group-specific quantile coefficients

Besides group identification, another objective of the paper is to improve the estimation efficiency of group-specific quantile coefficient estimation by pooling information across units in the same groups. Based on the estimated group

Table 3
The percentages of times that single quantiles and MQ were chosen by the clustering consensus criterion across 500 simulations in Models 1–3 with $N = 100$ and $T = 50$.

Model	Error	Single-quantile-based					MQ
		$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	
1	Normal	13.8	10.4	11	8.8	12.2	43.8
	$t(3)$	5.8	13	13.4	9.4	6	52.4
2	Normal	96.4	2.8	0.8	0	0	0
	$t(3)$	81.4	13.6	2	0	0	3
3	Normal	7.2	10.4	7.2	9.8	9.4	56
	$t(3)$	5.8	13	13.4	9.4	6	52.4

Table 4
The integrated mean squared error (IMSE) for the unit-specific and group-specific estimators of $\beta_i(0.5)$ based on the median, MQ, EOQ and mean clustering. Values in the parentheses are standard errors. All quantities in the table are multiplied by 100.

Model	Error	Unit-specific	Median	MQ	EOQ	Mean
1	Normal	10.82	6.81	5.70	6.21	5.47
		(1.59)	(1.92)	(1.81)	(2.02)	(1.65)
	$t(3)$	13.68	8.75	8.08	8.65	11.60
		(2.08)	(2.14)	(2.12)	(2.21)	(1.98)
2	Normal	23.05	15.33	10.76	5.31	12.52
		(3.31)	(3.50)	(3.39)	(2.74)	(3.27)
	$t(3)$	28.91	19.03	15.60	10.24	21.78
		(4.17)	(3.96)	(3.94)	(3.83)	(3.59)
3	Normal	1.04	0.21	0.15	0.17	0.21
		(0.15)	(0.10)	(0.08)	(0.09)	(0.08)
	$t(3)$	1.30	0.33	0.26	0.28	0.53
		(0.20)	(0.18)	(0.14)	(0.15)	(0.14)

Table 5
The average misclassification rates (MR) in percentages for Models 3–6 and different combinations of (N,T). Values in the parentheses are standard errors.

Model	p	$N = 50$			$N = 100$			$N = 200$		
		$T = 50$	$T = 75$	$T = 100$	$T = 50$	$T = 75$	$T = 100$	$T = 50$	$T = 75$	$T = 100$
3	2	14.3	5.7	2.9	13.1	5.7	2.7	12.1	5.4	2.7
		(7.9)	(3.7)	(2.6)	(5.2)	(2.5)	(1.8)	(2.8)	(1.7)	(1.2)
4	3	16.5	7.0	3.2	15.1	6.2	3.1	13.8	6.1	2.9
		(8.3)	(4.9)	(2.7)	(5.5)	(2.5)	(1.8)	(3.1)	(1.8)	(1.3)
5	5	20.9	9.0	4.0	19.1	7.7	3.6	17.0	7.5	3.6
		(9.2)	(5.4)	(3.4)	(6.8)	(3.2)	(2.1)	(3.9)	(2.1)	(1.4)
6	8	25.3	11.7	5.2	24.0	10.0	4.8	22.2	9.6	4.6
		(8.9)	(6.8)	(3.5)	(7.6)	(3.8)	(2.3)	(5.5)	(2.5)	(1.7)

membership $\hat{g}_i$, we define the integrated mean squared error (IMSE) as $(Np)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{p}\{\hat{\beta}_{\hat{g}_i,j}(\tau) - \beta_{g_i^0,j}(\tau)\}^2$, where $\beta_{i,j}$ denotes the jth element of $\boldsymbol{\beta}_i$. Table 4 summarizes the IMSE for $\boldsymbol{\beta}_{g_i}(0.5)$ based on the clustering from the single-quantile method at $\tau = 0.5$ (Median), MQ, EOQ and the mean-based method. For comparison, we also include the IMSE of the unit-specific estimator $\tilde{\boldsymbol{\beta}}_i(\tau)$ with each unit forming a separate group. Results at other quantile levels are similar and thus are omitted. The slope estimators based on the groups identified by the proposed quantile clustering methods (Median, MQ and EOQ) are clearly more efficient than the unit-specific estimators, and they have comparable or higher efficiency than the mean-based estimator in all scenarios considered.

5.3. Performance for cases with different (p, N, T)

Following one reviewer's suggestion, we assess the performance of the proposed method for cases with different combinations of p, N and T. Based on Model 3, we construct Models 4–6 by including one, three and six additional superfluous $N(0, 1)$ predictors with zero coefficients, giving rise to $p = 3, 5$ and 8, respectively. Table 5 summarizes the misclassification rates of the quantile-based method at $\tau = 0.3$ across 500 simulations for $N = 50, 100, 200$ and $T = 50, 75, 100$ in Models 3–6. Results show that the misclassification rate decreases for larger N and T, while an increase in T has more impact. Not surprisingly, the inclusion of predictors with zero or homogeneous effects will reduce the accuracy of clustering.

Table 6
The percentages of times that $\hat{G}$ is chosen by the CVa procedure among 100 simulations for Model 3 at $\tau = 0.3$. The true number of groups is $G = 3$.

$\hat{G}$	$N = 100$			$N = 200$			$N = 300$		
	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$
2	5	4	2	3	0	0	0	0	1
3	76	92	98	89	98	99	90	98	99
4	16	4	0	7	2	1	9	2	0
5	3	0	0	1	0	0	1	0	0

Table 7
The average instability measure $100 \times \hat{s}(G)$ of MQ and the mean-based method based on CVa with 500 cross validations for the economic growth data. Values in the parentheses are standard errors.

Method	$G = 2$	$G = 3$	$G = 4$	$G = 5$
MQ	15.54 (0.72)	20.48 (0.60)	17.18 (0.47)	16.36 (0.39)
Mean	18.28 (0.78)	27.07 (0.79)	21.27 (0.62)	17.04 (0.54)

5.4. Performance of CVa for choosing G

We consider Model 3 to assess the performance of CVa for choosing G at $\tau = 0.3$ with $N = 100, 200, 300$ and $T = 100, 150, 200$. For each scenario, we implement the CVa procedure for 100 times with the splitting ratio as $M = \lfloor 0.45N \rfloor$, and choose G among $\{2, 3, 4, 5\}$ that gives the smallest average clustering instability. Table 6 summarizes the percentages of the selected $\hat{G}$. Results suggest that CVa performs well for selecting the number of groups, and the performance gets better with larger N and T. The number of within-unit replicates T appears to have more impact on the selection accuracy than the number of units N.

6. Application to growth regression

Growth regression is a major tool for studying the economic growth across countries (Barro, 1991). Most studies in the economic growth literature assumed homogeneous models, which, although convenient, were found to be unrealistic in empirical studies; see for instance Barro and Sala-I-Martin (1992), Canova and Marcet (1995), Maddala and Wu (2000), and Durlauf et al. (2001). Some studies found that there may exist "convergence clubs", that is, groups of countries with similar steady states that can be characterized by the same model, and identifying such groups can assist researchers in making generalized hypothesis; see Durlauf and Johnson (1995) and Canova (2004).

In this paper, we apply the proposed method to analyze the GPD data from the Penn World Tables (PWT) v6.2 (Heston et al., 2006) to group countries and identify the group-specific parameters. Our analysis focuses on 99 countries that have complete GDP observations from 1965 to 2003. We consider the following quantile regression model:

$$Q_\tau(y_{it}|\boldsymbol{x}_{it}) = \alpha_i(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}_{g_i}(\tau), i = 1, \ldots, 99, t = 1, \ldots, 38,$$

where $y_{i,t}$ is the log per-capita GDP of the ith country at time t with $t = 0$ corresponding to the year of 1965, $\alpha_i(\tau)$ is the country-specific fixed effect, $\boldsymbol{x}_{it} = (y_{i,t-1}, t)'$, $\boldsymbol{\beta}_{g_i}(\tau) = (\beta_{g_i,1}(\tau), \beta_{g_i,2}(\tau))'$ denote the group-specific slope coefficients, and $g_i \in \{1, \ldots, G\}$ is the unknown group membership. A similar growth regression model was also considered in Lin and Ng (2012) for mean analysis, and we focus on the analysis of conditional quantiles.

We apply the proposed MQ method across five quantiles $\{0.3, 0.4, 0.5, 0.6, 0.7\}$, and the mean-based clustering method in Lin and Ng (2012). We apply the CVa procedure to choose the number of groups among $G \in \{2, 3, 4, 5\}$. Table 7 summarizes the average instability measure $\hat{s}(G)$ and the corresponding standard error based on 500 cross validations with the splitting ratio $M = \lfloor 0.45N \rfloor$. For the mean method, $\hat{s}(G)$ at $G = 5$ is significantly smaller than those at $G = \{2, 3, 4\}$, and for the MQ method, $\hat{s}(G)$ at $G = 5$ and 2 are similar and both are significantly smaller than those at $G = \{3, 4\}$. Therefore, following the guideline as suggested in Section 4.2, we choose $G = 5$ in the subsequent analysis.

We list the countries clustered to the five groups by the MQ and mean-based clustering methods in Tables S3–S4 of the supplement. Based on the MQ method, groups 1–5 include 11, 20, 28, 28 and 12 countries, respectively. The groups identified by the mean-based method are more unbalanced with group sizes of 2, 29, 48, 14 and 6.

The residual plot suggests that the data exhibits some heteroscedasticity across the lag variable; see Fig. 1 for the plot of residuals from the median regression with subgroups identified by the MQ method. For such data with heteroscedasticity, using the MQ method can help us obtain a complete picture about the evolution of the conditional distribution and separate groups with heterogeneous slopes at one or more quantile levels.

To further understand the clustering results from the MQ method, we plot in Fig. 2 the estimated group-specific slopes $\hat{\beta}_{g,1}(\tau)$ and $\hat{\beta}_{g,2}(\tau)$ for the identified five groups across nine deciles. The shaded bands represent the 95% pointwise confidence bands, constructed by using the asymptotic normality and the estimation of the asymptotic variance in

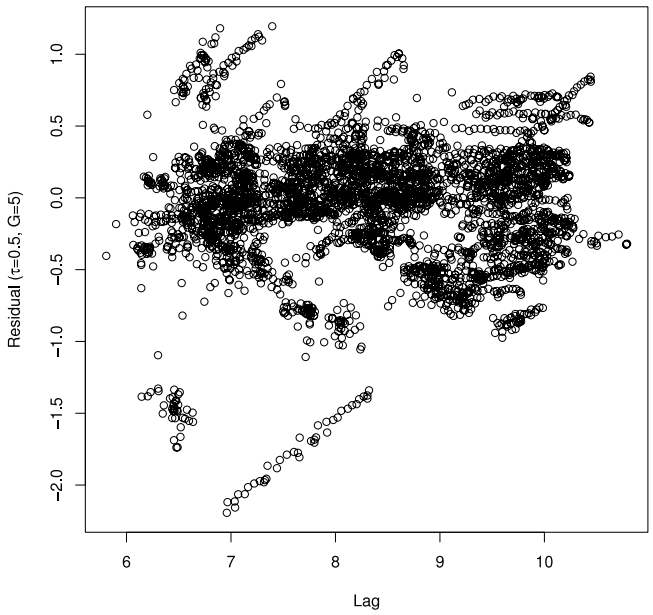
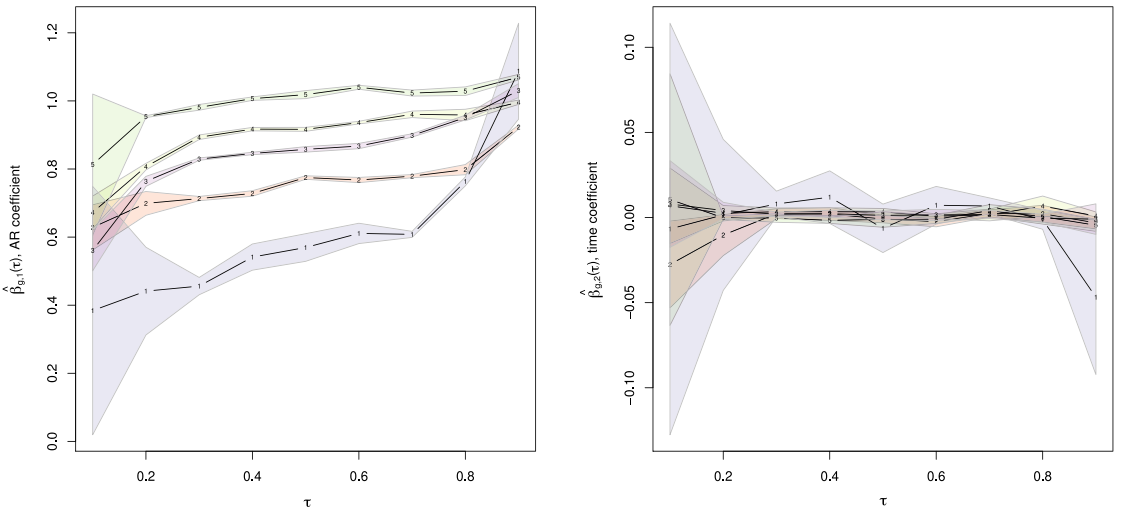


Fig. 1. Residuals from the median regression based on five subgroups identified by the MQ method against $y_{i,t-1}$ (lag).



(a) AR coefficients (b) Time coefficients

Fig. 2. The estimated group-specific quantile slopes across τ for the five groups identified by the MQ method, where $\hat{\beta}_{g,1}(\tau)$ and $\hat{\beta}_{g,2}(\tau)$ are the estimated coefficients for the lag and time variables at quantile level τ, respectively.

Corollary 1. The variance estimation is based on the asymptotic equivalency of $\hat{\boldsymbol{\beta}}_g(\tau)$ with the oracle estimator, and is likely to underestimate the variability in this empirical study of finite samples. Fig. 2 shows that the five groups differ mainly in terms of the lag coefficient. After accounting for the lag effect, the time slope appears to be homogeneous and not significant across τ. The estimated autoregressive (AR) coefficients $\hat{\beta}_{g,1}(\tau)$ show clear difference across five groups. Group five has the largest AR coefficients, and this group includes the fast growing countries/districts such as Hong Kong, Singapore, Taiwan and Thailand. Groups four and three have modest AR coefficients, and they consist of mostly high-income economies including 14 of the 20 original OECD countries. Group one has the lowest AR coefficient, and it contains mostly poor countries including Burkina Faso, Ghana, Guinea-Bissau, Mali, Nepal, and Senegal. It is interesting to note that United States and Turkey are also clustered to group one. To confirm the results, we also examined the country-specific estimates, and found that the AR coefficient estimates for these two countries are around 0.6 across quantiles, which are lower than the other groups. In addition, for all five groups, the estimated quantile AR coefficients are increasing

across τ, suggesting that the autocorrelation of GDP is stronger at the right tail of the per-capita output distribution. In contrast, for the five groups identified by the mean-based method, the estimated mean AR coefficients (see Table S5 in the supplement) show little difference among groups 2–5, making it difficult to explain the clustering results.

7. Discussion

To account for the unknown fixed effects $\alpha_i(\tau)$, we consider a two-step estimator of the slopes based on a preliminary estimator $\tilde{\alpha}_i(\tau)$ obtained by separate regression on each unit. Alternatively, we can also iteratively update the group membership and all parameters $\{\alpha_i(\tau), \boldsymbol{\beta}_g(\tau), i = 1, \ldots, N, g = 1, \ldots, G\}$ jointly by using the fixed-effect quantile regression estimation method as in Kato et al. (2012). However, this one-step approach is computationally more intensive as each iteration would require estimating a large number of fixed effects, which is challenging when the group membership is unknown. In contrast, the two-step procedure treats the estimation of fixed effects and group-specific slopes separately, and thus leads to a simpler and numerically more stable algorithm. In addition, in the second step, we can update the group-specific slope $\boldsymbol{\beta}_g(\tau)$ by adopting the elegant minimum distance (MD) quantile regression estimator in Galvao and Wang (2015), which is essentially a weighted average of unit-specific quantile slope estimators $\tilde{\boldsymbol{\beta}}_i(\tau)$ for units in the same group. However, this method requires estimating the covariance matrix of $\tilde{\boldsymbol{\beta}}_i(\tau)$ for each i, which involves the unknown conditional density function that is difficult to estimate well especially for small T. Our investigation shows that the algorithms based on the MD and the proposed two-step estimators are both computationally more efficient than the one-step approach, but the MD method tends to give higher misclassification rate than the proposed two-step algorithm in finite samples. Numerical results for the alternative procedures can be found in Section S2.2 of the supplement.

On the other hand, one limitation of the two-step approach is that the estimation of these (nuisance) fixed effects would increase the variability of the slope estimator. For panel models with homogeneous effects, Koenker (2004) and Lamarche (2010) considered penalized quantile regression estimators by shrinking the fixed effects towards constant. We can adopt this shrinkage idea and iteratively update the estimation of the penalized fixed effects, and that of the group membership and slope parameters. This approach may reduce the additional variability in the group-specific slope estimator, and we defer it to future study.

Acknowledgments

The authors would like to thank two anonymous reviewers and the editor for constructive comments and helpful suggestions. This research is supported by National Science Foundation grant DMS-1712760, the OSR-2015-CRG4-2582 grant from KAUST, the National Natural Science Foundation of China grants 11671096, 11690013 and 11731011, a fellowship from China Scholarship Council, the Key Laboratory for Applied Statistics of MOE, Northeast Normal University 130028849, and the IR/D program from the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Technical proofs and additional numerical results

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.04.005.

References

Ando, T., Bai, J., 2016. Panel data models with grouped factor structure under unknown group membership. J. Appl. Econometrics 31 (1), 163–191.
Ando, T., Bai, J., 2017. Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. J. Amer. Statist. Assoc.
Angrist, J., Chernozhukov, V., Fernández-Val, I., 2006. Quantile regression under misspecification, with an application to the US wage structure. Econometrica 74 (2), 539–563.
Barro, R.J., 1991. Economic growth in a cross section of countries. Q. J. Econ. 106 (2), 407–443.
Barro, R., Sala-I-Martin, X., 1992. Convergence. J. Political Econ. 100 (2), 223–251.
Bondell, H.D., Reich, B.J., 2008. Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. Biometrics 64 (1), 115–123.
Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. Econometrica 83 (3), 1147–1184.
Bottou, L., Bengio, Y., 1995. Convergence properties of the k-means algorithms. In: Advances in Neural Information Processing Systems. pp. 585–592.
Browning, M., Carro, J., 2007. Heterogeneity and microeconometrics modeling. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress. In: Econometric Society Monographs, Cambridge: Cambridge University Press, pp. 47–74.
Burnside, C., 1996. Production function regressions, returns to scale, and externalities. J. Monetary Econ. 37 (2), 177–201.
Canay, I.A., 2011. A simple approach to quantile regression for panel data. Econom. J. 14 (3), 368–386.
Canova, F., 2004. Testing for convergence clubs in income per capita: a predictive density approach. Internat. Econom. Rev. 45 (1), 49–77.
Canova, F., Marcet, A., 1995. The Poor Stay Poor: Non-Convergence Across Countries and Regions. Economics Working Papers 137, Department of Economics and Business, Universitat Pompeu Fabra.
Chamberlain, G., 1994. Quantile regression, censoring, and the structure of wages. In: Sims, C. (Ed.), Advances in Econometrics: Sixth World Congress. In: Econometric Society Monographs, Cambridge: Cambridge University Press, pp. 171–210,

Chetverikov, D., Larsen, B., Palmer, C., 2016. IV quantile regression for group-level treatments, with an application to the distributional effects of trade. Econometrica 84 (2), 809–833.
Durlauf, S.N., Johnson, P.A., 1995. Multiple regimes and cross-country growth behaviour. J. Appl. Econometrics 10 (4), 365–384.
Durlauf, S.N., Kourtellos, A., Minkin, A., 2001. The local solow growth model. Eur. Econ. Rev. 45 (4), 928–940.
Galvao, A.F., 2011. Quantile regression for dynamic panel data with fixed effects. J. Econometrics 164 (1), 142–157.
Galvao, A.F., Juhl, T., Montes-Rojas, G., Olmo, J., 2017. Testing slope homogeneity in quantile regression panel data with an application to the cross-section of stock returns. J. Financ. Econ. nbx016.
Galvao, A.F., Kato, K., 2017. Quantile regression methods for longitudinal data. In: Koenker, R., Chernozhukov, V., He, X., Peng, L. (Eds.), Handbook of Quantile Regression. CRC Press, pp. 363–380.
Galvao, A.F., Montes-Rojas, G., 2010. Penalized quantile regression for dynamic panel data. J. Statist. Plann. Inference 140 (11), 3476–3497.
Galvao, A.F., Montes-Rojas, G., 2015. On bootstrap inference for quantile regression panel data: A Monte Carlo study. Econometrics 3 (3), 654–666.
Galvao, A.F., Wang, L., 2015. Efficient minimum distance estimator for quantile regression fixed effects panel data. J. Multivariate Anal. 133, 1–26.
Harding, M., Lamarche, C., 2009. A quantile regression approach for estimating panel data models using instrumental variables. Econom. Lett. 104 (3), 133–135.
Heston, A., Summers, R., Aten, B., 2006. Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices. University of Pennsylvania, Philadelphia.
Hsiao, C., Tahmiscioglu, A.K., 1997. A panel analysis of liquidity constraints and firm investment. J. Amer. Statist. Assoc. 92 (438), 455–465.
Kato, K., Galvao, A.F., Montes-Rojas, G.V., 2012. Asymptotics for panel quantile regression models with individual effects. J. Econometrics 170 (1), 76–91.
Ke, Z.T., Fan, J., Wu, Y., 2015. Homogeneity pursuit. J. Amer. Statist. Assoc. 110 (509), 175–194.
Ke, Y., Li, J., Zhang, W., 2016. Structure identification in panel data analysis. Ann. Statist. 44 (3), 1193–1233.
Koenker, R., 2004. Quantile regression for longitudinal data. J. Multivariate Anal. 91 (1), 74–89.
Lamarche, C., 2010. Robust penalized quantile regression estimation for panel data. J. Econometrics 157 (2), 396–408.
Lee, K., Pesaran, M., Smith, R., 1997. Growth and convergence in a multi-country empirical stochastic solow model. J. Appl. Econometrics 12 (4), 357–392.
Li, R., Li, X.-Q., Wang, G., 2015. Global optimal and minimal solutions to K-means cluster analysis. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p. 274.
Lin, C.-C., Ng, S., 2012. Estimation of panel data models with parameter heterogeneity when group membership is unknown. J. Econ. Methods 1 (1), 42–55.
Ma, S., Huang, J., 2017. A concave pairwise fusion approach to subgroup analysis. J. Amer. Statist. Assoc. 112 (517), 410–423.
Maddala, G., Wu, S., 2000. Cross-country growth regressions: problems of heterogeneity, stability and interpretation. Appl. Econ. 32 (5), 635–642.
Monti, S., Tamayo, P., Mesirov, J., Golub, T., 2003. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach. Learn. 52 (1–2), 91–118.
Oka, T., Qu, Z., 2011. Estimating structural changes in regression quantiles. J. Econometrics 162 (2), 248–267.
Pollard, D., 1981. Strong consistency of *k*-means clustering. Ann. Statist. 9 (1), 135–140.
Qu, Z., 2008. Testing for structural change in regression quantiles. J. Econometrics 146 (1), 170–184.
Rosen, A.M., 2012. Set identification via quantile restrictions in short panels. J. Econometrics 166 (1), 127–137.
Rubin, D.B., Wu, Y.N., 1997. Modeling schizophrenic behavior using general mixture components. Biometrics 53 (1), 243–261.
Shen, X., Huang, H.-C., 2010. Grouping pursuit through a regularization solution surface. J. Amer. Statist. Assoc. 105 (490), 727–739.
Su, L., Shi, Z., Phillips, P.C., 2016. Identifying latent structures in panel data. Econometrica 84 (6), 2215–2264.
Sun, Z., Rosen, O., Sampson, A.R., 2007. Multivariate Bernoulli mixture models with application to postmortem tissue studies in schizophrenia. Biometrics 63 (3), 901–909.
Tang, C., Monteleoni, C., 2016. Convergence rate of stochastic k-means. arXiv preprint arXiv:1610.04900.
Tang, X., Qu, A., 2016. Mixture modeling for longitudinal data. J. Comput. Graph. Statist. 25 (4), 1117–1137.
Wang, J., 2010. Consistent selection of the number of clusters via crossvalidation. Biometrika 97 (4), 893–904.
Yu, P., 2013. Integrated Quantile Threshold Regression and Distributional Threshold Effects. Tech. Rep., mimeo, Department of Economics, University of Auckland.
Zhang, L., Wang, H.J., Zhu, Z., 2017. Composite change point estimation for bent line quantile regression. Ann. Inst. Statist. Math. 69 (1), 145–168.
Zhu, Y., Shen, X., Pan, W., 2013. Simultaneous grouping pursuit and feature selection over an undirected graph. J. Amer. Statist. Assoc. 108 (502), 713–725.