Panel Quantile Regression with Time-Invariant Rank

Xin Liu*

November 3, 2020

Click here for the most recent version

Abstract

I propose a quantile-based random coefficient panel data framework to study heterogeneous causal effects. The heterogeneity depends on unobservables, as opposed to heterogeneity for which we can add interaction terms. This connects it to other structural quantile models. My approach uses panel data to address "endogeneity," meaning dependence between the explanatory variables and the random coefficients. The random coefficient vector depends on an unobserved, scalar, time-invariant "rank" variable, in which outcomes are monotonic at a particular point. I develop the theory first in a simplified model and then extend results to a more general model. First, I establish identification and uniformly consistent estimation. Second, I use a Dirichlet approach to establish small-n confidence sets or uniform confidence bands of the coefficient function. Third, I establish asymptotic normality of the coefficient estimator in the simplified model by applying the functional delta method to the empirical process. This facilitates a bootstrap confidence interval for the coefficient estimator at each specific rank. Finally, I illustrate the proposed methods by examining the causal effect of a country's oil wealth on its political violence and military defense spending.

JEL classification: C23, C21

Keywords: panel data, structural model, unobserved heterogeneity, heterogeneous causal effect, endogeneity, random coefficient, rank variable.

1 Introduction

Panel data models are popular in applied research because they can address endogeneity in structural models without using instruments. They achieve identification of causal effects by controlling for unobserved time-invariant heterogeneity, since we observe multiple time periods for each individual. The term "individual" may broadly refer to a person, country, firm, school, etc.

*Email: xliu@mail.missouri.edu. Mail: Department of Economics, University of Missouri, 909 University Ave, 118 Professional Bldg, Columbia, MO 65211-6040, United States.

Much research is interested in not only the causal effect of variables on the mean outcome, but also the effect on distributional features, such as (conditional) quantiles. Quantile regression is appealing since it can account for unobserved heterogeneity and capture heterogeneous effects.

Recently, there have been a growing number of studies on panel quantile regression. Such studies are challenging because the conditional quantile operator is nonlinear. In the standard fixed effects model, the unobserved individual effect U usually correlates with the regressor vector $\mathbf{X}$, so pooled OLS results in biased estimates of the structural parameter vector $\boldsymbol{\beta}$. This $\boldsymbol{\beta}$ is identified by first-differencing or demeaning to get rid of the unobserved individual effect. However, we cannot first-difference or demean in quantile models due to the nonlinearity (i.e., the difference of quantiles is not the quantile of the difference).

I propose a structural random coefficient model that addresses endogeneity with panel data and captures heterogeneous causal effects as in quantile models. Quantile models can be interpreted as random coefficient models in a cross-sectional setting (Koenker, 2005, Section 2.6). The linear quantile regression model

$$\mathrm{Q}_\tau(Y \mid \mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}(\tau) \tag{1}$$

is generated by the structural random coefficient model

$$Y = \mathbf{X}'\boldsymbol{\beta}(U) \tag{2}$$

with $U \sim \text{Unif}(0,1)$ given the exogeneity assumption $U \perp\!\!\!\perp \mathbf{X}$ and the monotonicity condition that $\mathbf{X}'\boldsymbol{\beta}(u)$ is increasing in u. With panel data, I can relax the independence restriction $U \perp\!\!\!\perp \mathbf{X}$ to allow unrestricted dependence between U and $\mathbf{X}$.

To represent the unobserved heterogeneity, I use the "rank variable" idea from Chernozhukov and Hansen's (2005) instrumental variable quantile regression model. The rank variable determines the individual's ranking (quantile) in the outcome distribution if hypothetically every individual were made to have the same explanatory variable values, like U in (2). The standard fixed effects panel models use an additively separable individual characteristic to represent the unobserved time-invariant heterogeneity. Instead, I allow the unobserved time-invariant individual characteristic to enter nonseparably as the rank variable, similar to (2):

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta}(U_i) + V_{it}, \tag{3}$$

where the additional idiosyncratic error V_{it} is assumed exogenous. Different from many studies that assume the individual rank variable is iid across time, the time-invariant rank variable in this paper makes more economic intuition in many cases.

My structural model includes the standard fixed effects (FE) model as a special case. If only the intercept (not slope coefficients) depends on the rank variable U_i, then $\boldsymbol{\beta}(U_i)' = (\beta_0(U_i), \beta_1, \beta_2, \ldots) \equiv (\alpha_i, \boldsymbol{\beta}'_{-1})$, and writing $\mathbf{X}' = (1, \mathbf{X}_{-1})$, then (3) becomes the traditional linear structural FE model

$$Y_{it} = \mathbf{X}'_{-1,it}\boldsymbol{\beta}_{-1} + \alpha_i + V_{it}. \tag{4}$$

In (4), the structural effect of $\mathbf{X}$ on Y is homogeneous across individuals. In contrast, my model in (3) allows a heterogeneous (individual-specific) structural effect of $\mathbf{X}$ on Y, through the dependence of $\boldsymbol{\beta}(U_i)$ on U_i.

My model can be applied to a wide range of economic settings. For example, consider estimating a Cobb–Douglas production function given a panel dataset of farm i over time periods t (seasons or years). The productivity is heterogeneous across farms, due to soil quality, labor skill, machine productivity, etc. If we consider the general productivity of each farm as its latent rank variable, we could write the generalized Cobb–Douglas production function as in (3), where Y_{it} is the log output of farm i at time t; $\mathbf{X}_{it}$ includes an intercept and the log of the inputs, such as labor and capital; U_i denotes farm i's general productivity; and the idiosyncratic disturbance V_{it} denotes some external shock outside the farm's control, such as rainfall. This model is more general than the traditional Cobb–Douglas production function analysis, which assumes the productivity of labor and capital are identical across farms.

I make four primary contributions. First, I contribute a new identification strategy for heterogeneous causal effects in both a simplified model and a long panel version of the model in (3). The structural coefficients measure the causal effects of explanatory variables vector $\mathbf{X}$ on the outcome Y for individuals in the population with rank variable value $U = \tau$. Identification is achieved by assuming endogeneity is due only to time-invariant (not idiosyncratic) unobservables, along with a weaker version of the monotonicity assumption of Chernozhukov and Hansen (2005).

Second, I establish uniform consistency of the coefficient estimators over all individual rank values under a certain rate relation between n and T (the cross-sectional and time series dimensions of the data). I first estimate the individual-specific coefficients, relying on established time series OLS consistency results. Then I implement a sorting approach on a hypothetical outcome at certain $\mathbf{x}^*$ with the estimated individual coefficients. Under a certain rate condition between n and T, the order of the rank variables can be recovered through the order of the hypothetical outcomes. Therefore, the estimator of coefficient at any specific rank value is determined and consistently estimated.

Third, I contribute a novel inference approach to build a uniform confidence band for the coefficient function using a Dirichlet method, based on the fact that the finite-sample uniform order statistics jointly follow an ordered Dirichlet distribution. If each individual's coefficient were known, then the unobserved individual rank variable would be the only source of uncertainty. More generally, there is also estimation error. In the general model, the individual coefficient has estimation error, which essentially comes from the idiosyncratic error. The uncertainty comes from both the rank variable and the individual coefficient estimation. For the coefficient as a function of rank, I construct confidence sets both pointwise and joint over n rank values, which I extend to a uniform confidence band if the coefficient is an increasing function of rank.

Finally, I propose a bootstrap inference method for the coefficient estimator at any rank value. I establish the asymptotic normality of the coefficient function estimator uniformly over all rank values, by applying the functional delta-method to the empirical quantile process. I establish bootstrap consistency results in a simplified model to provide the intuition and theoretic grounds for using the bootstrap inference method in the general model. Based on the uniform asymptotic normality and the bootstrap consistency results in the simplified model, I use a cluster-robust resampling method to build a bootstrap standard error and bootstrap confidence interval for the causal parameters at any specific rank value.

I also apply my new methodology in an empirical application examining the causal effect of oil wealth on a country's military spending. Many researchers (e.g., Besley and Persson, 2011; Bohn and Deacon, 2000; Collier and Hoeffler, 1998, 2004; Fearon and Laitin, 2003) have investigated the macro-level "oil-fuels-war" hypothesis. Various methods are proposed, and they arrive at different conclusions. I use my new methods to estimate the causal effect of oil wealth for countries at any specific rank, interpreting rank as a general military spending tendency. I build confidence sets using the Dirichlet method and construct a bootstrap confidence interval for the causal effect of oil wealth. Although there is some heterogeneity, I arrive at the same conclusion that oil wealth has no significant effect on military spending, as found by Cotet and Tsui (2013) assuming the effect is homogeneous.

These contributions raise additional research questions that I am pursuing in separate papers. One question relates to partial identification. When the data set has missing values or the number of time series observations is very small, point identification fails. I plan to characterize upper and lower bounds for the coefficient function in these situations. Another research question concerns a bias–variance tradeoff, motivated by empirical applications. If covariates lack variation for each individual, then the variance from the time series estimation can be very large. Grouping individuals with similar covariate values may be able to reduce the standard error and improve efficiency, if the increase in bias is small enough.

Paper structure and notation Section 2 reviews the literature. Section 3 presents identification results. Section 4 defines the estimator and shows uniform consistency. Section 5 derives confidence sets and uniform confidence bands for the coefficient function based on the Dirichlet approach. Section 6 derives the uniform asymptotic normality of the coefficient function estimator. It discusses bootstrap consistency and presents bootstrap confidence intervals. Section 7 provides empirical illustrations. Section 8 concludes and discusses future research.

Acronyms used include those for fixed effects (FE) and quantile regression (QR). Notationally, random and non-random vectors are respectively typeset as, e.g., $\mathbf{X}$ and $\mathbf{x}$, while random and non-random scalars are typeset as X and x (with the exception of certain constants like the time series dimension T), and random and non-random matrices as $\underline{\mathbf{X}}$ and $\underline{\mathbf{x}}$; $\mathbb{1}\{\cdot\}$ is the indicator function. The Dirichlet distribution with parameters $a_1, \ldots, a_K$ is written $\mathrm{Dir}(a_1, \ldots, a_K)$, the beta distribution $\mathrm{Beta}(a, b)$, and the uniform distribution $\mathrm{Unif}(a, b)$; in some cases these stand for random variables following such distributions. Let $\rightsquigarrow$ denote weak convergence. Let δ_x denote the Dirac measure.

2 Literature

2.1 Quantile regression

Quantile regression is first introduced by Koenker and Bassett (1978). A variety of approaches have been developed to address endogeneity in quantile models. This includes the instrumental variable quantile regression methods of Chernozhukov and Hansen (2005, 2006, 2008), the local quantile treatment effect of Abadie, Angrist, and Imbens (2002) (and see Melly and Wüthrich, 2017, for a survey), and the control function methods for triangular structural models in Chesher (2003), Lee (2007), and Imbens and Newey (2009), among others. Chernozhukov and Hansen (2005) introduce the "rank variable" idea to represent unobserved heterogeneity in quantile models. They achieve identification of structural heterogeneous effects by restricting the rank variable and imposing a monotonicity assumption on the conditional outcome as a function of the rank variable.

2.2 Panel quantile regression

A variety of approaches have been proposed to study quantile regression with panel data, mainly in two categories.

First, a series of papers study the panel fixed effects quantile regression (FE-QR), including Koenker (2004), Kato, Galvao, and Montes-Rojas (2012), Canay (2011), and Galvao and Wang (2015), among others. They assume the conditional quantile function is additive in an individual effect and a linear function of the explanatory variables. The individual effect cannot be consistently estimated in a short panel because the number of parameters is proportional to n, and the individual effect cannot be differenced out due to the nonlinearity of the conditional quantile operator. Consequently, these FE-QR studies assume a large-T (long panel) framework, i.e., $T \to \infty$. Koenker (2004) first proposes a penalized estimator in panel quantile regression. It assumes the individual effect is a location shift that is the same at all quantiles. Kato, Galvao, and Montes-Rojas (2012) generalize the individual effect to be a function of the quantile. They establish large-sample theory for the FE-QR estimator when both n and T go to infinity. Canay (2011) proposes a two-step estimator that first estimates the individual effect and then runs quantile regression after subtracting the individual effect from the outcome variable. It treats the idiosyncratic error as the rank variable that follows a standard uniform distribution, and it is assumed to be iid across both individual and time dimension. Frumento, Bottai, and Fernández-Val (2020) extend Canay's (2011) model to a two-level quantile regression with the level-one quantile regression the same as Canay's (2011) model and the level-two quantile regression modeling the scalar-valued individual fixed effect as a quantile regression of the time-invariant characteristics.

There are several extensions of these FE-QR models. Galvao (2011) studies a dynamic FE-QR model. Galvao and Kato (2016) provide a smoothed FE-QR estimator. Harding and Lamarche (2009) use an instrumental variables approach to study the panel quantile regression with endogenous regressors. Rosen (2012) derives set identification in the FE-QR model with a short panel.

Second, the other approach to study panel quantile regression is based on the correlated random coefficient model (CRC-QR). As shown by Koenker (2005, Section 2.6), a quantile regression model can be equivalently written as a random coefficient model with a monotonicity condition and an independence assumption between the explanatory variables and the unobserved heterogeneity (rank variable) in the cross-sectional setting. With the availability of panel data, the independence assumption can be relaxed. The correlated random coefficient model starts from Chamberlain (1982). Abrevaya and Dahl (2008) study the effect of birth inputs on birthweight using quantile estimation in a random coefficient short panel model. They allow the regressors to be correlated with the unobserved individual characteristics. Arellano and Bonhomme (2016) assume a random coefficient panel data model and achieve a linear interpretation of the conditional quantile function under certain conditions. They assume the coefficient is a function of the idiosyncratic error, which follows a standard uniform distribution and is iid in both individual and time dimensions. Graham, Hahn, Poirier, and Powell (2018) propose a partially varying coefficient quantile regression model. They show this model can be derived from a random coefficient model with coefficients depending on the regressors and the unobserved individual effect. They assume the unobserved individual effect conditionally follows a standard uniform distribution and is iid in both individual and time dimensions.

There are some other studies on quantile effects in nonlinear panel data models. Chernozhukov, Fernández-Val, Hahn, and Newey (2013) study the average treatment effect and quantile treatment effects in nonlinear panel models. They treat time as randomly assigned. Under this time homogeneity condition, they establish identification, partial identification, and estimation of average and quantile treatment effects with discrete-valued regressors. Powell (2016) studies panel quantile regression with nonadditive fixed effects in an instrumental variable framework. It allows dependence between the instruments and the fixed effect. It assumes the coefficients are a function of both the fixed effect and rank variable, which is assumed iid over both individuals and time.

2.3 Random coefficient panel model

The random coefficient panel model is an alternative to the fixed effect panel model. The fixed effect model considers an additive unobserved individual heteorgeneity, whereas the random coefficient model use the individual random coefficient to caputure the unobserved individual heterogeneity.

The marginal effect of a variable might be varying over time and be different for different individuals depending on many factors. If all the factors are caputured in the model, then the constant slope model might be reasonable. However, many variables are unobserved. Therefore, a varying coefficient model is more reasonable to caputure the unobserved heterogeneity.

A variety of random coefficient panel models have been discussed, depending on different assumptions imposed on the random coefficient. The classical random-coefficient panel model assumes the random coefficient is a constant plus an idiosyncratic error. The average causal effect in this model will be the same as the causal effect in the fixed-coefficient model. Arellano and Bonhomme (2012) assume the individual-specific coefficient is a random draw from an unrestricted conditional distribution given the regressors and identify the distributional characteristics (variance and conditional density) of the random coefficient. The correlated random coefficient (CRC) model assumes the random coefficient as a function of the regressors and/or the unobservables. Fernández-Val and Lee (2013) first use each individual's time series observations to estimate each individual-specific coefficient, then incorporate the individual-specific unobserved heterogeneity in a GMM framework. It includes the correlated random coefficient model with random coefficient depending on the individual-specific unobserved heterogeneity as a special case. When the random coefficient is assumed to be a function of the rank variable, it becomes the panel quantile models. The existing panel quantile studies assume the rank variable is iid over time and over individual. My paper considers all the unobserved variables summarized into one scalar variable normalized on $(0, 1)$, which is the rank variable. I assume the rank variable is iid only over individuals and is time-invariant.

3 Identification

This section presents the panel quantile models and identification results. First, I introduce a simplified model to develop intuition. Second, I give identification results for the main model.

3.1 Simplified model point identification

Consider a simple random coefficient panel data structural model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta}(U_i), \quad i = 1, \dots, n,\ t = 1, \dots, T, \tag{5}$$

where Y_{it} is the dependent variable for individual i at time t, $\mathbf{X}_{it}$ is the corresponding K-vector of explanatory variables, and U_i is the unobserved individual-specific rank variable normalized to $U_i \sim \text{Unif}(0, 1)$. This $\mathbf{X}_{it}$ usually includes a constant, so the first component of $\boldsymbol{\beta}(U_i)$ is the individual-specific intercept, whereas the standard FE model writes the individual-specific intercept separately. The dependence between U_i and $\mathbf{X}_{it}$ is unrestricted. The function $\boldsymbol{\beta}(\cdot)\colon \mathbb{R} \mapsto \mathbb{R}^K$ is deterministic. Different individuals have different coefficient vectors $\boldsymbol{\beta}(U_i)$ depending on their U_i. For example, $\boldsymbol{\beta}(0.5)$ is the structural coefficient vector for an individual with median level $U_i = 0.5$. The $\boldsymbol{\beta}(U_i)$ is a random coefficient that can vary across individuals, but the source of the randomness is restricted to be a scalar U_i.

The K-vector coefficient $\boldsymbol{\beta}(U_i)$ as a function of the unobserved individual component captures the heterogeneity in response.

The number of time periods T is assumed to be fixed and greater than or equal to the number of random coefficients K. When $T = K$, we are able to identify all individual coefficients $\boldsymbol{\beta}(U_i)$ perfectly under a rank condition on $\mathbf{X}_{it}$. When $T > K$, having more time periods observations will add no additional information to identification because all the extra time periods' observations fit perfectly on the same line as identified with K time periods' observations.

Assumption A1 (iid sampling)**.** Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT})'$ and $\underline{\mathbf{X}}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})'$. The observables $(\mathbf{Y}_i, \underline{\mathbf{X}}_i)$ are assumed to be independent and identical distributed (iid) across $i = 1, \ldots, n$.

Assumption A2 (Time-invariant rank variable)**.** (i) The unobserved time-invariant individual rank variable U_i is iid across i. It follows a standard uniform distribution, $U_i \overset{iid}{\sim} \text{Unif}(0, 1)$. (ii) The relationship between $\underline{\mathbf{X}}_i$ and U_i is unrestricted.

Assumption A3 (Monotonicity)**.** There exists known K-vector $\mathbf{x}^*$ in the support of $\mathbf{X}_{it}$ such that the map $u \mapsto \mathbf{x}^{*\prime}\boldsymbol{\beta}(u)$ is strictly increasing on $[0, 1]$.

Assumption A4 (Continuity)**.** At $\mathbf{X} = \mathbf{x}^*$, the function $u \mapsto \mathbf{x}^{*\prime}\boldsymbol{\beta}(u)$ is continuous in u on (0,1).

Assumption A5 (Rank condition)**.** For each individual i, the observed $T \times K$ explanatory variable matrix $\underline{\mathbf{X}}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})'$ has (full column) rank K with probability one.

A2 assumes the rank variable is time-invariant and is iid only across i. A2 is different from most existing studies that assume the individual rank is iid both across i and across t. A3 is similar to the monotonicity assumption in Chernozhukov and Hansen (2005). A3 is a weaker version that assumes monotonicity only at a single $\mathbf{x}$ value. A4 guarantees a unique solution for identification. A5 ensures that for each individual, we are able to identify their individual-specific coefficient $\boldsymbol{\beta}_i = \boldsymbol{\beta}(U_i)$ using their T observations. A5 requires the K variables are linearly independent. If $\mathbf{X}$ includes an intercept term, then it cannot include any other time-invariant variable. Moreover, A5 implies there is no "stayer" for more than $(T - K)$ periods in the data set.

Under A1–A5, we can identify all the individual coefficients from their time series dimension variation. We can also identify the coefficient at any rank value.

Theorem 1 (Identification in simplified model)**.** *Suppose A1–A5 hold.*

(i) For each representative individual i, its individual coefficient $\boldsymbol{\beta}(U_i)$ is uniquely determined by its observations $(\mathbf{Y}_i, \underline{\mathbf{X}}_i)$, with probability one. That is, the random matrix $(\mathbf{Y}_{T\times 1}, \underline{\mathbf{X}}_{T\times K})$ uniquely determines the random vector $\boldsymbol{\beta}(U)$.

(ii) For any $\tau \in (0, 1)$, there exists a $\boldsymbol{\beta}(\tau)$ in the support of $\boldsymbol{\beta}(U)$ that uniquely satisfies

$$\mathrm{P}(\mathbf{x}^{*\prime}\boldsymbol{\beta}(U) \leq \mathbf{x}^{*\prime}\boldsymbol{\beta}(\tau)) = \tau.$$

Generally, $\boldsymbol{\beta}(\tau)$ represents the marginal effect of $\mathbf{X}$ on Y for the individual whose rank variable value is $U = \tau$. It captures the heterogeneous effect across individuals.

Additionally, given $\mathbf{X} = \mathbf{x}^*$, $\boldsymbol{\beta}(\tau)$ can be interpreted as the marginal effect of $\mathbf{X}$ on the τ-quantile structural function introduced by Imbens and Newey (2009). If A3 is strengthed to hold at all $\mathbf{x}$ in the support of $\mathbf{X}$, as in Chernozhukov and Hansen (2005), then $\mathbf{x}'\boldsymbol{\beta}(U)$ is the τ-quantile structural function, the same as identified by Chernozhukov and Hansen (2005).

3.2 Large-T point identification

I extend the simplified model to a general large-T model by adding idiosyncratic disturbances.

Consider the structural model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta}(U_i) + V_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \tag{6}$$

where Y_{it}, $\mathbf{X}_{it}$, and U_i are the same as before, and V_{it} is the idiosyncratic disturbance. Again, this model is a random coefficient model that can capture heterogeneous effects across individuals with endogeneity and unobserved heterogeneity.

Assumption A6 (Idiosyncratic disturbance)**.** The idiosyncratic disturbance V_{it} satisfies

$$\mathbb{E}(\mathbf{X}_{it}V_{it}) = \mathbf{0}, \text{ for all } i = 1, \ldots, n \text{ and } t = 1, \ldots, T, \tag{7}$$

and each individual-specific coefficient $\boldsymbol{\beta}(U_i)$ is identified.

For example, A6 is satisfied in the special case of covariance stationarity with invertible matrix $\mathrm{E}[\mathbf{X}_{it}\mathbf{X}_{it}']$.

The general model includes the same assumptions about the rank variable, monotonicity, and continuity as in the simplified model. Assumption A6 replaces A5 for identification of individual coefficients, after which the argument is the same for identification of the coefficients at different ranks τ.

Theorem 2 (Identification in general model)**.** *In the general model, under A1–A4 and A6, the coefficients are identified at any rank value. That is, for any $\tau \in (0, 1)$, there exists $\boldsymbol{\beta}(\tau)$ in the support of $\boldsymbol{\beta}(U)$ uniquely satisfying*

$$\mathrm{P}\big(\mathbf{x}^{*\prime}\boldsymbol{\beta}(U) \le \mathbf{x}^{*\prime}\boldsymbol{\beta}(\tau)\big) = \tau.$$

Again, the parameter $\boldsymbol{\beta}(\tau)$ is interpreted as the marginal effect of $\mathbf{X}$ on outcome variable Y for an individual in the population with rank variable value $U = \tau$.

3.3 Relation to other models

3.3.1 Nonseparable idiosyncratic error

The model in (6) has an additive idiosyncratic error, outside the coefficient function. Alternatively, we can assume an even more general model with nonseparable idiosyncratic error,

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta}(U_i, V_{it}), \quad i = 1, \ldots, n, \; t = 1, \ldots, T. \tag{8}$$

There are two ways to connect the nonseparable and additive idiosyncratic error models, (8) and (6) respectively. One is to consider model (6) as a special case of model (8), in which the idiosyncratic error only enters the intercept, not the slope coefficients, and it does so additively. That is, writing β_j for the jth component of vector $\boldsymbol{\beta}$, let $\beta_1(U_i, V_{it}) = \gamma_1(U_i) + V_{it}$, $\beta_k(U_i, V_{it}) = \gamma_k(U_i)$ for $k = 2, \ldots, K$, and $\mathbf{X}_{it} = (1, \mathbf{X}_{-1,it}')'$, so

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta}(U_i, V_{it}) = \gamma_1(U_i) + V_{it} + \sum_{k=2}^{K} X_{kit}\gamma_k(U_i) = \mathbf{X}_{it}'\boldsymbol{\gamma}(U_i) + V_{it}.$$

The other way to connect the two models is to assume an additive idiosyncratic error inside the coefficient function, $\boldsymbol{\beta}(U_i, V_{it}) = \boldsymbol{\beta}(U_i + V_{it})$, and make a Taylor expansion of $\boldsymbol{\beta}(U_i + V_{it})$ at $\boldsymbol{\beta}(U_i)$. If $\boldsymbol{\beta}(\cdot)$ is differentiable and V_{it} is relatively small, then $\boldsymbol{\beta}(U_i + V_{it}) \approx \boldsymbol{\beta}(U_i) + \boldsymbol{\beta}'(U_i)V_{it}$. This is an exact equality in the special case of linear $\boldsymbol{\beta}(\cdot)$. Thus,

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta}(U_i + V_{it}) \approx \mathbf{X}_{it}'\boldsymbol{\beta}(U_i) + \mathbf{X}_{it}'[\boldsymbol{\beta}'(U_i)V_{it}].$$

The second term on the right-hand side satisfies the linear projection property when V_{it} is mean-zero and mean independent of $\mathbf{X}_{it}$ and U_i,

$$\mathrm{E}\{\mathbf{X}_{it}\mathbf{X}_{it}'[\boldsymbol{\beta}'(U_i)V_{it}]\} = \mathrm{E}\{\mathbf{X}_{it}\mathbf{X}_{it}'[\boldsymbol{\beta}'(U_i)]\overbrace{\mathrm{E}[V_{it} \mid \mathbf{X}_{it}, U_i]}^{=0}\} = \mathbf{0}.$$

Thus, model (8) is approximately the same as model (6) with redefined idiosyncratic error.

3.3.2 Standard FE model

Moreover, the standard FE panel structural model can be considered as a special case of the general model studied in this paper. If we restrict the slope to be constant across individuals, let $\mathbf{X}' = (1, \mathbf{X}_{-1})$ include an intercept term, and let $\alpha_i \equiv \beta_1(U_i)$ and $\beta_k(U_i) = \beta_k$ for $k = 2, \ldots, K$ with $\boldsymbol{\beta} \equiv (\beta_2, \ldots, \beta_K)'$, then the general model (6) becomes

$$Y_{it} = \mathbf{X}_{-1,it}'\boldsymbol{\beta} + \alpha_i + V_{it}.$$

4 Estimation

4.1 Simplified model point estimation

First, I describe the coefficient estimator at any rank value in the simplified model. Then, I show uniform consistency of this estimator. These results develop intuition and intermediate theoretical results that are useful later for the full model.

4.1.1 Order statistics of rank variable

Consider order statistics of the individual rank variable U_i. Let $U_{n:k}$ denote the kth order statistic in a sample of size n: $U_{n:k}$ is the kth-smallest value among $\{U_1, \ldots, U_n\}$. Assumption A2 assumes $U_i \overset{iid}{\sim} \text{Unif}(0,1)$, which results in the following properties from Wilks (1962, pp. 236–238):

$$U_{n:k} \sim \text{Beta}(k, n+1-k), \tag{9}$$

$$(U_{n:1}, U_{n:2} - U_{n:1}, \ldots, U_{n:n} - U_{n:n-1}, 1 - U_{n:n}) \sim \text{Dir}(\overbrace{1, \ldots, 1}^{n+1}), \tag{10}$$

a Dirichlet distribution with all $n+1$ parameters equal to 1.

Although the individual rank variable U_i is unobserved, we know the joint distribution of its order statistics from (10), which proves useful for inference. Assumption A3 enables us to uncover the order of the unobserved individual ranks through the order of observable values. Specifically, at $\mathbf{X} = \mathbf{x}^*$, the ordering of the hypothetical outcome variables $Y_i^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_i)$ represents the ordering of the unobserved individual rank variables U_i:

$$Y_{n:1}^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_{n:1}) < Y_{n:2}^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_{n:2}) < \cdots < Y_{n:n}^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_{n:n}). \tag{11}$$

4.1.2 Estimation

In the simplified model, we can perfectly estimate the coefficient parameter $\boldsymbol{\beta}(U_i)$ from data, although the value U_i remains unknown. Without loss of generality, assume $T = K$. Under the full rank condition A5,

$$\boldsymbol{\beta}(U_i) = \underline{\mathbf{X}}_i^{-1}\mathbf{Y}_i, \quad i = 1, \ldots, n. \tag{12}$$

Given the n individual coefficients $\boldsymbol{\beta}(U_i)$ for $i = 1, \ldots, n$, for any $\tau \in (0,1)$, my coefficient estimator at rank $U = \tau$ is

$$\hat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big), \tag{13}$$

where $U_{n:\lceil \tau n \rceil}$ is the $\lceil \tau n \rceil$th order statistic, and $\lceil \tau n \rceil$ denotes the ceiling function of τn, i.e., the least integer greater than or equal to τn. Because $0 < \tau < 1$, $1 \le \lceil \tau n \rceil \le n$. Although the U_i values are unobserved, their order can be discovered through the order of hypothetical outcome values Y_i^* at $\mathbf{X} = \mathbf{x}^*$ under A3, which determines $\boldsymbol{\beta}(U_{n:k})$.

Define the permutation $\sigma(\cdot)$ to associate the individual "i" with its order "k":

$$\sigma(k) \equiv \{i : U_{n:k} = U_i\}. \tag{14}$$

So we can write the coefficient $\boldsymbol{\beta}(U_{n:k}) = \boldsymbol{\beta}\big(U_{\sigma(k)}\big)$, and the estimator is $\hat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}\big(U_{\sigma(\lceil \tau n \rceil)}\big)$. The permutation $\sigma(\cdot)$ can be learned from the Y_i^* ordering because $Y_{n:k}^* = Y_{\sigma(k)}^*$.

Specifically, the estimation of $\boldsymbol{\beta}(\tau)$ consists of four steps.

Step 1: For each individual, use its T time period observations to (perfectly) estimate its individual coefficient: $\boldsymbol{\beta}(U_i) = \underline{\mathbf{X}}_i^{-1}\mathbf{Y}_i,\ i = 1, \ldots, n$.

Step 2: Given all perfectly estimated individual coefficients $\boldsymbol{\beta}(U_1), \boldsymbol{\beta}(U_2), \ldots, \boldsymbol{\beta}(U_n)$ from Step 1, compute the n corresponding outcome values Y^* at $\mathbf{X} = \mathbf{x}^*$,

$$Y_i^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_i),\ i = 1, \ldots, n. \tag{15}$$

Step 3: Sort the outcome values $\{Y_i^*\}_{i=1}^n$ in increasing order. The order of the Y^* represents the order of the U: the permutation $\sigma(\cdot)$ from (14) satisfies $Y_{n:k}^* = Y_{\sigma(k)}^*$ for $k = 1, \ldots, n$.

Step 4: Using the $\boldsymbol{\beta}(U_i)$ from Step 1 and the permutation $\sigma(\cdot)$ from Step 3, the coefficient estimator at rank τ is

$$\hat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}\big(U_{\sigma(\lceil n\tau \rceil)}\big). \tag{16}$$

4.1.3 Large-n uniform consistency

This subsection shows the estimator in the simplified model is uniformly consistent over $0 < \tau < 1$.

Assumption A7 (Uniform continuity)**.** Each element of the coefficient $\boldsymbol{\beta}(\cdot)$ is a uniformly continuous function on $(0, 1)$. That is, writing $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_K(\cdot))'$, for any $k = 1, \ldots, K$,

$$\forall \epsilon > 0, \exists\, \delta > 0 \text{ such that } \forall x, y \in (0,1),\ |x - y| \le \delta \implies |\beta_k(x) - \beta_k(y)| < \epsilon.$$

A7 is a stronger condition than A4. With the uniform continuous coefficient function assumed in A7 and the uniform consistency of the uniform sample quantile function from Shorack and Wellner (1986, Thm. 3, p. 95), I establish the uniform consistency of the coefficient function estimator $\hat{\boldsymbol{\beta}}(\cdot)$ on $(0, 1)$.

Theorem 3 (Uniform consistency in simplified model)**.** *Under A1–A3, A5, and A7, the coefficient function estimator uniformly converges to the true coefficient function on* $(0,1)$*:*

$$\sup_{\tau\in(0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| \xrightarrow{a.s.} 0 \quad \textit{as } n \to \infty.$$

4.2 Large-T point estimation

I now describe the estimator in the general large-T model and show consistency. I assume that each individual's coefficient vector estimator is consistent as $T \to \infty$, based on well-established time series theories. I show the uniform consistency of the estimator as both T and n go to infinity under a certain rate condition.

Estimation consists of two steps. First, the individual coefficients $\hat{\boldsymbol{\beta}}(U_i)$ are estimated for each i. Second, the order of the n unobserved individual ranks U_i is estimated based on the estimated coefficients, and the estimator $\hat{\boldsymbol{\beta}}(\tau)$ at any rank $\tau \in (0,1)$ is the estimated coefficient at the estimated $\lceil n\tau \rceil$th order statistic.

One of the major differences in estimation between the general model and the simplified model is that instead of perfectly estimating the true individual coefficients in the simplified model, the estimated individual coefficients now have estimation error in general model. This might affect the ordering of the fitted outcome values $\hat{Y}_i^*$, implying an imprecise ordering of the unobserved rank values U_i.

4.2.1 Time series OLS consistency

Let $\hat{\boldsymbol{\beta}}(U_i)$ denote the time series OLS (TS-OLS) estimator for individual i's coefficient $\boldsymbol{\beta}(U_i)$,

$$\hat{\boldsymbol{\beta}}(U_i) = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{it}\mathbf{X}_{it}' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{it} Y_{it} \right). \tag{17}$$

There are many large-sample results for the TS-OLS estimator under various assumptions. For example, Hamilton (1994, Ch. 8.2) lists five different scenarios under which the TS-OLS estimator is consistent. Relying on these well-established results, I assume that the TS-OLS estimator of each individual-specific coefficient is consistent as $T \to \infty$.

Assumption A8 (TS-OLS consistency)**.** (i) The TS-OLS coefficient estimator is consistent for all individuals.

$$\hat{\boldsymbol{\beta}}_T(U_i) \xrightarrow{p} \boldsymbol{\beta}(U_i) \text{ as } T \to \infty,\ \forall i = 1,\ldots,n. \tag{18}$$

(ii) The convergence rate is T^{κ} for some positive $\kappa > 0$:

$$\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i) = O_p\big(T^{-\kappa}\big) \text{ as } T \to \infty,\ \forall i = 1,\ldots,n. \tag{19}$$

(iii) As $n \to \infty$, the TS-OLS coefficient estimates are uniformly consistent in probability, i.e., for any positive $\epsilon > 0$,

$$\mathrm{P}\left(\sup_{1 \le i \le n} \left\| \hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i) \right\| > \epsilon \right) \to 0 \text{ as } n, T \to \infty. \tag{20}$$

The norm in A8(iii) is taken as the supremum norm for simplicity, but it can be interpreted as any L_p norm due to the equivalence of norms in finite-dimensional space. Assumption A8 implicitly restricts how fast T must grow with respect to n. A sufficient n/T rate relation is given in the uniform consistency result of Theorem 5.

4.2.2 Large-T estimation

Define the estimated permutation $\hat{\sigma}(\cdot)$ to associate the individual i with its fitted outcome value order k:

$$\hat{\sigma}(k) \equiv \left\{ i : \hat{Y}^*_{n:k} = \hat{Y}_i \right\}. \tag{21}$$

The estimated permutation $\hat{\sigma}(k)$ depends on the data and the $\mathbf{x}^*$ value. Later Proposition 4 shows the estimated order equals the true order with probability approaching one under a certain rate relation between n and T, so $\mathrm{P}(\hat{\sigma}(\cdot) = \sigma(\cdot)) \to 1$.

Specifically, the estimation of $\boldsymbol{\beta}(\tau)$ consists of four steps.

Step 1: For each individual, use its T time period observations to estimate its individual coefficient: $\hat{\boldsymbol{\beta}}(U_i) = (\underline{\mathbf{X}}_i' \underline{\mathbf{X}}_i)^{-1} (\underline{\mathbf{X}}_i' \mathbf{Y}_i)$, $i = 1, \ldots, n$.

Step 2: Given the estimated individual coefficients $\hat{\boldsymbol{\beta}}(U_1), \hat{\boldsymbol{\beta}}(U_2), \ldots, \hat{\boldsymbol{\beta}}(U_n)$ from Step 1, compute the n fitted outcome values $\hat{Y}^*$ at $\mathbf{X} = \mathbf{x}^*$,

$$\hat{Y}^*_i = \mathbf{x}^{*\prime} \hat{\boldsymbol{\beta}}(U_i), \; i = 1, \ldots, n. \tag{22}$$

Step 3: Sort the fitted outcome values $\{\hat{Y}^*_i\}_{i=1}^n$ in increasing order: $\hat{Y}^*_{n:1} \le \hat{Y}^*_{n:2} \le \cdots \le \hat{Y}^*_{n:n}$. The order represents the estimated order of the U_i: the estimated permutation $\hat{\sigma}(\cdot)$ satisfies $\hat{Y}^*_{n:k} = Y^*_{\hat{\sigma}(k)}$ for $k = 1, \ldots, n$.

Step 4: Using the $\hat{\boldsymbol{\beta}}(U_i)$ from Step 1 and the estimated permutation $\hat{\sigma}(\cdot)$ from Step 3, the estimator at rank τ is

$$\hat{\boldsymbol{\beta}}(\tau) = \hat{\boldsymbol{\beta}}\big(U_{\hat{\sigma}(\lceil n\tau \rceil)}\big). \tag{23}$$

4.2.3 Large-T, large-n uniform consistency

The estimator in the general model is uniformly consistent over $0 < \tau < 1$, under the following additional assumption.

Assumption A9. (Extended monotonicity condition) Let $g(u) = \mathbf{x}^{*\prime}\boldsymbol{\beta}(u)$. The derivative function $g'(u)$ has a positive lower bound L on $(0,1)$: $\inf_{0<u<1} g'(u) \geq L > 0$.

The monotonicity condition A3 assumes the function $g(u) = \mathbf{x}^{*\prime}\boldsymbol{\beta}(u)$ is strictly increasing on $(0,1)$. Assumption A9 strengthens this slightly, so $g(u)$ has a strictly positive lower bound of its derivative. It prevents the situation that the function $g(u)$ is almost flat on some intervals; for example, $g(u) = \sin(u\pi/2)$ satisfies A3 because $g'(u) > 0$ for all $0 < u < 1$ but violates A9 because $\lim_{u\uparrow 1} g'(u) = 0$. This restriction is not a strong assumption in the sense that it still allows an individual element function of the vector-valued coefficient $\boldsymbol{\beta}(\cdot)$ to be non-monotone or almost flat.

Let $Z_i = \|\hat{\boldsymbol{\beta}}_{T,i} - \boldsymbol{\beta}_i\|$ denote a random variable based on the TS-OLS estimator. The Z_i are iid over i from A1.

Assumption A10. For all $T > T_0$, $\mathrm{E}(T^{\kappa} Z_i) < \infty$, where $T_0 \in \mathbb{N}$ is some constant.

Assumption A10 helps establish the stochastic order of $\max_{1\leq i\leq n}\|\hat{\boldsymbol{\beta}}_{T,i} - \boldsymbol{\beta}_i\|$ in terms of n, which then translates to T using a specified n/T rate relation.

Proposition 4 shows a sufficient n/T rate relation that guarantees the true order of the U_i can be discovered from the order of $\hat{Y}_i^*$. That is, Proposition 4 implies the estimated permutation is asymptotically the same as the true permutation.

Proposition 4 (Estimated permutation consistency)**.** *Suppose A1–A4 and A6–A10 hold. Under the rate relation* $n = o\big(T^{\kappa/(3+\delta)}\big)$*, for some small positive* $\delta > 0$*,*

$$\mathrm{P}\left(\frac{\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}}{2} > \max_{1\leq i\leq n} |\hat{Y}^*_{T,i} - Y^*_i|\right) \to 1 \text{ as } n, T \to \infty. \tag{24}$$

The coefficient estimator defined based on the estimated permutation is uniformly consistent over $(0,1)$.

Theorem 5 (Uniform consistency in general model)**.** *Under A1–A3 and A6–A10, as* n *and* T *go to infinity with the rate relation* $n = o(T^{\kappa/(3+\delta)})$*, the estimated coefficient function uniformly converges to the true coefficient function,*

$$\sup_{\tau\in(0,1)} \left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right\| = o_p(1).$$

5 Inference

In this section, I provide a novel method to obtain confidence sets or uniform confidence bands for the coefficient function $\boldsymbol{\beta}(\cdot)$ on $(0,1)$. My approach uses the ordered Dirichlet distribution of uniform order statistics to capture uncertainty in the unobserved rank variable values. It sheds light on the shape of the coefficient function, showing patterns in the heterogeneous effects across individuals.

First, I derive results in the simplified model to develop intuition and intermediate theoretical results. Specifically, I derive confidence sets for points on the coefficient function as well as a uniform confidence band if the coefficient function is assumed monotonic. Then, I generalize these to the general model.

Notationally, the scalar $\beta(\cdot)$ refers to any single component of the vector-valued $\boldsymbol{\beta}(\cdot)$.

5.1 Dirichlet method in simplified model

In the simplified model, we perfectly estimate all the individual coefficients. There is no uncertainty in the vertical direction in the coefficient function graph. The uncertainty comes from the random rank variable U in the horizontal direction. Therefore, the confidence interval for individual coefficient $\boldsymbol{\beta}(U_{n:k})$ is horizontal, instead of a conventional vertical one.

5.1.1 Pointwise confidence interval

This subsection presents individual pointwise confidence intervals. These are depicted as the red intervals in the left graph in Figure 1.

Define the kth horizontal interval as

$$\mathcal{I}_k = \{(x, y) \mid a_k \leq x \leq b_k; y = \beta(U_{n:k})\}, \tag{25}$$

where a_k and b_k are the $\alpha/2$-quantile and $(1-\alpha/2)$-quantile of the Beta$(k, n+1-k)$ distribution. Using the uniform order statistic property in (9), $\mathrm{P}(a_k \leq U_{n:k} \leq b_k) = 1-\alpha$, so the interval $\mathcal{I}_k$ is an equal-tailed $100(1-\alpha)\%$ confidence interval in the sense that it contains $\beta(U_{n:k})$ with probability $1-\alpha$, with equal probability of $\beta(U_{n:k})$ being too far to the left or right. Having fixed endpoints (a_k, b_k) and random $\beta(U_{n:k})$ is different than the usual case of a random interval containing a fixed point, but the interpretation of frequentist coverage probability is otherwise the same. It can also be interpreted that the coefficient function $\beta(\cdot)$ crosses the interval $\mathcal{I}_k$ with $1-\alpha$ probability.

Each single interval $\mathcal{I}_k$ $(k = 1, \ldots, n)$ includes the corresponding individual-specific coefficient $\beta(U_{n:k})$ with exact probability $1-\alpha$, even with small n. However, as usual, these n separate intervals $\{\mathcal{I}_k\}_{k=1}^n$ have less than $1-\alpha$ joint coverage probability. In the next subsection, I construct confidence sets that can cover the coefficient function with the desired joint coverage probability.

5.1.2 Joint confidence sets and uniform confidence bands

First, I construct confidence sets for the n individual coefficients $\beta(U_{n:1}), \ldots, \beta(U_{n:n})$ jointly, based on results from Goldman and Kaplan (2018). Second, I restrict the coefficient function to be monotone[1] and establish uniform confidence bands for the whole coefficient function.

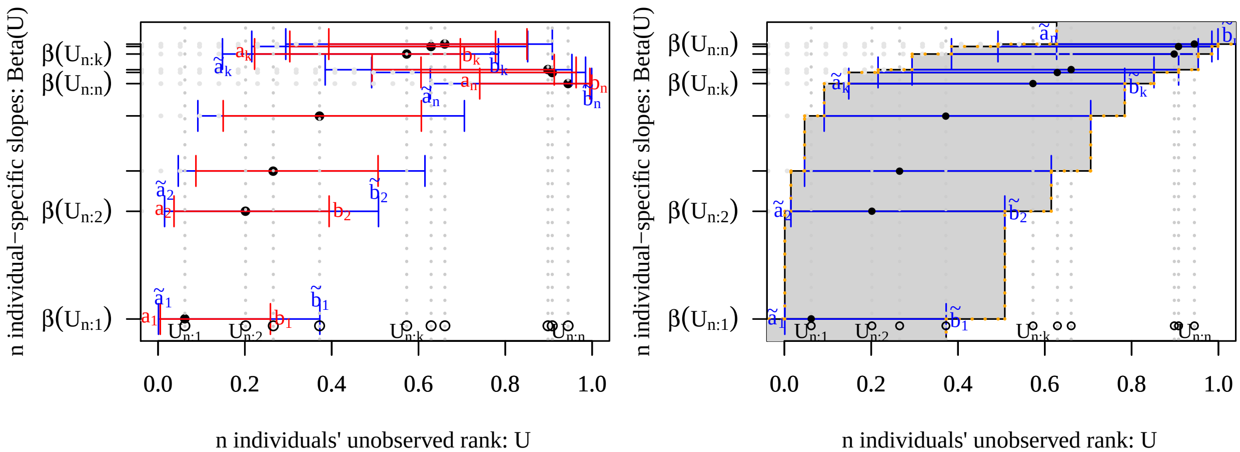


Figure 1: Left: pointwise confidence intervals (red) and joint confidence sets (blue) for the n points on the generic coefficient function in Case 1. Right: uniform confidence band for the monotone coefficient function $\beta(\cdot)$ in Case 2.

Case 1: Generic coefficient function and confidence sets

Goldman and Kaplan (2018) provide joint confidence intervals for the n jointly ordered Dirichlet-distributed uniform order statistics $(U_{n:1}, \ldots, U_{n:n})$. That is, given any $\alpha \in (0, 1)$,

$$\mathrm{P}(\tilde{a}_1 < U_{n:1} < \tilde{b}_1 \text{ and } \cdots \text{ and } \tilde{a}_n < U_{n:n} < \tilde{b}_n) = 1 - \alpha, \tag{26}$$

where $\tilde{a}_k$ is the $\tilde{\alpha}/2$-quantile of the Beta$(k, n+1-k)$ distribution; $\tilde{b}_k$ is the $(1-\tilde{\alpha}/2)$-quantile of the Beta$(k, n+1-k)$ distribution; and $\tilde{\alpha}$ is theoretically from Theorem 5 in Goldman and Kaplan (2018). This $\tilde{\alpha}$ can be analytically approximated from Fact 6 in Goldman and Kaplan (2018),

$$\tilde{\alpha} = \exp\{-c_1(\alpha) - c_2(\alpha)\sqrt{\ln \ln n} - c_3(\alpha)[\ln n]^{c_4(\alpha)}\}, \tag{27}$$

where $c_1(\alpha) = -2.75 - 1.04\ln(\alpha)$, $c_2(\alpha) = 4.76 - 1.20\alpha$, $c_3(\alpha) = 1.15 - 2.39\alpha$ and $c_4(\alpha) = -3.96 + 1.72\alpha^{0.171}$.

[1] The monotone coefficient function in this section is different from the monotonicity condition in Assumption A3. The monotonicity condition A3 is monotonicity of the inner product of the vector $\mathbf{X} = \mathbf{x}^*$ and the vector-valued coefficient $\boldsymbol{\beta}(\cdot)$ as a function of the rank variable, whereas here the monotone coefficient function is referring to each scalar-valued coefficient function $\beta(\cdot)$ itself being monotone.

Define the n intervals in the coefficient function graph

$$\tilde{\mathcal{I}}_k = \{(x, y) \mid \tilde{a}_k \leq x \leq \tilde{b}_k; y = \beta(U_{n:k})\}, \quad k = 1, \ldots, n, \tag{28}$$

where the $\tilde{a}_k$ and $\tilde{b}_k$ are the same as in (26). These $\tilde{\mathcal{I}}_k$ are depicted as the blue intervals in the left graph in Figure 1. They jointly cover the n individual coefficients with at least $1-\alpha$ probability.

Theorem 6 (Confidence set for generic coefficient function in simplified model)**.** *Under A1–A5, the joint intervals $\cup_{k=1}^n \tilde{\mathcal{I}}_k$ form a confidence set for the n individual coefficient points on the generic coefficient function at a $100(1-\alpha)\%$ confidence level:*

$$1 - \alpha \leq \mathrm{P}\Big(\{\cup_{k=1}^n (U_{n:k}, \beta(U_{n:k}))\} \subseteq \{\cup_{k=1}^n \tilde{\mathcal{I}}_k\}\Big).$$

It can be interpreted that the coefficient function $\beta(\cdot)$ crosses all n intervals $\cup_{k=1}^n \tilde{\mathcal{I}}_k$ jointly with probability at least $1-\alpha$. Equality is achieved when the n individual coefficients are distinct, in which case the coefficient function crosses the joint intervals $\cup_{k=1}^n \tilde{\mathcal{I}}_k$ with exact probability $1-\alpha$.

Generalization to vector-valued coefficient function

We can generalize the confidence sets for the n coefficient estimates on the scalar-valued coefficient function $\beta(\cdot)$ to that on the vector-valued coefficient function $\boldsymbol{\beta}(\cdot)$. The former can be considered as one element function of the coefficient in the latter case.

The uncertainty of the vector-valued coefficient function in the simplified model only comes from the uncertainty of the individual rank variable. First, in the simplified model, there is no vertical uncertainty from estimation error of the individual coefficients. The only uncertainty comes from the horizontal direction about the individual rank order statistics $U_{n:k}$. Second, the model assumes that each element of the random coefficient vector $\boldsymbol{\beta}(U)$ depends on the same scalar U. There is no more uncertainty about the full vector than any individual component. Both reduce to the uncertainty about the same scalar order statistics $U_{n:k}$.

Consider a K-vector coefficient function $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_K(\cdot))'$. We can construct the confidence sets on each of the element coefficient functions $\beta_1(\cdot), \ldots, \beta_K(\cdot)$. That is, we draw the n intervals as the confidence sets in each of the K graphs with axes $u \mapsto \beta_1(u)$, $u \mapsto \beta_2(u)$, …, $u \mapsto \beta_K(u)$. The horizontal endpoints $\{(\tilde{a}_j, \tilde{b}_j)\}_{j=1}^n$ of the n intervals in the K graphs are the same. The only difference is the vertical axis values in the K graphs.

Consider the lth element function of the vector-valued coefficient. Define the n intervals

$$\tilde{\mathcal{I}}_k^l = \{(x, y) \mid \tilde{a}_k \leq x \leq \tilde{b}_k; y = \beta_l(U_{n:k})\}, \; k = 1, \ldots, n,$$

where the $\tilde{a}_k$ and $\tilde{b}_k$ are the same as in (26). From Theorem 6, the confidence set $\cup_{k=1}^n \tilde{\mathcal{I}}_k^l$ covers the n individual coefficients on the lth element coefficient function with probability $1-\alpha$ at least.

Let the "vertical" dimension denote the K-vector coefficient. Define the K-dimensional intervals

$$\tilde{\boldsymbol{\mathcal{I}}}_k = \Big\{(x, \mathbf{y}) \mid \tilde{a}_k \leq x \leq \tilde{b}_k; \mathbf{y} = (\beta_1(U_{n:k}), \ldots, \beta_K(U_{n:k}))'\Big\}. \tag{29}$$

The confidence set $\cup_{k=1}^n \tilde{\boldsymbol{\mathcal{I}}}_k$, which consists of n K-dimensional intervals, jointly covers the n vector-valued points $\boldsymbol{\beta}(U_{n:1}), \ldots, \boldsymbol{\beta}(U_{n:n})$ on the vector-valued coefficient function with probability $1-\alpha$ at least.

Corollary 7 (Confidence set for vector-valued generic coefficient function in simplified model)**.** *Under A1–A5, with a vector-valued coefficient function $\boldsymbol{\beta}(\cdot)$, using the definitions in* (26)*,* (27)*, and* (29)*, $\cup_{k=1}^n \tilde{\boldsymbol{\mathcal{I}}}_k$ is a* $100(1-\alpha)\%$ *confidence set for the n vector-valued individual coefficients $\boldsymbol{\beta}(U_{n:1}), \ldots, \boldsymbol{\beta}(U_{n:n})$:*

$$1 - \alpha \leq \mathrm{P}\Big(\{\cup_{k=1}^n (U_{n:k}, \boldsymbol{\beta}(U_{n:k}))\} \subseteq \Big\{\cup_{k=1}^n \tilde{\boldsymbol{\mathcal{I}}}_k\Big\}\Big).$$

The intuition for Corollary 7 is that each element of the vector coefficient function depends on the same scalar U; $U_{n:k}$ is the only source of uncertainty of each element coefficient function. For a vector-valued coefficient function, we can construct the confidence set on each element coefficient function at the same time. It has the property that the first element function $\beta_1(\cdot)$ crosses its confidence set $\cup_{k=1}^n \tilde{\mathcal{I}}_k^1$ if and only if the second element function $\beta_2(\cdot)$ crosses its confidence set $\cup_{k=1}^n \tilde{\mathcal{I}}_k^2$ if and only if the third element function $\beta_3(\cdot)$ crosses its confidence set, etc. Each element coefficient function going through its corresponding n-interval confidence set is the same as the vector-valued coefficient function going through the n K-dimensional intervals.

Case 2: Monotonic coefficient function and uniform confidence bands

Assume the coefficient function is strictly monotone. Monotonicity enables us to build a uniform confidence band for the whole coefficient function based on the n individual coefficients. For simplicity, I focus on a scalar-valued coefficient function. The results can be generalized to vector-valued coefficient functions as well.

With monotonicity, we are able to bound the coefficient function value between two

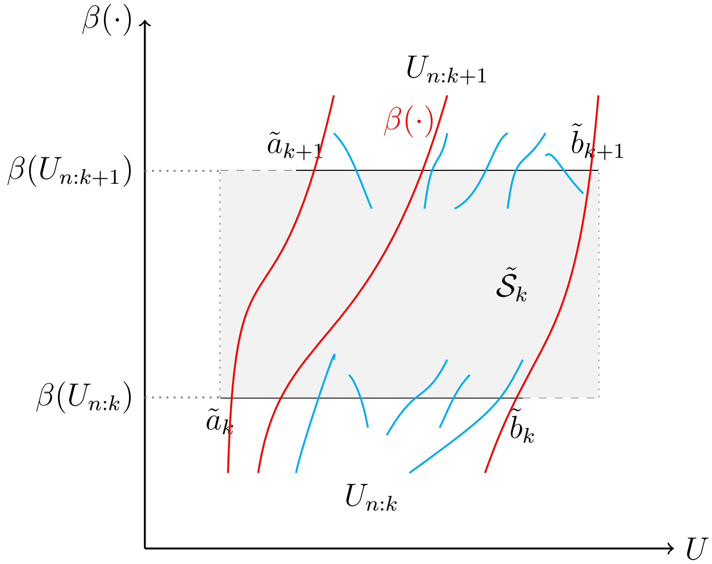


Figure 2: Comparison between Case 1 and Case 2. In Case 2, the red nondecreasing curves that are within the shaded gray area indicate the possible coefficient functions consistent with the two intervals $\tilde{\mathcal{I}}_k$ and $\tilde{\mathcal{I}}_{k+1}$. In Case 1, we only know that the possible coefficient function (denoted as the blue curves) crosses $\tilde{\mathcal{I}}_k$ and $\tilde{\mathcal{I}}_{k+1}$; there is no information about the coefficient function for the value between $\beta(U_{n:k})$ and $\beta(U_{n:k+1})$.

observed coefficients $\beta(U_i)$ and $\beta(U_j)$. That is, a monotonic coefficient function guarantees that for any value u in between the U_i and U_j values, the coefficient function value $\beta(u)$ is between the two observed coefficient values $\beta(U_i)$ and $\beta(U_j)$.

Figure 2 illustrates the difference between the confidence sets in the generic coefficient function case and the uniform confidence bands in the monotone coefficient function case. Without loss of generality, assume the coefficient function $\beta(\cdot)$ is increasing. Consider two adjacent rank variable order statistics and their coefficients $\beta(U_{n:k})$ and $\beta(U_{n:k+1})$. When the coefficient function is unrestricted, we only know that (with high probability) the coefficient function goes through the two horizontal intervals $\tilde{\mathcal{I}}_k$ and $\tilde{\mathcal{I}}_{k+1}$. The coefficient function can cross the two horizontal intervals from any direction, either from top-left to bottom-right or from bottom-left to top-right. In contrast, when the coefficient function is increasing, the two intervals $\tilde{\mathcal{I}}_k$ and $\tilde{\mathcal{I}}_{k+1}$ can be strengthened to a confidence band for the coefficient function. That is, for all rank values between $U_{n:k}$ and $U_{n:k+1}$, the coefficient value $\beta(\cdot)$ has to between $\beta(U_{n:k})$ and $\beta(U_{n:k+1})$. That gives the rectangle $\mathcal{S}_k$ defined later. Additionally, it restricts a monotonic coefficient function to cross all horizontal intervals from the same direction: either all from bottom-left to top-right or all from top-left to bottom-right.

To construct the band, first construct the $n-1$ rectangles $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{n-1}$,

$$\mathcal{S}_k = [\tilde{a}_k, \tilde{b}_{k+1}] \times [\beta(U_{n:k}), \beta(U_{n:k+1})]$$

$$= \{(x,y) \mid \tilde{a}_k \le x \le \tilde{b}_{k+1}; \beta(U_{n:k}) \le y \le \beta(U_{n:k+1})\}, \quad k = 1, \ldots, n-1, \tag{30}$$

and thenconstruct two open-ended areas $\mathcal{S}_0$ and $\mathcal{S}_n$,

$$\mathcal{S}_0 = \left[-\infty, \tilde{b}_1\right] \times [-\infty, \beta(U_{n:1})] = \{(x,y) \mid x \le \tilde{b}_1; y \le \beta(U_{n:1})\},$$
$$\mathcal{S}_n = [\tilde{a}_n, +\infty] \times [\beta(U_{n:n}), +\infty] = \{(x,y) \mid x \ge \tilde{a}_n; y \ge \beta(U_{n:n})\},$$

where $\tilde{a}_k$ and $\tilde{b}_k$ are the same as in (26). Define the total area $\mathcal{S}$ as

$$\mathcal{S} = \cup_{k=0}^{n} \mathcal{S}_k. \tag{31}$$

The area $\mathcal{S}$, which is composed of the $n-1$ rectangles plus two open-ended areas, provides a uniform confidence band for the monotone coefficient function $\beta(\cdot)$. It is depicted as the gray area in the right graph in Figure 1.

Theorem 8 (Uniform confidence band for monotone coefficient function in simplified model)**.** *Under A1–A5, the area $\mathcal{S}$ is a uniform confidence band for the strictly monotone coefficient function at a $100(1-\alpha)\%$ confidence level:*

$$1 - \alpha = \mathrm{P}(\{(u,y) : y = \beta(u), 0 \le u \le 1\} \subseteq \mathcal{S}).$$

The vertical intervals of the uniform confidence band collapse as n goes to infinity because $(U_{n:k+1} - U_{n:k}) \sim \text{Beta}(1,n) \xrightarrow{p} 0$ for any $k = 1, \ldots, n-1$ and $\beta(\cdot)$ is continuous. Although not as immediatley obvious, the horizontal intervals collapse as well as $n \to \infty$, since each interval for the random variable $U_{n:k} \sim \text{Beta}(k, n+1-k)$ shrinks to a point $U_{n:\lceil n\tau \rceil} \xrightarrow{p} \tau$. Therefore, the uniform confidence band collapses in both horizontal and vertical directions to an infinitesimal small point, so it traces out the true coefficient function $\beta(\cdot)$.

5.2 Dirichlet method in general model

This subsection extends the Dirichlet inference method to the large-T general model. First, I establish the pointwise confidence set for each individual coefficient. Second, I build joint confidence sets for the n individual coefficients.

In the general model, we no longer observe the exact individual coefficient parameters. Instead, we observe n individual TS-OLS coefficient estimates $\hat{\beta}(U_{n:k})$ for $k = 1, \ldots, n$. Now, there are two dimensions of uncertainty in the coefficient function graph. One is from the uncertainty in the rank variable $U_{n:k}$, since it is an unobserved random variable following a beta distribution. The other source of uncertainty is from the TS-OLS estimation of coefficient, which essentially comes from the idiosyncratic error V_{it}. Therefore, the confidence set is made up of rectangles in the general model, in contrast to the horizontal intervals in the simplified model.

Assumption A11 (Inference for TS-OLS individual coefficients)**.** Given n individuals, for any $\alpha \in (0,1)$, we can obtain confidence interval $\mathcal{T}_i^{\alpha}$ for each individual i's coefficient $\boldsymbol{\beta}_i$ with asymptotic coverage probability $1-\alpha$ as $T \to \infty$:

$$\lim_{T\to\infty} \mathrm{P}(\boldsymbol{\beta}_i \in \mathcal{T}_i^{\alpha}) = 1-\alpha.$$

The confidence interval $\mathcal{T}_i^{\alpha}$ depends implicitly on the TS-OLS estimator $\hat{\boldsymbol{\beta}}_i$. For example, when the time dimension observations are iid and regular OLS results apply ($\kappa = 1/2$ in A8), we can write the asymptotic 95% confidence interval as $\mathcal{T}_i^{0.05} = \hat{\beta}(U_i) \pm 1.96\hat{\sigma}/\sqrt{T}$, for some standard deviation estimator $\hat{\sigma}$ that is consistent for the asymptotic variance of $\sqrt{T}(\hat{\beta}_i - \beta_i)$. The parameter α denotes nominal coverage probability $1-\alpha$ of the confidence interval $\mathcal{T}_i^{\alpha}$.

Another complication is that TS-OLS estimation error can make the estimated permutation $\hat{\sigma}(\cdot)$ from (21) not equal the true permutation $\sigma(\cdot)$ in finite samples. However, Proposition 4 shows that they are equal asymptotically with probability approaching one. Thus, at least asymptotically, this source of error has a negligible effect on coverage probability.

5.2.1 Pointwise confidence set

Define the rectangle

$$\mathcal{W}_k^{(\alpha_1,\alpha_2)} \equiv [a_k^{\alpha_1}, b_k^{\alpha_1}] \times \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2} = \left\{(x,y) \mid a_k^{\alpha_1} \le x \le b_k^{\alpha_1}; y \in \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}\right\},\ k = 1,\ldots,n, \tag{32}$$

where $a_k^{\alpha_1}$ and $b_k^{\alpha_1}$ are from (25) with coverage probability $1-\alpha_1$, and $\mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}$ is defined through the estimated permutation with asymptotic coverage probability $1-\alpha_2$. The rectangle provides a pointwise confidence set as described in Theorem 9.

Theorem 9 (Pointwise confidence set for individual coefficient in the fixed-n, large-T general model)**.** *Let A1–A4, A6, and A8–A11 hold. For any $\alpha \in (0,1)$ and any constant $w \in (0,1)$, let $\alpha_1 = w\alpha$ and $\alpha_2 = (1-w)\alpha$. Then, the rectangle $\mathcal{W}_k^{(\alpha_1,\alpha_2)}$ contains the point $(U_{n:k}, \beta(U_{n:k}))$ with asymptotic coverage probability bounded below by $1-\alpha$ as $T \to \infty$: $\lim_{T\to\infty} \mathrm{P}\{(U_{n:k}, \beta(U_{n:k})) \in \mathcal{W}_k^{(\alpha_1,\alpha_2)}\} \ge 1-\alpha$ for any $k = 1,\ldots,n$.*

The pointwise confidence sets each help quantify uncertainty about a single point on the function $\beta(\cdot)$. They are smaller than the rectangles in the joint confidence sets and uniform confidence band introduced below, but they may not properly capture the overall shape of the coefficient function.

5.2.2 Joint confidence set

Parallel to Section 5.1.2 for the simplified model, this subsection provides joint confidence sets for n points on the coefficient function in the general model.

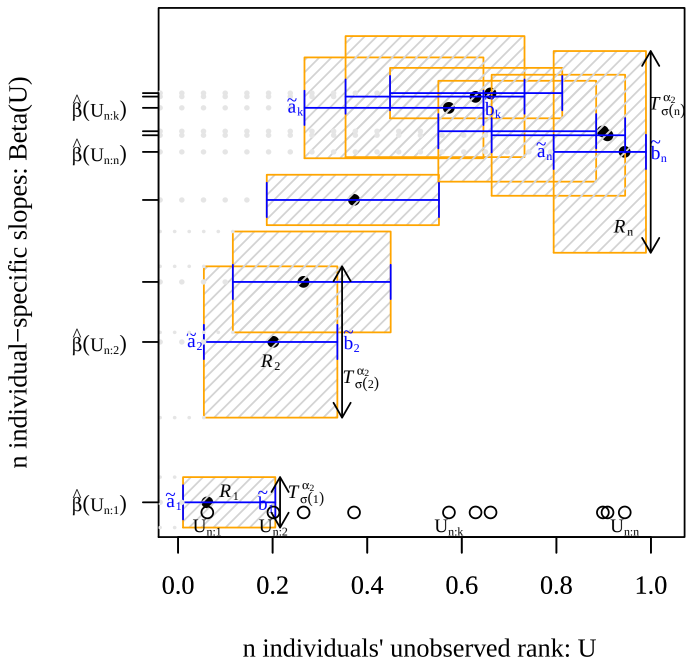


Figure 3: Confidence set for n points on the generic coefficient function in general model.

Given some constant $\alpha_1, \alpha_2 \in (0, 1)$, define the rectangles in the coefficient function graph

$$\mathcal{R}_k^{(\alpha_1,\alpha_2)} \equiv [\tilde{a}_k^{\alpha_1}, \tilde{b}_k^{\alpha_1}] \times \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2} = \left\{(x, y) \mid \tilde{a}_k^{\alpha_1} \leq x \leq \tilde{b}_k^{\alpha_1}; y \in \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}\right\}, \; k = 1, \ldots, n, \tag{33}$$

where $\tilde{a}_k^{\alpha_1}$ and $\tilde{b}_k^{\alpha_1}$ are from (26) with (joint) coverage probability $1 - \alpha_1$, and $\mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}$ is defined through the estimated permutation $\hat{\sigma}(\cdot)$. Define the total area $\mathcal{R}^{(\alpha_1,\alpha_2)}$ as

$$\mathcal{R}^{(\alpha_1,\alpha_2)} \equiv \cup_{k=1}^n \mathcal{R}_k^{(\alpha_1,\alpha_2)}. \tag{34}$$

It is depicted as the n gray boxes in Figure 3. It provides a joint confidence set as described in Theorem 10.

Theorem 10 (Joint confidence set for generic coefficient function in the fixed-n, large-T general model)**.** *Under A1–A4, A6, and A8–A11, in the general model, given n individuals, the area $\mathcal{R}^{(\alpha_1,\alpha_2)}$ is a confidence set for the n individual coefficients on the generic coefficient function. Specifically, for any $\alpha \in (0, 1)$, with $\alpha_1 = \alpha/2$ and $\alpha_2 = 1 - (1 - \alpha/2)^{1/n}$, the confidence set $\mathcal{R}^{(\alpha_1,\alpha_2)}$ has asymptotic coverage probability bounded below by $1 - \alpha$ as*

$T \to \infty$:

$$\lim_{T\to\infty} \mathrm{P}\big(\{\cup_{k=1}^{n}(U_{n:k}, \beta(U_{n:k}))\} \subseteq \mathcal{R}^{(\alpha_1,\alpha_2)}\big) \geq 1-\alpha.$$

Generally, the α_1 and α_2 in Theorem 10 can be generalized to any values that satisfy $\alpha_1 + 1 - (1-\alpha_2)^n = \alpha$, with $0 < \alpha_1, \alpha_2 < 1$. For example, let $\alpha_1 = w\alpha$ and $\alpha_2 = 1-(1-(1-w)\alpha)^{1/n}$, where w is a constant $0 < w < 1$. It satisfies $\alpha_1 + 1 - (1-\alpha_2)^n = \alpha$, and the confidence set $\mathcal{R}^{(\alpha_1,\alpha_2)}$ still has at least $1-\alpha$ asymptotic coverage probability.

The α_1 and α_2 relationship can be used to adjust the width and height of the confidence set rectangles. As α_1 increases, the confidence level $1-\alpha_1$ decreases, so each horizontal interval gets narrower. As α_2 decreases, the confidence level $1-\alpha_2$ increases, so each vertical interval gets taller. Therefore, if the n rectangles of the confidence set are initially wide and short, we can increase α_1 and decrease α_2 to narrow the width and increase the height. Conversely, if the n rectangles are initially narrow and tall, we can decrease α_1 and increase α_2 to widen and shorten the rectangles.

The confidence set consisting of n boxes in Figure 3 will shrink to n horizontal intervals as in the left figure in Figure 1 as $T \to \infty$. This is because the horizontal intervals are the joint confidence set for the n uniform order statistics. The horizontal interval length depends on n, but not on the time dimension T. The vertical intervals depend on both n and T, since they have to satisfy the joint coverage of all n individual coefficients. In the fixed-n, large-T general model, the horizontal interval is fixed as $T \to \infty$, whereas the vertical interval height shrinks to zero.

If the coefficient function is monotone, then the confidence set in Figure 3 extends to a uniform confidence band, parallel to the extension from the left graph to the right graph in Figure 1 for the simplified model. For example, if $\beta(\cdot)$ is increasing, instead of two separate rectangles $\mathcal{R}_k^{(\alpha_1,\alpha_2)}$ and $\mathcal{R}_{k+1}^{(\alpha_1,\alpha_2)}$, the unifrom confidence band expands to the box starting from the bottom-left corner of $\mathcal{R}_k^{(\alpha_1,\alpha_2)}$ to the top-right corner of $\mathcal{R}_{k+1}^{(\alpha_1,\alpha_2)}$.

6 Bootstrap inference

I show the uniform asymptotic normality of the coefficient function estimator in both the simplified model and the general model. Then, I provide bootstrap confidence intervals and show bootstrap consistency.

6.1 Asymptotic normality

I establish the uniform asymptotic normality of the coefficient function estimator by applying the functional delta-method to the well-known Donsker's theorem, assuming differentiability of the coefficient function.

Donsker's theorem (e.g., van der Vaart, 1998, Thm. 19.3) states the asymptotic Gaussianity of the empirical distribution function. Letting F_n be the sequence of the empirical distribution function indexed by sample size n, assuming iid sampling from distribution function F, then

$$\sqrt{n}(F_n - F) \rightsquigarrow \mathbb{G} \circ F, \quad \text{in } D[-\infty, +\infty], \quad \text{as } n \to \infty, \tag{35}$$

where $\mathbb{G}$ is the standard Brownian bridge.[2] The limit process is mean zero and has covariance function $\text{Cov}[\mathbb{G}_F(u), \mathbb{G}_F(v)] = F(u) \wedge F(v) - F(u)F(v)$. The $D[s,t]$ denotes the Skorohod space of all cadlag functions on the interval $[s,t] \subset \bar{\mathbb{R}}$.

Assumption A12. Consider the K-vector coefficient function $\boldsymbol{\beta}\colon (0,1) \mapsto \mathbb{R}^K$. Assume each of its element function $\beta_k\colon (0,1) \mapsto \mathbb{R}$, $k = 1, \ldots, K$, is differentiable with uniformly continuous and bounded derivative.

Assumption A12 is a regularity condition that assumes the differentiability of the coefficient function. Applying the fundamental property in Lemma 18 to a standard uniform distribution function, we get the uniform asymptotic normality of the coefficient function estimator in the simplified model.

Theorem 11 (Uniform asymptotic normality of coefficient function estimator in simplified model)**.** *Under A1–A5 and A12 in the simplified model, for any element coefficient function $k = 1, \ldots, K$, we have the uniform asymptotic normality of the coefficient function estimator*

$$\sqrt{n}\big(\hat{\beta}_k(\cdot) - \beta_k(\cdot)\big) \rightsquigarrow \beta_k' \mathbb{G}, \quad \textit{in } \ell^\infty(0,1), \ \textit{as } n \to \infty.$$

The limit process is Gaussian with mean zero and covariance function $\beta_k'(u)\beta_k'(v)(u \wedge v - uv)$, $u, v \in (0,1)$.

The symbol $\ell^\infty(T)$ denotes the collection of all bounded functions $f\colon T \mapsto \mathbb{R}$ with the norm $\|f\|_\infty = \sup_{t \in T} f$.

Corollary 12 applies Theorem 11 at a fixed τ to get the pointwise asymptotic normality of the coefficient estimator at any specific rank.

Corollary 12 (Pointwise asymptotic normality of coefficient estimator in simplified model)**.** *Under A1–A5 and A12 in the simplified model, for any $\tau \in (0,1)$,*

$$\sqrt{n}\big(\hat{\beta}_k(\tau) - \beta_k(\tau)\big) \xrightarrow{d} \beta_k'(\tau)\mathbb{G}(\tau), \ \textit{as } n \to \infty,$$

[2] The standard Brownian bridge process $\mathbb{G}$ on the unit interval $[0,1]$ is mean zero with covariance function $\text{Cov}[\mathbb{G}(u), \mathbb{G}(v)] = u \wedge v - uv$, where $u \wedge v$ denotes $\min\{u, v\}$.

where $\beta_k'(\tau)\mathbb{G}(\tau)$ *is a Gaussian random variable with mean zero and variance* $\tau(1-\tau)(\beta_k'(\tau))^2$*.*

The uniform asymptotic normality of the coefficient function estimator also holds in the general model under the same n/T rate relation for its uniform consistency. In the general model, the TS-OLS individual coefficient estimates do not exactly equal to the true coefficient parameters, but the order of the rank variables can still be discovered under the rate condition. Under the same rate condition, the TS-OLS estimation error goes away even after scaling by $\sqrt{n}$. Thus, the coefficient function estimator has the same limit as in the simplified model.

Theorem 13 (Uniform asymptotic normality in general model)**.** *Under A1–A4, A6–A10, and A12 in the general model, as* $n, T \to \infty$ *with rate relation* $n = o\big(T^{\kappa/(3+\delta)}\big)$*, the coefficient function estimator converges to a Gaussian limit,*

$$\sqrt{n}\big(\hat{\beta}_k(\cdot) - \beta_k(\cdot)\big) \rightsquigarrow \beta_k'\mathbb{G}, \quad \textit{in } \ell^\infty(0,1),$$

where the limit is mean zero and has covariance function $\beta_k'(u)\beta_k'(v)(u \wedge v - uv)$*,* $u, v \in (0,1)$*.*

6.2 Bootstrap in the simplified model

In this subsection, I first show bootstrap consistency, which provides the theoretic grounds for the bootstrap method suggested later. Then, I construct a bootstrap confidence interval for the coefficient at any specific rank value. Finally, I present a cluster-robust bootstrap algorithm.

Generally, let $\mathbb{P}_n$ be the empirical measure for a size n iid sample $X_1, X_2, \ldots, X_n$ from probability measure $\mathbb{P}$:

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i},$$

where δ_x is the Dirac measure, whose mass is concentrated at a point x on the domain of Borel sets in $\mathbb{R}$. This $\mathbb{P}_n$ is the probability measure for the empirical distribution function F_n,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \le x\}.$$

Let $\mathcal{F}$ be the set of indicator functions $\mathcal{F} = \big\{\mathbb{1}\{(-\infty, t]\} : t \in \bar{\mathbb{R}}\big\}$. It is known that each such indicator function is measurable and that $\mathcal{F}$ is a Donsker class for any underlying measure $\mathbb{P}$ of $X_1, \ldots, X_n$. Donsker's theorem can alternatively be written as

$$\sqrt{n}(\mathbb{P}_n - \mathbb{P}) \rightsquigarrow \mathbb{H}, \text{ in } \ell^\infty(\mathcal{F}), \text{ as } n \to \infty, \tag{36}$$

where the limit process $\mathbb{H}$ is the same as $\mathbb{G} \circ F$ previously, with $\mathbb{G}$ being the standard Brownian bridge.

Consider the standard bootstrap method. Given a sample $X_1, X_2, \ldots, X_n$, the bootstrap sample $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n$ is an iid sample drawn from the original sample $X_1, X_2, \ldots, X_n$ with replacement. It treats the empirical measure $\mathbb{P}_n$ as the true measure for bootstrap resampling. Let $\hat{\mathbb{P}}_n$ denote the bootstrap empirical measure,

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^{n} M_{ni} \delta_{X_i}, \tag{37}$$

where M_{ni} denotes the number of times X_i is redrawn in the size n bootstrap sample. Conditional on the data, the random vector $(M_{n1}, \ldots, M_{nn})$ follows a multinomial distribution with parameters n and $(1/n, \ldots, 1/n)$.

The bootstrap empirical process is

$$\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n).$$

It is well known that the bootstrap empirical process converges to the same limit process as in Donsker's theorem.

Lemma 14 (The bootstrap empirical process convergence, as a special case of van der Vaart and Wellner (1996) Theorem 3.6.13)**.** *Conditional on almost all data sequences $X_1, X_2, \ldots$, the bootstrap empirical process converges to a limit process,*

$$\hat{\mathbb{G}}_n \rightsquigarrow \mathbb{H}, \ \textit{in } \ell^\infty(\mathcal{F}), \ \textit{as } n \to \infty,$$

where the limit is the same as in (36)*.*

6.2.1 Bootstrap consistency

Definition 1 (Validity of bootstrap)**.** A bootstrap method is "valid" if the following two sufficient conditions are met. First, the centered and scaled estimator asymptotically converges to some distribution,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathrm{N}(0, V), \ \text{as } n \to \infty. \tag{I}$$

Second, conditional on almost all sequences $X_1, X_2, \ldots$ drawn iid from the true CDF F, the centered and scaled bootstrap estimator asymptotically converges to the same limiting distribution,

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \mid (X_1, \ldots, X_n) \xrightarrow{d} \mathrm{N}(0, V), \ \text{as } n \to \infty, \tag{II}$$

where θ_0 is the true parameter, $\hat{\theta}$ is the estimator for θ_0 from the sample, and $\hat{\theta}^*$ is the bootstrap estimator.

To show bootstrap validity for the coefficient estimator in this paper, I show the first condition is met by Theorem 11. Applying the delta-method to Lemma 14, I show the second condition is met.

Lemma 15 (Bootstrap consistency)**.** *Let F be a distribution function with compact support $[s,t]$ and be continuously differentiable on its support with strictly positive derivative f. Let F_n be the empirical distribution of an iid sample from F with size n. Let $\hat{F}_n$ be the bootstrap empirical distribution of an iid sample from F_n with size n. Let $\beta_k(\cdot)$ denote the kth element of the vector-valued coefficient function $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_k(\cdot), \ldots, \beta_K(\cdot))$. Under A12, for all $k = 1, \ldots, K$, conditional on almost all sequences $X_1, X_2, \ldots$, the bootstrap empirical process converges to a Gaussian process,*

$$\sqrt{n}\Big(\beta_k \circ \hat{F}_n^{-1} - \beta_k \circ F_n^{-1}\Big) \rightsquigarrow \beta_k'\big(F^{-1}\big)\frac{\mathbb{G}}{f(F^{-1})}, \text{ in } \ell^\infty(0,1), \text{ as } n \to \infty, \tag{38}$$

where the limit process is the same as in Lemma 18.

Let $\hat{\beta}_k^*(\cdot)$ denote the coefficient estimator as a function of the rank value τ uniformly over $(0,1)$ from the bootstrapped sample. Let $\hat{\beta}_k(\cdot)$ denote the coefficient estimator as a function of rank value τ uniformly over $(0,1)$ from the original sample. Applying Lemma 15 to the standard uniform distribution function, we have the conditional convergence of the bootstrap coefficient function estimator uniformly over all τ on $(0,1)$ in the simplified model.

Theorem 16 (Bootstrap consistency in simplified model)**.** *Under A1–A5 and A12 in the simplified model,*

$$\sqrt{n}\big(\hat{\beta}_k^*(\cdot) - \hat{\beta}_k(\cdot)\big) \mid (X_1, \ldots, X_n) \rightsquigarrow \beta_k'\mathbb{G}, \quad \text{in } \ell^\infty(0,1), \text{ as } n \to \infty,$$

where the limit process is the same as in Theorem 11.

Theorem 16 establishes bootstrap consistency in the simplified model. This provides a suggestive ground for using bootstrap methods in the general model in practice.

6.2.2 Bootstrap confidence interval

In this section, I compute a bootstrap standard error and construct a bootstrap confidence interval for the coefficient estimator at any rank value.

For any fixed rank value τ, we can obtain the coefficient estimates $\hat{\beta}_k(\tau)$, denoted as $\hat{\beta}_\tau^k$, as in Section 4.2.2. For simplicity of notation, I suppress the τ and k, using $\hat{\beta}$ to denote the coefficient estimates at rank value τ for the kth element of the coefficient function.

I use Algorithm 3 of Chernozhukov, Fernández-Val, and Melly (2013) to estimate the standard error based on the bootstrap interquartile range,

$$\widehat{\mathrm{SE}}_{bs} = \frac{\hat{\beta}^*_{0.75} - \hat{\beta}^*_{0.25}}{z_{0.75} - z_{0.25}}, \tag{39}$$

where z_τ denote the τ-quantile of a standard normal distribution, and $\hat{\beta}^*_{0.75}$ and $\hat{\beta}^*_{0.25}$ are the 0.75- and 0.25-quantile of the bootstrapped estimators $\{\hat{\beta}^b, b = 1, \ldots, B\}$ among B bootstrap replications. The bootstrap standard deviation is $\hat{\sigma}_{bs} = \sqrt{n}\widehat{\mathrm{SE}}_{bs}$. As $n \to \infty$, the bootstrapped standard deviation converges in probability to the real-world standard deviation of the $\sqrt{n}$-scaled estimator, i.e., $\hat{\sigma}_{bs} \xrightarrow{p} \sigma$.

Construct the bootstrap confidence interval as

$$[\hat{\beta} - \mathrm{CV}^{bs}_{1-\alpha}\widehat{\mathrm{SE}}_{bs}, \hat{\beta} + \mathrm{CV}^{bs}_{1-\alpha}\widehat{\mathrm{SE}}_{bs}], \tag{40}$$

where $\hat{\beta}$ is the estimator from the given dataset, the standard error estimator $\widehat{\mathrm{SE}}_{bs}$ is derived from (39), and the bootstrap critical value $\mathrm{CV}^{bs}_{1-\alpha}$ is defined as the $1-\alpha$ quantile of the $|\hat{\beta}^b - \hat{\beta}|/\widehat{\mathrm{SE}}_{bs}$ among the B bootstrap replications,

$$\mathrm{CV}^{bs}_{1-\alpha} = (1-\alpha)\text{-quantile of } \left\{|\hat{\beta}^b - \hat{\beta}|/\widehat{\mathrm{SE}}_{bs}, b = 1, \ldots, B\right\}. \tag{41}$$

The bootstrap confidence interval has asymptotic coverage probability $1-\alpha$. From bootstrap consistency, $\sqrt{n}(\hat{\beta}^b - \hat{\beta})$ converges in distribution to the same limit as $\sqrt{n}(\hat{\beta} - \beta)$, conditional on almost all sequences of data. Further, $\hat{\sigma}_{bs} \xrightarrow{p} \sigma_{bs}$. Thus,

$$\frac{\hat{\beta}^b - \hat{\beta}}{\widehat{\mathrm{SE}}_{bs}} = \frac{\sqrt{n}\left(\hat{\beta}^b - \hat{\beta}\right)}{\hat{\sigma}_{bs}}$$

converges to the same limit distribution as $\sqrt{n}(\hat{\beta} - \beta)/\sigma_{bs}$. Based on the uniform asymptotic normality showed in Section 6.1, $\sqrt{n}(\hat{\beta} - \beta)/\sigma_{bs}$ converges to the standard normal distribution. Therefore, the bootstrap critical value $\mathrm{CV}^{bs}_{1-\alpha}$ asymptotically converges the $1-\alpha$ quantile of the absolute value of the standard normal distribution. The bootstrap confidence interval $\hat{\beta} \pm \mathrm{CV}^{bs}_{1-\alpha}\widehat{\mathrm{SE}}_{bs}$ asymptotically converges to the interval $\hat{\beta} \pm \mathrm{CV}^{\mathrm{N}(0,1)}_{1-\alpha}\mathrm{SE}_\beta$.

6.2.3 Bootstrap algorithm:

I draw the bootstrap sample using the panel nonparametric bootstrap method suggested in Hansen (2020). I randomly draw individuals with replacement and keep the corresponding time series for each individual, keeping the time series dependence structure. This procedure essentially comes from the clustered bootstrap method in Cameron, Gelbach, and Miller (2008). It is also a special case of the two-step clustered bootstrapping method in Davison and Hinkley (1997). Davison and Hinkley (1997) suggest drawing individuals with replacement in the first step and resampling the time series data either with or without replacement in the second step. They show that the second-stage drawing without replacement is better than that with replacement.

Given any $\tau \in (0,1)$, we can construct the bootstrap confidence interval for the true coefficient β at this rank value τ.[3] The bootstrap method algorithm is as follows.

Step 1: Given a panel dataset $\{(Y_{it}, \mathbf{X}_{it}),\ i = 1, \dots, n,\ t = 1, \dots, T\}$, compute the original estimator $\hat{\beta}$.

Step 2: Draw the bootstrap sample $\left\{\left(Y_{it}^b, \mathbf{X}_{it}^b\right),\ i = 1, \dots, n,\ t = 1, \dots, T\right\}$. Use the bootstrapped sample to compute the bootstrap estimator $\hat{\beta}^b$.

Step 3: Repeat the last step, bootstrapping for $b = 1, \dots, B$.

Step 4: Compute bootstrap standard error $\widehat{\mathrm{SE}}_{bs} = \dfrac{\hat{\beta}^*_{0.75} - \hat{\beta}^*_{0.25}}{z_{0.75} - z_{0.25}}$.

Step 5: Compute the bootstrap critical value $\mathrm{CV}^{bs}_{1-\alpha}$ as the $(1-\alpha)$-quantile of the Studentized $\left\{|\hat{\beta}^b - \hat{\beta}|/\widehat{\mathrm{SE}}_{bs}, b = 1, \dots, B\right\}$.

Step 6: Compute the bootstrap confidence interval $[\hat{\beta} - \mathrm{CV}^{bs}_{1-\alpha}\widehat{\mathrm{SE}}_{bs}, \hat{\beta} + \mathrm{CV}^{bs}_{1-\alpha}\widehat{\mathrm{SE}}_{bs}]$.

7 Empirical illustration

To apply the proposed method to an empirical example, I study the causal effect of a country's oil wealth on its military defense burden using the data from Cotet and Tsui (2013). There are many studies debating the effect of oil wealth on political conflict. On one hand, resource scarcity triggers conflict, so resource abundance should mitigate regime instability. On the other hand, political instability in the Middle East and some developing countries with rich oil resources makes people arrive at the "oil-fuels-war" conclusion. The political violence in a regime can be measured in multiple ways, for example, the onset of civil war. Additionally, political conflict may not necessarily lead to regime shift or war.

[3] For simplicity of notation, I suppress the τ and denote the estimator as $\hat{\beta}$.

Military defense burden is also considered as a barrier to political conflict. My empirical example studies the effect of oil wealth on military defense spending.

First, I provide point estimates of the effect of oil wealth on military defense spending (as a ratio of GDP) for countries at various rank levels τ. The rank variable represents a country's unobserved general propensity of having military spending. Second, I present bootstrap confidence intervals for the effects. Finally, I show joint confidence sets for the oil wealth effect as a function of unobserved military spending propensity using the Dirichlet method.

Cotet and Tsui (2013) use the standard fixed-effect method to study the effect of oil wealth on political violence. They provide several measures of oil wealth, such as oil reserves or oil discoveries, and several measures related to political violence, such as the onset of civil war or the military defense spending as a ratio of GDP. They find that after controlling for the country fixed effect, oil wealth has no effect on civil conflict onset or military spending. However, the effect is significant among the subsample of non-democratic countries. The estimated effect has opposite signs between the democratic and non-democratic subsamples. Instead of effect heterogeneity varying by the observable democracy variable, I investigate effect heterogeneity across the unobserved military spending rank variable.

Consider the structural model

$$\log(\mathit{DEFENSE}_{it}) = \beta(U_i) \times \log(\mathit{OILWEALTH}_{it}) + \mathbf{X}_{it}'\boldsymbol{\gamma}(U_i) + V_{it}. \tag{42}$$

The outcome variable is measured as (the log of the percentage points of) the ratio of military defense spending to GDP. The variable of interest is $\log(\mathit{OILWEALTH}_{it})$, the log of dollar-valued inflation-adjusted oil wealth per capita. The parameter $\beta(U_i)$ is the causal effect of oil wealth on military spending. The unobserved country rank variable U_i denotes country i's general propensity to have military spending. The effect can differ across countries with different U_i. The covariate vector $\mathbf{X}_{it}$ contains country characteristics including economic growth and population. V_{it} is the idiosyncratic error, which is assumed exogenous so that TS-OLS is consistent. The goal is to estimate $\beta(\tau)$ for various τ and to provide inference on the coefficient function $\beta(\cdot)$.

The data is from Cotet and Tsui (2013). My sample includes 45 countries over years 1988–2003.[4] The military defense burden outcome variable is measured as a log of percentage points. For example, if a country's military spending is 3.2% of its GDP, then its outcome

[4] I clean out observations with missing values on the outcome variable and explanatory variables as well as "stayers" according to A5. The original dataset has 222 countries from 1920–2008. My 45 countries are listed in Appendix B.3 with their time-averaged democracy levels. The sample used in this section is different from that used in Cotet and Tsui (2013) and does not represent the world. However, the sample selection issue is beyond the scope of this paper and left to future work.

variable value is log(3.2). Oil wealth is measured as log oil value per capita.[5] For example, a country with 6 million dollars per capita oil reserve wealth is counted as log(6,000,000) = 6.8 as its $\log(OILWEALTH_{it})$ value. The effect $\beta(U_i)$ measures the elasticity given a 1% increase in oil wealth per capita, i.e., the percent effect on the percentage points of the military spending to GDP ratio.

The two other covariates included in my application are the country's economic growth and population density. Economic growth is measured as the country's annual GDP growth rate. For example, if a country's annual GDP growth is 3%, then the value of economic growth is 0.03. The population density is measured by the log of the country's total population.[6]

Table 3 presents the coefficient estimates at rank values from $\tau = 0.1$ to 0.9 and their bootstrap standard errors, bootstrap critical values, and bootstrap confidence intervals. At $\tau = 0.5$, the coefficient is -0.18, which means for a country who is at the median propensity for military spending, a 1% increase of oil wealth per capita causes a 0.18% decrease in the percentage points of the military spending as a ratio of GDP. The effect is not statistically significant at all rank levels.

Table 3 shows the estimated coefficient function is nonlinear. For the countries who are very likely to have political violence (high τ), if they have extra oil discoveries, it will decrease their military spending ratio, meaning decreasing the military power to prevent civil war or political violence. For the countries who are least prone to military spending (low τ), having extra oil wealth increases their military spending ratio, meaning increasing the barrier of civil war. However, the overall pattern of the coefficient function is not very clear in this example. This is expected due to the limited number of observations, which causes large standard errors and imprecise estimation.

The resampling-based bootstrap critical value and bootstrap standard error converge to their true bootstrap-world limits as the number of bootstrap replications B increases. Table 3 reports their limit values (i.e., for large B) at all rank values. The bootstrap confidence interval is wider than the traditional confidence interval, whose derivation is based on the standard normal distribution. The asymptotic normal distribution might be a poor approximation to capture the finite-sample distribution of the estimator. This helps illustrate the importance of using a bootstrap critical value instead of the traditional critical value 1.64 for building 90% confidence intervals.

Figure 4 illustrates the joint confidence set for the 45 countries' causal effects of oil wealth

[5] Cotet and Tsui (2013) instead use the log of oil wealth per capita *divided by* 100. My coefficient estimates can be multiplied by 100 to compare directly with those of Cotet and Tsui (2013).

[6] Cotet and Tsui (2013) use log of the country's population divided by 100 as the population density.

Table 1: Empirical results for effect of oil wealth on military defense spending.

τ	Slope Estimates	Bootstrap S.E.	Bootstrap Critical Value	90% Bootstrap CI Lower Bound	90% Bootstrap CI Upper Bound
0.1	0.18	0.13	4.41	−0.38	0.75
0.2	−0.43	0.32	2.13	−1.11	0.26
0.3	0.06	0.23	2.15	−0.43	0.54
0.4	−0.14	0.20	1.91	−0.53	0.25
0.5	−0.18	0.25	1.72	−0.60	0.25
0.6	0.12	0.22	2.39	−0.39	0.64
0.7	−0.19	0.22	1.99	−0.64	0.25
0.8	−0.01	0.24	1.82	−0.44	0.42
0.9	−1.01	0.40	3.68	−2.47	0.46
OLS	0.04	0.04	1.64	−0.02	0.11

900 bootstrap replications.

on military defense spending using the Dirichlet method. Each dot in this graph denotes a point estimate for one country. The vertical value of each dot is the country's estimated effect of oil wealth on military spending. I label the highest (0.75) and lowest (−1.11) estimates. The horizontal value of each dot is its corresponding estimated rank variable value. Each dot is contained in a rectangle, which denotes the confidence region for each country's coefficient estimate. The n rectangles form a joint confidence set for all 45 points on the function $\beta(\cdot)$, using a 90% confidence level.

Figure 4 shows that most of the confidence rectangles are relatively informative. However, for a few, the vertical height is very large. This is because the TS-OLS standard error estimates for the corresponding countries are very big, partly from the very limited time dimension ($T = 15$) in this example. With more time series observations, the standard errors would decrease, and the confidence rectangles would become smaller and more precise. In future work, I will pursue grouping similar individuals to reduce the TS-OLS standard error and improve inference.

8 Conclusion

In this paper, I propose a structural quantile-based random coefficient panel model with the goal of discovering heterogeneous causal effects while accounting for endogeneity. I combine the idea of quantile regression and random coefficient models by introducing a time-invariant rank variable and considering the coefficient vector as a function of the rank variable. I develop several theoretical results in both a simplified model and a general long-

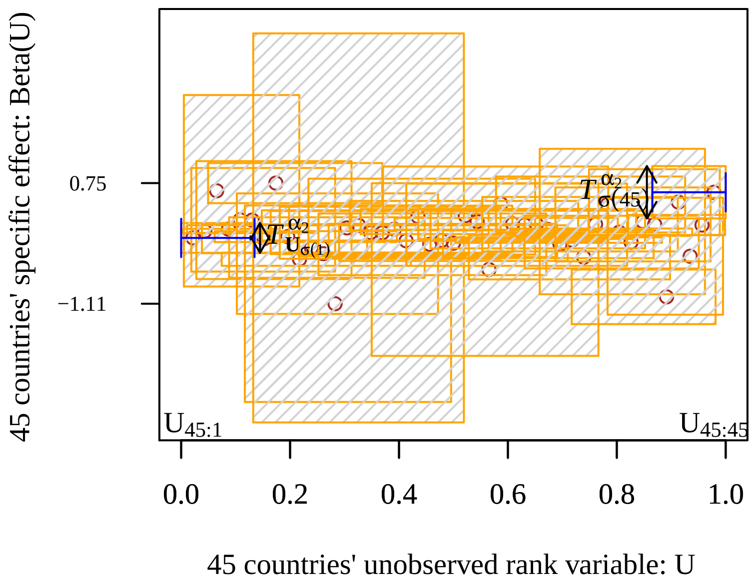


Figure 4: Joint confidence set for $\beta(\cdot)$.

panel model. First, I establish identification and uniformly consistent estimation based on a sorting method. Second, I propose a novel inference method, building confidence sets or uniform confidence bands for the coefficient function based on a Dirichlet approach. Third, I establish the uniform asymptotic normality of the coefficient estimator and its bootstrap consistency. This facilitates a bootstrap confidence interval for the coefficient estimator at each specific rank.

In future work, I plan to study multiple extensions. First, I consider the partial identification. This could occur in a few cases, such as when T is less than K, or some individuals have missing values or lack variation in their time series data. In the case that we cannot obtain point identification, a worst-case bound of the coefficient estimator might help. Second, when the dataset lacks variation, I consider grouping similar individuals to create more variation, hoping to choose groups in an optimal way.

A Proofs

A.1 Proof of Theorem 1: identification in simplified model

Proof. (i) Under A5, without loss of generality assume $T = K$ and thus $\underline{\mathbf{X}}_i$ is a squared matrix. (If $T > K$, simply select K rows, which has a row rank K, out of $\underline{\mathbf{X}}_i$. Let $\underline{\mathbf{X}}_i$ be the new $K \times K$ squared matrix.) We can identify all the individual-specific coefficients from its time dimension variations; with probability one, $\underline{\mathbf{X}}_i^{-1}$ exists, so

$$\boldsymbol{\beta}(U_i) = \underline{\mathbf{X}}_i^{-1}\mathbf{Y}_i. \tag{43}$$

(ii) Based on A3 the monotonicity of $\mathbf{x}^{*\prime}\boldsymbol{\beta}(u)$ in u and A2 the standard uniform distribution of U,

$$\begin{aligned} \mathrm{P}(\mathbf{x}^{*\prime}\boldsymbol{\beta}(U) \leq \mathbf{x}^{*\prime}\boldsymbol{\beta}(\tau)) &= \mathrm{P}(U \leq \tau) \quad \text{By } A3 && (44) \\ &= \tau. \quad \text{By } A2 && (45) \end{aligned}$$

Therefore, $\boldsymbol{\beta}(\tau)$ is a solution to $\mathrm{P}(\mathbf{x}^{*\prime}\boldsymbol{\beta}(U) \leq \mathbf{x}^{*\prime}\boldsymbol{\beta}(\tau)) = \tau$. The uniqueness comes from the fact that $\mathbf{x}^{*}\boldsymbol{\beta}(u)$ is strictly increasing in u from A3 and that $\mathbf{x}^{*}\boldsymbol{\beta}(u)$ is continuous in u from A4. □

A.2 Proof of Theorem 2: identification in general model

Proof. The proof will be the same as the Theorem 1 proof of (ii) under A1–A4. □

A.3 Proof of Theorem 3: uniform consistency in simplified model

Proof. Under A1–A3 and A5 in the simplified model, every individual's coefficient parameter can be perfectly estimated.

$$\hat{\boldsymbol{\beta}}(U_i) = \boldsymbol{\beta}(U_i) = \underline{\mathbf{X}}_i^{-1}\mathbf{Y}_i.$$

Define the estimator $\hat{\boldsymbol{\beta}}(\tau)$ of the true coefficient at any rank $\tau \in (0,1)$ as the $\lceil \tau n \rceil$th order statistic of the n individual coefficients, $\boldsymbol{\beta}(U_1), \boldsymbol{\beta}(U_2), \ldots, \boldsymbol{\beta}(U_n)$,

$$\hat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big).$$

Let $F_U(\cdot)$ denote the distribution function of the standard uniform distributed random variable U. The function $F_U(\cdot)$ is continuous on $[0,1]$. It can be written out explicitly as

$$F_U(u) = \begin{cases} 0, & u \leq 0 \\ u, & 0 \leq u \leq 1 \\ 1, & u \geq 1. \end{cases} \tag{46}$$

Denote the τ-quantile of F_U as

$$Q_U(\tau) = F_U^{-1}(\tau) \equiv \inf\{x : F_U(x) \geq \tau\}. \tag{47}$$

Both the distribution function and the quantile function of the standard uniform distributed random variable U are identity functions on $(0, 1)$.

$$F_U(u) = u, \text{ for all } u \in (0,1); \tag{48}$$

$$Q_U(\tau) = \tau, \text{ for all } \tau \in (0,1). \tag{49}$$

Thus, $F_U = I$ and $Q_U = I$, where I is the identity function. Define the empirical distribution function of a size n sample of the standard uniform distributed $U_1, \ldots, U_n$ as

$$F_{n,U}(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{U_i \le u\}. \tag{50}$$

Define the sample quantile function as the inverse of the empirical distribution function

$$\hat{Q}_U(\tau) \equiv F_{n,U}^{-1}(\tau) = \inf\{x : F_{n,U}(x) \ge \tau\}. \tag{51}$$

It is well known that the uniform sample quantile function is uniformly convergent. (See Shorack and Wellner (1986, Thm. 3, p. 95))

$$\sup_{\tau\in(0,1)} \left|F_{n,U}^{-1}(\tau) - I(\tau)\right| \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty. \tag{52}$$

Moreover, the standard uniform distributed U has property that

$$F_{n,U}^{-1}(\tau) = U_{n:\lceil \tau n \rceil}. \tag{53}$$

Therefore,

$$\sup_{\tau\in(0,1)} \left|U_{n:\lceil \tau n\rceil} - \tau\right| = \sup_{\tau\in(0,1)} \left|F_{n,U}^{-1}(\tau) - I(\tau)\right| \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty. \tag{54}$$

That is, the sequence of functions $g_n(\tau) = U_{n:\lceil \tau n\rceil}$ uniformly converges to function $g(\tau) = \tau$ on $(0, 1)$. Equivalently, with probability 1,

$$\forall \epsilon > 0, \ \exists N_\epsilon \in \mathbb{N} \text{ such that } \ \forall n > N_\epsilon, \ \forall \tau \in (0,1), \ \left|U_{n:\lceil \tau n\rceil} - \tau\right| < \epsilon. \tag{55}$$

The uniform continuity assumption A7 states that for all $k = 1, \ldots, K$, $\forall \epsilon > 0, \exists\ \delta_\epsilon > 0$, such that $\forall x, y \in (0, 1), |x - y| \le \delta_\epsilon \implies |\beta_k(x) - \beta_k(y)| < \epsilon$. Therefore, for all $k = 1, \ldots, K$,

$$\forall \epsilon > 0, \exists N_{\delta_\epsilon} \in \mathbb{N} \text{ such that } \ \forall n > N_{\delta_\epsilon} \text{ and } \ \forall \tau \in (0,1), \ \left|U_{n:\lceil \tau n\rceil} - \tau\right| < \delta_\epsilon, \text{with probability 1,}$$

then

$$\left|\beta_k\big(U_{n:\lceil \tau n\rceil}\big) - \beta_k(\tau)\right| < \epsilon. \tag{56}$$

The sequence of functions $\beta_k\big(U_{n:\lceil \tau n\rceil}\big)$ uniformly converges almost surely to the corresponding function $\beta_k(\tau)$ as n goes to infinity. For all $k = 1, \ldots, K$,

$$\sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| = \sup_{\tau\in(0,1)}\left|\beta_k\big(U_{n:\lceil \tau n\rceil}\big) - \beta_k(\tau)\right| \xrightarrow{a.s.} 0, \text{ as } \ n \to \infty. \tag{57}$$

To incorporate the uniform convergence of all the scalar-valued coefficient element function $\beta_k(\cdot)$ into that of the overall vector-valued coefficient function $\boldsymbol{\beta}(\cdot)$, consider the supremum norm at first.

$$\left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right\| = \left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right\|_\infty = \max_{1\le k\le K}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|.$$

On the one hand,

$$\sup_{\tau\in(0,1)}\max_{1\le k\le K}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| \ge \sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|, \text{ for each } k = 1, \ldots K.$$

Thus,

$$\sup_{\tau\in(0,1)}\max_{1\le k\le K}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| \ge \max_{1\le k\le K}\sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|.$$

On the other hand,

$$\sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| \ge \left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|, \text{ for all } 0 < \tau < 1.$$

Thus,

$$\max_{1\le k\le K}\sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| \ge \left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|, \text{ for all } k = 1, \ldots, K \text{ and } 0 < \tau < 1.$$

Since $\sup_{\tau\in(0,1)}\max_{1\le k\le K}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|$ is the smallest value v such that

$$v \ge \left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|, \text{ for all } k = 1, \ldots, K \text{ and all } 0 < \tau < 1,$$

it should be no greater than other value who satisfy such condition, especially

$$\max_{1\le k\le K}\sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| \ge v = \sup_{\tau\in(0,1)}\max_{1\le k\le K}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|.$$

Altogether,

$$\sup_{\tau\in(0,1)}\max_{1\le k\le K}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right| = \max_{1\le k\le K}\sup_{\tau\in(0,1)}\left|\hat{\beta}_k(\tau) - \beta_k(\tau)\right|, \tag{58}$$

Therefore,

$$
\begin{aligned}
\sup_{\tau\in(0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| &= \sup_{\tau\in(0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\|_\infty \\
&= \sup_{\tau\in(0,1)} \max_{1\le k\le K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right| \\
&= \max_{1\le k\le K} \sup_{\tau\in(0,1)} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right| \\
&= \sup_{\tau\in(0,1)} \left| \hat{\beta}_{k^*}(\tau) - \beta_{k^*}(\tau) \right|, \text{ for some } k^* \text{ in } 1,\ldots,K \\
&\xrightarrow{a.s.} 0, \text{ as } \ n\to\infty, \text{ by } (57)\ .
\end{aligned} \tag{59}
$$

The vector-valued coefficient function estimator uniformly converges to the true coefficient function on $(0,1)$.

The uniform consistency of the vector-valued coefficient function estimator can be generalized to any norm, since the coefficient vector is in a finite dimension and all the norms are equivalent on a finite dimensional space. For example, consider a L_p-norm.

$$
\left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| = \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\|_{L_p} = \left(\sum_{k=1}^{K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right|^p \right)^{1/p} .
$$

$$
\begin{aligned}
\sup_{\tau\in(0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| &= \sup_{\tau\in(0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\|_{L_p} = \sup_{\tau\in(0,1)} \left(\sum_{k=1}^{K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right|^p \right)^{1/p} \\
&\le \sup_{\tau\in(0,1)} \left(K \max_{1\le k\le K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right|^p \right)^{1/p} \\
&= K^{1/p} \sup_{\tau\in(0,1)} \left(\max_{1\le k\le K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right|^p \right)^{1/p} \\
&= K^{1/p} \sup_{\tau\in(0,1)} \left(\left[\max_{1\le k\le K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right| \right]^p \right)^{1/p} \\
&= K^{1/p} \sup_{\tau\in(0,1)} \max_{1\le k\le K} \left| \hat{\beta}_k(\tau) - \beta_k(\tau) \right| \\
&= K^{1/p} \sup_{\tau\in(0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\|_\infty \\
&\xrightarrow{a.s.} 0, \text{ as } \ n\to\infty, \text{ by } (59)\ .
\end{aligned} \tag{60}
$$

Therefore,

$$\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right\| = \sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right\|_{L_p} \xrightarrow{a.s.} 0, \text{ as } \ n\to\infty. \tag{61}$$

The vector-valued coefficient function estimator again uniformly converges to the true coefficient function under L_p norm. □

A.4 Proof of Proposition 4: consistency of the estimated permutation

Proof. Let W denote the event that

$$W=\left\{\frac{\min\{Y^*_{n:2}-Y^*_{n:1},\ldots,Y^*_{n:n}-Y^*_{n:n-1}\}}{2} > \max_{1\le i\le n}\left|\hat{Y}^*_{T,i}-Y^*_i\right|\right\}. \tag{62}$$

In the event, the left hand side random variable $\min\{Y^*_{n:2}-Y^*_{n:1},\ldots,Y^*_{n:n}-Y^*_{n:n-1}\}$ only relates with n rate. The right hand side random variable $\max_{1\le i\le n}\left|\hat{Y}^*_{T,i}-Y^*_i\right|$ relates with both n rate and T rate. We want to show there exists a sequence of small $\epsilon_n>0$ such that

$$\mathrm{P}\left(\frac{\min\{Y^*_{n:2}-Y^*_{n:1},\ldots,Y^*_{n:n}-Y^*_{n:n-1}\}}{2\|\mathbf{x}^*\|} > \epsilon_n > \frac{\left|\hat{Y}^*_{T,i}-Y^*_i\right|}{\|\mathbf{x}^*\|},\ \forall i=1,\ldots,n\right)\to 1 \tag{63}$$

as $n,T\to\infty$ under the n and T rate relation. This is a sufficient but not necessary condition for (24).

Let

$$A=\left\{\frac{\min\{Y^*_{n:2}-Y^*_{n:1},\ldots,Y^*_{n:n}-Y^*_{n:n-1}\}}{2\|\mathbf{x}^*\|} > \epsilon_n\right\} \tag{64}$$

and

$$B=\left\{\epsilon_n > \frac{\left|\hat{Y}^*_{T,i}-Y^*_i\right|}{\|\mathbf{x}^*\|},\ \forall i=1,\ldots,n\right\}. \tag{65}$$

I want to show

$$\mathrm{P}(A)\to 1 \text{ as } n\to\infty,$$

and for small $\delta>0$,

$$\mathrm{P}(B)\to 1 \text{ as } n,T\to\infty \text{ with } n=o\big(T^{\kappa/(3+\delta)}\big).$$

If these both hold, then

$$\mathrm{P}(A \text{ and } B)\ge \mathrm{P}(A)-\mathrm{P}(\bar{B})\to 1,\ \text{ as } n,T\to\infty,\ \text{ with } n=o\big(T^{\kappa/(3+\delta)}\big), \tag{66}$$

where $\bar{B}$ is the complement of B.

If $\epsilon_n = c_0 n^{-2-\delta}$ for small $\delta > 0$, then $\mathrm{P}(A) \to 1$ as $n \to \infty$:

$$\begin{aligned}
1 \geq \mathrm{P}(A) &= \mathrm{P}\big(\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\} > \epsilon_n \cdot 2\|\mathbf{x}^*\|\big) \\
&= 1 - \mathrm{P}\big(\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\} \leq \epsilon_n \cdot 2\|\mathbf{x}^*\|\big) \\
&\geq 1 - \left\{1 - \left[1 - (n+1)\frac{\epsilon_n \cdot 2\|\mathbf{x}^*\|}{L}\right]^n\right\} \quad \text{by (83)} \\
&= [1 - (n+1)\epsilon_n C]^n, \quad \text{for some positive constant } C = 2\|\mathbf{x}^*\|/L, \\
&= \left[1 - (n+1)n^{-2-\delta} C_1\right]^n \quad \text{for constant } C_1 \equiv c_0 C \\
&\to 1 \text{ as } n \to \infty \text{ by Proposition 17.}
\end{aligned} \tag{67}$$

Notice

$$\begin{aligned}
\mathrm{P}(B) &= \mathrm{P}\left(\|\mathbf{x}^*\| c_0 n^{-2-\delta} > \max_{1\leq i\leq n}\left|\hat{Y}^*_{T,i} - Y^*_i\right|\right) \\
&\geq \mathrm{P}\left(\|\mathbf{x}^*\| c_0 n^{-2-\delta} > \|\mathbf{x}^{*\prime}\| \max_{1\leq i\leq n}\left\|\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\right\|\right) \quad \text{by (73)} \\
&= \mathrm{P}\left(c_0 n^{-2-\delta} > \max_{1\leq i\leq n}\left\|\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\right\|\right)
\end{aligned}$$

It remains to show $\mathrm{P}(B) \to 1$ as $n, T \to \infty$ under the n/T rate relation. It is sufficient to show $\max_{1\leq i\leq n}\|\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\| = o_p(n^{-2-\delta})$ as $n, T \to \infty$ under the n/T rate relation.

Let $Z_i \equiv \|\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\|$. By A10 and the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n Z_i = T^{-\kappa}\left(\frac{1}{n}\sum_{i=1}^n T^\kappa Z_i\right) = T^{-\kappa}[\mathrm{E}(T^\kappa Z_i) + o_p(1)] = O_p\big(T^{-\kappa}\big).$$

Thus,

$$\max_{1\leq i\leq n} Z_i \leq \sum_{i=1}^n Z_i = n\frac{1}{n}\sum_{i=1}^n Z_i = nO_p(T^{-\kappa}) = O_p(nT^{-\kappa}).$$

The n/T rate relation $n = o\big(T^{\kappa/(3+\delta)}\big)$ implies $n^{3+\delta}/T^\kappa \to 0$ as $n, T \to \infty$, or equivalently $T^{-\kappa} = o(n^{-3-\delta})$ as $n, T \to \infty$. Therefore, $\max_{1\leq i\leq n}\|\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\| = O_p(nT^{-\kappa}) = o_p(n^{-2-\delta})$. □

Proposition 17 proves a property which is used in the proof of Proposition 4.

Proposition 17. *For some small positive $\delta > 0$,*

(i)

$$\left[1 - (n+1)\frac{2\|\mathbf{x}^*\|}{L} c_0 n^{-3-\delta}\right]^{n^2} \to 1, \ \textit{as } n \to \infty; \tag{68}$$

(ii)

$$\left[1-(n+1)\frac{2\|\mathbf{x}^*\|}{L}c_0 n^{-3}\right]^{n^2} \to e^{-C}, \ \ as\ n \to \infty; \tag{69}$$

(iii)

$$\left[1-(n+1)\frac{2\|\mathbf{x}^*\|}{L}c_0 n^{-3+\delta}\right]^{n^2} \to 0, \ \ as\ n \to \infty, \tag{70}$$

where c_0 is a constant, and $C = \frac{2\|\mathbf{x}^*\|}{L}c_0$ is a constant.

Proof. We prove (i) here. The proof of (ii) and (iii) can be obtained in a similar way. The proof of (ii) and (iii) are identical to (i) until the last equality of computing $\lim_{n\to\infty} \ln a$, at which the results differ depending on $\delta > 0, \delta = 0$, or $\delta < 0$ in the derivation of (i).

Let

$$a = \left[1-(n+1)\frac{2\|\mathbf{x}^*\|}{L}c_0 n^{-3-\delta}\right]^{n^2}, \tag{71}$$

then

$$\ln a = n^2 \ln\big[1-(n+1)Cn^{-3-\delta}\big], \tag{72}$$

where $C = (2\|\mathbf{x}^*\|/L)c_0$ is some constant. The limit of $\ln a$ is

$$\begin{aligned}
\lim_{n\to\infty} \ln a &= \lim_{n\to\infty} \frac{\ln\big[1-(n+1)Cn^{-3-\delta}\big]}{n^{-2}}, \\
&= \lim_{n\to\infty} \frac{-\frac{1}{1-(n+1)Cn^{-3-\delta}}(-C)\big[(2+\delta)n^{-3-\delta}+(3+\delta)n^{-4-\delta}\big]}{-2n^{-3}} \quad \text{by L'Hospital's Rule} \\
&= \lim_{n\to\infty} \frac{-\frac{C}{2}\frac{1}{n^{\delta}}\big[2+\delta+\frac{3+\delta}{n}\big]}{1-\frac{C(n+1)}{n^{3+\delta}}} \\
&= 0.
\end{aligned}$$

The last equality is because $\frac{1}{n^{\delta}} \to 0$, $\big[2+\delta+\frac{3+\delta}{n}\big] \to 2+\delta$, and $\Big[1-\frac{C(n+1)}{n^{3+\delta}}\Big] \to 1$, as $n \to \infty$. Therefore, by the continuous mapping theorem,

$$\lim_{n\to\infty} a = \lim_{n\to\infty} e^{\ln a} = e^{\lim_{n\to\infty} \ln a} = e^0 = 1. \qquad \square$$

A.5 Proof of Theorem 5: uniform consistency in general model

Proof. **Main idea of proof**

To obtain the consistency of the coefficient estimator at any rank in the large-T large-n general model, it requires to ensure two things. One is that in the fixed-n large-T setting, the estimated n individual coefficients should be consistent via its time-dimension observations. The other is that the ordering of the predicted outcome values $\hat{Y}^*$ should correctly represent the ordering of the unobserved rank variable values.

The first requirement is met by Assumption A8. The second requirement is a non-trivial work to show. Let Y^* denote the true outcome values at $\mathbf{X} = \mathbf{x}^*$ computed using the unknown true coefficient. Let $\hat{Y}^*$ denote the fitted outcome values at $\mathbf{X} = \mathbf{x}^*$ computed using the estimated individual coefficients. Assumption A3 indicates the ordering of outcome values Y^* can represent the ordering of unobserved rank variable U. However, we can only obtain the fitted outcome values $\hat{Y}^*$ instead of the true outcome values Y^* in the general model. Since the estimated individual coefficient has estimation error in the general model, the ordering of the fitted outcome values $\hat{Y}^*$ might be different from the ordering of true outcome values Y^*. If so, the ordering of $\hat{Y}^*$ cannot represent the ordering of rank variable U.

The idea is to show the fitted outcome value $\hat{Y}^*$ is close enough to the true outcome value Y^*. Technically, we hope to show the fitted outcome value $\hat{Y}^*$ is within some ϵ-ball of the true value Y^*, with radius less than the smallest value of all $Y^*_{n:k+1} - Y^*_{n:k}$, $k = 1, \ldots, n-1$. The ϵ-ball between $\hat{Y}^*$ and Y^* measures the time-dimension estimation error and it is related with the convergence rate of T. The $\{Y^*_{n:k+1} - Y^*_{n:k}\}_{k=1}^{n-1}$'s are random variables with respect to the individual dimension n and they are related with the convergence rate of n. We hope to find certain rate relation between T and n, at which the ϵ-ball shrinks faster than the smallest value of all $Y^*_{n:k+1} - Y^*_{n:k}$, $k = 1, \ldots, n-1$ does in order to guarantee the order of fitted outcome values $\hat{Y}^*$ represent the order of the true outcome values Y^* and the order of the unobserved rank variable values U.

A.5.1 Time dimension

Let $\hat{\boldsymbol{\beta}}_T(U_i)$ denote the TS-OLS estimator computed using its T time periods data for individual i. From A8 (i), we know $\forall\ \epsilon > 0$, with probability approaching 1,

$$\begin{aligned}
\left|\hat{Y}^*_{T,i} - Y^*_i\right| &= \left|\mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}_T(U_i) - \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_i)\right| \\
&= \left|\mathbf{x}^{*\prime}\left(\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\right)\right| \\
&\leq \|\mathbf{x}^{*\prime}\| \left\|\hat{\boldsymbol{\beta}}_T(U_i) - \boldsymbol{\beta}(U_i)\right\| \text{ by the Cauchy} - \text{Schwarz inequality} \qquad (73) \\
&< \|\mathbf{x}^{*\prime}\|\epsilon, \qquad (74)
\end{aligned}$$

where $\|\cdot\|$ denotes the L_2 norm. That is, for any vector $\mathbf{a} = (a_1, \ldots, a_K)' \in \mathbb{R}^K$, $\|\mathbf{a}\| = \left(\sum_{i=1}^K a_i^2\right)^{1/2}$.

A.5.2 Individual dimension

Consider the individual dimension random variable

$$\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}.$$

Each element $Y^*_{n:k+1} - Y^*_{n:k}$, $k = 1, \ldots, n-1$, is a random variable.

$$Y^*_{n:k+1} - Y^*_{n:k} = \mathbf{x}^{*\prime}(\boldsymbol{\beta}(U_{n:k+1}) - \boldsymbol{\beta}(U_{n:k})), k = 1, \ldots, n-1. \tag{75}$$

Take the first order Taylor expansion

$$\boldsymbol{\beta}(U_{n:k+1}) - \boldsymbol{\beta}(U_{n:k}) = \boldsymbol{\beta}'(\tilde{u}) \underbrace{(U_{n:k+1} - U_{n:k})}_{\sim \text{Beta}(1,n)}, \tag{76}$$

where $\boldsymbol{\beta}'(\tilde{u})$ is a $K \times 1$ vector for some $0 < \tilde{u} < 1$. Since $(U_{n:k+1} - U_{n:k}) \xrightarrow{p} 0$, each random variable $(Y^*_{n:k+1} - Y^*_{n:k}) \xrightarrow{p} 0$, as $n \to \infty$. Under A9, $\mathbf{x}^{*\prime}\boldsymbol{\beta}'(\tilde{u})$ has a positive lower bound L. Therefore,

$$\begin{aligned} Y^*_{n:k+1} - Y^*_{n:k} &= \mathbf{x}^{*\prime}(\boldsymbol{\beta}(U_{n:k+1}) - \boldsymbol{\beta}(U_{n:k})) \\ &= \underbrace{\mathbf{x}^{*\prime}\boldsymbol{\beta}'(\tilde{u})}_{\geq L \ \text{by} A9}(U_{n:k+1} - U_{n:k}) \\ &\geq L(U_{n:k+1} - U_{n:k}). \end{aligned} \tag{77}$$

Furthermore,

$$\mathrm{P}\big(Y^*_{n:k+1} - Y^*_{n:k} \leq x\big) \leq \mathrm{P}(L(U_{n:k+1} - U_{n:k}) \leq x) = \mathrm{P}\Big(U_{n:k+1} - U_{n:k} \leq \frac{x}{L}\Big), \tag{78}$$

and

$$\begin{aligned} \mathrm{P}\big(\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\} \leq x\big) &= \mathrm{P}\big(\cup_{k=1}^{n-1}\{Y^*_{n:k+1} - Y^*_{n:k} \leq x\}\big) \\ &\leq \mathrm{P}\Big(\cup_{k=1}^{n-1}\Big\{U_{n:k+1} - U_{n:k} \leq \frac{x}{L}\Big\}\Big) \\ &= \mathrm{P}\Big(\min\{U_{n:2} - U_{n:1}, \ldots, U_{n:n} - U_{n:n-1}\} \leq \frac{x}{L}\Big). \end{aligned}$$

The order of convergence for random variable $\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}$ is bounded by the order of convergence for random variable $\min\{U_{n:2} - U_{n:1}, \ldots, U_{n:n} - U_{n:n-1}\}$.

Let $D_k = U_{n:k+1} - U_{n:k}$, $k = 1, \ldots, n-1$. Let $D_0 = U_{n:1}$ and $D_n = 1 - U_{n:n}$. We know from equation (10) that $D_0, \ldots, D_n$ jointly follows the Dirichlet distribution with $n+1$

parameters 1.

$$(D_0, \dots, D_n) \sim \mathrm{Dir}\,\underbrace{(1, \dots, 1)}_{n+1}.$$

Then we can use the distribution function of $\min\{D_0, \dots, D_n\}$ as the upper bound for the distribution function of the random variable $\min\{D_1, \dots, D_{n-1}\}$

$$\mathrm{P}(\min\{D_1, \dots, D_{n-1}\} \le x) \le \mathrm{P}(\min\{D_0, \dots, D_n\} \le x). \tag{79}$$

We can derive the distribution function of $\min\{D_0, \dots, D_n\}$, that for any $x \in \left[0, \frac{1}{n+1}\right]$,

$$\begin{aligned}\mathrm{P}(\min\{D_0, \dots, D_n\} \le x) &= 1 - \mathrm{P}(D_0 > x, \text{ and } \dots, \text{ and } D_n > x)\\ &= 1 - [1 - (n+1)x]^n.\end{aligned} \tag{80}$$

$\mathrm{P}(\min\{D_0, \dots, D_n\} \le x) = 0$, for $x \le 0$; $\mathrm{P}(\min\{D_0, \dots, D_n\} \le x) = 1$, for $x \ge \frac{1}{n+1}$.

The last equality in (80) is based on the fact that

$$\begin{aligned}&\mathrm{P}(D_0 > x, \text{ and } \dots, \text{ and } D_n > x)\\ &= \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_x^{1-x-d_0-d_1-\dots-d_{n-2}} f_{\mathrm{Dir}}\left(d_0, d_1, \dots, d_{n-1}, 1 - \sum_{k=0}^{n-1} d_k\right) dd_{n-1} \cdots dd_0\\ &= \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_x^{1-x-d_0-d_1-\dots-d_{n-2}} n! \;\, dd_{n-1} \cdots dd_0\\ &= n! \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_x^{1-x-d_0-d_1-\dots-d_{n-2}} 1 \;\, dd_{n-1} \cdots dd_0\\ &= n! \frac{[1-(n+1)x]^n}{n!} = [1-(n+1)x]^n.\end{aligned} \tag{81}$$

where $f_{\mathrm{Dir}}\big(d_0, d_1, \dots, d_{n-1}, 1 - \sum_{k=0}^{n-1} d_k\big)$ denote the Dirichlet density function with $(n+1)$ parameter $(1, \dots, 1)$.

$$f_{\mathrm{Dir}}\left(d_0, d_1, \dots, d_{n-1}, 1 - \sum_{k=0}^{n-1} d_k; \underbrace{1, \dots, 1}_{n+1}\right) = \frac{\left[\prod_{k=0}^{n-1} d_k^{1-1}\right](1 - \sum_{k=0}^{n-1} d_k)^{1-1}}{\left[\frac{\prod_{k=0}^{n} \Gamma(1)}{\Gamma(n+1)}\right]} = \Gamma(n+1) = n!, \tag{82}$$

and the gamma function with any positive integer n is $\Gamma(n) = (n-1)!$.

The second to last equality in (81) is based on a calculus derivation using changing of variables.

$$\int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_x^{1-2x-d_0-d_1-\dots-d_{n-3}} \int_x^{1-x-d_0-d_1-\dots-d_{n-2}} 1 \;\, dd_{n-1}\, dd_{n-2} \cdots dd_0$$

$$
\begin{aligned}
&= \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \\
&\qquad \int_x^{1-3x-d_0-d_1-\ldots-d_{n-4}} \int_x^{1-2x-d_0-d_1-\ldots-d_{n-3}} (1-2x-d_0-d_1-\ldots-d_{n-2})\ dd_{n-2}\, dd_{n-3} \cdots dd_0 \\
&\overset{\text{Let } s_2=1-2x-d_0-d_1-\ldots-d_{n-2}}{=} - \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \\
&\qquad \int_x^{1-3x-d_0-d_1-\ldots-d_{n-4}} \int_{1-3x-d_0-d_1-\ldots-d_{n-3}}^{0} s_2\ ds_2\, dd_{n-3} \cdots dd_0 \\
&= \frac{1}{2} \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_x^{1-3x-d_0-d_1-\ldots-d_{n-4}} (1-3x-d_0-d_1-\ldots-d_{n-3})^2\, dd_{n-3} \cdots dd_0 \\
&\overset{\text{Let } s_3=1-3x-d_0-d_1-\ldots-d_{n-3}}{=} -\frac{1}{2} \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_{1-4x-d_0-d_1-\ldots-d_{n-4}}^{0} s_3^2\, ds_3\, dd_{n-4} \cdots dd_0 \\
&= \frac{1}{2\cdot 3} \int_x^{1-nx} \int_x^{1-(n-1)x-d_0} \int_x^{1-(n-2)x-d_0-d_1} \cdots \int_x^{1-4x-d_0-d_1-\ldots-d_{n-5}} (1-4x-d_0-d_1-\ldots-d_{n-4})^3\, dd_{n-4} \cdots dd_0 \\
&= \ldots \\
&= \frac{1}{(n-1)!} \int_x^{1-nx} (1-nx-d_0)^{n-1}\, dd_0 \\
&\overset{\text{Let } s=1-nx-d_0}{=} -\frac{1}{(n-1)!} \int_{1-(n+1)x}^{0} s^{n-1}\, ds \\
&= \frac{1}{n!}\, s^n \big|_0^{1-(n+1)x} \\
&= \frac{[1-(n+1)x]^n}{n!}.
\end{aligned}
$$

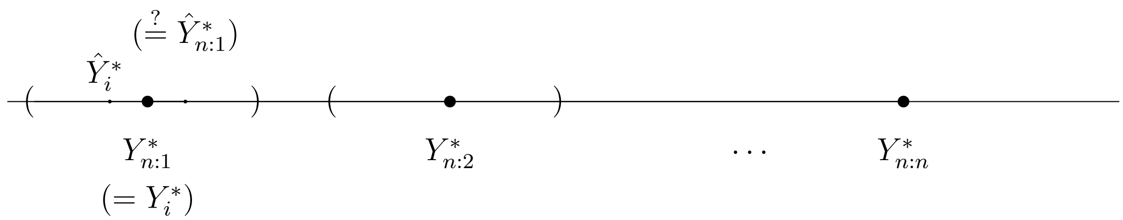
For the special case $x = 0$,

$$
\int_0^1 \int_0^{1-d_0} \int_0^{1-d_0-d_1} \cdots \int_0^{1-d_0-d_1-\ldots-d_{n-2}} f_{\text{Dir}}\left(d_0, d_1, \ldots, d_{n-1}, 1 - \sum_{k=0}^{n-1} d_k\right) dd_{n-1} \cdots dd_0 = n! \frac{1}{n!} = 1.
$$

Altogether, we can derive the upper bound for the distribution function of the random variable $\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}$. For $x \in \left[0, \frac{L}{n+1}\right]$,

$$
\begin{aligned}
\mathrm{P}\Big(\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\} \le x\Big) &\le \mathrm{P}\left(\min\{U_{n:2} - U_{n:1}, \ldots, U_{n:n} - U_{n:n-1}\} \le \frac{x}{L}\right) \\
&\le 1 - [1-(n+1)x/L]^n. \qquad (83)
\end{aligned}
$$

A.5.3 Correct ordering under rate relation



Proposition 4 implies that when n and T are sufficiently large with the rate relation $n = o(T^{\kappa/(3+\delta)})$, the order of fitted outcome values $\hat{Y}^*$ can represent the order of the true outcome values Y^*, which further represents the order of the unobserved rank variable values U. Therefore,

$$\hat{Y}^*_{n:k} = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}(U_{n:k}), k = 1, \ldots, n. \tag{84}$$

The order of the fitted outcome values indicates the order of the rank variable values. Therefore, the estimator $\hat{\boldsymbol{\beta}}(\tau)$, defined as the coefficient estimates associated with the $\lceil n\tau \rceil$th order statistic of the fitted outcome values, is actually the coefficient estimates for the person with the $\lceil n\tau \rceil-$th order statistic of rank variable.

$$\hat{\boldsymbol{\beta}}(\tau) = \hat{\boldsymbol{\beta}}\big(U_{n:\lceil n\tau \rceil}\big). \tag{85}$$

A.5.4 Consistency

Next we are going to show the estimator is uniform consistent in probability under the n/T rate relation. That is to show, as $n \to \infty, T \to \infty$, and $n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big)$, for any $\eta > 0$,

$$\mathrm{P}\left(\sup_{\tau \in (0,1)} \left\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| > \eta \right) \to 0. \tag{86}$$

Let A denote the event which will result in the correct ordering. (It is the same as event W previously.)

$$\begin{aligned} A = \Bigg\{ &\frac{\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}}{2} > \left|\hat{Y}^*_{T,1} - Y^*_1\right|, \text{ and } \ldots, \\ &\ldots, \text{ and } \frac{\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}}{2} > \left|\hat{Y}^*_{T,n} - Y^*_n\right| \Bigg\} \\ = \Bigg\{ &\frac{\min\{Y^*_{n:2} - Y^*_{n:1}, \ldots, Y^*_{n:n} - Y^*_{n:n-1}\}}{2} > \max_{1 \le i \le n} \left|\hat{Y}^*_{T,i} - Y^*_i\right| \Bigg\}. \end{aligned} \tag{87}$$

From Proposition 4, we know under the n/T rate relation, the event A eventually occurs with probability 1, i.e., as $n, T \to \infty$, with $n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big)$,

$$\mathrm{P}(A) \to 1. \tag{88}$$

Let $\bar{A}$ denote the complement of event A. Then $\mathrm{P}(\bar{A}) \to 0$, as $n, T \to \infty$ with $n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big)$.

For any $\eta > 0$,

$$\mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right\| > \eta\right) \le \underbrace{\mathrm{P}(\bar{A})}_{\to 0} + \underbrace{\mathrm{P}(A)}_{\to 1}\mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right\| > \eta \mid A\right). \tag{89}$$

In Appendix A.5.3 we showed under event A, the order of fitted outcome values $\hat{Y}^*$ correctly represents the order of unobserved rank variable. Then

$$\begin{aligned}
0 &\le \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right\| > \eta \mid A\right)\\
&= \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(\tau)\right\| > \eta \mid A\right)\\
&= \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil}) + \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(\tau)\right\| > \eta \mid A\right)\\
&\le \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| + \sup_{\tau\in(0,1)}\left\|\boldsymbol{\beta}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(\tau)\right\| > \eta \mid A\right)\\
&= \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\boldsymbol{\beta}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(\tau)\right\| > \eta - \sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| \mid A\right)\\
&= \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| \ge \eta \mid A\right)\\
&\quad + \mathrm{P}\Bigg(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| < \eta \ \text{ and}\\
&\qquad\qquad \sup_{\tau\in(0,1)}\left\|\boldsymbol{\beta}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(\tau)\right\| > \eta - \sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| \mid A\Bigg)\\
&\le \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| \ge \eta \mid A\right)\\
&\quad + \mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\boldsymbol{\beta}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(\tau)\right\| > \eta - \sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| > 0 \mid A\right).
\end{aligned} \tag{90}$$

Let K denote the event $K = \left\{\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(U_{n:\lceil\tau n\rceil}) - \boldsymbol{\beta}(U_{n:\lceil\tau n\rceil})\right\| \ge \eta\right\}$. A8 (iii) assumes

$\mathrm{P}(K) \to 0$, as $n \to \infty, T \to \infty$. Then,

$$\mathrm{P}(K \mid A) = \frac{\mathrm{P}(K \text{ and } A)}{\mathrm{P}(A)} \leq \frac{\overbrace{\mathrm{P}(K)}^{\to 0}}{\underbrace{\mathrm{P}(A)}_{\to 1}} \to 0, \text{ as } n \to \infty, T \to \infty, \text{ with } n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big). \tag{91}$$

Let $\gamma = \eta - \sup_{\tau \in (0,1)} \Big\| \hat{\boldsymbol{\beta}}\big(U_{n:\lceil \tau n \rceil}\big) - \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big) \Big\| > 0$. Denote the event

$$E = \left\{ \sup_{\tau \in (0,1)} \big\| \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big) - \boldsymbol{\beta}(\tau) \big\| > \gamma \right\}.$$

From the proof of Theorem 3 (see (57)), we know

$$\mathrm{P}(E) \to 0, \quad \text{as } n \to \infty. \tag{92}$$

The equality $\mathrm{P}(E) = \mathrm{P}(E \mid A)\,\mathrm{P}(A) + \mathrm{P}\big(E \mid \bar{A}\big)\,\mathrm{P}\big(\bar{A}\big)$ implies

$$\mathrm{P}(E \mid A) = \frac{\mathrm{P}(E) - \mathrm{P}\big(E \mid \bar{A}\big)\,\mathrm{P}\big(\bar{A}\big)}{\mathrm{P}(A)}. \tag{93}$$

Since $\mathrm{P}(E) \to 0$, $\mathrm{P}(A) \to 1$, and $\mathrm{P}\big(\bar{A}\big) \to 0$ under the n/T rate relation,

$$\mathrm{P}(E \mid A) \to 0, \quad \text{as } n \to \infty, T \to \infty, \text{ with } n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big). \tag{94}$$

Therefore,

$$\begin{aligned}
0 &\leq \mathrm{P}\left(\sup_{\tau \in (0,1)} \Big\| \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \Big\| > \eta \mid A \right) \\
&\leq 1 \cdot \underbrace{\mathrm{P}\left(\sup_{\tau \in (0,1)} \Big\| \hat{\boldsymbol{\beta}}\big(U_{n:\lceil \tau n \rceil}\big) - \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big) \Big\| \geq \eta \mid A \right)}_{= \mathrm{P}(K \mid A) \to 0, \text{ as } n \to \infty, T \to \infty, \text{ with } n = o\left(T^{\frac{\kappa}{3+\delta}}\right) \text{ by (91)}} \\
&\quad + \underbrace{\mathrm{P}\left(\sup_{\tau \in (0,1)} \big\| \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big) - \boldsymbol{\beta}(\tau) \big\| > \eta - \sup_{\tau \in (0,1)} \Big\| \hat{\boldsymbol{\beta}}\big(U_{n:\lceil \tau n \rceil}\big) - \boldsymbol{\beta}\big(U_{n:\lceil \tau n \rceil}\big) \Big\| > 0 \mid A \right)}_{= \mathrm{P}(E \mid A) \to 0, \text{ as } n \to \infty, T \to \infty, \text{ with } n = o\left(T^{\frac{\kappa}{3+\delta}}\right) \text{ by (94)}} \\
&\to 0.
\end{aligned} \tag{95}$$

Thus,

$$\mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right\|>\eta\mid A\right)\to 0,\ \text{as } n\to\infty, T\to\infty,\ \text{with } n=o\left(T^{\frac{\kappa}{3+\delta}}\right). \tag{96}$$

And altogether in (89)

$$\begin{aligned}\mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right\|>\eta\right) &\le \underbrace{\mathrm{P}(\bar{A})}_{\to 0}+\underbrace{\mathrm{P}(A)}_{\to 1}\underbrace{\mathrm{P}\left(\sup_{\tau\in(0,1)}\left\|\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right\|>\eta\mid A\right)}_{\to 0}\\ &\to 0. \end{aligned}\tag{97}$$

Therefore, the coefficient function estimator is uniformly consistent to the true coefficient function on $(0,1)$ as $n\to\infty$, $T\to\infty$, with n and T rate relation $n=o(T^{\kappa/(3+\delta)})$. □

A.6 Proof of Theorem 6 : Dirichlet method in simplified model—confidence set of generic coefficient function

Proof.

$$\begin{aligned}&\mathrm{P}\Big(\{(U_{n:1},\beta(U_{n:1})),\ldots,(U_{n:n},\beta(U_{n:n}))\}\subseteq\Big\{\cup_{k=1}^{n}\tilde{\mathcal{I}}_k\Big\}\Big)\\ &=\mathrm{P}\Big(\{(U_{n:1},\beta(U_{n:1})),\ldots,(U_{n:n},\beta(U_{n:n}))\}\subseteq\Big\{\cup_{k=1}^{n}\{(x,y)\mid\tilde{a}_k\le x\le\tilde{b}_k;y=\beta(U_{n:k})\}\Big\}\Big)\\ &\ge\mathrm{P}\Big((U_{n:1},\beta(U_{n:1}))\in\tilde{\mathcal{I}}_1,\ \text{and}\ \ldots,(U_{n:n},\beta(U_{n:n}))\in\tilde{\mathcal{I}}_n\Big)\\ &=\mathrm{P}\Big(U_{n:1}\in(\tilde{a}_1,\tilde{b}_1),\ \text{and}\ ,\ldots,\ \text{and}\ U_{n:n}\in(\tilde{a}_n,\tilde{b}_n)\Big)\\ &=1-\alpha\ \text{by (26)}.\end{aligned}$$

The equality establishes when the n individual coefficients $\beta(U_{n:1}),\ldots,\beta(U_{n:n})$ are distinct. □

A.7 Proof of Corollary 7: Confidence set for vector-valued generic coefficient function in simplified model

Proof. Consider each element of the vector-valued coefficient function, $\beta_l(\cdot)$, for $l=1,\ldots,K$. From Theorem 6,

$$1-\alpha\le\mathrm{P}\Big(\{\cup_{k=1}^{n}(U_{n:k},\beta_l(U_{n:k}))\}\subseteq\Big\{\cup_{k=1}^{n}\tilde{\mathcal{I}}_k^l\Big\}\Big). \tag{98}$$

Since the uncertainty only comes from the $U_{n:1}, \ldots, U_{n:n}$, the n points $\cup_{k=1}^{n}(U_{n:k}, \beta_l(U_{n:k}))$ on first element function is in the confidence set $\cup_{k=1}^{n}\tilde{\mathcal{I}}_k^1$ if and only if the n points $\cup_{k=1}^{n}(U_{n:k}, \beta_2(U_{n:k}))$ on second element function is in the confidence set $\cup_{k=1}^{n}\tilde{\mathcal{I}}_k^2$ etc. if and only if the n points $\cup_{k=1}^{n}(U_{n:k}, \beta_l(U_{n:k}))$ on lth element function is in the confidence set $\cup_{k=1}^{n}\tilde{\mathcal{I}}_k^l$. Let event A_l denote $A_l = \left\{\cup_{k=1}^{n}(U_{n:k}, \beta_l(U_{n:k})) \subseteq \cup_{k=1}^{n}\tilde{\mathcal{I}}_k^l\right\}$. Therefore, $A_1 \Leftrightarrow A_2 \Leftrightarrow \ldots \Leftrightarrow A_K$. Then

$$\begin{aligned} \mathrm{P}\Big(\{\cup_{k=1}^{n}(U_{n:k}, \boldsymbol{\beta}(U_{n:k}))\} \subseteq \left\{\cup_{k=1}^{n}\tilde{\boldsymbol{\mathcal{I}}}_k\right\}\Big) &= \mathrm{P}(A_1 \text{ and } A_2, \ldots, \text{ and } A_K) \\ &= \mathrm{P}(A_1) \\ &\geq 1-\alpha. \end{aligned}$$

This holds with equality when the n individual coefficients $\beta_k(U_{n:1}), \ldots, \beta_k(U_{n:n})$ are distinct for each of the element coefficients $k = 1, \ldots, K$. □

A.8 Proof of Theorem 8: Dirichlet method in simplified model—uniform confidence band of monotone coefficient function

Proof. Under A1–A5, with a strict monotone coefficient function, Theorem 6 shows

$$1-\alpha = \mathrm{P}\Big(\{\cup_{k=1}^{n}(U_{n:k}, \beta(U_{n:k}))\} \subseteq \left\{\cup_{k=1}^{n}\tilde{\mathcal{I}}_k\right\}\Big).$$

Let event A denote

$$A = \left\{\{\cup_{k=1}^{n}(U_{n:k}, \beta(U_{n:k}))\} \subseteq \left\{\cup_{k=1}^{n}\tilde{\mathcal{I}}_k\right\}\right\}. \tag{99}$$

We know

$$\begin{aligned} A &= \left\{\{(U_{n:1}, \beta(U_{n:1})), \ldots, (U_{n:n}, \beta(U_{n:n}))\} \subseteq \left\{\cup_{k=1}^{n}\{(x,y) \mid \tilde{a}_k \leq x \leq \tilde{b}_k; y = \beta(U_{n:k})\}\right\}\right\} \\ &= \left\{(U_{n:1}, \beta(U_{n:1})) \in \tilde{\mathcal{I}}_1, \text{ and } \ldots, \text{ and } (U_{n:n}, \beta(U_{n:n})) \in \tilde{\mathcal{I}}_n\right\}. \end{aligned} \tag{100}$$

Under the additional assumption that the coefficient function is monotone, we know the event A implies the event $\{\{(u,y) : y = \beta(u), 0 \leq u \leq 1\} \subseteq \mathcal{S}\}$. Conversely, suppose the coefficient function is all within $\mathcal{S}$, from the fact that the rectangle S_k and S_{k+1} overlap on the interval $\tilde{\mathcal{I}}_{k+1}$, there is no way that the coefficient function is inside the area $\mathcal{S}$ but outside the interval $\tilde{\mathcal{I}}_{k+1}$. And in the simplified model, all the coefficients are perfectly observed. That implies the $(U_{n:k+1}, \beta(U_{n:k+1}))$ is inside the interval $\tilde{\mathcal{I}}_{k+1}$.

Altogether, under the monotone coefficient function condition, the event A and the event

$\{\{(u,y) : y = \beta(u), 0 \le u \le 1\} \subseteq \mathcal{S}\}$ are equivalent. Therefore,

$$\mathrm{P}(\{(u,y) : y = \beta(u), 0 \le u \le 1\} \subseteq \mathcal{S}) = \mathrm{P}(A) = 1 - \alpha. \tag{101}$$

The area $\mathcal{S}$ has exact coverage probability $1-\alpha$ for the coefficient function $\beta(\cdot)$. □

A.9 Proof of Theorem 9: Dirichlet method in general model—pointwise confidence set of individual coefficient on the generic coefficient function in fixed-n, large-T setting

Proof. Consider an individual whose individual rank variable is the kth smallest value in the size n sample. That is, the individual has rank variable $U_{n:k}$. Let A and B denote the event $A = \{U_{n:k} \in [a_k^{\alpha_1}, b_k^{\alpha_1}]\}$ and $B = \left\{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(k)}^{\alpha_2}\right\}$. We know from (25) that $\mathrm{P}(A) = 1 - \alpha_1$ and from the A11 that $\lim_{T\to\infty} \mathrm{P}(B) = 1 - \alpha_2$.

Then

$$\begin{aligned}
\mathrm{P}\Big((U_{n:k}, \beta(U_{n:k})) \in [a_k^{\alpha_1}, b_k^{\alpha_1}] \times \mathcal{T}_{\sigma(k)}^{\alpha_2}\Big) &= \mathrm{P}\Big(U_{n:k} \in [a_k^{\alpha_1}, b_k^{\alpha_1}] \text{ and } \beta(U_{n:k}) \in \mathcal{T}_{\sigma(k)}^{\alpha_2}\Big) \\
&= \mathrm{P}(A \cap B) \\
&= \mathrm{P}(A) - \mathrm{P}\big(A \cap \bar{B}\big) \\
&\ge \mathrm{P}(A) - \mathrm{P}\big(\bar{B}\big).
\end{aligned} \tag{102}$$

Thus,

$$\begin{aligned}
\lim_{T\to\infty} \mathrm{P}\Big((U_{n:k}, \beta(U_{n:k})) \in [a_k^{\alpha_1}, b_k^{\alpha_1}] \times \mathcal{T}_{\sigma(k)}^{\alpha_2}\Big) &\ge \mathrm{P}(A) - \lim_{T\to\infty} \mathrm{P}\big(\bar{B}\big) \\
&= 1 - \alpha_1 - \alpha_2 \\
&= 1 - w\alpha - (1-w)\alpha = 1 - \alpha.
\end{aligned}$$

From Proposition 4, we know the true order of rank variable can be revealed by the order of the fitted outcome values with probability approaching 1, when n goes to infinity slower than T does with the rate relation $n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big)$. In the fixed-$n$, large-$T$ setting, it is automatically satisfied that as T goes to infinity, the fitted outcome values $\hat{Y}_i^*$ become close enough to the true outcome values Y_i^*, smaller than all the increments $Y_{n:k}^* - Y_{n:k-1}^*$ that do not shrink as $T \to \infty$ when n is fixed. Therefore, the order of fitted outcome values can asymptotically represent the true order of rank variable, and the estimated permutation will equal to the true permutation with probability approaching one:

$$\mathrm{P}(\hat{\sigma}(k) = \sigma(k), \forall k = 1, \ldots, n) \to 1, \text{ as } T \to \infty, \text{ and } n \text{ fixed}. \tag{103}$$

We can conservatively bound the coverage probability of the vertical interval based on the estimated permutation,

$$\mathrm{P}\Big(\beta(U_{n:k}) \in \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}\Big) \geq \mathrm{P}\Big(\beta(U_{n:k}) \in \mathcal{T}_{\sigma(k)}^{\alpha_2}\Big) - \underbrace{\mathrm{P}(\hat{\sigma}(\cdot) \neq \sigma(\cdot))}_{\to 0,\text{ from (103)}} \tag{104}$$

Therefore,

$$\begin{aligned}
\lim_{T\to\infty} \mathrm{P}\Big((U_{n:k}, \beta(U_{n:k})) \in \mathcal{W}_k^{(\alpha_1,\alpha_2)}\Big) &= \lim_{T\to\infty} \mathrm{P}\Big(U_{n:k} \in [a_k^{\alpha_1}, b_k^{\alpha_1}] \text{ and } \beta(U_{n:k}) \in \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}\Big) \\
&\geq \lim_{T\to\infty} \mathrm{P}\Big(U_{n:k} \in [a_k^{\alpha_1}, b_k^{\alpha_1}] \text{ and } \beta(U_{n:k}) \in \mathcal{T}_{\sigma(k)}^{\alpha_2}\Big) + o(1) \\
&\geq 1 - \alpha. \qquad (105)
\end{aligned}$$

That is, the confidence set $\mathcal{W}_k^{(\alpha_1,\alpha_2)}$ has at least $1-\alpha$ asymptotic coverage probability for the individual coefficient. □

A.10 Proof of Theorem 10: Dirichlet method in general model—joint confidence sets for all n individual coefficients on the generic coefficient function in fixed-n, large-T setting

Proof. For any $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$, denote the event $A_0, \ldots, A_n$ as

$$\begin{aligned}
A_0 &\equiv \Big\{U_{n:1} \in [\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}] \text{ and } \cdots \text{ and } U_{n:n} \in [\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}]\Big\}, \\
A_k &\equiv \Big\{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(k)}^{\alpha_2}\Big\},\ k = 1, \ldots, n.
\end{aligned}$$

From (26), $\mathrm{P}(A_0) = 1 - \alpha_1$ and $\mathrm{P}\big(\bar{A}_0\big) = 1 - \mathrm{P}(A_0) = \alpha_1$. Define events B_i

$$B_i \equiv \{\beta(U_i) \in \mathcal{T}_i^{\alpha_2}\},\ i = 1, \ldots, n.$$

Assumption A11 implies

$$\mathrm{P}(B_i) \to 1 - \alpha_2, \text{ as } T \to \infty. \tag{106}$$

Denote the event

$$B \equiv A_1 \cap \cdots \cap A_n.$$

Then

$$\begin{aligned}
\mathrm{P}(B) &= \mathrm{P}(A_1 \cap \cdots \cap A_n) \\
&= \mathrm{P}\big(\cap_{k=1}^n \{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(k)}^{\alpha_2}\}\big) \\
&= \mathrm{P}\big(\cap_{i=1}^n \{\beta(U_i) \in \mathcal{T}_i^{\alpha_2}\}\big)
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{P}(B_1 \cap \cdots \cap B_n) \\
&= \mathrm{P}(B_1) \cdots \mathrm{P}(B_n) \ \text{ since the } B_i \text{ are independent across individuals} \\
&\to (1-\alpha_2)^n, \ \text{ as } T \to \infty.
\end{aligned}
$$

Thus, $\mathrm{P}(\bar{B}) \to 1 - (1-\alpha_2)^n$ as $T \to \infty$.

Altogether,

$$
\begin{aligned}
&\mathrm{P}\Big(\Big\{U_{n:1} \in [\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}] \text{ and } \cdots U_{n:n} \in [\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}]\Big\} \cap \Big\{\cap_{k=1}^n \Big\{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(n)}^{\alpha_2}\Big\}\Big\}\Big) \\
&= \mathrm{P}(A_0 \cap A_1 \cap \cdots \cap A_n). \\
&= \mathrm{P}(A_0 \cap B) \\
&\geq 1 - \mathrm{P}\big(\bar{A}_0\big) - \mathrm{P}\big(\bar{B}\big).
\end{aligned}
$$

Asymptotically,

$$
\begin{aligned}
&\lim_{T\to\infty} \mathrm{P}\Big(\Big\{U_{n:1} \in [\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}] \text{ and } \cdots U_{n:n} \in \Big[\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}\Big]\Big\} \cap \Big\{\cap_{k=1}^n \Big\{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(n)}^{\alpha_2}\Big\}\Big\}\Big) \\
&\geq 1 - \mathrm{P}\big(\bar{A}_0\big) - \lim_{T\to\infty} \mathrm{P}\big(\bar{B}\big) \\
&= 1 - \alpha_1 - (1 - (1-\alpha_2)^n) \\
&= 1 - \alpha/2 - \alpha/2 = 1 - \alpha.
\end{aligned}
$$

Proposition 4 implies that as $T \to \infty$ and n staying fixed, the estimated permuation equals the true permuation with probability approaching one. Thus, the probability of the n individual coefficients inside the area $\mathcal{R}^{(\alpha_1,\alpha_2)}$ can be bounded below,

$$
\begin{aligned}
&\mathrm{P}\big(\{\cup_{i=1}^n (U_i, \beta(U_i))\} \subseteq \mathcal{R}^{(\alpha_1,\alpha_2)}\big) \\
&= \mathrm{P}\Big(\{\cup_{k=1}^n (U_{n:k}, \beta(U_{n:k}))\} \subseteq \Big\{\cup_{k=1}^n \mathcal{R}_k^{(\alpha_1,\alpha_2)}\Big\}\Big) \\
&= \mathrm{P}\Big(\{(U_{n:1}, \beta(U_{n:1})), \ldots, (U_{n:n}, \beta(U_{n:n}))\} \subseteq \Big\{\cup_{k=1}^n \Big\{\Big[\tilde{a}_k^{\alpha_1}, \tilde{b}_k^{\alpha_1}\Big] \times \mathcal{T}_{\hat{\sigma}(k)}^{\alpha_2}\Big\}\Big\}\Big) \\
&\geq \mathrm{P}\Big(U_{n:1} \in \Big[\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}\Big] \text{ and } \ldots, U_{n:n} \in \Big[\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}\Big] \text{ and } \beta(U_{n:1}) \in \mathcal{T}_{\hat{\sigma}(1)}^{\alpha_2} \text{ and } \ldots, \beta(U_{n:n}) \in \mathcal{T}_{\hat{\sigma}(n)}^{\alpha_2}\Big) \\
&= \mathrm{P}\Big(\Big\{U_{n:1} \in \Big[\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}\Big] \text{ and } \ldots, U_{n:n} \in \Big[\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}\Big]\Big\} \cap \Big\{\cap_{k=1}^n \Big\{\beta(U_{n:k}) \in \mathcal{T}_{\hat{\sigma}(n)}^{\alpha_2}\Big\}\Big\}\Big) \\
&\geq \mathrm{P}\Big(\Big\{U_{n:1} \in \Big[\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}\Big] \text{ and } \ldots, U_{n:n} \in \Big[\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}\Big]\Big\} \cap \Big\{\cap_{k=1}^n \Big\{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(n)}^{\alpha_2}\Big\}\Big\}\Big) \\
&\quad - \underbrace{\mathrm{P}(\hat{\sigma}(\cdot) \neq \sigma(\cdot))}_{\to 0, \text{ as } T\to\infty, \text{ and } n \text{ fixed}} . \qquad (107)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\lim_{T\to\infty} \mathrm{P}\big(\{\cup_{i=1}^{n}(U_i, \beta(U_i))\} \subseteq \mathcal{R}^{(\alpha_1,\alpha_2)}\big) \\
&\geq \lim_{T\to\infty} \mathrm{P}\Big(\Big\{U_{n:1} \in \Big[\tilde{a}_1^{\alpha_1}, \tilde{b}_1^{\alpha_1}\Big] \text{ and } \ldots, U_{n:n} \in \Big[\tilde{a}_n^{\alpha_1}, \tilde{b}_n^{\alpha_1}\Big]\Big\} \cap \Big\{\cap_{k=1}^{n}\Big\{\beta(U_{n:k}) \in \mathcal{T}_{\sigma(n)}^{\alpha_2}\Big\}\Big\}\Big) \\
&\geq 1-\alpha. \qquad (108)
\end{aligned}
$$

Thus, the confidence set $\mathcal{R}^{(\alpha_1,\alpha_2)}$ with $\alpha_1 = \alpha/2$ and $\alpha_2 = 1-(1-\alpha/2)^{1/n}$ has asymptotic coverage probability bounded below by $1-\alpha$. □

A.11 Lemma 18: asymptotic normality of empirical process

Assume a general version of A12 about the differentiability of the coefficient function. Under this assumption, we build the uniform asymptotic normality of the empirical process as the coefficient function composited with any general distribution function. This provide the intermediate results for Theorem 11, which is the special case that applies to standard uniform distribution.

Assumption A13. Consider the K-vector coefficient function $\boldsymbol{\beta} : (s,t) \subset \mathbb{R} \mapsto \mathbb{R}^K$. Assume each of its element function $\beta_k : (s,t) \subset \mathbb{R} \mapsto \mathbb{R}$, $k = 1, \ldots, K$, is differentiable with uniform continuous and bounded derivative.

Lemma 18 (Asymptotic normality of empirical process)**.** *Let F be a distribution function that have compact support $[s,t]$ and be continuously differentiable on its support with strictly positive derivative f. Let F_n be the empirical distribution of an iid sample from F with size n. Suppose A13 holds. Then the sequence of empirical processes $\sqrt{n}(\beta_k \circ F_n^{-1} - \beta_k \circ F^{-1})$ converges in distribution in the space $\ell^\infty(0,1)$ to a limit Gaussian process*

$$
\sqrt{n}\big(\beta_k \circ F_n^{-1} - \beta_k \circ F^{-1}\big) \rightsquigarrow \beta_k'\big(F^{-1}\big)\frac{\mathbb{G}}{f(F^{-1})}, \ \textit{in } \ell^\infty(0,1), \ \textit{as } n \to \infty, \qquad (109)
$$

where $\mathbb{G}$ is the standard Brownian bridge. The limit process is Gaussian with mean zero and covariance function

$$
\beta_k'\big(F^{-1}(u)\big)\beta_k'\big(F^{-1}(v)\big)\frac{u \wedge v - uv}{f(F^{-1}(u))f(F^{-1}(v))}, \ u, v \in \mathbb{R}.
$$

Proof. Let $D[s,t]$ denote the Skorohod space of all cadlag functions on the interval $[s,t] \subset \bar{\mathbb{R}}$ with uniform norm. Let $C[s,t]$ denote the space of all continuous functions on $[s,t]$. Let $\ell^\infty(T)$ denote the collection of all bounded functions $f : T \mapsto \mathbb{R}$ with the norm $\|f\|_\infty = \sup_{t\in T} f$.

Let $\mathbb{D}_\phi$ be the set of all distribution functions with measures concentrated on $[s,t]$. Van der Vaart and Wellner (1996, Lemma 3.9.23(ii)) show that the inverse map $G \mapsto G^{-1}$ as a map $\phi : \mathbb{D}_\phi \subset D[s,t] \mapsto \ell^\infty(0,1)$ is Hadamard-differentiable at F tangentially to $C[s,t]$. The derivative is

$$\phi'_F(\alpha) = -(\alpha/f) \circ F^{-1} = -\frac{\alpha \circ F^{-1}}{f(F^{-1})},$$

for any function $\alpha \in D[s,t]$.

Consider the map $\psi : \ell^\infty(0,1) \mapsto \ell^\infty(0,1)$ given by $\psi(A)(x) = \beta_k(A(x)) = \beta_k \circ A(x)$. The domain of this map is the set of elements of $\ell^\infty(0,1)$ that take their values in the domain of β_k. Let $\mathbb{D}_\psi = \{A : A \in \ell^\infty(0,1) : s < A < t\}$. By van der Vaart and Wellner (1996, Lemma 3.9.25), we know the map $A \mapsto \beta_k \circ A$ as a map $\mathbb{D}_\psi \subset \ell^\infty(0,1) \mapsto \ell^\infty(0,1)$ is Hadamard-differentiable at every $A \in \mathbb{D}_\psi$. The derivative is $\psi'_A(\alpha) = \beta'_k(A(x))\alpha(x)$.

Therefore, the map $F^{-1} \mapsto \beta_k \circ F^{-1}$ as a map $\psi : \mathbb{D}_\psi \subset \ell^\infty(0,1) \mapsto \ell^\infty(0,1)$ is Hadamard-differentiable at every $F^{-1} \in \mathbb{D}_\psi$. The derivative is $\psi'_{F^{-1}}(\alpha) = \beta'_k(F^{-1}(x))\alpha(x)$, for function $\alpha \in \ell^\infty(0,1)$.

By the chain rule, (See van der Vaart and Wellner (1996, Lemma 3.9.3).) the map $\psi \circ \phi : \mathbb{D}_\phi \mapsto \ell^\infty(0,1)$ is Hadamard-differentiable at F tangentially to $C[s,t]$, with derivative

$$\big(\psi'_{\phi(F)} \circ \phi'_F\big)(\alpha) = (\psi'_{F^{-1}} \circ \phi'_F)(\alpha) = \beta'_k\big(F^{-1}\big)\phi'_F(\alpha) = \beta'_k\big(F^{-1}\big)\big(-(\alpha/f) \circ F^{-1}\big),$$

for any function $\alpha \in D[s,t]$. That is, at any $p \in (0,1)$, the derivative is

$$\big(\psi'_{\phi(F)} \circ \phi'_F\big)(\alpha)(p) = \beta'_k\big(F^{-1}(p)\big)\left(-\frac{\alpha(F^{-1}(p))}{f(F^{-1}(p))}\right).$$

Applying the functional delta-methods (See van der Vaart and Wellner (1996, Theorem 3.9.4)) to the Donsker's theorem and the above result that the map $\psi \circ \phi : \mathbb{D}_\phi \mapsto \ell^\infty(0,1)$ is Hadamard-differentiable at F tangentially to $C[s,t]$,

$$\begin{aligned}
\sqrt{n}\big(\beta_k \circ F_n^{-1} - \beta_k \circ F^{-1}\big) &= \sqrt{n}((\psi \circ \phi)(F_n) - (\psi \circ \phi)(F)) \\
&\rightsquigarrow (\psi \circ \phi)'_F(\mathbb{G} \circ F) \\
&= -\beta'_k\big(F^{-1}\big)\frac{(\mathbb{G} \circ F) \circ F^{-1}}{f(F^{-1})} \\
&= -\beta'_k\big(F^{-1}\big)\frac{\mathbb{G}}{f(F^{-1})} \\
&= \beta'_k\big(F^{-1}\big)\frac{\mathbb{G}}{f(F^{-1})}, \text{ in } \ell^\infty(0,1), \qquad (110)
\end{aligned}$$

where $\mathbb{G}$ is the standard Brownian bridge. The last equality is due to the symmetry of the

Gaussian process $\mathbb{G}$. The limit process is Gaussian with mean zero and covariance function

$$\beta_k'\big(F^{-1}(u)\big)\beta_k'\big(F^{-1}(v)\big)\frac{u \wedge v - uv}{f(F^{-1}(u))f(F^{-1}(v))},\ u, v \in \mathbb{R}. \qquad \square$$

A.12 Proof of Theorem 11: uniform asymptotic normality of coefficient function estimator in simplified model

Proof. Consider the standard uniform distribution F_U with domain $(s, t) = (0, 1)$. It is continuously differentiable on $(0, 1)$ with a strictly positive constant derivative $f_U = 1$ on $(0, 1)$. Therefore, under A12, from Lemma 18,

$$\sqrt{n}\big(\beta_k \circ F_{n,U}^{-1} - \beta_k \circ F_U^{-1}\big) \rightsquigarrow \beta_k'\big(F_U^{-1}\big)\mathbb{G} = \beta_k'\mathbb{G},\ \text{in } \ell^\infty(0, 1). \tag{111}$$

The last equality is due to F_U is an identity function on $(0, 1)$. The limit is Gaussian process with mean zero and covariance function

$$\beta_k'\big(F_U^{-1}(u)\big)\beta_k'\big(F_U^{-1}(v)\big)(u \wedge v - uv) = \beta_k'(u)\beta_k'(v)(u \wedge v - uv),\ u, v \in \mathbb{R}.$$

Under A1–A5 in the simplified model, we get the perfect coefficient estimation $\hat{\beta}_k(\tau) = \beta_k\big(U_{n:\lceil \tau n \rceil}\big)$ for all τ on $(0, 1)$. Since $U_{n:\lceil \tau n \rceil} = F_{n,U}^{-1}(\tau)$ we have $\hat{\beta}_k = \beta_k \circ F_{n,U}^{-1}$ on $(0, 1)$. Also, $F_U^{-1}(\tau) = \tau$. Therefore,

$$\begin{aligned}\sqrt{n}\Big(\hat{\beta}_k(\cdot) - \beta_k(\cdot)\Big) &= \sqrt{n}\big(\beta_k \circ F_{n,U}^{-1} - \beta_k \circ F_U^{-1}\big) \\ &\rightsquigarrow \beta_k'\big(F_U^{-1}\big)\mathbb{G} = \beta_k'\mathbb{G},\ \text{in } \ell^\infty(0, 1), \end{aligned} \tag{112}$$

where the limit is Gaussian process with mean zero and covariance function $\beta_k'(u)\beta_k'(v)(u \wedge v - uv)$, $u, v \in \mathbb{R}$. $\square$

A.13 Proof of Corollary 12: pointwise asymptotic normality of coefficient estimator in simplified model

Proof. It is straightforward to apply Theorem 11 the specific point $\tau \in (0, 1)$. $\square$

A.14 Proof of Theorem 13: uniform asymptotic normality of coefficient function estimator in general model

Proof. Similar as shown in the proof in Theorem 11, when applying Lemma 18 to the standard uniform distribution F_U, under A12,

$$\sqrt{n}\big(\beta_k \circ F_{n,U}^{-1} - \beta_k \circ F_U^{-1}\big) \rightsquigarrow \beta_k'\big(F_U^{-1}\big)\mathbb{G} = \beta_k'\mathbb{G}, \text{ in } \ell^\infty(0,1). \tag{113}$$

The limit process is Gaussian with mean zero and covariance function

$$\beta_k'\big(F_U^{-1}(u)\big)\beta_k'\big(F_U^{-1}(v)\big)(u \wedge v - uv) = \beta_k'(u)\beta_k'(v)(u \wedge v - uv), \ u, v \in \mathbb{R}.$$

Under A1–A4 and A6–A10 and the n/T rate relation $n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big)$, we know the ordering of rank variable can be correctly discovered, as shown in the proof of Theorem 5; The individual with the kth ordering in the true outcome values $Y^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}(U_{n:k})$ at $X = x^*$ is also the same individual who has the kth ordering in the fitted outcome values $\hat{Y}^* = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}(U_{n:k})$ at $X = x^*$. Specifically Assumption A8 (ii) assumes the gap between the estimated individual coefficient and the true coefficient in terms of T order.

$$\hat{\beta}_k(U_i) - \beta_k(U_i) = O_p\big(T^{-\kappa}\big), \text{ as } T \to \infty, \text{ for } \kappa > 0. \tag{114}$$

Therefore, the coefficient estimator can be written as

$$\begin{aligned}\hat{\beta}_k(\tau) = \hat{\beta}_k\big(U_{n:\lceil \tau n \rceil}\big) &= \hat{\beta}_k\big(U_{\sigma(\lceil \tau n \rceil)}\big) \\ &= \beta_k\big(U_{\sigma(\lceil \tau n \rceil)}\big) + O_p\big(T^{-\kappa}\big) \\ &= \beta_k\big(U_{n:\lceil \tau n \rceil}\big) + O_p\big(T^{-\kappa}\big).\end{aligned} \tag{115}$$

Then,

$$\begin{aligned}\sqrt{n}\Big(\hat{\beta}_k(\cdot) - \beta_k(\cdot)\Big) &= \sqrt{n}\big(\beta_k \circ F_{n,U}^{-1} + O_p\big(T^{-\kappa}\big) - \beta_k \circ F_U^{-1}\big) \\ &= \sqrt{n}\big(\beta_k \circ F_{n,U}^{-1} - \beta_k \circ F_U^{-1}\big) + n^{1/2}O_p\big(T^{-\kappa}\big) \\ &\rightsquigarrow \beta_k'\big(F_U^{-1}\big)\mathbb{G} = \beta_k'\mathbb{G}, \text{ in } \ell^\infty(0,1), \text{ as } n \to \infty, T \to \infty,\end{aligned} \tag{116}$$

since $n^{1/2}O_p(T^{-\kappa}) \xrightarrow{p} 0$ under the rate relation $n = o\Big(T^{\frac{\kappa}{3+\delta}}\Big)$. That is, $0 \le \frac{n}{T^{2\kappa}} \le \frac{n}{T^{\kappa/(3+\delta)}} \to 0$, as $n \to \infty, T \to \infty$, for $\kappa > 0$, $\delta > 0$.[7] □

[7] The δ here is the same as it is introduced in the proof of Theorem 5.

A.15 Proof of Lemma 15: consistency of bootstrap empirical process

Proof. Let F be a distribution function that have compact support $[s,t]$ and be continuously differentiable on its support with strictly positive derivative f. Let $\mathbb{P}$ denote its probability measure. Let F_n be the empirical distribution of an iid sample from F with size n. Let $\mathbb{P}_n$ denote its probability measure.

Let $\hat{F}_n$ be the bootstrap empirical distribution of an iid sample from F_n with size n. Let $\hat{\mathbb{P}}_n$ denote its probability measure.

Let $\mathbb{D}_\phi$ be the set of all distribution functions with measures concentrated on $[s,t]$. Let map $\psi : \ell^\infty(0,1) \mapsto \ell^\infty(0,1)$ be $\psi(A)(x) = \beta_k(A(x)) = \beta_k \circ A(x)$. Let $\phi : \mathbb{D}_\phi \subset D[s,t] \mapsto \ell^\infty(0,1)$ be the inverse map $\phi(G) = G^{-1}$. In the Lemma 18 proof, we have shown by chain rule the map $\psi \circ \phi : \mathbb{D}_\phi \mapsto \ell^\infty(0,1)$ is Hadamard-differentiable at $\mathbb{P}$ tangentially to $C[s,t]$ under A12.

As Lemma 14 stated, which is proved in van der Vaart and Wellner (1996, Theorem (3.6.13)), conditional on almost all the random dataset $X_1, \dots, X_n$, the sequence of bootstrap empirical process converges in distribution to some limit process $\mathbb{H}$

$$\sqrt{n}\Big(\hat{\mathbb{P}}_n - \mathbb{P}_n\Big) \mid (X_1, \dots, X_n) \rightsquigarrow \mathbb{H}. \tag{117}$$

The limit process $\mathbb{H}$ is the same as the limit of the empirical process $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$, which is $\mathbb{G} \circ F$ previously with $\mathbb{G}$ being the standard Brownian bridge. The Lemma 14 is the same as the equation (3.9.9) in van der Vaart and Wellner (1996), which is a condition used to apply the delta method. That is,

$$\sup_{h \in BL_1(\mathbb{D})} \Big| \mathrm{E}_M\, h\Big(\sqrt{n}\Big(\hat{\mathbb{P}}_n - \mathbb{P}_n\Big)\Big) - \mathrm{E}\, h(\mathbb{H}) \Big| \to 0, \tag{118}$$

where $BL_1(\mathbb{D})$ is the set of all functions $h : \ell^\infty(\mathcal{F}) \mapsto [0,1]$ on $\mathbb{D}$ with Lipschitz norm bounded by 1. The E_M denotes the expectation is with respect to the random vector $\mathbf{M} = (M_{n1}, \dots, M_{nn})$. The random vector $\mathbf{M}$ follows a multinomial distribution with parameter $\big(n; \frac{1}{n}, \dots, \frac{1}{n}\big)$.[8] The left-hand side is conditional on (almost all) the (random) dataset $X_1, \dots, X_n$, since $\mathbb{P}_n = \mathbb{P}_n(X_1, \dots, X_n)$ and $\hat{\mathbb{P}}_n = \hat{\mathbb{P}}_n(X_1, \dots, X_n, \mathbf{M})$.

By van der Vaart and Wellner (1996, Theorem 3.9.11), applying the delta-method to the

[8] Here we consider the standard bootstrap. It can be viewed as a special case of the weighted bootstrap, where $\mathbf{M}$ is exchangeable nonnegative random vector. The generalization to weighted bootstrap is showed in Appendix B.2.

bootstrap empirical process with the Hadamard-differentiable function $\psi \circ \phi$,

$$\sup_{h \in BL_1(\mathbb{D})} \left| \mathrm{E}_M\, h\Big(\sqrt{n}\Big(\psi \circ \phi(\hat{\mathbb{P}}_n) - \psi \circ \phi(\mathbb{P}_n)\Big)\Big) - \mathrm{E}\, h\big((\psi \circ \phi)'_{\mathbb{P}}(\mathbb{H})\big) \right| \to 0. \tag{119}$$

That is, conditional on almost all the random dataset, the bootstrap empirical process of the $\psi \circ \phi$ transformed functions converges in distribution to a limit process

$$\sqrt{n}\Big((\psi \circ \phi)\Big(\hat{\mathbb{P}}_n\Big) - (\psi \circ \phi)(\mathbb{P}_n)\Big) \rightsquigarrow (\psi \circ \phi)'_{\mathbb{P}}(\mathbb{H}). \tag{120}$$

Denote the inverse map ϕ of a probability measure as the inverse of the distribution function. That is, $\phi(\mathbb{P}) = F^{-1}$. Thus, $\phi(\mathbb{P}_n) = F_n^{-1}$ and $\phi\Big(\hat{\mathbb{P}}_n\Big) = \hat{F}_n^{-1}$. The equation (120) can be written as

$$\sqrt{n}\Big(\beta_k \circ \hat{F}_n^{-1} - \beta_k \circ F_n^{-1}\Big) \rightsquigarrow \big(\beta_k \circ F^{-1}\big)'_F(\mathbb{G} \circ F), \tag{121}$$

where $\mathbb{G}$ is the standard Brownian bridge. The limit process has been derived in Lemma 18. That is, conditional on almost all the random draw of data,

$$\sqrt{n}\Big(\beta_k \circ \hat{F}_n^{-1} - \beta_k \circ F_n^{-1}\Big) \rightsquigarrow \beta_k'\big(F^{-1}\big)\frac{\mathbb{G}}{f(F^{-1})}, \text{ in } \ell^\infty(0,1). \qquad \square$$

A.16 Proof of Theorem 16: consistency of bootstrap estimation function in simplified model

Proof. Under A1–A5 in the simplified model, we have perfect coefficient estimation $\hat{\beta}_k(\tau) = \beta_k\big(U_{n:\lceil \tau n\rceil}\big)$ for all τ on $(0,1)$. Since $U_{n:\lceil \tau n\rceil} = F_{n,U}^{-1}(\tau)$, the coefficient function estimator can be written as a composition $\hat{\beta}_k = \beta_k \circ F_{n,U}^{-1}$.

Let $(Y_{it}^b, X_{it}^b), i = 1, \ldots, n,\ t = 1, \ldots, T$ denote the bootstrap sample. Denote it in a matrix form $\mathbf{Y}_i^b = (Y_1, \ldots, Y_T)$ and $\underline{\mathbf{X}}_i^b = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})$. For each bootstrap individual, we can perfectly estimate its coefficient parameter $\boldsymbol{\beta}^b(U_i) = \big(\underline{\mathbf{X}}_i^b\big)^{-1}\mathbf{Y}_i^b$.

Let $\big(U_1^b, \ldots, U_n^b\big)$ denote the unobserved rank variable associated with each individual for the bootstrap sample. The $\big(U_1^b, \ldots, U_n^b\big)$ is implicitly drawn from $(U_1, \ldots, U_n)$ with probability measure $\mathbb{P}_n$. Each unobserved rank variable in the original sample follows a standard uniform distribution $U_i \sim$ Unif $(0,1)$. Let $U_{n:\lceil n\tau\rceil}^b$ denote the $\lceil n\tau\rceil$th order statistic of the bootstrap sample $\big(U_1^b, \ldots, U_n^b\big)$. We have $U_{n:\lceil n\tau\rceil}^b = \hat{F}_{n,U}^{-1}(\tau)$, where $\hat{F}_{n,U}$ is the distribution function for the bootstrap sample $\big(U_1^b, \ldots, U_n^b\big)$.

Given the data $(Y_{it}, \mathbf{X}_{it})$ for $i = 1, \ldots, n,\ t = 1, \ldots, T$, implementing the estimation process in the simplified model, we obtain the n individuals coefficients parameters in the bootstrap sample $\boldsymbol{\beta}\big(U_1^b\big), \ldots, \boldsymbol{\beta}\big(U_n^b\big)$, since $\boldsymbol{\beta}\big(U_i^b\big) = \boldsymbol{\beta}^b(U_i)$. Based on A3, the order of

$Y^{b*} = \mathbf{x}^{*\prime}\boldsymbol{\beta}^b$ represents the order of U^b. Thus, by the definition of coefficient estimator in the simplified model, the coefficient associated with the $\lceil n\tau \rceil$th ordering of Y^{b*} is the the coefficient at the $\lceil n\tau \rceil$th order of the rank variable values $\left(U_1^b, \ldots, U_n^b\right)$. That is, $\hat{\boldsymbol{\beta}}^b(\tau) = \boldsymbol{\beta}\left(U_{n:\lceil n\tau \rceil}^b\right)$. Therefore, $\hat{\boldsymbol{\beta}}^b(\tau) = \boldsymbol{\beta}\left(U_{n:\lceil n\tau \rceil}^b\right) = \boldsymbol{\beta}\left(\hat{F}_{n,U}^{-1}(\tau)\right)$. The bootstrap coefficient function estimator can be written as the composition of two functions. For each element of the bootstrap coefficient function $\hat{\beta}_k^*(\cdot)$, we have $\hat{\beta}_k^*(\cdot) = \beta_k \circ \hat{F}_{n,U}^{-1}$.

Therefore, conditional on almost all the random data, the bootstrap empirical process converges to a limit process

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_k^*(\cdot) - \hat{\beta}_k(\cdot)\right) &= \sqrt{n}\left(\beta_k \circ \hat{F}_{n,U}^{-1} - \beta_k \circ F_{n,U}^{-1}\right) \\ &\rightsquigarrow \beta_k'\left(F_U^{-1}\right)\frac{\mathbb{G}}{f_U\left(F_U^{-1}\right)}, \text{by Lemma 15} \\ &= \beta_k'\mathbb{G}, \text{ in } \ell^\infty(0,1). \end{aligned}$$

□

B Supplemental details and proofs

B.1 Background about empirical process

Let $D[a,b]$ denote the Skorohod space of all cadlag functions on the interval $[a,b] \subset \bar{\mathbb{R}}$ with uniform norm. Let $C[a,b]$ denote the space of all continuous function on $[a,b]$. Let $\ell^\infty(T)$ denote the collection of all bounded functions $f : T \mapsto \mathbb{R}$ with the norm $\|f\|_\infty = \sup_{t\in T} f$.

B.1.1 Donsker Theorem

Let F_n be the sequence of the empirical distribution function of an iid sampling $X_1, X_2, \ldots$ with size n from distribution function F. The Donsker's theorem is

$$\sqrt{n}(F_n - F) \rightsquigarrow \mathbb{G} \circ F, \quad \text{in } D[-\infty, +\infty] \tag{122}$$

where $\mathbb{G}$ is the standard Brownian bridge with mean zero and variance function $\mathrm{Cov}[\mathbb{G}(s), \mathbb{G}(t)] = s \wedge t - st$.[9]

For the standard uniform distribution, the proceeding Donsker's theorem results become

$$\sqrt{n}(F_{n,U} - F_U) \rightsquigarrow \mathbb{G} \circ F_U = \mathbb{G}, \quad \text{in } D[0,1] \tag{123}$$

since F_U is the identity function on $[0,1]$.

[9] The right-hand side has mean zero and variance function $\mathrm{Cov}[\mathbb{G}_F(s), \mathbb{G}_F(t)] = F(s) \wedge F(t) - F(s)F(t)$.

B.1.2 Inverse map

Let $0 < p < 1$. Let $\mathbb{D}_\phi$ be the collection of all nondecreasing functions. Let the distribution function $F \in \mathbb{D}_\phi$ be differentiable at $\xi_p \in (a, b)$ such that $F(\xi_p) = p$ with strictly positive derivative. By the Lemma 3.9.20 in van der Vaart and Wellner (1996), the inverse map $G \mapsto G^{-1}$ as a map $\phi : \mathbb{D}_\phi \subset D[a, b] \mapsto \mathbb{R}$ is Hadamard-differentiable at F tangentially to the set of $\mathbb{D}_0$ that are the set of functions $\alpha \in D[a, b]$ that are continuous at $\xi_p \equiv F^{-1}(p)$. The derivative is $\phi_F(\alpha) = -\alpha(\xi_p)/F'(\xi_p) = -(\alpha/f)(\xi_p)$.

B.1.3 Empirical quantile: pointwise central limit theorem

Apply both delta-method (See van der Vaart and Wellner (1996, Theorem 3.9.4)) and the inverse map results (van der Vaart and Wellner (1996, Lemma 3.9.20)) to the Donsker's theorem, we have the pointwise central limit theorem for the empirical quantile function. (See van der Vaart and Wellner (1996, Example 3.9.21).) That is, for any $q \in (0, 1)$,

$$\sqrt{n}\big(F_n^{-1}(p) - F^{-1}(p)\big) \rightsquigarrow \phi'_F(\mathbb{G} \circ F) = -\frac{\mathbb{G}(p)}{f(F^{-1}(p))}. \tag{124}$$

For the standard uniform distribution function F_U and its sequence of empirical distribution function $F_{n,U}$, the pointwise empirical quantile convergence is, for any $0 < \tau < 1$,

$$\sqrt{n}\big(F_{n,U}^{-1}(\tau) - F_U^{-1}(\tau)\big) = \sqrt{n}\big(U_{n:\lceil \tau n \rceil} - \tau\big) \rightsquigarrow \mathbb{G}(\tau), \tag{125}$$

since its density function f_U is uniformly 1 on $(0, 1)$ and the standard Brownian bridge is symmetric around zero, $\mathbb{G}(\tau) = -\mathbb{G}(\tau)$.

B.1.4 Empirical quantile processes: uniform central limit theorem

More general uniform central limit theorem of the sequence of empirical quantile processes can be obtained through the Hadamard-differentiability of the inverse map (See van der Vaart and Wellner (1996, Lemma 3.9.23 (i))) and the delta-method.

Let $0 < p < q < 1$. Let F be continuously differentiable on the interval $[a, b] = [F^{-1}(p) - \epsilon, F^{-1}(q) + \epsilon]$ for some $\epsilon > 0$, with positive derivative f. Then the inverse map $G \mapsto G^{-1}$ as $\phi : \mathbb{D}_1 \subset D[a, b] \mapsto \ell^\infty[p, q]$ is Hadamard-differentiable at F tangentially to $C[a, b]$, by van der Vaart and Wellner (1996, Lemma 3.9.23 (i)). Applying the delta-method to the inverse map, together with the Donsker's theorem, it yields the empirical quantile process convergence. (See van der Vaart and Wellner (1996, Example 3.9.24))

$$\sqrt{n}\big(F_n^{-1} - F^{-1}\big) \rightsquigarrow -\frac{\mathbb{G} \circ F(F^{-1})}{f(F^{-1})} = -\frac{\mathbb{G}}{f(F^{-1})} \quad \text{in } \ell^\infty[p, q]. \tag{126}$$

For the uniform distribution function, the proceeding results yields a uniform convergence of the empirical quantile process.

$$\sqrt{n}\big(F_{n,U}^{-1} - F_U^{-1}\big) \rightsquigarrow -\mathbb{G} = \mathbb{G} \ \text{ in } \ell^\infty(0,1). \tag{127}$$

That is the same as the equation (61) in Shorack and Wellner (1986, Section 3.1 Theorem 1).

B.2 Generalization to the weighted empirical bootstrap process

The standard bootstrap in this paper can be generalized to the weighted bootstrap.

Given a sample $X_1, X_2, \dots, X_n$, the bootstrap sample $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m$ is an iid sampling drawn from the original sample $X_1, X_2, \dots, X_n$ with replacement, It treats the empirical measure $\mathbb{P}_n$ as the true measure for bootstrap sampling. Let $\hat{\mathbb{P}}_m$ denote the bootstrap empirical measure,

$$\hat{\mathbb{P}}_m = \frac{1}{m}\sum_{i=1}^{m} M_{ni}\delta_{X_i} \tag{128}$$

where M_{mi} denote the number of times X_i is redrawn in the size m bootstrap sample. It is the standard bootstrap, if $m = n$. It is the m out of n bootstrap, if $m < n$. The bootstrap empirical process is

$$\hat{\mathbb{G}}_{n,m} = \sqrt{m}\Big(\hat{\mathbb{P}}_m - \mathbb{P}_n\Big) = \sqrt{m}\left[\frac{1}{m}\sum_{i=1}^{n} M_{mi}\delta_{X_i} - \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}\right] = \frac{1}{\sqrt{m}}\sum_{i=1}^{n}\Big(M_{mi} - \frac{m}{n}\Big)\delta_{X_i}, \tag{129}$$

with $\sum_{i=1}^{n} M_{mi} = m$ and $M_{mi} \in \{0, 1, \dots, m\}$.

More generally, consider the interchangeable weighted bootstrap. For each n, let $(W_{n1}, \dots, W_{nn})$ be an exchangeable, nonnegative random vector.[10] Let $\hat{\mathbb{P}}_n$ denote the weighted bootstrap empirical measure,

$$\hat{\mathbb{P}}_n = \frac{1}{n}\sum_{i=1}^{n} W_{ni}\delta_{X_i}. \tag{130}$$

It includes both the standard bootstrap and m out of n bootstrap.[11] With the exchangeable weighted bootstrap measure, we can write the bootstrap empirical distribution as $\hat{F}_n$.

$$\hat{F}_n(x) = \frac{1}{n_B}\sum_{i=1}^{n_B} \mathbb{1}\{\hat{X}_i \le x\} \tag{131}$$

[10] W_{ni} can be non-integer valued.

[11] By van der Vaart and Wellner (1996, Example 3.6.10 and Example 3.6.11), the standard bootstrap is the case $(W_{n1}, \dots, W_{nn}) \sim \text{Multinomial}(n; 1/n, \dots, 1/n)$ with $c = 1$; the m out of n bootstrap is the case $(W_{n1}, \dots, W_{nn}) \sim \sqrt{\frac{n}{m}}\text{Multinomial}(m; 1/n, \dots, 1/n)$ with $c = 1$.

where $n_B = \sum_{i=1}^n W_{ni}$ is the size of bootstrap sample and $\hat{X}_i$ denotes the bootstrap sample. The empirical bootstrap process is

$$\hat{\mathbb{G}}_n = \sqrt{n}\Big(\hat{\mathbb{P}}_n - \bar{W}_n\mathbb{P}_n\Big), \tag{132}$$

where $\bar{W}_n = 1/n\sum_{i=1}^n W_{ni}$. In the standard bootstrap that n variables are resampled, $\bar{W}_n = 1$ and the empirical process is $\hat{\mathbb{G}}_n = \sqrt{n}\Big(\hat{\mathbb{P}}_n - \mathbb{P}_n\Big)$.

Assumption A14 (Exchangeable weighted bootstrap weights from van der Vaart and Wellner (1996) (Section 3.6.2))**.** For each n, the exchangeable, nonnegative random weights vector $(W_{n1}, \dots, W_{nn})$ satisfies

$$\sup_n \big\|W_{n1} - \bar{W}_n\big\|_{2,1} < \infty, \tag{i}$$

$$\frac{1}{\sqrt{n}} E \max_{1\le i\le n}\big|W_{ni} - \bar{W}_n\big| \xrightarrow{p} 0, \tag{ii}$$

$$\frac{1}{n}\sum_{i=1}^n \big(W_{ni} - \bar{W}_n\big)^2 \xrightarrow{p} c^2 > 0, \tag{iii}$$

where $\bar{W}_n = 1/n\sum_{i=1}^n W_{ni}$; $\|\cdot\|_{2,1}$ means the $L_{2,1}$ norm of a random variable η, $\|\eta\|_{2,1} = \int_0^\infty \sqrt{\mathrm{P}(|\eta| > x)}dx$; c is some constant.

We can rescale the weights $(W_{n1}, \dots, W_{nn})$ by dividing c. Then the rescaled weights satisfy A14 with c=1. Under A14, we have the uniform convergence of the bootstrap empirical process.

Lemma 19 (The weighted bootstrap empirical process convergence, same as van der Vaart and Wellner (1996) Theorem 3.6.13)**.** *Under A14, the weighted bootstrap empirical process converges to a limit process.* $\hat{\mathbb{G}}_n \rightsquigarrow \mathbb{H}$, *as* $n \to \infty$.

B.3 45 countries in empirical example and their democracy level

See Table 2.

B.4 Bootstrap confidence interval at 95% level and bootstrap critical value in empirical example

See Table 3.

Table 2: The 45 countries in empirical example and their average democracy level indexed between 0 and 1.

Country	Dem	Country	Dem	Country	Dem	Country	Dem	Country	Dem
DZA	0.20	CAN	1.00	HUN	0.43	NLD	1.00	SYR	0.21
AGO	0.28	TCD	0.24	IND	0.93	NGA	0.45	THA	0.47
AZE	0.21	CHL	0.63	IDN	0.35	NOR	1.00	ARE	0.10
ARG	0.46	CHN	0.14	IRN	0.21	PAK	0.54	TUN	0.17
AUS	1.00	COL	0.77	ITA	0.82	PER	0.59	TUR	0.69
BHR	0.05	DNK	1.00	KWT	0.09	ROM	0.33	EGY	0.27
BOL	0.52	ECU	0.63	MYS	0.75	RUS	0.72	GRB	1.00
BRA	0.54	FXX	0.87	MEX	0.37	SAU	0.00	USA	1.00
CMR	0.19	DEU	0.59	OMN	0.08	SDN	0.31	VEN	0.66

Use 0.5 as the cutoff to divide the democratic and non-democratic countries, as Cotet and Tsui (2013) did, there are 21 democratic countries and 24 countries as non-democratic countries in the empirical example.

References

Abadie, Alberto, Joshua Angrist, and Guido Imbens. 2002. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica* 70 (1):91–117. URL `https://www.jstor.org/stable/2692164`.

Abrevaya, Jason and Christian M. Dahl. 2008. "The Effects of Birth Inputs on Birthweight." *Journal of Business & Economic Statistics* 26 (4):379–397. URL `https://doi.org/10.1198/073500107000000269`.

Arellano, Manuel and Stéphane Bonhomme. 2012. "Identifying Distributional Characteristics in Random Coefficients Panel Data Models." *The Review of Economic Studies* 79 (3):987–1020. URL `https://doi.org/10.1093/restud/rdr045`.

———. 2016. "Nonlinear panel data estimation via quantile regressions." *Econometrics Journal* 19 (3):C61–C94. URL `https://doi.org/10.1111/ectj.12062`.

Besley, Timothy and Torsten Persson. 2011. "The Logic of Political Violence." *Quarterly Journal of Economics* 126 (3):1411–1445. URL `https://doi.org/10.1093/qje/qjr025`.

Bohn, Henning and Robert T. Deacon. 2000. "Ownership Risk, Investment, and the Use of Natural Resources." *American Economic Review* 90 (3):526–549. URL `https://doi.org/10.1257/aer.90.3.526`.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3):414–427. URL `https://doi.org/10.1162/rest.90.3.414`.

Canay, Ivan A. 2011. "A simple approach to quantile regression for panel data." *Econometrics Journal* 14 (3):368–386. URL `https://www.jstor.org/stable/23116937`.

Chamberlain, Gary. 1982. "Multivariate regression models for panel data." *Journal of Econometrics* 18 (1):5–46. URL `https://doi.org/10.1016/0304-4076(82)90094-X`.

Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn, and Whitney Newey. 2013. "Average and Quantile Effects in Nonseparable Panel Models." *Econometrica* 81 (2):535–580. URL `https://www.jstor.org/stable/23524292`.

Table 3: The estimates of the effect oil wealth per capita of on military defense spending as ratio of GDP, its standard error, and 95% CI at ranks 0.1 to 0.9.

τ	Slope Estimates	Bootstrap S.E.	Bootstrap Critical Value	95% Bootstrap CI LowerBound	UpperBound
0.1	0.18	0.13	4.41	−0.38	0.75
0.2	−0.43	0.32	3.63	−1.60	0.75
0.3	0.06	0.23	5.20	−1.11	1.23
0.4	−0.14	0.20	4.39	−1.03	0.75
0.5	−0.18	0.25	2.47	−0.79	0.43
0.6	0.12	0.22	3.27	−0.58	0.83
0.7	−0.19	0.22	2.76	−0.81	0.42
0.8	−0.01	0.24	1.98	−0.47	0.46
0.9	−1.01	0.40	3.68	−2.47	0.46
OLS	0.04	0.04	1.96	−0.03	0.12

700 bootstrap times. The bootstrap critical values are the converged limit values as bootstrap times increases.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81 (6):2205–2268. URL `https://www.jstor.org/stable/23524318`.

Chernozhukov, Victor and Christian Hansen. 2005. "An IV Model of Quantile Treatment Effects." *Econometrica* 73 (1):245–261. URL `https://www.jstor.org/stable/3598944`.

———. 2006. "Instrumental quantile regression inference for structural and treatment effect models." *Journal of Econometrics* 132 (2):491–525. URL `https://doi.org/10.1016/j.jeconom.2005.02.009`.

———. 2008. "Instrumental variable quantile regression: A robust inference approach." *Journal of Econometrics* 142 (1):379–398. URL `https://doi.org/10.1016/j.jeconom.2007.06.005`.

Chesher, Andrew. 2003. "Identification in Nonseparable Models." *Econometrica* 71 (5):1405–1441. URL `https://www.jstor.org/stable/1555507`.

Collier, Paul and Anke Hoeffler. 1998. "On economic causes of civil war." *Oxford Economic Papers* 50 (4):563–573. URL `https://doi.org/10.1093/oep/50.4.563`.

———. 2004. "Greed and grievance in civil war." *Oxford Economic Papers* 56 (4):563–595. URL `https://doi.org/10.1093/oep/gpf064`.

Cotet, Anca M. and Kevin K. Tsui. 2013. "Oil and Conflict: What Does the Cross Country Evidence Really Show?" *American Economic Journal: Macroeconomics* 5 (1):49–80. URL `https://doi.org/10.1257/mac.5.1.49`.

Davison, A. C. and D. V. Hinkley. 1997. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. URL `https://doi.org/10.1017/CBO9780511802843`.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (1):75–90. URL `http://www.jstor.org/stable/`

3118222.
Fernández-Val, Iván and Joonhwah Lee. 2013. “Panel data models with nonadditive unobserved heterogeneity: Estimation and inference.” *Quantitative Economics* 4 (3):453–481. URL https://doi.org/10.3982/QE75.
Frumento, Paolo, Matteo Bottai, and Iván Fernández-Val. 2020. “Parametric Modeling of Quantile Regression Coefficient Functions with Longitudinal Data.” Working paper, available at https://arxiv.org/abs/2006.00160.
Galvao, Antonio F. 2011. “Quantile regression for dynamic panel data with fixed effects.” *Journal of Econometrics* 164 (1):142–157. URL https://doi.org/10.1016/j.jeconom.2011.02.016.
Galvao, Antonio F. and Kengo Kato. 2016. “Smoothed quantile regression for panel data.” *Journal of Econometrics* 193 (1):92–112. URL https://doi.org/10.1016/j.jeconom.2016.01.008.
Galvao, Antonio F. and Liang Wang. 2015. “Efficient minimum distance estimator for quantile regression fixed effects panel data.” *Journal of Multivariate Analysis* 133:1–26. URL https://doi.org/10.1016/j.jmva.2014.08.007.
Goldman, Matt and David M. Kaplan. 2018. “Comparing distributions by multiple testing across quantiles or CDF values.” *Journal of Econometrics* 206 (1):143–166. URL https://doi.org/10.1016/j.jeconom.2018.04.003.
Graham, Bryan S., Jinyong Hahn, Alexandre Poirier, and James L. Powell. 2018. “A quantile correlated random coefficients panel data model.” *Journal of Econometrics* 206 (2):305–335. URL https://doi.org/10.1016/j.jeconom.2018.06.004.
Hamilton, James D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.
Hansen, Bruce E. 2020. “Econometrics.” URL https://www.ssc.wisc.edu/~bhansen/econometrics. Textbook draft.
Harding, Matthew and Carlos Lamarche. 2009. “A quantile regression approach for estimating panel data models using instrumental variables.” *Economics Letters* 104 (3):133–135. URL https://doi.org/10.1016/j.econlet.2009.04.025.
Imbens, Guido W. and Whitney K. Newey. 2009. “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity.” *Econometrica* 77 (5):1481–1512. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7108.
Kato, Kengo, Antonio F. Galvao, and Gabriel V. Montes-Rojas. 2012. “Asymptotics for panel quantile regression models with individual effects.” *Journal of Econometrics* 170 (1):76–91. URL https://doi.org/10.1016/j.jeconom.2012.02.007.
Koenker, Roger. 2004. “Quantile regression for longitudinal data.” *Journal of Multivariate Analysis* 91 (1):74–89. URL https://doi.org/10.1016/j.jmva.2004.05.006.
———. 2005. *Quantile Regression*, *Econometric Society Monographs*, vol. 38. Cambridge University Press. URL https://doi.org/10.1017/CBO9780511754098.
Koenker, Roger and Gilbert Bassett, Jr. 1978. “Regression Quantiles.” *Econometrica* 46 (1):33–50. URL https://www.jstor.org/stable/1913643.
Lee, Sokbae. 2007. “Endogeneity in quantile regression models: A control function approach.” *Journal of Econometrics* 141 (2):1131–1158. URL https://doi.org/10.1016/j.jeconom.2007.01.014.
Melly, Blaise and Kaspar Wüthrich. 2017. “Local Quantile Treatment Effects.” In *Handbook of Quantile Regression*, edited by Roger Koenker, Victor Chernozhukov, Xuming He, and

Limin Peng. CRC/Chapman-Hall, 145–164. URL https://www.routledgehandbooks.com/doi/10.1201/9781315120256.

Powell, David. 2016. "Quantile Regression with Nonadditive Fixed Effects." Working paper available at https://sites.google.com/site/davidmatthewpowell/quantile-regression-with-nonadditive-fixed-effects.

Rosen, Adam M. 2012. "Set identification via quantile restrictions in short panels." *Journal of Econometrics* 166 (1):127–137. URL https://doi.org/10.1016/j.jeconom.2011.06.011.

Shorack, Galen R. and Jon A. Wellner. 1986. *Empirical Processes with Applications to Statistics*. New York: John Wiley & Sons.

van der Vaart, Aad W. 1998. *Asymptotic Statistics*. Cambridge: Cambridge University Press. URL https://books.google.com/books?id=UEuQEM5RjWgC.

van der Vaart, Aad W. and Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. New York: Springer. URL https://doi.org/10.1007/978-1-4757-2545-2.

Wilks, Samuel Stanley. 1962. *Mathematical Statistics*. New York: Wiley. URL https://www.worldcat.org/oclc/602895854.