DYNAMIC HETEROGENEOUS DISTRIBUTION REGRESSION PANEL MODELS, WITH AN APPLICATION TO LABOR INCOME PROCESSES*

IVÁN FERNÁNDEZ-VAL, WAYNE YUAN GAO, YUAN LIAO, AND FRANCIS VELLA

Abstract. We consider estimation of a dynamic distribution regression panel data model with heterogeneous coefficients across units. The objects of interest are functionals of these coefficients including linear projections on unit level covariates. We also consider predicted actual and stationary distributions of the outcome variable. We investigate how changes in initial conditions or covariate values affect these objects. Coefficients and their functionals are estimated via fixed effect methods, which are debiased to deal with the incidental parameter problem. We propose a cross-sectional bootstrap method for uniformly valid inference on function-valued objects. This avoids coefficient re-estimation and is shown to be consistent for a large class of data generating processes. We employ PSID annual labor income data to illustrate various important empirical issues we can address. We first predict the impact of a reduction in income on future income via hypothetical tax policies. Second, we examine the impact on the distribution of labor income from increasing the education level of a chosen group of workers. Finally, we demonstrate the existence of heterogeneity in income mobility, which leads to substantial variation in individuals' incidences to be trapped in poverty. We also provide simulation evidence confirming that our procedures work well.

Keywords: distribution regression, individual heterogeneity, panel data, uniform inference, labor income dynamics, incidental parameter problem, poverty traps

1. Introduction

Empirical studies increasingly feature analyses of data comprising repeated observations on the same or similar units. While the most common example is panel data, many of its attractive features are found in other data structures, such as network and spatial data. From an econometric perspective, the availability of panel data naturally accommodates a treatment of time invariant unit-specific heterogeneity (see,

Date: This draft: February 10, 2022.
* We thank Matt Hong, Koen Jochmans, Dennis Kristensen, Allan Timmermann, Xi Wang, and seminar participants at 2021 IESR Microeconometrics Workshop, 2022 Econometric Society North America Winter Meeting, Tsinghua and York for comments.

for example, Mundlak 1978) and also provides internal instruments in the presence of time varying endogeneity (see, for example, Hausman and Taylor 1981, Arellano and Bond 1991). It also facilitates the estimation of dynamic relationships within unit, or contemporaneous relationships between units. However, a feature of the panel data literature is its limited treatment of parameter heterogeneity. Although the random coefficient panel model allows heterogeneous coefficients between units, and some recent developments that we discuss below incorporate heterogeneous coefficients within units, there are relatively few studies that incorporate heterogeneous coefficients both between and within units.[1]

We consider panel models with coefficient heterogeneity between and within units, using a dynamic distribution regression model with heterogeneous coefficients. This model captures within unit heterogeneous relationships between outcome and covariates through function-valued coefficients, and between unit heterogeneity by allowing the coefficients to vary across units in an unrestricted fashion. The objects of interest are functionals of the coefficients including linear projections on individual covariates and predicted distributions. We also consider the impact on these objects from manipulating the values of the initial conditions of the outcome or the covariates. We consider both one-period-ahead and stationary counterfactual distributions to measure the short and long term effects of these changes.

Our proposed estimator employs fixed effect methods, which allow an unrestricted relationship between the unobserved unit-specific heterogeneity, the covariates, and the initial conditions. Estimation and inference consists of four steps. First, we estimate unit-specific coefficients by distribution regression exploiting the time series dimension of the panel. Second, we estimate the functionals of interest using the plug-in method. It is necessary to debias the resulting estimates to account for the incidental parameter problem (Neyman and Scott, 1948). Third, we construct plug-in estimators of quantiles and quantile effects of the counterfactual distributions. Fourth, we perform inference using a cross-sectional bootstrap method which resamples with replacement the estimated coefficients of the units and avoids the computationally expensive first-step estimation. We show how to construct confidence bands and test hypotheses for the quantiles and quantile effects, which are uniformly valid over a prespecified region of quantile indexes.

[1]Exceptions include Chetverikov et al. (2016), Okui and Yanagi (2019), Zhang et al. (2019) and Chen (2021), which are discussed in the literature review.

We derive novel inferential theory of wider interest for estimating functionals. The novelty lies in the unknown degree of heterogeneity that may affect both the rate of convergence and the asymptotic distribution, making them unknown and *continuously varying* across different assumptions on the heterogeneity. We identify an important problem with traditional analytical plug-in methods in performing inference in models with heterogeneous coefficients. We show these methods are very sensitive to the degree of heterogeneity as measured by the variance of the coefficients unexplained by the covariates. Formally, we establish that analytical methods break down in data generating processes where there is coefficient homogeneity or, more broadly, when the degree of heterogeneity is sufficiently small relative to the sample size. Both the rate of convergence and the asymptotic distribution of the estimated quantities are unknown and may vary depending on the unknown degree of heterogeneity. However, we prove that a simple cross-sectional bootstrap method is uniformly valid for a large class of data generating processes including the case of homogeneous coefficients.

Our methodology is applicable to a wide range of settings and we employ it to examine labor income dynamics. This is a well studied topic, starting with Champernowne (1953), Hart (1976), Shorrocks (1976) and Lillard and Willis (1978), among others. We apply our model to the Panel Study of Income Dynamics (PSID) data to perform experiments corresponding to various counterfactual analyses. First, we consider how a reduction in annual income in a given year, implemented via a flat or progressive tax, affects future annual labor income. We find that the predicted effect on the cross-sectional distribution of labor income after one period varies substantially after we account for heterogeneity in the level and persistence of income. Interestingly, our model predicts significantly smaller effects than do models that impose homogeneous effects. Second, we consider a hypothetical scenario that assigns 12 years of schooling to those individuals who have not completed high school. We find important short and long run distributional effects as it increases the incomes of those in the lower tails of the one-period ahead and stationary labor income distributions. However, it has little effect on their upper tails. This exercise, which cannot be analyzed using traditional homogeneous autoregressive models, illustrates the importance of individual characteristics in earnings dynamics. We also investigate a number of issues related to heterogeneity that have implications for poverty and income inequality. We uncover substantial cross-sectional heterogeneity in the level and persistence of annual labor income and identify the responsible individual characteristics. We show that this heterogeneity has implications for an individual's tendency to remain below or above specific quantiles of the income distribution.

1.1. **Relationship with existing literature.** The literature examining labor income processes has typically focused on allocating the total error variances into transitory and permanent components. A summary is provided in Moffitt and Zhang (2018) and two important recent innovations are Arellano et al. (2017) and Hu et al. (2019). The first examined nonlinear persistence in the permanent component and how it varies over the earnings distribution. The second allowed for a flexible representation of the distributions of both components. Our approach is not intended to supersede these methodologies. Rather, we examine earnings dynamics to illustrate how we can complement these earlier studies. However, the approach most similar to ours is Arellano et al. (2017). While that paper also focused on the impact of earnings on consumption, an important feature is the treatment of the persistence in the earnings process. They considered a dynamic earnings process with nonlinear persistence that can vary by location in the earnings distribution. Our approach incorporates a generalized linear process which not only varies by location in the earnings distribution but also across workers. Moreover, we allow persistence to be a function of both observed and unobserved individual characteristics. Our analysis of income mobility and persistence relies on a representation of the model as a discrete Markov chain when labor income is treated as discrete. Champernowne (1953) and Shorrocks (1976) previously used Markov chain representations of the labor income process to analyze the same issue. We allow unrestricted heterogeneity across workers by estimating a separate Markov chain for each worker.[2] Finally, Hirano (2002) and Gu and Koenker (2017) estimated autoregressive labor income processes using flexible semiparametric Bayesian methods.

The model we consider differs from the traditional random coefficients model of Swamy (1970), Hsiao and Pesaran (2008), Arellano and Bonhomme (2012), Fernández-Val and Lee (2013) and Su et al. (2016), among others, as we allow for heterogeneous coefficients both between and within units. It is more flexible than existing distribution and quantile regression models with fixed effects that allow the intercepts to vary across units but restrict the slopes to be homogeneous. See, for example, Koenker (2004), Galvao (2011), Galvao and Kato (2016), Kato et al. (2012), Arellano and Weidner (2017), and Chernozhukov et al. (2018). Chetverikov et al. (2016),

[2]Lillard and Willis (1978) considered an alternative method to separate permanent and transitory income and incorporate worker heterogeneity using a parametric linear panel model.

Okui and Yanagi (2019), Zhang et al. (2019) and Chen (2021) are the closest papers to ours. Okui and Yanagi (2019) provided methods to estimate distributions of heterogeneous moments such as means, autocovariances and autocorrelations. The model and objects considered there are very different from ours. Zhang et al. (2019) proposed a quantile regression grouped panel data model with heterogeneous coefficients, but where the distribution of the coefficients is restricted to have finite support. Chetverikov et al. (2016) and Chen (2021) develop models similar to ours. They targeted projections of the model coefficients as the objects of interest, but, unlike here, did not consider counterfactual distributions. Chetverikov et al. (2016), Zhang et al. (2019) and Chen (2021) focused on models with strictly exogenous covariates, which rule out dynamic models that include lagged outcomes as covariates. Moreover, their methodology is also based on quantile regression. Distribution regression has several appealing features in our setting including: (i) It deals with continuous, discrete and mixed outcomes without modification, and (ii) it yields simple analytical forms for the functionals of interest. In this sense, we extend the use of the distribution regression of Foresi and Peracchi (1995) and Chernozhukov et al. (2013) to panel models with random coefficients.

Bias correction methods based on large-T asymptotic approximations for fixed effects estimators of dynamic and nonlinear panel models were studied in Nickell (1981), Phillips and Moon (1999), Hahn and Newey (2004), Fernández-Val (2009), Hahn and Kuersteiner (2011), Dhaene and Jochmans (2015), and Fernández-Val and Weidner (2016), among others. We refer the readers to Arellano and Hahn (2007) and Fernández-Val and Weidner (2018) for recent reviews. We extend these debiasing methods to functionals of the coefficients such as projections and counterfactual distributions. The cross-sectional bootstrap was previously used for panel data as a resampling scheme that preserves the dependence in the time series dimension, e.g., Kapetanios (2008), Kaffo (2014), and Gonçalves and Kaffo (2015). We demonstrate that it also has robustness properties in models with heterogeneous coefficients.

1.2. **Outline.** The rest of the paper is organized as follows. Section 2 presents the model and objects of interest. Section 3 discusses estimation and inference including an issue with standard inference on models with heterogeneous coefficients, which is solved with the use of a cross-sectional bootstrap scheme. We present the empirical application in Section 4. Section 5 establishes asymptotic theory for our estimation and inference methods. Section 6 reports simulation evidences. Proofs and additional results are gathered in the Appendix.

2. The model and objects of interest

2.1. **The model.** We observe a panel data set $\{(y_{it}, \boldsymbol{x}_{it}) : 1 \leq i \leq N, 1 \leq t \leq T\}$, where i typically indexes observational units and t time periods. The scalar variable y_{it} represents the outcome or response of interest, which can be continuous, discrete or mixed; and $\boldsymbol{x}_{it}$ is a d_x-vector of covariates, which includes a constant, lagged outcome values, and other predetermined covariates denoted by $\boldsymbol{v}_{it}$, that is

$$\boldsymbol{x}_{it} = (1, y_{i(t-1)}, ..., y_{i(t-L)}, \boldsymbol{v}_{it}')'.$$

Let $\mathcal{F}_{it}$ be a filtration over t that includes $\boldsymbol{x}_{it}$ and any time invariant variable for unit i. We model the distribution of y_{it} conditional on $\mathcal{F}_{it}$ as, for any $y \in \mathbb{R}$,

$$\Pr(y_{it} \leq y \mid \mathcal{F}_{it}) = \Lambda(-\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(y)), \quad 1 \leq t \leq T, \quad 1 \leq i \leq N, \tag{2.1}$$

where $\Lambda : \mathbb{R} \mapsto [0, 1]$ is a known, strictly increasing link function[3] (e.g., the standard normal or logistic distribution CDF), and $y \mapsto -\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(y)$ is increasing almost surely (a.s). The model is a distribution regression model for panel data with heterogeneous coefficients. We allow the coefficient vector $\boldsymbol{\beta}_i(y)$ to vary both between i and within i over y. For example, in the empirical application, the intercept is a fixed effect that measures the level of the distribution, whereas the coefficient of lagged labor income measures persistence. Both level and persistence coefficients are heterogeneous between and within workers. The model also embodies a Markov-type condition for each individual as only the first L lags of the outcome and contemporaneous values of the other covariates determine the conditional distribution of y_{it}.[4] It also imposes an index restriction on the effect of $\boldsymbol{x}_{it}$. This restriction can be considered mild as the coefficient $\boldsymbol{\beta}_i(y)$ varies with i and y, and can be further weakened by replacing $\boldsymbol{x}_{it}$ by $T(\boldsymbol{x}_{it})$, where T is a vector of transformations of $\boldsymbol{x}_{it}$. Our theory would still apply provided that T is known and has a fixed dimension.

The heterogeneous distribution regression (HDR) model in (2.1) encompasses other commonly used models. For example, the homogeneous location-shift model

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \sigma\varepsilon_{it}, \quad \varepsilon_{it} \mid \mathcal{F}_{it} \sim \Lambda,$$

is a special case of HDR with $\boldsymbol{\beta}_i(y) = (\boldsymbol{\beta} - \boldsymbol{e}_1 y)/\sigma$, where $\boldsymbol{e}_1$ is a unitary d_x-vector with a one in the first position. This model imposes that all components of $\boldsymbol{\beta}_i(y)$ are homogeneous across i and only the intercept can vary across y. Another important

[3] We could allow Λ to vary across i and y, but we do not pursue those extensions here.

[4] Lagged values of the covariates can be included in $\boldsymbol{v}_{it}$.

case is the homogeneous location-shift model with fixed effects

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \sigma\varepsilon_{it}, \quad \varepsilon_{it} \mid \mathcal{F}_{it} \sim \Lambda.$$

This is a special case of HDR with $\boldsymbol{\beta}_i(y) = [\boldsymbol{\beta} - \boldsymbol{e}_1(y+\alpha_i)]/\sigma$. It is more flexible than the location-shift model as the intercept is heterogeneous across i, but it imposes strong homogeneity restrictions compared to HDR. The cross-sectional version of the distribution regression model of Foresi and Peracchi (1995) and Chernozhukov et al. (2013) imposes the restriction $\boldsymbol{\beta}_i(y) = \boldsymbol{\beta}(y)$, which allows for heterogeneity within the distribution but not between units. We refer to this as the homogeneous DR model. The panel distribution regression model of Chernozhukov et al. (2018) imposes $\boldsymbol{\beta}_i(y) = \boldsymbol{\beta}(y) + \boldsymbol{e}_1\alpha_i(y)$, which allows for heterogeneity in the intercept across i, but restricts the slopes to be homogeneous between units.

When y_{it} is continuous, the HDR has the following representation as an implicit nonseparable model by the probability integral transform

$$-\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(y_{it}) = \varepsilon_{it}, \quad \varepsilon_{it} \mid \mathcal{F}_{it} \sim \Lambda.$$

The rank of the error ε_{it}, $\Lambda(\varepsilon_{it})$, can be interpreted as the unobserved ranking of the observation y_{it} in the conditional distribution. The previous representation reduces to the homogeneous location-shift models described above by imposing suitable restrictions on $\boldsymbol{\beta}_i(y)$.

2.2. **Objects of interest.** We are interested in the following types of objects.

2.2.1. *Projections of $\boldsymbol{\beta}_i(y)$ on Covariates.* Let $\boldsymbol{w}_i, \boldsymbol{z}_i \in \mathcal{F}_{i1}$ denote time invariant covariates such that $\dim(\boldsymbol{w}_i) \geq \dim(\boldsymbol{z}_i)$ and $\mathbb{E}(\boldsymbol{w}_i\boldsymbol{z}_i')$ have full column rank. Consider the instrumental variable projection of $\boldsymbol{\beta}_i(y)$ on $\boldsymbol{z}_i$:

$$\boldsymbol{\beta}_i(y) = \boldsymbol{\theta}(y)\boldsymbol{z}_i + \boldsymbol{\gamma}_i(y), \quad \mathbb{E}(\boldsymbol{\gamma}_i(y) \mid \boldsymbol{w}_i) = 0, \tag{2.2}$$

which covers the standard linear projection by setting $\boldsymbol{w}_i = \boldsymbol{z}_i$. This object is useful for exploring which covariates are associated with the heterogeneity in $\boldsymbol{\beta}_i(y)$ across i, where we allow these associations to vary within the distribution as indexed by y. For example, in the empirical application, we explore whether initial income, education, race and year of birth are associated with differences in the level and persistence of labor income at different locations of the income distribution.

2.2.2. *Cross-sectional Distributions.*

Actual and predicted distributions: By iterating expectations, the cross-sectional distribution of the observed outcome at time t can be written in terms of the model coefficients as

$$F_t(y) := \mathbb{E}_t 1\{y_{it} \leq y\} = \mathbb{E}_t \mathbb{E}(1\{y_{it} \leq y\} \mid \mathcal{F}_{it}) = \mathbb{E}_t \Lambda(-\boldsymbol{x}_{it}' \boldsymbol{\beta}_i(y)), \tag{2.3}$$

where the expectation $\mathbb{E}_t$ is taken with respect to the joint cross-sectional distribution of the variables in $\mathcal{F}_{it}$ at time t.

This representation serves several purposes. First, it is the basis for a specification test of the model where an estimator of (2.3) is compared with the cross-sectional empirical distribution of y_{it}. Second, in pure dynamic models where $\boldsymbol{x}_{it}$ only includes lagged values of y_{it}, we can construct one-period-ahead predicted distributions by setting $t = T + 1$. These distributions are useful for forecasting. Third, we can analyze dynamics of the distribution of y_{it} over time. In the empirical application, for example, we analyze labor income mobility and the persistence of poverty traps. Fourth, we can consider the impact of interventions by comparing the counterfactual distribution after some intervention with the actual distribution.

Counterfactual distributions: From (2.3), we can construct counterfactual distributions by manipulating the covariates $\boldsymbol{x}_{it}$ and/or the coefficients $\boldsymbol{\beta}_i(y)$, that is

$$G_t(y) = \mathbb{E}_t \Lambda(-h_{it}(\boldsymbol{x}_{it})' \boldsymbol{\beta}_i^g(y)), \tag{2.4}$$

where h_{it} is a possibly data dependent transformation, and

$$\boldsymbol{\beta}_i^g(y) = \boldsymbol{\theta}(y) g(\boldsymbol{z}_i) + \boldsymbol{\gamma}_i(y) = \boldsymbol{\beta}_i(y) + \boldsymbol{\theta}(y)[g(\boldsymbol{z}_i) - \boldsymbol{z}_i],$$

for a known transformation g of the time invariant covariates $\boldsymbol{z}_i$.

We provide examples of h_{it} and g in the context of the empirical application. Starting with h_{it}, we can study the effect of a proportional reduction of labor income in the previous year using the transformation

$$h_{it}(\boldsymbol{x}_{it}) = (1, y_{i(t-1)} + \log(1 - \kappa))', \tag{2.5}$$

where $y_{i(t-1)}$ is measured in logarithmic scale and κ is the tax rate. This can be interpreted as a proportional or flat tax. Another example where the transformation h_{it} is data dependent is a progressive reduction of labor income in the previous year depending on the ranking in the distribution. For example, this could be implemented as

$$h_{it}(\boldsymbol{x}_{it}) = \Big(1, y_{i(t-1)} + \log(1 - \frac{\kappa_{it}}{2})\Big)', \quad \kappa_{it} = \mathbb{E}_t(1\{y_{i(t-1)} \leq y\})\big|_{y = y_{i(t-1)}}, \tag{2.6}$$

which can be interpreted as a progressive tax where the tax rate is half of the ranking of the worker in the distribution. These tax exercises are interesting as we can evaluate their impact on future labor income operating through the inherent dynamics.

Turning to g, we can consider a hypothetical scenario at time t that increases the number of years of schooling to 12 for those workers with less than 12 years of schooling. If $\boldsymbol{z}_i = (z_{1i}, \boldsymbol{z}'_{-1,i})'$, where z_{1i} is the observed years of schooling of worker i and $\boldsymbol{z}_{-1,i}$ includes the remaining components of $\boldsymbol{z}_i$. This counterfactual scenario can be implemented via the transformation

$$g(\boldsymbol{z}_i) = (\max(z_{1i}, 12), \boldsymbol{z}_{-1,i}). \tag{2.7}$$

$G_t(y)$ would then represent the counterfactual distribution at t. Another example is

$$g(\boldsymbol{z}_i) = (z_{1i} + 1, \boldsymbol{z}_{-1,i}),$$

which corresponds to an additional year of schooling for all workers.

2.3. **Stationary Distributions:** Assume that the process $\{y_{i1}, \dots, y_{iT}\}$ is ergodic for each i, y_{it} is discrete with support $\mathcal{Y}_i = \{y_i^1 < \cdots < y_i^K\}$, which might be different for each unit, and the only covariate is the first lag of the outcome, i.e. $\boldsymbol{x}_{it} = (1, y_{i(t-1)})'$. The conditional distribution can now be represented by a time-homogeneous K-state Markov chain and the stationary distribution can be characterized from the transition matrix of the Markov chain. The method can be extended to include more lags of the outcome at the cost of more cumbersome notation.

For each i, let $\boldsymbol{P}_i$ be the $K \times K$ transition matrix. The typical element of this matrix can be expressed as the following functional of the model:

$$P_{i,jk} = \Pr(y_{it} = y_i^j \mid y_{i(t-1)} = y_i^k, \mathcal{F}_{it}) = \Lambda\left(-\boldsymbol{x}_i^{k\prime}\boldsymbol{\beta}_i(y_i^j)\right) - 1(j > 1)\Lambda\left(-\boldsymbol{x}_i^{k\prime}\boldsymbol{\beta}_i(y_i^{j-1})\right), \tag{2.8}$$

where $\boldsymbol{x}_i^k = (1, y_i^k)'$. By standard theory for Markov Chains, see, e.g., (Hamilton, 2020, p. 684), the ergodic probabilities $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{iK})$ are

$$\boldsymbol{\pi}_i = (\boldsymbol{A}_i'\boldsymbol{A}_i)^{-1}\boldsymbol{A}_i'\boldsymbol{e}_{K+1}, \quad \boldsymbol{A}_i = \begin{pmatrix} \boldsymbol{I}_K - \boldsymbol{P}_i \\ \boldsymbol{1}' \end{pmatrix},$$

where $\boldsymbol{I}_K$ is the identity matrix of size K, $\boldsymbol{1}$ is a K-vector of ones, and $\boldsymbol{e}_{K+1}$ is the $(K+1)$th column of $\boldsymbol{I}_{K+1}$. The cross-sectional stationary actual distribution is now

$$F_\infty(y) = \mathbb{E}[F_{i,\infty}(y)], \quad F_{i,\infty}(y) = \sum_{k: y_i^k \leq y} \pi_{ik},$$

where $F_{i,\infty}$ is a step function with steps at the elements of $\mathcal{Y}_i$.

Stationary counterfactual distributions can be formed by replacing $\boldsymbol{\beta}_i(y_i^j)$ by $\boldsymbol{\beta}_i^g(y_i^j)$ in (2.8). That is

$$P_{i,jk}^g = \Lambda\left(-\boldsymbol{x}_i^{k\prime}\boldsymbol{\beta}_i^g(y_i^j)\right) - 1(j > 1)\Lambda\left(-\boldsymbol{x}_i^{k\prime}\boldsymbol{\beta}_i^g(y_i^{j-1})\right).$$

We denote the resulting cross-sectional stationary distribution as G_∞. We do not consider transformations h_{it} as they would produce the stationary distribution F_∞. Note that changes in $y_{i(t-1)}$ do not affect the stationary distribution by the ergodicity assumption.

2.3.1. *Quantile effects.* We consider quantiles of the actual and counterfactual cross-sectional distributions, and define quantile effects as the difference between them. Given a univariate distribution F, the quantile (left-inverse) operator is

$$\phi(F,\tau) := \inf\{y \in \mathbb{R} : F(y) \geq \tau\}, \quad \tau \in [0,1].$$

We apply this operator to the cross-sectional distributions defined above to obtain the quantile effects of interest as

$$\mathsf{QE}_t(\tau) := \phi(G_t,\tau) - \phi(F_t,\tau), \quad \mathsf{QE}_\infty(\tau) := \phi(G_\infty,\tau) - \phi(F_\infty,\tau), \quad \tau \in [0,1].$$

These quantile effects measure the short and long term impacts of the hypothetical policies at different parts of the outcome distribution. They are unconditional or marginal as they are based on comparisons between counterfactual and actual marginal distributions.

3. Estimation and Inference Methods

3.1. **Estimators.** We employ a three-stage estimation procedure where the first step obtains the model coefficients, the second constructs their functionals, and the third calculates quantile effects. The coefficients are estimated by HDR applied separately to the time series dimension of each unit, with debiasing to address the incidental parameter problem. The functionals of the coefficients are estimated using the plug-in method. The estimators of the distributions are debiased. The estimators of the projection coefficients do not need to be debiased as these projections are linear functionals of the model coefficients. The quantile effects are estimated by applying the generalized inverse operator of Chernozhukov et al. (2010).

3.1.1. *First stage: Model coefficients.* We start by obtaining the DR estimator of $\boldsymbol{\beta}_i(y)$, that is

$$\widetilde{\boldsymbol{\beta}}_i(y) = \arg\max_{\beta \in \mathbb{R}^{d_x}} Q_{y,i}(\beta), \quad y \in \mathcal{Y}_i, \quad i = 1, ..., N,$$

where

$$Q_{y,i}(\beta) = \sum_{t=1}^{T} 1\{y_{it} \leq y\}\Lambda(-\boldsymbol{x}_{it}'\beta) + \sum_{t=1}^{T} 1\{y_{it} > y\}[1 - \Lambda(-\boldsymbol{x}_{it}'\beta)],$$

and $\mathcal{Y}_i$ is the set of observed values of the outcome for unit i, i.e. $\mathcal{Y}_i = \{y_{i1}, \ldots, y_{iT}\}$. If Λ is the standard normal or logistic link, these are logit or probit estimators that can be computed using standard software. We then obtain $\widetilde{\boldsymbol{\beta}}_i(y)$ for other values of y noting that $y \mapsto \widetilde{\boldsymbol{\beta}}_i(y)$ is a vector of step functions with steps at the elements of $\mathcal{Y}_i$.

Two complications arise: $\widetilde{\boldsymbol{\beta}}_i(y)$ is well-defined only if $y \in [\underline{y}_i, \overline{y}_i)$, where $\underline{y}_i = \min_{1 \leq t \leq T} y_{it}$ and $\overline{y}_i = \max_{1 \leq t \leq T} y_{it}$, and, when $\widetilde{\boldsymbol{\beta}}_i(y)$ is well-defined, it has bias of order $O(T^{-1})$. Let $N_0(y)$ be the number of indexes i for which $y < \underline{y}_i$, $N_1(y)$ be the number of indexes i for which $y \geq \overline{y}_i$, and $N_{01}(y) = N - N_0(y) - N_1(y)$, that is the number of indexes i for which $\widetilde{\boldsymbol{\beta}}_i(y)$ exists. Without loss of generality we rearrange the index i such that that $\widetilde{\boldsymbol{\beta}}_i(y)$ exists for all $i = 1, \ldots, N_{01}(y)$. We show below how to adjust the plug-in estimators of the functionals to incorporate the units $i > N_{01}(y)$.

Due to the incidental parameter bias, we should debias $\widetilde{\boldsymbol{\beta}}_i(y)$ when T is of moderate size relative to N. Plug-in estimators of nonlinear functionals based on debiased estimators are easier to debias than those based on the initial estimators. We debias $\widetilde{\boldsymbol{\beta}}_i(y)$ using analytical methods. That is

$$\widehat{\boldsymbol{\beta}}_i(y) = \widetilde{\boldsymbol{\beta}}_i(y) - \frac{\widehat{B}_{i,T}(y)}{T}, \quad i = 1, \ldots, N_{01}(y), \tag{3.1}$$

where $\widehat{B}_{i,T}(y)$ is a consistent estimator of the bias of $\widetilde{\boldsymbol{\beta}}_i(y)$ of order $O(T^{-1})$. The specific expressions of the bias and its estimator are presented in the Appendix, where we also consider alternative debiasing methods based on Jackknife (Dhaene and Jochmans, 2015). While our theory applies to both analytical and Jackknife methods, we focus on analytical methods because they have less demanding data requirements and perform better in our numerical simulations.

3.1.2. *Second stage: Functionals.* We provide estimators for all the functionals of interest. Denote the second derivative of the link function by $\ddot{\Lambda}$.

Projections of Coefficients. A plug-in estimator of $\boldsymbol{\theta}(y)$ corresponds to applying two-stage least squares to (2.2) replacing $\boldsymbol{\beta}_i(y)$ by $\widehat{\boldsymbol{\beta}}_i(y)$. This yields,

$$\widehat{\boldsymbol{\theta}}(y) = \sum_{i=1}^{N_{01}(y)} \widehat{\boldsymbol{\beta}}_i(y)\widehat{\boldsymbol{z}}_i(y)' \left(\sum_{i=1}^{N_{01}(y)} \widehat{\boldsymbol{z}}_i(y)\widehat{\boldsymbol{z}}_i(y)' \right)^{-1}, \tag{3.2}$$

where

$$\widehat{\boldsymbol{z}}_i(y) := \sum_{j=1}^{N_{01}(y)} z_j \boldsymbol{w}_j' \left(\sum_{j=1}^{N_{01}(y)} \boldsymbol{w}_j \boldsymbol{w}_j' \right)^{-1} \boldsymbol{w}_i.$$

When $\boldsymbol{w}_i = \boldsymbol{z}_i$, the estimator simplifies to the OLS estimator with $\widehat{\boldsymbol{z}}_i(y) = \boldsymbol{z}_i$.

Actual and Counterfactual Distributions. The plug-in estimators of the actual and counterfactual distributions are

$$\begin{aligned} \widehat{F}_t(y) &= \frac{1}{N} \sum_{i=1}^{N_{01}(y)} \Lambda(-\boldsymbol{x}_{it}'\widehat{\boldsymbol{\beta}}_i(y)) + \frac{N_1(y)}{N} - \frac{\widehat{B}(y)}{T}, \\ \widehat{G}_t(y) &= \frac{1}{N} \sum_{i=1}^{N_{01}(y)} \Lambda(-h(\boldsymbol{x}_{it})'\widehat{\boldsymbol{\beta}}_i^g(y)) + \frac{N_1(y)}{N} - \frac{\widehat{B}_G(y)}{T}, \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_i^g(y) &= \widehat{\boldsymbol{\beta}}_i(y) + \widehat{\boldsymbol{\theta}}(y)[g(\boldsymbol{z}_i) - \boldsymbol{z}_i], \\ \widehat{B}(y) &= \frac{1}{2}\frac{1}{N} \sum_{i=1}^{N_{01}(y)} \operatorname{tr}\left(\ddot{\Lambda}(-\boldsymbol{x}_{it}'\widehat{\boldsymbol{\beta}}_i(y))\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\widehat{\Sigma}_i(y)^{-1} \right) \\ \widehat{B}_G(y) &= \frac{1}{2}\frac{1}{N} \sum_{i=1}^{N_{01}(y)} \operatorname{tr}\left(\ddot{\Lambda}(-h(\boldsymbol{x}_{it})'\widehat{\boldsymbol{\beta}}_i^g(y))h(\boldsymbol{x}_{it})h(\boldsymbol{x}_{it})'\widehat{\Sigma}_i(y)^{-1} \right). \end{aligned}$$

Here $\widehat{B}(y)$ and $\widehat{B}_G(y)$ are estimators of the first-order bias coming from the non-linearity of F_t and G_t as a functional of $\boldsymbol{\beta}(y)$, and $\widehat{\Sigma}_i(y)^{-1}$ is an estimator of the asymptotic variance-covariance matrix of $\sqrt{T}(\widetilde{\boldsymbol{\beta}}_i(y) - \boldsymbol{\beta}_i(y))$. For units for which $\widehat{\boldsymbol{\beta}}_i(y)$ is not well-defined we set $\Lambda(-\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(y)) = \Lambda(-h(\boldsymbol{x}_{it})'\boldsymbol{\beta}_i^g(y)) = 1$ if $y < \underline{y}_i$ and $\Lambda(-\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(y)) = \Lambda(-h(\boldsymbol{x}_{it})'\boldsymbol{\beta}_i^g(y)) = 0$ if $y \geq \overline{y}_i$.

Stationary Distributions. We start from a preliminary plug-in estimator of $\boldsymbol{P}_i$ by the empirical transition matrix, which we modify to enforce that all the entries are non-negative and the rows add to one. More precisely, we define the $K \times K$ matrix $\widehat{\boldsymbol{Q}}_i$ with typical element

$$\widehat{Q}_{i,jk} = 1(j = K) + 1(j < K)\Lambda\left(-\boldsymbol{x}_i^{k\prime}\widehat{\beta}_i(y_i^j)\right). \tag{3.4}$$

For each row of $\widehat{\boldsymbol{Q}}_i$, we sort (rearrange) the elements in increasing order to form the matrix $\check{\boldsymbol{Q}}_i$ with typical element $\check{Q}_{i,jk}$. We then construct the empirical transition matrix $\widehat{\boldsymbol{P}}_i$ with typical element

$$\widehat{P}_{i,jk} = \check{Q}_{i,jk} - 1(j > 1)\check{Q}_{i,(j-1)k}.$$

The empirical ergodic probabilities $\widehat{\boldsymbol{\pi}}_i = (\widehat{\pi}_{i1}, \dots, \widehat{\pi}_{iK})$ are now

$$\widehat{\boldsymbol{\pi}}_i = (\widehat{\boldsymbol{A}}_i'\widehat{\boldsymbol{A}}_i)^{-1}\widehat{\boldsymbol{A}}_i'\boldsymbol{e}_{K+1} - \frac{1}{T}\widehat{B}_{\boldsymbol{\pi}_i}, \quad \widehat{\boldsymbol{A}}_i = \begin{pmatrix} \boldsymbol{I}_K - \widehat{\boldsymbol{P}}_i \\ \boldsymbol{1}' \end{pmatrix},$$

where $\widehat{B}_{\boldsymbol{\pi}_i}$ is an estimator of the bias coming from the nonlinearity of $\boldsymbol{\pi}_i$ as a functional of $(\boldsymbol{\beta}_i(y_i^1), \dots, \boldsymbol{\beta}_i(y_i^K))$. We give the expression of $\widehat{B}_{\boldsymbol{\pi}_i}$ in the Appendix.

The estimator of the stationary distribution is

$$\widehat{F}_\infty(y) = \frac{1}{N}\sum_{i=1}^N \widehat{F}_{i,\infty}(y), \quad \widehat{F}_{i,\infty}(y) = \sum_{k: y_i^k \le y} \widehat{\pi}_{ik}.$$

Estimators of stationary counterfactual distributions can be formed by replacing $\widehat{\boldsymbol{\beta}}_i(y_i^j)$ by $\widehat{\boldsymbol{\beta}}_i^g(y_i^j)$ and modifying the bias estimator, $\widehat{B}_{\boldsymbol{\pi}_i}$, in (3.3). The modified expression of the estimator of the bias is given in the Appendix. The resulting estimator of G_∞ is denoted by $\widehat{G}_\infty$.

3.1.3. *Third stage: Quantile effects.* The estimators of the short and long term quantile effects are:

$$\widehat{\mathsf{QE}}_t(\tau) = \widetilde{\phi}(\widehat{G}_t, \tau) - \widetilde{\phi}(\widehat{F}_t, \tau), \quad \widehat{\mathsf{QE}}_\infty(\tau) = \widetilde{\phi}(\widehat{G}_\infty, \tau) - \widetilde{\phi}(\widehat{F}_\infty, \tau), \tag{3.5}$$

where $\widetilde{\phi}$ is the generalized inverse or rearrangement operator

$$\widetilde{\phi}(F, \tau) = \int_0^\infty 1\{F(y) \le \tau\}\mathrm{d}y - \int_{-\infty}^0 1\{F(y) \ge \tau\}\mathrm{d}y.$$

which monotonizes $y \mapsto F(y)$ before applying the inverse operator.

3.2. **Inference.** To begin we highlight an important problem with standard analytical plug-in methods where the heterogeneous coefficients are estimated via fixed effect approaches. We show that these methods are not uniformly valid with respect to the degree of heterogeneity as measured by the variance of the coefficients. We propose a cross-sectional bootstrap scheme that has good computational properties and prove its uniform validity over a large class of data generating processes in Section 5.4.1.

3.2.1. *Inference problem.* While the inference problem affects all the functionals we consider, we illustrate it via a simple example that abstracts from other complications such as the need of debiasing. Consider the model

$$y_{it} = \beta_i + e_{it}, \quad \mathbb{E}(e_{it} \mid \beta_i) = 0, \quad \mathbb{E}(\beta_i) = \theta,$$

where we allow $\mathsf{Var}(\beta_i) \in [0, C]$ to be on a compact support, with zero as an admissible value. This class of data generating processes captures different degrees of heterogeneity that might arise in empirical applications. For simplicity, we assume e_{it} and β_i are both i.i.d. sequences in both i and t and mutually independent. The estimator of θ is

$$\widehat{\theta} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_i, \quad \widehat{\beta}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} = \beta_i + \frac{1}{T}\sum_{t=1}^{T} e_{it}.$$

The goal is to make inference about θ based on $\widehat{\theta}$ that remains uniformly valid over $\mathsf{Var}(\beta_i) \in [0, C]$.

Let $\overline{\beta} = \sum_{i=1}^{N} \beta_i / N$. The asymptotic distribution of $\widehat{\theta}$ is determined by two components:

$$\widehat{\theta} - \theta = (\widehat{\theta} - \overline{\beta}) + (\overline{\beta} - \theta),$$

where

$$\widehat{\theta} - \overline{\beta} \;=\; \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}, \quad \overline{\beta} - \theta = \frac{1}{N}\sum_{i=1}^{N}(\beta_i - \mathbb{E}(\beta_i)).$$

While both terms admit central limit theorems, they may have different rates of convergence. The rate of convergence of $\overline{\beta} - \theta$ depends on the degree of heterogeneity, $\mathsf{Var}(\beta_i)$, which is unknown. All we know is that it is supported on a compact set $[0, C]$ for some $C > 0$, with zero as an admissible boundary. This makes the final rate of convergence and the associated asymptotic distribution unknown. To illustrate this, consider two special extreme cases:

(1) *Strong heterogeneity*: It has been customary in the literature to assume that $\mathsf{Var}(\beta_i)$ is bounded away from zero, such that $\overline{\beta} - \theta = O_P(N^{-1/2})$. Then this term dominates in the expansion, yielding

$$\sqrt{N}(\widehat{\theta} - \theta) \to^d \mathcal{N}(0, \mathsf{Var}(\beta_i)).$$

(2) *Weak heterogeneity*: If $\mathsf{Var}(\beta_i)$ is *near the zero boundary*, such that when treated as a sequence, it decays faster than $O(T^{-1})$, then $\widehat{\theta} - \overline{\beta}$ becomes the

dominating term, yielding

$$\sqrt{NT}(\widehat{\theta} - \theta) \to^d \mathcal{N}(0, \mathsf{Var}(e_{it})).$$

We refer to this case as "weak heterogeneity" as it covers not only when β_i is homogeneous, but also when the degree of heterogeneity is small relative to the sample size as formalized by $\mathsf{Var}(\beta_i) = o(T^{-1})$. This case can be relevant in many empirical applications where the degree of heterogeneity is unknown and the time dimension is only moderately large.

It can be also shown that any degree of heterogeneity in between the above two extreme cases would lead to an unknown rate of convergence $\widehat{\theta} - \theta = O_P(\xi_{NT})$ where $\xi_{NT} \in [(NT)^{-1/2}, N^{-1/2}]$.

The unknown rate of convergence has consequences for the properties of standard inferential methods. Note that

$$\mathsf{Var}(\widehat{\theta}) = \frac{1}{NT}\mathsf{Var}(e_{it}) + \frac{1}{N}\mathsf{Var}(\beta_i). \tag{3.6}$$

A common method to estimate this variance is to plug in sample analogs of $\mathsf{Var}(e_{it})$ and $\mathsf{Var}(\beta_i)$. This procedure, however, does not provide uniformly valid inference. To understand the issue, consider the estimation of $\mathsf{Var}(\beta_i)$. If β_i were known, it could have been estimated by

$$\widetilde{\mathsf{Var}}(\beta_i) := \frac{1}{N}\sum_{i=1}^{N}(\beta_i - \overline{\beta})^2. \tag{3.7}$$

Replacing β_i with its consistent estimator $\widehat{\beta}_i$, we obtain

$$\widehat{\mathsf{Var}}(\beta_i) := \frac{1}{N}\sum_{i=1}^{N}(\widehat{\beta}_i - \widehat{\theta})^2.$$

Then $\widehat{\mathsf{Var}}(\beta_i) - \mathsf{Var}(\beta_i)$ has the following decomposition:

$$\underbrace{\frac{1}{N}\sum_{i=1}^{N}[(\widehat{\beta}_i - \widehat{\theta})^2 - (\beta_i - \overline{\beta})^2]}_{\beta\text{- estimation error}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N}[(\beta_i - \overline{\beta})^2 - \mathsf{Var}(\beta_i)]}_{\text{LLN- error}}, \tag{3.8}$$

where "LLN- error" refers to the error associated with the law of large numbers.

The main issue is that the β- estimation error cannot be controlled uniformly over $\mathsf{Var}(\beta_i) \in [0, C]$. Note that

$$\widehat{\beta}_i - \widehat{\theta} = \beta_i - \overline{\beta} + (\bar{e}_i - \bar{e}), \quad \bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it}, \quad \bar{e} = \frac{1}{N}\sum_{i=1}^{N}\bar{e}_i.$$

This leads to, if $T = o(N)$,

$$\beta\text{- estimation error} = \frac{1}{N}\sum_i (\bar{e}_i - \bar{e})^2 + \frac{2}{N}\sum_i (\bar{e}_i - \bar{e})(\beta_i - \overline{\beta}) \asymp O_P(T^{-1}).$$

The β-estimation error is an incidental parameter bias whose order does not adapt to $\mathsf{Var}(\beta_i)$, leading to first order bias of $\widehat{\mathsf{Var}}(\widehat{\theta})$ in the weak heterogeneity case. Thus, the estimation error $|\widehat{\mathsf{Var}}(\widehat{\theta}) - \mathsf{Var}(\widehat{\theta})|$ is lower bounded by an order of $O_P((NT)^{-1})$, which is *not* negligible when $\sqrt{NT}(\widehat{\theta} - \theta) \to^d \mathcal{N}(0, \mathsf{Var}(e_{it}))$. Consequently, the usual plug-in variance estimator using $\widehat{\mathsf{Var}}(\widehat{\theta})$ would lead to an asymptotically incorrect inference. To see this, note that the confidence interval will be distorted by a quantity of the same order as the length of the interval, that is

$$\mathsf{CI}_{1-p}(\theta) = \widehat{\theta} \pm z_{1-p/2}\sqrt{\widehat{\mathsf{Var}}(\widehat{\theta})} = \widehat{\theta} \pm z_{1-p/2}\sqrt{\mathsf{Var}(\widehat{\theta}) + O_P((NT)^{-1})},$$

where z_p is the p-quantile of the standard normal. The two terms inside the square root are of the same order since $\mathsf{Var}(\widehat{\theta}) = O((NT)^{-1})$, leading to incorrect coverage even asymptotically,

$$\Pr(\theta \in \mathsf{CI}_{1-p}(\theta)) = 1 - p + O(1).$$

Alternatively, ignoring $\mathsf{Var}(\beta_i)$ by setting $\widehat{\mathsf{Var}}(\beta_i) = 0$ would result in asymptotic under-coverage unless we are in the weak heterogeneity case. We can conclude that the plug-in method is not uniformly valid over $\mathsf{Var}(\beta_i) \in [0, C]$.

A simple solution in this example is to estimate $\mathsf{Var}(\widehat{\theta})$ by

$$\widehat{\mathsf{Var}}(\widehat{\theta}) = \frac{1}{N}\widehat{\mathsf{Var}}(\widehat{\beta}_i),$$

i.e., omit the first term of (3.6) in the plug-in estimator. This is an appropriate estimator since

$$N(\widehat{\mathsf{Var}}(\widehat{\theta}) - \mathsf{Var}(\widehat{\theta})) = \underbrace{\frac{1}{N}\sum_i [(\bar{e}_i - \bar{e})^2 - \mathsf{Var}(\bar{e}_i)] + \frac{2}{N}\sum_i (\bar{e}_i - \bar{e})(\beta_i - \overline{\beta})}_{\beta\text{- estimation error}} + \text{LLN- error}$$

automatically adapts to the rate of convergence of $\widehat{\theta}$. The key is that the recentering by $\mathsf{Var}(\bar{e}_i) = \mathsf{Var}(e_{it})/T$ reduces the order of the first term of the β- estimation error. Note that the LLN-error is of a higher order regardless of the magnitude of $\mathsf{Var}(\beta_i) \in [0, C]$. For example, the LLN- error$= 0$ if $\mathsf{Var}(\beta_i) = 0$ because $\beta_i = \overline{\beta}$ almost surely. In the next section, we propose a bootstrap method that is also robust to the degree of heterogeneity and is convenient for simultaneous inference on function-valued parameters.

3.2.2. *The cross-sectional bootstrap.* We develop a simple cross-sectional bootstrap scheme that is uniformly valid over a large class of data generating processes that include both weak and strong heterogeneity. We introduce the method in the context of the example from the previous section and provide implementation algorithms for the functionals of interest in our model in Appendix A. The formal theoretical results on the validity of cross-sectional bootstrap are given in Section 5.4.1.

The cross-sectional bootstrap is based on resampling with replacement of the estimated coefficients $\widehat{\beta}_i$ instead of the observations y_{it}. We call this a cross-sectional bootstrap because it is equivalent to resampling the entire time series $\{y_{i1}, \dots, y_{iT}\}$ of each cross-sectional unit. Let $\{\widehat{\beta}_i^* : i = 1, ..., N\}$ be random sample with replacement from $\{\widehat{\beta}_i : i = 1, ..., N\}$. The bootstrap draw of $\widehat{\theta}$ is

$$\widehat{\theta}^* = \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}_i^*.$$

We approximate the asymptotic distribution of $\widehat{\theta} - \theta$ by the bootstrap distribution of $\widehat{\theta}^* - \widehat{\theta}$. If q_p is the p-quantile of the bootstrap distribution of $|\widehat{\theta}^* - \widehat{\theta}|/s^*$, where s^* is the bootstrap standard deviation of $\widehat{\theta}^*$, then the p-confidence interval for θ is

$$\mathsf{CI}_p(\theta) = \widehat{\theta} \pm q_p s^*.$$

This procedure is very simple, but perhaps surprisingly, leads to the desired uniform coverage. To see this, note that the bootstrap variance of $\widehat{\theta}^*$ is

$$\frac{1}{N^2} \sum_{i=1}^{N} (\widehat{\beta}_i - \widehat{\theta})^2 = \frac{1}{N} \widehat{\mathsf{Var}}(\widehat{\beta}_i),$$

which, as we have shown above, is an estimator of $\mathsf{Var}(\widehat{\theta})$ that adapts automatically to the degree of heterogeneity.

Figure 3.1 provides a numerical comparison of analytical and cross-sectional bootstrap estimators of the standard deviation of $\widehat{\theta}$ using a design where $e_{it} \sim \mathcal{N}(0, 1)$, $\beta_i \sim \mathcal{N}(\theta, \mathsf{Var}(\beta_i))$, $\mathsf{Var}(\beta_i) \in \{0, 0.1, \dots, 1\}$, $\theta = 1$, $N = 100$, and $T = 10$. It reports the (true) standard deviation of $\widehat{\theta}$, based on $\mathsf{Var}(\widehat{\theta}) = \mathsf{Var}(e_{it})/(NT) + \mathsf{Var}(\beta_i)/N$, as a function of $\mathsf{Var}(\beta_i)$; together with averages over $5,000$ simulations of the following estimators:

(1) Standard plug-in: based on

$$\widehat{\mathsf{Var}}(\widehat{\theta}) = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \widehat{\beta}_i)^2 + \frac{1}{N^2} \sum_{i=1}^{N} (\widehat{\beta}_i - \widehat{\theta})^2.$$

This estimator is labeled as "over".

(2) Plug-in that omits the heterogeneity in β_i: based on the first term of the previous expression. This estimator is labeled as "under".

(3) Cross-sectional bootstrap standard deviation based on $1,000$ draws.

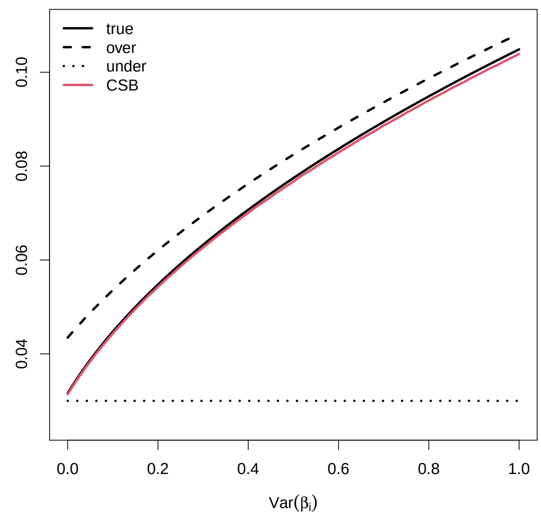

FIGURE 3.1. Comparison of analytical and cross-sectional bootstrap estimators of standard deviation of $\widehat{\theta}$ in this example.

We find that the standard analytical plug-in estimator overestimates the standard error for any degree of heterogeneity, whereas the analytical plug-in estimator that omits the heterogeneity in β_i underestimates the standard error in the presence of any heterogeneity. The mean of cross-sectional bootstrap estimator is very close to the standard error uniformly for all the degrees of heterogeneity considered, as predicted by the asymptotic theory.

3.2.3. *Simultaneous Inference.* The bootstrap algorithms for the model functionals presented in Appendix A are designed to construct confidence bands that cover the functionals simultaneously over the region of points of interest. For example, if we are interested in the scalar function $y \mapsto \xi(y)$ over $y \in \mathcal{Y}$, the asymptotic p-confidence band $\mathsf{CI}_p(\xi(y)) := [\widehat{\xi}_l(y), \widehat{\xi}_u(y)]$ is defined by the data dependent end-point functions $y \mapsto \widehat{\xi}_l(y)$ and $y \mapsto \widehat{\xi}_u(y)$ that satisfy

$$\Pr\left(\widehat{\xi}_l(y) \leq \xi(y) \leq \widehat{\xi}_u(y), y \in \mathcal{Y}\right) \to p \text{ as } N, T \to \infty.$$

We illustrate in Section 4 how this confidence bands can be used to test multiple hypotheses about the sign and shape of the functionals. Pointwise confidence intervals are special cases by setting the region $\mathcal{Y}$ to include only one point.

4. The Dynamics of Labor Income

4.1. **Data.** We employ data from the Panel Study of Income Dynamics for the years 1967 to 1996 (PSID, 2020). The sample selection is the same as in Hu et al. (2019) which restricts the sample to male heads of household working a minimum of 40 weeks.[5] We drop the worker-year observations where labor income is above the 99 sample percentile or below the 1 sample percentile, and keep workers observed for a minimum of 15 years. This selection results in an unbalanced panel with 1,629 workers and 33,338 worker-year observations.

The variables used in the analysis include measures of labor income, years of schooling, number of children, marital status, year of birth, survey year and an indicator for the worker being white. The years of schooling variable is constructed from the categorical variable highest grade completed with the following equivalence: 0-5 grades = 5 years, 6-8 grades = 7 years, 9-11 grades = 10 years, 12 grades = 12 years, some college = 14 years, and college degree = 16 years. Following the literature on labor income processes, we construct the outcome, y_{it}, as the residuals of the pooled regression of the logarithm of annual real labor income in 1996 US dollars, deflated by the CPI-U-RS price deflator, on indicators for marital status, number of children, year of birth and survey year. We refer to these residuals as labor income.

4.2. **Model coefficients.** We estimate the HDR model (2.1) with $\boldsymbol{x}_{it} = (1, y_{i,t-1})'$. We denote the model coefficients by $\boldsymbol{\beta}_i(y) = (\alpha_i(y), \rho_i(y))'$ and their bias corrected estimates by $\widehat{\boldsymbol{\beta}}_i(y) = (\widehat{\alpha}_i(y), \widehat{\rho}_i(y))'$, where we refer to $y \mapsto \alpha_i(y)$ as the intercept or level function and $y \mapsto \rho_i(y)$ as the slope or persistence function. These estimates are obtained using (3.1). The left panel of Figure 4.1 (Between Median) plots the kernel density of the estimated slope function $\widehat{\rho}_i(y)$ at a fixed value of y corresponding to the sample median of y_{it} pooled across workers and years. We find substantial heterogeneity between workers in this parameter. The density of the persistence coefficient includes both positive and negative values corresponding to positive and negative state dependencies in the labor income process at the median. The right panel of Figure 4.1 (Within Median) plots the pointwise sample median of the function $y \mapsto \widehat{\rho}_i(y)$ over a region $\mathcal{Y}$ that includes all the sample percentiles of the sample

[5]This sample is commonly employed in this literature as it represents full time full year workers.

values of y_{it} pooled across workers and years. The function is plotted with respect to the probability level of the sample percentile to facilitate interpretation. We find substantial heterogeneity in the slope within the distribution of the median worker. The slope is increasing with the percentile level indicating higher persistence parameter at the upper tail of the distribution. The two figures combined illustrate the existence of substantial heterogeneity in income dynamics both between and within workers.

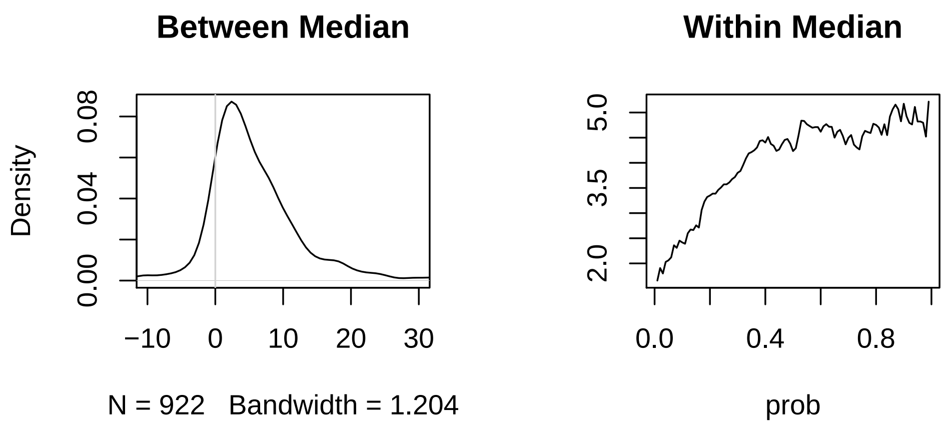

FIGURE 4.1. The left panel plots the cross-sectional density of $\widehat{\rho}_i(y)$ when y is fixed to the sample median of y_{it} in the pooled sample; the right panel plots the cross-sectional pointwise median of the function $y \mapsto \widehat{\rho}_i(y)$.

4.3. **Goodness of Fit.** To assess the model's performance, Figure 4.2 compares the empirical distributions of y_{it} in 1981 and 1991 with the corresponding distributions predicted by the HDR model. We find that the model provides a remarkably close fit to the empirical distribution for all the values of y, including the tails.

4.4. **Projections of Coefficients.** We obtain projections of the estimated coefficients to explore if specific worker characteristics are associated with the heterogeneity in the level and persistence of labor income between workers. We apply (3.2) with $\boldsymbol{z}_i$ including a constant, the initial labor income, number of years of schooling, a white indicator and year of birth, and $\boldsymbol{w}_i = \boldsymbol{z}_i$.

Figure 4.3 reports the estimates and 90% confidence bands of the projection coefficient function $y \mapsto \boldsymbol{\theta}(y)$ for education over a region $\mathcal{Y}$ that includes all the sample percentiles of the pooled sample of y_{it} with probability levels $\{0.10, 0.11, \ldots, 0.90\}$, plotted with respect to these probability levels. We find education level is associated

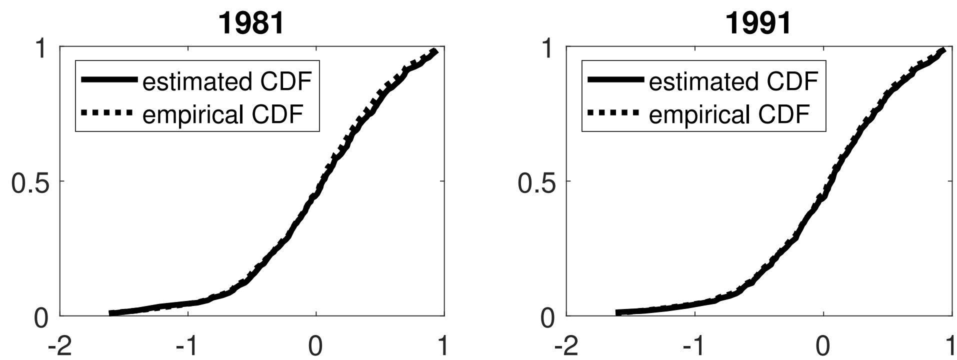

FIGURE 4.2. Empirical and predicted actual distributions, F_t, $t \in \{1981, 1991\}$.

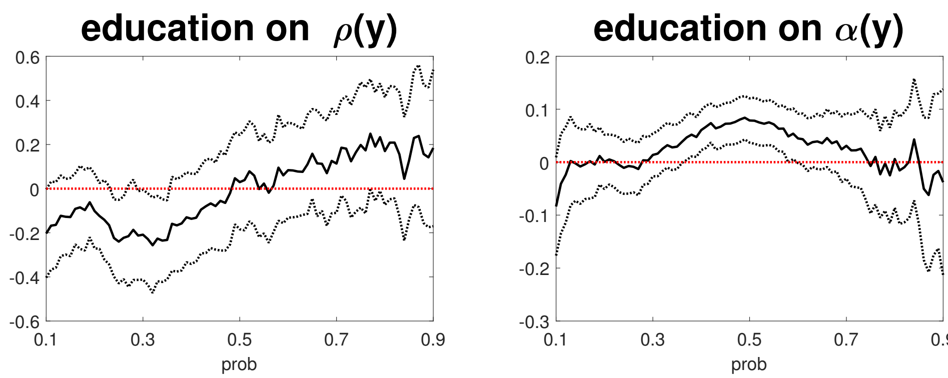

FIGURE 4.3. Projection coefficients of $\boldsymbol{\beta}_i(y) = (\rho_i(y), \alpha_i(y))$ on worker education levels. The confidence bands are obtained by cross-sectional bootstrap using Algorithm A.1 with $B = 500$.

with coefficient heterogeneity at some locations of the distribution. For example, the persistence parameter $\rho_i(y)$ is negatively associated with education at the bottom of the distribution, whereas the level parameter $\alpha_i(y)$ is positively associated with education in the middle of the distribution. The effect of education on $\rho_i(y)$ is increasing with y, although this pattern should be interpreted carefully as the function is not very precisely estimated, as reflected by the width of the confidence band.

4.5. **The Impact of Tax Policies.** An important implication of the HDR representation of labor income is that an individual's location in the income distribution in a specific time period partially depends on his location in previous periods. Moreover, the nature of this dependence varies by worker. This indicates that a shock to current labor income will determine the path of future income. To illustrate the presence and heterogeneity of this dependence we examine the impact on future income resulting from a negative shock to initial income. We implement the shock through two hypothetical tax policies corresponding to a proportional tax of 25 percent and the progressive tax between 0 and 50 percent on labor income in 1985.[6] We interpret this as a partial equilibrium analysis in that we change the level of initial income but keep all other aspects of the model constant. Specifically, we estimate the counterfactual distribution (2.4) for the transformations h_{it} given in (2.5) with $\kappa = 0.25$ and (2.6). Each transformation yields a counterfactual distribution of labor income in $t = 1986$. We also estimate the actual distribution and the corresponding quantile effects. We compare the estimates from the proposed HDR model with estimates obtained from the homogenous location-shift, homogenous location-shift with fixed effects and homogenous DR models described in Section 2.1.

The parameters of the location-shifts models are estimated by least squares; the parameters of the homogeneous DR model are estimated by distribution regression with Λ equal to the standard logistic distribution.

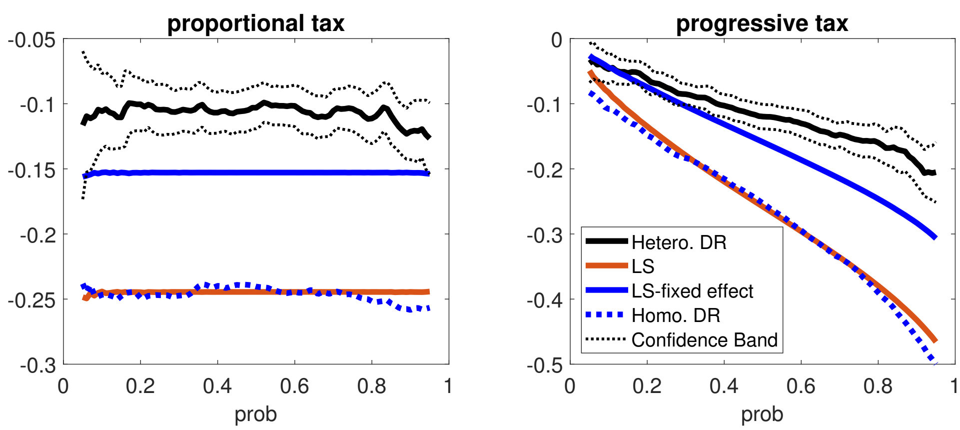

FIGURE 4.4. Quantile effects of counterfactual tax policies. Left panel: proportional tax; right panel: progressive tax. Hetero.DR refers to the proposed approach; LS refers to the homogeneous location-shift model; LS-fixed effect additionally adds fixed effects; Homo.DR refers to the homogeneous DR model. The confidence band for the estimation of the heterogeneous DR model (the proposed approach) is also plotted.

Figure 4.4 reports estimates and 90% confidence bands of the quantile effects for the proportional tax in the left panel and for the progressive tax in the right panel, together with the estimates obtained from the alternative models. The confidence bands are computed using Algorithm A.3 with $p = .90$, $B = 500$ and $\mathcal{T} = \{.05, .06, \ldots, .95\}$.

[6]We choose 1985 as the base year because it is the year with the largest number of observations in the dataset.

The estimates of the proportional tax show that the fully homogeneous location-shift and DR models predict that the tax reduces next period income almost in a one-for-one basis throughout the distribution. The model with fixed effects lowers the effect to about 15%, whereas the HDR model further ameliorates it to about 10%. The confidence band shows that there is no evidence of heterogeneous effects across the distribution. The comparison of the estimated effects from the progressive tax from each of the models reveals that allowing for heterogeneity again reduces the impact of the tax. However, the progressive nature of the tax produces heterogeneous effects across the distribution.

For both taxes, the confidence bands of the HDR model do not fully cover the estimates of the other three models. This comparison provides the basis of a specification test. The results in this plot are sufficient to formally reject the restrictions imposed by the alternative models.

4.6. **Dynamic Aspects of Relative Poverty.** We now analyze labor income mobility and the existence of "relative poverty" traps. We evaluate the probability of remaining in lower locations of the residual distribution noting that we refer to this as relative poverty as we acknowledge that the total income level may not be below the poverty line. We do so via the model from Section 2.3, where the conditional distribution is represented by a discrete Markov chain. We set the states for each worker as the observed values of y_{it}, that is $\mathcal{Y}_i := \{y_{it} : t = 1, ..., T\}$ and $K = T$.

Following Hu et al. (2019), we consider the following probabilities to describe mobility:

$$P_i(p, q, h) := \Pr(y_{i(t+h)} < y_p \mid y_{it} < y_q, \mathcal{F}_{it}), \quad i = 1, ..., N,$$

where y_p and y_q are the p-quantile and q-quantile of the distribution of labor income. These probabilities correspond to the following experiment: If we exogenously set labor income below y_q at time t, then $P_i(p, q, h)$ is the probability labor income is below y_p after h years.[7] For example, if we define the poverty line as the 10-percentile, then $P_i(0.1, 0.1, 5)$ is the probability that worker i would remain in poverty after 5 years if he falls below the poverty line due to, for example, a negative income shock.

[7]The probability $P_i(p, q, h)$ is identified if y_{it} is observed below y_p for some t. We restrict the sample to workers that satisfy this condition in the sample period to estimate these probabilities.

Our model allows the probabilities $P_i(p,q,h)$ to be heterogeneous across workers. To summarize this heterogeneity, we can examine the average probability:

$$\bar{P}(p,q,h) = \frac{1}{N}\sum_{i=1}^{N} P_i(p,q,h).$$

For instance, $\bar{P}(0.3, 0.1, 1)$ is the probability that a randomly chosen worker is below the 30-percentile if the previous year he was below the 10-percentile. We also examine quantiles of the probabilities such as:

$$Q_\tau(p,q,h)$$

which denotes the τ-quantile of $\{P_i(p,q,h) :, i = 1, ..., N\}$ for fixed (p,q,h). For example, $Q_{0.25}(0.3, 0.1, 1)$ is the first quartile of the probability that a worker is below the 30-percentile if the previous year he was below the 10-percentile.

The upper panel of Figure 4.6 plots $p \mapsto \bar{P}(p,q,h)$ for $p \in [0, 0.5]$, $q \in \{0.1, 0.25, 0.5\}$ and $h \in \{1, 2, 5\}$. We find heterogeneity with respect to the initial condition that vanishes with time due to the ergodicity of the process. The probability that a randomly selected worker remains below the 10-percentile after one year is more than 50%, whereas this probability decreases by about half if the worker was initially below the median. This difference in probabilities reduces after two years and almost vanishes after five years. The lower panel of Figure 4.6 plots $p \mapsto Q_\tau(p,q,h)$ for $p \in [0, 0.5]$, $q = 0.1$, $h \in \{1, 2, 5\}$ and $\tau \in \{0.1, 0.5, 0.9\}$. We uncover significant heterogeneity across workers that is hidden in the analysis of the mean worker. Even after 5 periods the deciles of the probability of remaining below the 10-percentile range from 0 to more than 0.9. This illustrates the importance of accounting for heterogeneity in understanding labor income risk.

Let $h_i(p)$ denote the recurrence time of y_p, that is, starting from $\{y_{it} < y_p\}$, the number of years h until the first occurrence of $\{y_{i(t+h)} > y_p\}$. For example, if $y_{0.10}$ is the poverty line, $h_i(0.10)$ is a random variable that measures the number of years that worker i takes to escape from poverty. Then,

$$\Pr(h_i(p) = h) = \Pr(y_{i(t+h)} > y_p, y_{i(t+h-1)} < y_p, ..., y_{i(t+1)} < y_p \mid y_{it} < y_p, \mathcal{F}_{it}),$$

which can be expressed as a functional of the parameters of the HDR model. Another interesting quantity is

$$H_i(p) = \sum_h h \Pr(h_i(p) = h),$$

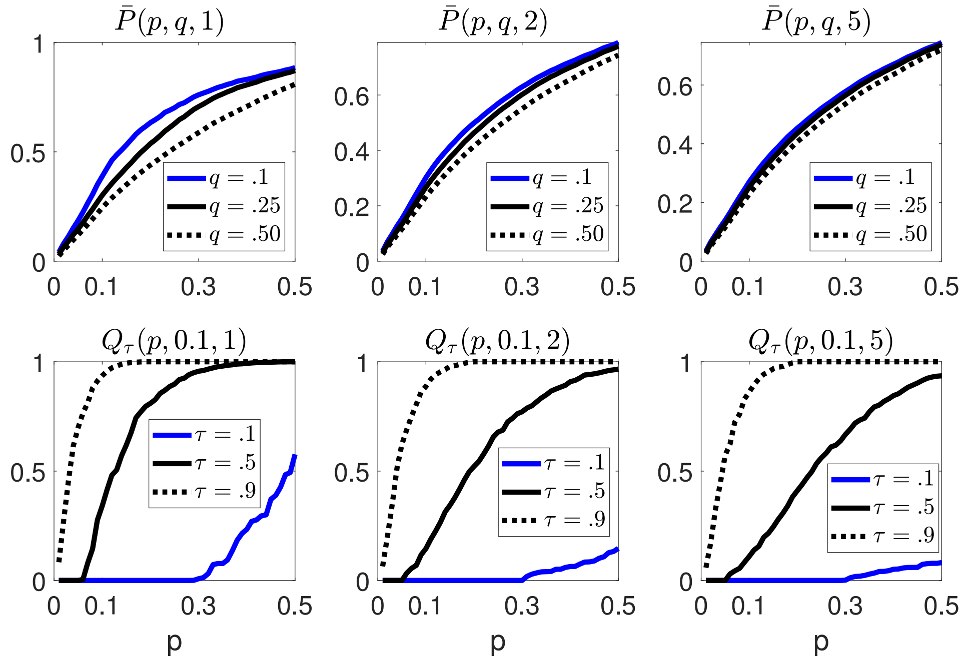

FIGURE 4.5. Means and quantiles of probabilities of income mobility. The upper panels report $\bar{P}(p, q, h)$ and the lower panels report $Q_\tau(p, q, h)$.

which gives the expected recurrence time for each individual. In the previous example, $H_i(0.10)$ gives the expected number of years that worker i would take to escape from poverty. Figure 4.6 plots a histogram of the estimated $H_i(0.10)$. More than 60% of the workers would escape from the poverty in two or less years, but about 10% of the workers would stay for more than 20 years. Table 4.6 reports several quantiles of the estimated $H_i(0.1)$ for groups stratified by education and race. We find substantial heterogeneity between workers associated with education and race. Whereas the deciles of the expected recurrence time range from 1 to 7 years for workers with at least high school, the corresponding value of 176 years indicates there are more than 10% of workers with less than high school that would never escape poverty. The distribution of the expected recurrence time also differs by race. The upper decile of the expected recurrence time is about 20 years higher for nonwhite than for white workers. This heterogeneity in the persistence of poverty has clear implications for the design of poverty alleviation policies. As they employ a different sample to ours and employ a different definition of "relative poverty" we do not directly compare these results to Lillard and Willis (1978). However, in addition to confirming the dependence in labor income documented in their study, we illustrate the remarkable difficulty facing some workers in escaping relative poverty.

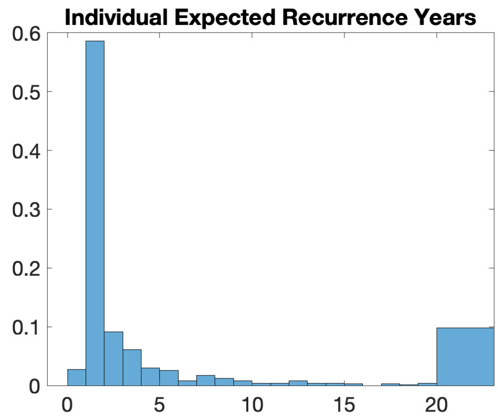

FIGURE 4.6. Histogram of expected recurrence time out-of-poverty in years, $H_i(0.10)$

TABLE 4.1. Quantiles of expected recurrence time out-of-poverty in years, $H_i(0.10)$, by education and racial groups

	Quantiles				
	0.10	0.25	0.50	0.75	0.90
All	1.00	1.00	1.47	3.63	19.45
Edu$<$ 12 years	1.00	1.35	2.92	9.75	175.8
Edu$\geq$ 12 years	1.00	1.00	1.20	2.39	7.37
White	1.00	1.00	1.27	3.12	13.88
non-White	1.00	1.11	1.81	5.52	33.91

4.7. **The Impact of Completing High School.** We now evaluate a hypothetical scenario in which workers with less than 12 years of schooling are assigned a high school degree (12 years of schooling). This also reflects a form of partial equilibrium analysis as the model parameters and the income distribution are based on the pre-intervention setting and we do not allow for possible general equilibrium effects. In particular, we assume that the resulting distribution is (2.4) with $h(\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}$ and g defined in (2.7). We set the values of $y_{i(t-1)}$ to the observed values in 1985 and assume that the change occurs at the beginning of 1986. We estimate the actual and counterfactual distributions in 1986 using (3.3), and the short and long term quantile effects using (3.5). To estimate the stationary distributions, we set the states for each worker in the Markov chain to the observed values of y_{it}, that is $\mathcal{Y}_i := \{y_{it} : t = 1, ..., T\}$ and $K = T$.

Figure 4.7 reports estimates and 90% confidence bands of QE_t in the left panel and QE_∞ in the right panel. The confidence bands are computed using Algorithm A.3 with $p = .90$, $B = 500$ and $\mathcal{T} = \{.05, .06, \dots, .95\}$. We find this intervention has heterogeneous effects across the distribution. The lower tail increases by around 7.5% after one year to almost 15% in the long run, whereas there is very little effect at the upper tail both in the short and long run. The confidence bands show that the results at the lower tail are statistically significant and allow us to formally reject the hypothesis of constant effects across the distribution. The magnitudes of the effects are economically noteworthy given the policy affects a relatively small fraction of the population. The results indicate that the increase in education for those with lower levels of education shifts the bottom tail of the labor income distribution of the entire population. This supports the commonly held policy view that increasing education of the lowly educated will reduce the level of inequality. There is no evidence of movements in the distribution at higher levels of labor income.

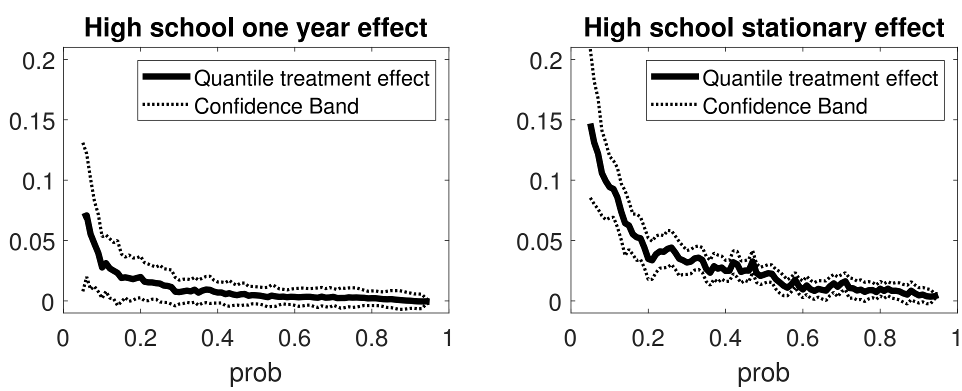

FIGURE 4.7. Quantile effects of counterfactual high school policy.

5. Asymptotic Theory

This section develops asymptotic theory for the estimators of the functionals of interest. We start by introducing some notation. Recall that the loss function for the estimation of the coefficients is: $Q_{y,i}(b) = T^{-1} \sum_{t=1}^{T} q_{y,it}(b)$, where

$$q_{y,it}(b) = 1\{y_{it} \leq y\}\Lambda(-\boldsymbol{x}_{it}'b) + 1\{y_{it} > y\}[1 - \Lambda(-\boldsymbol{x}_{it}'b)].$$

Let

$$
\begin{aligned}
\psi_{it}(y) &= \nabla q_{y,it}(\boldsymbol{\beta}_i(y)) \\
\varpi_{it}^d(y) &= \nabla^d q_{y,it}(\boldsymbol{\beta}_i(y)) - \mathbb{E}\nabla^d q_{y,it}(\boldsymbol{\beta}_i(y)), \quad d = 1, 2, 3. \\
\mathbb{A}_{1i}(y) &= [\mathbb{E}\nabla^2 q_{y,it}(\boldsymbol{\beta}_i(y))]^{-1}, \quad \mathbb{A}_{2i}(y) = \mathbb{E}\nabla^3 q_{y,it}(\boldsymbol{\beta}_i(y)),
\end{aligned}
$$

where all terms are defined using the true $\boldsymbol{\beta}_i(y)$. Specifically, when β is a vector, the third order derivative matrix $\nabla^3 q(\beta)$ is a $d_\beta \times d_\beta^2$ matrix, defined as $(\nabla B_1(\beta), ..., \nabla B_{d_\beta}(\beta))$, where $B_j(\beta)$ is the $d_\beta \times d_\beta$ Jacobian of the j th row of $\nabla^2 q$, here $d_\beta := \dim(\beta)$.

5.1. **Sampling.** The following assumptions relate to the properties of the sampling process.

Recall that $\mathcal{F}_{i1} \subset ... \subset \mathcal{F}_{iT}$ is the sequence of filtrations over time that include covariates and any time invariant variables for unit i.

Assumption 5.1 (Cross-section dimension)**.** *(i)* $\mathbb{E}(\gamma_i(y_1)|\boldsymbol{w}_i, \psi_{it}(y_2)) = 0$ *for any* y_1, y_2 *and* $i = 1, ..., N$.

(ii) The filtrations $\mathcal{F}_{iT}$ *are independent across* $i = 1, ..., N$.

(iii) $\{(Y_{it}, \boldsymbol{x}_{it}, \boldsymbol{w}_t) : t \leq T\}$ *are identically distributed across* $i = 1, ..., N$.

Assumption 5.2 (Time series dimension)**.** *There are universal constants* $C, c > 0$ *such that almost surely,*

$$
\begin{aligned}
&\max_{i \leq N} \mathbb{E}\left[\sup_{y \in \mathcal{Y}} \|\frac{1}{\sqrt{T}}\sum_t \varpi_{it}^d(y)\|^{8+c}\right] < C, \\
&\max_{i \leq N} \mathbb{E}\left[\sup_{|y_1 - y_2| \leq \epsilon} \|\frac{1}{\sqrt{T}}\sum_t \varpi_{it}^d(y_1) - \frac{1}{\sqrt{T}}\sum_t \varpi_{it}^d(y_2)\|^8\right] < C\epsilon^2,
\end{aligned}
$$

for $d = 1, 2, 3$.

Assumption 5.2 imposes conditions regarding serial dependence. We impose two high level conditions regarding the empirical process for weakly dependent data. It requires some primitive conditions, e.g., mixing conditions, so that $\{(Y_{it}, \boldsymbol{x}_{it}) : t \leq T\}$ is serially weakly dependent.

5.2. **Projections of Coefficients.** The main result of this section is to show that $\widehat{\boldsymbol{\theta}}(y) - \boldsymbol{\theta}(y)$ converges to a Gaussian process.

We start by defining the covariance kernel of the limiting process of $\widehat{\theta} - \theta$. For a given integer $M > 0$, let $Y_M = (y_1, ..., y_M)'$ be an arbitrary M-dimensional vector on

$\otimes_{i=1}^{M}\mathcal{Y}$. Let $S_{wz} := C_1^{-1}C_2(C_2'C_1^{-1}C_2)^{-1}$ where $C_1 = \mathbb{E}\boldsymbol{w}_i\boldsymbol{w}_i'$, $C_2 = \mathbb{E}\boldsymbol{w}_i\boldsymbol{z}_i'$, and

$$\begin{aligned}
V_\psi(y_k, y_l) &= \mathbb{E}\left\{(S_{wz}'\boldsymbol{w}_i\boldsymbol{w}_i'S_{wz}) \otimes \left[\mathbb{A}_{1i}(y_k)\mathbb{E}(\frac{1}{T}\sum_{s,t\le T}\psi_{it}(y_k)\psi_{it}(y_l)'|\boldsymbol{w}_i)\mathbb{A}_{1i}(y_l)\right]\right\}. \\
V_\gamma(y_k, y_l) &= \mathbb{E}\left\{(S_{wz}'\boldsymbol{w}_i\boldsymbol{w}_i'S_{wz}) \otimes \mathbb{E}(\gamma_i(y_k)\gamma_i(y_l)' \mid \boldsymbol{w}_i)\right\} \\
\Sigma_{NT}(y_k, y_l) &= \frac{1}{NT}V_\psi(y_k, y_l) + \frac{1}{N}V_\gamma(y_k, y_l) \\
\Sigma_{NT}(y) &= \Sigma_{NT}(y, y). \qquad (5.1)
\end{aligned}$$

The covariance kernel is now given by the limit of the elements of the following $M \times M$ matrix:

$$H_{\eta,NT} = (H_{\eta,NT}(y_k, y_l))_{M\times M}$$

where

$$H_{\eta,NT}(y_k, y_l) = \frac{\eta'\Sigma_{NT}(y_k, y_l)\eta}{[\eta'\Sigma_{NT}(y_k)\eta]^{1/2}[\eta'\Sigma_{NT}(y_l)\eta]^{1/2}}$$

and $\eta \in \mathbb{R}^{\dim(\mathsf{vec}\theta)}$. We make the following assumptions about the covariance kernel:

Assumption 5.3 (Covariance kernel)**.** *For any $\eta \in \mathbb{R}^{\dim(\mathsf{vec}\theta)}$ and $\|\eta\| > c > 0$, any integer $M > 0$, and any M-dimensional vector $Y_M = (y_1, ..., y_M)'$ on $\otimes_{i=1}^{M}\mathcal{Y}$, there is an $M \times M$ matrix H_η, such that almost surely,*

$$\lim_{N,T\to\infty} H_{\eta,NT} = H_\eta. \qquad (5.2)$$

In addition, there is $c_{Y_M,\eta} > 0$ such that

$$\lambda_{\min}(H_\eta) > c_{Y_M,\eta}. \qquad (5.3)$$

Here $c_{Y_M,\eta}$ may depend on Y_M, M and η.

Condition (5.3) is used to establish the finite dimensional distribution (f.i.d.i.) of $\eta'\mathsf{vec}(\widehat{\theta}(\cdot) - \theta(\cdot))$, which is required for a given Y_M, M and η. Therefore, the constant $c_{Y_M,\eta}$ is allowed to depend on these parameters. To show that Assumption 5.3 is reasonable even though the variance of $\gamma_i(y) = \boldsymbol{\beta}_i(y) - \boldsymbol{\theta}(y)\boldsymbol{z}_i$ may vary across y in the second-stage regression, we consider the following model:

$$\begin{aligned}
\gamma_i(y) &= \xi_{NT}(y)\bar{\gamma}_i(y), \quad \forall y \in \mathcal{Y}, \forall i \le N. \\
V_\gamma(y_k, y_l) &= \xi_{NT}(y_k)\xi_{NT}(y_l)V_{\bar{\gamma}}(y_k, y_l), \quad \inf_y \lambda_{\min}(V_{\bar{\gamma}}(y, y)) > c. \qquad (5.4)
\end{aligned}$$

Here $\xi_{NT}(y)$ is a bounded non-stochastic sequence that may converge to zero, whose rate depends on y; $\bar{\gamma}_i(y)$ is a random vector of "normalized" $\gamma_i(y)$, so $V_{\bar{\gamma}}(y, y)$ can be understood as a normalized covariance matrix. Hence the strength of $\gamma_i(y)$ is determined by the rate of convergence of $\xi_{NT}(y)$. Given this setting, consider the following special cases:

Case 1: $\xi_{NT}(y_k) = o(T^{-1/2})$ and $\xi_{NT}(y_l) \gg T^{-1/2}$. Here the explanatory power of $\boldsymbol{w}_i$ is strong for $\beta_i(y_k)$, but relatively weak for $\beta_i(y_l)$. Then

$$\lim_{N,T\to\infty} H_{\eta,NT}(y_k, y_l) = 0.$$

Note that the opposite case of $\xi_{NT}(y_k) \gg o(T^{-1/2})$ and $\xi_{NT}(y_l) = T^{-1/2}$ is also covered.

Case 2: Both $\xi_{NT}(y_k), \xi_{NT}(y_l) \gg T^{-1/2}$. Here the explanatory power of $\boldsymbol{w}_i$ is strong for both $\beta_i(y_k)$ and $\beta_i(y_l)$. Then

$$\lim_{N,T\to\infty} H_{\eta,NT}(y_k, y_l) = \lim_{N\to\infty} \frac{\eta' V_{\bar{\gamma}}(y_k, y_l)\eta}{[\eta' V_{\bar{\gamma}}(y_k, y_k)\eta]^{1/2}[\eta' V_{\bar{\gamma}}(y_l, y_l)\eta]^{1/2}},$$

where the limit of the right hand side is assumed to exist.

Case 3: Both $\xi_{NT}(y_k), \xi_{NT}(y_l) \ll T^{-1/2}$. Here the explanatory power of $\boldsymbol{w}_i$ is relatively weak for both $\beta_i(y_k)$ and $\beta_i(y_l)$. Then

$$\lim_{N,T\to\infty} H_{\eta,NT}(y_k, y_l) = \lim_{N\to\infty} \frac{\eta' V_{\psi}(y_k, y_l)\eta}{[\eta' V_{\psi}(y_k, y_k)\eta]^{1/2}[\eta' V_{\psi}(y_l, y_l)\eta]^{1/2}},$$

where the limit of the right hand side is assumed to exist.

So each element has a limit given on the right hand side. With sufficient variations (across y_k), one may assume the limit of the matrix $H_{\eta,NT}$ is non-degenerate that satisfies (5.3).

The following condition describes the continuity of some moment functions. For notational simplicity, we write

$$V_\gamma(y) := V_\gamma(y, y), \quad V_\psi(y) := V_\psi(y, y).$$

Assumption 5.4 (Continuity)**.** *There is a universal constant $C > 0$ such that for all $y_1, y_2 \in \mathcal{Y}$,*

$$\|V_\psi(y_1) - V_\psi(y_2)\| + \max_{i\le N} \|\mathbb{A}_{d,i}(y_1) - \mathbb{A}_{d,i}(y_2)\| < C|y_1 - y_2|, \quad d = 1, 2.$$

In addition, for all $\epsilon > 0$,

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\sup_{|y_1-y_2|<\epsilon} \frac{\|\gamma_i(y_1)\boldsymbol{w}_i' - \gamma_i(y_2)\boldsymbol{w}_i'\|^4}{M(y_1, y_2)^2}\right] < C\epsilon$$

$$\sup_{|y_1-y_2|<\epsilon} \frac{\|V_\gamma(y_1) - V_\gamma(y_2)\|}{M(y_1, y_2)} \le C\epsilon.$$

where $M(y_1, y_2) := \min\{\lambda_{\min}(V_\gamma(y_1)), \lambda_{\min}(V_\gamma(y_2))\}$.

Assumption 5.5 (Moment bounds). *There are universal constants* $C, c > 0$ *so that*

(i) For some $a > 0$,

$$\mathbb{E}\left[\sup_{y\in\mathcal{Y}}\left(\frac{\|\gamma_i(y)\boldsymbol{w}_i'\|}{\lambda_{\min}^{1/2}(V_\gamma(y))}\right)^4\right] < C.$$

(ii) Let Θ *be the parameter space for* $\{\beta_1(y), ..., \beta_N(y) : y \in \mathbb{R}\}$. *The following moment bounds hold:*

(a) $\max_{i\le N}\sup_{y\in\mathcal{Y}}[\|\mathbb{A}_{1i}(y)\| + \|\mathbb{A}_{2i}(y)\|] < C$

(b) $\sup_y \sup_{b\in\Theta}\max_{i\le N}[\|\nabla^3 Q_{y,i}(b)\| + \|\nabla^4 Q_{y,i}(b)\| + \|(\nabla^2 Q_{y,i}(b))^{-1}\|] = O_P(1)$

(c) $\max_{i\le N}\mathbb{E}\|\boldsymbol{w}_i\|^{8+c} < C$.

(iii) For all $y \in \mathcal{Y}$, *and all* $i = 1, ..., N$, *we have* $\min_{t\le T} y_{it} < y < \max_{t\le T} y_{it}$ *with probability approaching one.*

(iv) Let $S_{\psi,i}(y) = \mathsf{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_{it}(y)|\boldsymbol{w}_i\right)$. *Then almost surely,*

$$\min_i \inf_{y\in\mathcal{Y}} \lambda_{\min}(S_{\psi,i}(y)) > c.$$

In addition, all eigenvalues of C_1 *and* $C_2'C_2$ *are bounded away from zero and infinity, where* $C_1 = \mathbb{E}\boldsymbol{w}_i\boldsymbol{w}_i'$ *and* $C_2 = \mathbb{E}\boldsymbol{w}_i\boldsymbol{z}_i'$, *with* $\mathsf{rank}(C_2) \ge \dim(\boldsymbol{z}_i)$.

(v) $\frac{1}{T}\sum_t \mathbb{E}_i\boldsymbol{x}_{it}\boldsymbol{x}_{it}'$ *is of full rank for each* i, *where the expectation* $\mathbb{E}_i$ *is taken with respect to the joint density of* $(\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})$ *conditional on* $\mathcal{F}_{i1}$.

Condition (i) of this assumption requires that the fourth moment of $\gamma_i(y)$ is bounded by its second moment up to a constant, uniformly in y. To see the plausibility of this condition, again consider model (5.4). Then the left hand side of condition (i) becomes

$$\mathbb{E}\left[\sup_{y\in\mathcal{Y}}\left(\frac{\|\gamma_i(y)\boldsymbol{w}_i'\|^2}{\lambda_{\min}(V_\gamma(y))}\right)^2\right] = \frac{\frac{1}{N}\sum_{i=1}^N \mathbb{E}(\|\bar{\gamma}_i(y)\boldsymbol{w}_i'\|^4)}{\inf_{y\in\mathcal{Y}}\lambda_{\min}^2(V_{\bar{\gamma}}(y,y))},$$

which is upper bounded by a constant provided $\frac{1}{N}\sum_{i=1}^N \mathbb{E}(\|\bar{\gamma}_i(y)\boldsymbol{w}_i'\|^4) < C$. Other conditions of this assumption are standard. Condition (iii) requires that we only focus on $y \in \mathcal{Y}$ that are in the range of the observed outcomes. Finally, conditions (iv) and (v) of Assumption 5.5 identify the parameters $\theta(y)$ and $\boldsymbol{\beta}_i(y)$. To see this, note that the model implies

$$-\frac{1}{T}\sum_{t=1}^T \mathbb{E}_i\boldsymbol{x}_{it}\Lambda^{-1}\left(\Pr(y_{it}\le y|\mathcal{F}_{it})\right) = \left(\frac{1}{T}\sum_{t=1}^T \mathbb{E}_i\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)\boldsymbol{\beta}_i(y).$$

Inverting $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_i\boldsymbol{x}_{it}\boldsymbol{x}_{it}'$ leads to the identification of $\boldsymbol{\beta}_i(y)$. In addition, $\mathsf{rank}(C_2)\geq \dim(\boldsymbol{z}_i)$ implies the identification of $\boldsymbol{\theta}(y)$.

In the theorem below, L denotes the number of lags used for the Newey-West truncation for long-run variance, which is needed for analytical bias corrections.

Theorem 5.1. *Suppose $N=o(T^2)$ and $NL^2=o(T^3)$. Assumptions 5.1-5.5 hold. If $\boldsymbol{\beta}_i(y)$ is estimated using Jackknife-debias, then we additionally assume Assumption C.1. For any η such that $\|\eta\|>c>0$,*

$$\frac{\eta'\mathsf{vec}(\widehat{\theta}(\cdot)-\theta(\cdot))}{[\eta'\Sigma_{NT}(\cdot)\eta]^{1/2}}\Rightarrow\mathbb{G}(\cdot)$$

where $\Sigma_{NT}(y)=\frac{1}{NT}V_\psi(y)+\frac{1}{N}V_\gamma(y)$ and $\mathbb{G}(\cdot)$ is a centered Gaussian process with a covariance function $H(y_k,y_l)$ as the (k,l) element of H_η.

5.3. **Counterfactual distributions and quantile effects.** For a generic estimator $\widehat{F}(y)$ of $F(y)$, which may be one of the cross-sectional distributions that we discussed earlier, one can show that it has the following expansion

$$\widehat{F}(y)-F(y)=\frac{1}{N}\sum_{i=1}^{N}[\frac{1}{\sqrt{T}}d_{\psi,i}(y)+d_{\gamma,i}(y)]+o_P(\zeta_{NT}(y))$$

where $\zeta_{NT}(y)=(NT)^{-1/2}+N^{-1/2}\mathsf{Var}_t(d_{\gamma,i})^{1/2}$, and the two leading terms $d_{\psi,i}(y)$ and $d_{\gamma,i}(y)$ are asymptotically independent, and respectively capture the sampling variation from the first-stage and second-stage. The quantile effects have similar expansions:

$$\widehat{\mathsf{QE}}(\tau)-\mathsf{QE}(\tau)=\frac{1}{N}\sum_{i=1}^{N}[\frac{1}{\sqrt{T}}p_{\psi,i}(\tau)+p_{\gamma,i}(\tau)]+o_P(\bar{\zeta}_{NT}(\tau)) \tag{5.5}$$

where $\bar{\zeta}_{NT}(\tau)=(NT)^{-1/2}+N^{-1/2}\mathsf{Var}_t(p_{\gamma,i}(\tau))^{1/2}$, and $p_{\psi,i}(\tau)$ and $p_{\gamma,i}(\tau)$ are zero-mean uncorrelated terms.

We make the following additional assumptions, which are assumed to hold for all $\widetilde{\boldsymbol{x}}_{it}\in\{\boldsymbol{x}_{it},h_{it}(\boldsymbol{x}_{it})\}$, i.e., either the original variable $\boldsymbol{x}_{it}$ or the counterfactual $h_{it}(\boldsymbol{x}_{it})$. The formal definitions of $(d_{\psi,i},d_{\gamma,i},p_{\psi,i},p_{\gamma,i})$ depend on the specific $F\in\{F_t,G_t,F_\infty,G_\infty\}$ and $\mathsf{QE}\in\{\mathsf{QE}_t,\mathsf{QE}_\infty\}$, which are given in the Appendix. We emphasize that F_t,G_t respectively denote the actual and counterfactual distributions at time t and F_∞,G_∞ respectively denote the actual and counterfactual stationary distributions.

Let $\dot{\Lambda}(s)=\frac{d}{ds}\Lambda(s)$ and $\ddot{\Lambda}(s)=\frac{d^2}{ds^2}\Lambda(s)$.

Assumption 5.6 (Moment bounds). *(i)* $\sup_s |\dot{\Lambda}(s)| + \sup_s |\ddot{\Lambda}(s)| < C$.

(ii) $\mathbb{E}[\psi_{it}(y_k)|\beta_i(y_l), \widetilde{\boldsymbol{x}}_{it}, \boldsymbol{z}_i, \boldsymbol{w}_i, \gamma_i(y_l)] = 0$ *for any* $y_k, y_l \in \mathcal{Y}$.

(iii) $\mathbb{E}_t\|\widetilde{\boldsymbol{x}}_{it}\|^8 + \mathbb{E}\|\boldsymbol{x}_{it}\|^8\|g(\boldsymbol{z}_i) - \boldsymbol{z}_i\|^8 < C$.

(iv)

$$\mathbb{E}_t \sup_y \left[\frac{d_{\gamma,i}(y)}{\mathsf{Var}_t(d_{\gamma,i}(y))^{1/2}}\right]^4 < C, \quad \inf_y \lambda_{\min}(\mathsf{Var}_t(d_{\psi,i}(y))) > c > 0.$$

Assumption 5.7 (Continuity). *(i) There are* $C > 0$ *and* $k \geq 4$*, for any* $y_1, y_2 \in \mathcal{Y}$*,*

$$\begin{aligned}
&\mathbb{E}_t|\ddot{\Lambda}(-\widetilde{\boldsymbol{x}}_{it}'\beta_i(y_1)) - \ddot{\Lambda}(-\widetilde{\boldsymbol{x}}_{it}'\beta_i(y_2))|^k\|\widetilde{\boldsymbol{x}}_{it}\|^{2k} \leq C|y_1 - y_2|^k \\
&\mathbb{E}_t|\dot{\Lambda}(-\widetilde{\boldsymbol{x}}_{it}'\beta_i(y_1)) - \dot{\Lambda}(-\widetilde{\boldsymbol{x}}_{it}'\beta_i(y_2))|^k\|\widetilde{\boldsymbol{x}}_{it}\|^{k} \leq C|y_1 - y_2|^k \\
&\mathbb{E}_t|\ddot{\Lambda}(-\boldsymbol{x}_{it}'\beta_i^g(y_1)) - \ddot{\Lambda}(-\boldsymbol{x}_{it}'\beta_i^g(y_2))|^k\|\boldsymbol{x}_{it}\|^{2k} \leq C|y_1 - y_2|^k \\
&\mathbb{E}_t|\dot{\Lambda}(-\boldsymbol{x}_{it}'\beta_i^g(y_1)) - \dot{\Lambda}(-\boldsymbol{x}_{it}'\beta_i^g(y_2))|^k\|\boldsymbol{x}_{it}\|^{k} \leq C|y_1 - y_2|^k \\
&\mathbb{E}_t|\dot{\Lambda}(-\boldsymbol{x}_{it}'\beta_i^g(y_1)) - \dot{\Lambda}(-\boldsymbol{x}_{it}'\beta_i^g(y_2))|\|\boldsymbol{x}_{it}[g(\boldsymbol{z}_i) - \boldsymbol{z}_i]'\| \leq C|y_1 - y_2|.
\end{aligned}$$

(ii) There is $C > 0$*, for all* $\epsilon > 0$*,*

$$\begin{aligned}
&\sup_{|y_1 - y_2| < \epsilon} \frac{|\mathsf{Var}_t(d_{\gamma,i}(y_1)) - \mathsf{Var}_t(d_{\gamma,i}(y_2))|}{M(y_1, y_2)} \leq C\epsilon \\
&\sup_{|\tau_1 - \tau_2| < \epsilon} \frac{|\mathsf{Var}_t(p_{\gamma,i}(\tau_1)) - \mathsf{Var}_t(p_{\gamma,i}(\tau_2))|}{M_2(\tau_1, \tau_2)} \leq C\epsilon \\
&\mathbb{E}_t\left[\sup_{|y_1 - y_2| < \epsilon} \frac{|d_{\gamma,i}(y_1) - d_{\gamma,i}(y_2)|^4}{M(y_1, y_2)^2}\right] < C\epsilon.
\end{aligned}$$

where $M(y_1, y_2) = \mathsf{Var}_t(d_{\gamma,i}(y_1))^{1/2}\mathsf{Var}_t(d_{\gamma,i}(y_2))^{1/2}$ *and* $M_2(\tau_1, \tau_2) = \mathsf{Var}_t(p_{\gamma,i}(\tau_1))^{1/2}$ $\mathsf{Var}_t(p_{\gamma,i}(\tau_2))^{1/2}$.

We present the notation of $(d_{\gamma,i}, d_{\psi,i}, p_{\gamma,i}, p_{\psi,i})$ for all objects of interest in the appendix. The theorems below additionally require Assumptions C.2, C.3, which are based on some additional notation for the stationary distribution. We present them in the appendix.

Theorem 5.2. *Suppose the assumptions of Theorem 5.1 and Assumptions 5.6-5.7, C.3 hold. Then for* $F \in \{F_t, G_t, F_\infty, G_\infty\}$ *and* $\widehat{F} \in \{\widehat{F}_t, \widehat{G}_t, \widehat{F}_\infty, \widehat{G}_\infty\}$*, we have*

$$\frac{\widehat{F}(\cdot) - F(\cdot)}{\upsilon_{NT}(\cdot)} \Rightarrow \mathbb{G}(\cdot)$$

where $v_{NT}^2(y) = \frac{1}{NT}\mathsf{Var}_t(d_{\psi,i}(y)) + \frac{1}{N}\mathsf{Var}_t(d_{\gamma,i}(y))$ *and* $\mathbb{G}(\cdot)$ *is a centered Gaussian process with covariance kernel function*

$$\lim_{N,T} \frac{v_{NT}^2(y_k, y_l)}{v_{NT}(y_k)v_{NT}(y_l)}, \quad v_{NT}^2(y_k, y_l) := \frac{1}{NT}\mathbb{E}_t d_{\psi,i}(y_k)d_{\psi,i}(y_l) + \frac{1}{N}\mathbb{E}_t d_{\gamma,i}(y_k)d_{\gamma,i}(y_l),$$

assuming that $\lim_{NT}$ *exists for each pair* (y_k, y_l).

Theorem 5.3. *Suppose the assumptions of Theorem 5.2 and Assumption C.2 hold. Assume also, for all* $F \in \{F_t, G_t, F_\infty, G_\infty\}$, F *is continuously differentiable, whose density (denoted by* $\dot{F}$*) satisfies* $\inf_\tau \inf_{|y-\phi(F,\tau)|<C} \dot{F}(y) > c$ *for some* $C, c > 0$.

Then for $\mathsf{QE} \in \{\mathsf{QE}_t, \mathsf{QE}_\infty\}$ *and* $\widehat{\mathsf{QE}} \in \{\widehat{\mathsf{QE}}_t, \widehat{\mathsf{QE}}_\infty\}$,

$$\frac{\widehat{\mathsf{QE}}(\cdot) - \mathsf{QE}(\cdot)}{J_{NT}(\cdot)} \Rightarrow \mathbb{G}_{\mathsf{QE}}(\cdot),$$

where $J_{NT}^2(y) := J_{NT}^2(y, y)$, *with*

$$J_{NT}^2(y_k, y_l) := \frac{1}{NT}\mathbb{E}_t p_{\psi,i}(y_k)p_{\psi,i}(y_l) + \frac{1}{N}\mathbb{E}_t p_{\gamma,i}(y_k)p_{\gamma,i}(y_l),$$

and $\mathbb{G}_{\mathsf{QE}}(\cdot)$ *is a centered Gaussian process with covariance kernel function*

$$\lim_{N,T} \frac{J_{NT}^2(y_k, y_l)}{J_{NT}(y_k)J_{NT}(y_l)},$$

assuming that $\lim_{NT}$ *exists for each pair* (y_k, y_l).

5.4. **Discussion of asymptotic behavior.** To discuss the asymptotic behavior of the estimators, we closely examine the spot counterfactual effect $\mathsf{QE} = \mathsf{QE}_t$, estimated by $\widehat{\mathsf{QE}} = \widehat{\mathsf{QE}}_t$. We illustrate the complications that arise in our context and the need for a new inference method that is uniformly valid.

The asymptotic properties of other estimators are very similar. In this case, expansion (5.5) holds, with two leading terms $\frac{1}{\sqrt{T}}p_{\psi,i}$ and $p_{\gamma,i}$. The first term arises from the effect of estimating $\boldsymbol{\beta}_i(y)$. The second term is due to the cross-sectional projection, and can be expressed as

$$\begin{aligned}
p_{\gamma,i}(\tau) &= \kappa^{II}(\tau) \cdot (a) + \kappa^{II}(\tau) \cdot (b) + \kappa^0(\tau) \cdot (c) \\
(a) &= \boldsymbol{w}_i' S_{wz} \bar{G}(y_1)\gamma_i(y_1) \\
(b) &= \Lambda(-\boldsymbol{x}_{it}'\beta_i^g(y_1)) - \mathbb{E}_t\Lambda(-\boldsymbol{x}_{it}'\beta_i^g(y_1)) \\
(c) &= \Lambda(-\boldsymbol{x}_{it}'\beta_i(y_0)) - \mathbb{E}_t\Lambda(-\boldsymbol{x}_{it}'\beta_i(y_0)),
\end{aligned}$$

where $y_1 = \phi(G_t, \tau)$ and $y_0 = \phi(F_t, \tau)$, and other related quantities such as $\kappa^{II}(\tau)$ and $\kappa^0(\tau)$ are given in the supplementary appendix, and for now we them as constants that do not affect the asymptotic behavior. The key feature of our asymptotic analysis is that we allow any or all of the three terms to be either equal to or arbitrarily close to zero, leading to the robustness on the magnitude of $\mathsf{Var}_t(p_{\gamma,i}(\tau))$. Robustness to (a) is equivalent to robustness to the explanatory power in the random coefficient model $\boldsymbol{\beta}_i(y) = \boldsymbol{\theta}(y)\boldsymbol{w}_i + \gamma_i(y)$, while being robust on either (b) or (c) admits cross-sectional homogeneous models as special cases. This may also vary across quantile levels τ. For instance, at some quantiles, the model might be homogeneous in which both (b) and (c) are exactly zero; at other quantiles, the model might be heterogeneous, leaving one or both of them being nonzero. In practice, the heterogeneity is unobservable, and we make no assumptions about it.

The weak convergence of Theorem 5.3 implies that for each fixed τ,

$$\frac{\widehat{\mathsf{QE}}(\tau) - \mathsf{QE}(\tau)}{J_{NT}(\tau)} \to^d \mathcal{N}(0,1).$$

Consider a local sequence $\xi_{NT}(\tau) \geq 0$ and represent

$$p_{\gamma,i}(\tau) = \xi_{NT}(\tau)\bar{p}_{\gamma,i}(\tau)$$

where $\mathsf{Var}_t(\bar{p}_{\gamma,i}(\tau)) > c > 0$. So $\xi^2_{NT}(\tau)$ is the local rate of $\mathsf{Var}_t(p_{\gamma,i}(\tau))$, and

$$\widehat{\mathsf{QE}}(\tau) - \mathsf{QE}(\tau) = O_P(\frac{1}{\sqrt{NT}} + \frac{\xi_{NT}(\tau)}{\sqrt{N}}).$$

If $\xi^2_{NT}(\tau) > c$ for some constant $c > 0$, then

$$\sqrt{N}\mathsf{Var}_t(p_{\gamma,i}(\tau))^{-1/2}[\widehat{\mathsf{QE}}(\tau) - \mathsf{QE}(\tau)] \to^d \mathcal{N}(0, I).$$

The effect of the first-stage time series is absorbed by the cross-sectional regression. This leads to the usual $\sqrt{N}$- rate of convergence for two-step panel regressions. If, however, $\xi^2_{NT}(\tau) = o(T^{-1})$, then

$$\sqrt{NT}\mathsf{Var}_t(p_{\psi,i}(\tau))^{-1/2}[\widehat{\mathsf{QE}}(\tau) - \mathsf{QE}(\tau)] \to^d \mathcal{N}(0, I).$$

This occurs when either the observed characteristic $\boldsymbol{w}_i$ has almost full explanatory power of $\theta_i(\tau)$ or the model is cross-sectionally heterogeneous at the quantile level τ. The effect of the first stage time series regression plays the leading role in the final estimator, and the rate of convergence is much faster.

While the above considers two special cases, $\xi_{NT}(y)$ can be any sequence on a compact set $[0, C]$ that includes 0 as an admissive boundary point. This results in possibly varying rates of convergence for $\widehat{\mathsf{QE}}(\tau) - \mathsf{QE}(\tau)$ at various values of τ and data generating processes. This suggests the need for a uniform inferential method.

5.4.1. *Uniform inference using cross-sectional bootstrap.* The following result proves the validity of cross-sectional bootstrap in our setting, uniformly over a large class of data generating processes with varying degrees of coefficient heterogeneity.

Theorem 5.4. *Suppose the assumptions of Theorem 5.1 hold for all probability sequences $\{P_T : T \geq 1\} \subset \mathcal{P}$, where the universal constants do not depend on the specific choice of P_T. Then uniformly for all $\{P_T : T \geq 1\} \subset \mathcal{P}$,*

(i) We have

$$P_T\left(\eta' \mathsf{vec}(\boldsymbol{\theta}(y)) \in \mathsf{CI}_a(y), \forall y \in \mathcal{Y}\right) \to 1 - a,$$

where $\mathsf{CI}_a(y) = \{m : |\widehat{\boldsymbol{\theta}}(y) - m| \leq q_a \widetilde{s}^*(y)\}$, and q_a and $\widetilde{s}^*$ are defined corresponding to $(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ using the cross-sectional bootstrap algorithm in Appendix A.

(ii) For $(F, \widehat{F}) \in \{(F_t, \widehat{F}_t), (G_t, \widehat{G}_t), (F_\infty, \widehat{F}_\infty), (G_\infty, \widehat{G}_\infty)\}$

$$P_T\left(F(y) \in \mathsf{CI}_a(y), \forall \tau \in \mathcal{Y}\right) \to 1 - a.$$

where $\mathsf{CI}_a(y) = \{m : |\widehat{F}(y) - m| \leq q_a \widetilde{s}^*(y)\}$, and q_a and $\widetilde{s}^*$ are defined corresponding to the specific $(F, \widehat{F})$ using the cross-sectional bootstrap algorithm in Appendix A.

(iii) For $(\mathsf{QE}, \widehat{\mathsf{QE}}) \in \{(\mathsf{QE}_t, \widehat{\mathsf{QE}}_t), (\mathsf{QE}_\infty, \widehat{\mathsf{QE}}_\infty)\}$

$$P_T\left(\mathsf{QE}(\tau) \in \mathsf{CI}_a(\tau), \forall \tau \in \mathcal{T}\right) \to 1 - a.$$

where $\mathsf{CI}_a(\tau) = \{m : |\widehat{\mathsf{QE}}(\tau) - m| \leq q_a \widetilde{s}^*(\tau)\}$, and q_a and $\widetilde{s}^*$ are defined corresponding to the specific $(\mathsf{QE}, \widehat{\mathsf{QE}})$ using the cross-sectional bootstrap algorithm in Appendix A.

6. Simulation Evidence

We report finite-sample performances of our methods using simulations for two objects of interest: the projection parameters and the counterfactual treatment effects. Our simulation results illustrate the importance of bias correction and the uniform validity of our inference methods. The online appendix presents further simulation results using a calibrated model based on the PSID dataset.

Consider the following dynamic distribution regression model,

$$\begin{aligned} \Pr(y_{it} \leq y \mid \mathcal{F}_{it}) &= \Phi(y_{i(t-1)}\beta_i(y)), \\ \beta_i(y) &= \theta(y) w_i + \theta(y)\bar{\gamma}_i, \quad \mathbb{E}(\bar{\gamma}_i \mid w_i) = 0. \end{aligned}$$

with

$$\theta(y) = 3 \ \mathrm{sgn}(y-2)(y-2)^2, \ \text{for } y \in \mathcal{Y}.$$

We set $\mathcal{Y} = \{1.7, 1.8, ..., 2.3\}$, where the two endpoints of $\mathcal{Y}$ are chosen to avoid the estimation of extreme quantiles. The marginal probabilities $\Pr(y_{it} < 1.7)$ and $\Pr(y_{it} >$

2.3) are both approximately 0.1. We generate the simulated data by independently drawing $(e_{it}, w_i, \bar{\gamma}_i)$ from:

$$e_{it} \sim \mathcal{N}(0,1), \quad w_i \sim \text{Uniform}(1.5, 2.5), \quad \bar{\gamma}_i \sim \text{Uniform}(-0.5, 0.5).$$

Finally, y_{it} is initialized by $y_{i0} \sim \text{Uniform}(0.52, 1.52)$, and iteratively generated via

$$y_{it} = \theta^{-1}\left(\frac{e_{it}}{y_{i(t-1)}(w_i + \bar{\gamma}_i)}\right).$$

The parameters of this DGP are chosen so that $y_{i(t-1)}(w_i + \bar{\gamma}_i) > 0$ for all t almost surely. Therefore, $\Pr(y_{it} \le y \mid \mathcal{F}_{it}) = \Phi(y_{i(t-1)}\beta_i(y))$ is satisfied.

Figure 6 plots the variance of $\gamma_i(y)$, the noise level of $\beta_i(y)$, across $y \in \mathcal{Y}$. By construction, $\mathsf{Var}(\gamma(y))$ degenerates at $y = 2$, and increases as y deviates from 2, which affects the rate of convergence for estimating $\theta(y)$. The right panel plots the true standard error of the estimator $\widehat{\theta}(y)$, along with three estimators: the proposed bootstrap standard error $se^*(y)$ and two additional plug-in estimators defined below. The plug-in methods are clearly not robust to changes in $\mathsf{Var}(\gamma(y))$ across y.

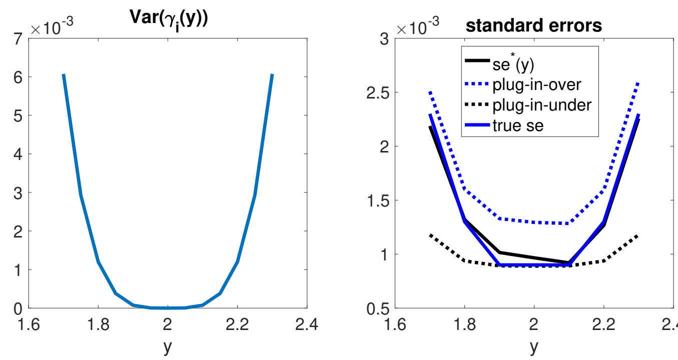

FIGURE 6.1. Left: $\mathsf{Var}(\gamma_i(y))$. Right: estimated and true standard errors of $\widehat{\theta}(y)$ for $y \in M$ in the dynamic DR model. In the right panel, we plot four "standard errors" for $y \in \mathcal{Y}$ under $N = T = 300$. The true standard error is calculated as the standard deviation of $\widehat{\theta}(y)$ from 1,000 simulations, while the other three are calculated using a fixed simulation of data.

6.1. **Coverage Probabilities of $\theta(y)$.** We examine the coverage properties of $\theta(y)$ and compare five inferential methods.

Proposed: The proposed uniform inference procedure using the interquartile range described in Remark A.1.

No-debias: This method does not debias, while all other steps are the same as the proposed method. We expect it to perform unsatisfactorily when $T \leq N$.

Conser-boot : This method replaces Steps 4-5 of Algorithm A.1 with:

Let q^*_τ be the $(1-\tau)$ th bootstrap quantile of

$$\left\{\sup_{y\in\mathcal{Y}} |\eta' \mathsf{vec}(\widehat{\theta}^*_b(y) - \widehat{\theta}(y))|\right\}_{b=1}^{B}$$

Compute the confidence band

$$[\eta' \mathsf{vec}(\widehat{\theta}(y)) - q^*_\tau, \eta' \mathsf{vec}(\widehat{\theta}(y)) + q^*_\tau].$$

Since the critical value q^*_τ is chosen for the worst case $y \in \mathcal{Y}$, we expect this method to be conservative.

Plug-in-over: This method plugs in the estimated standard error, while assumes the second stage regression error to be non-degenerate. Specifically, it estimates the two components $V_\psi(y)$ and $V_\gamma(y)$ in the standard error, and constructs confidence band:

$$[\eta' \mathsf{vec}(\widehat{\theta}(y)) - q_\tau(\eta' \widehat{\Sigma}_{NT}(y)\eta)^{1/2}, \eta' \mathsf{vec}(\widehat{\theta}(y)) + q_\tau(\eta' \widehat{\Sigma}_{NT}(y)\eta)^{1/2}],$$

where

$$\widehat{\Sigma}_{NT}(y) = \frac{1}{NT}\widehat{V}_\psi(y) + \frac{1}{N}\widehat{V}_\gamma(y).$$

As noted above, the estimation error of $\widehat{V}_\gamma(y)$ is not negligible when $V_\gamma(y)$ is near the boundary so this approach should have an over coveraging probability.

Plug-in-under: This method also plugs in the estimated standard error, but assumes that w_i fully explains $\beta_i(y)$, which is the standard treatment in the varying coefficient literature. Specifically, it replaces $\widehat{\Sigma}_{NT}(y)$ of the Plug-in-over method with

$$\widetilde{\Sigma}_{NT}(y) = \frac{1}{NT}\widehat{V}_\psi(y).$$

We expect the confidence band resulting from $\widetilde{\Sigma}_{NT}(y)$ would under-cover $\theta(y)$.

The last two "plug-in" procedures estimate $V_\psi(y)$ and $V_\gamma(y)$ by:

$$\begin{aligned}
\widehat{V}_\psi(y) &= \frac{1}{N}\sum_{i=1}^{N}(S'_{wz}\boldsymbol{w}_i\boldsymbol{w}'_i S_{wz}) \otimes \left[\widehat{\mathbb{A}}_{y,1i}\Xi(y)\widehat{\mathbb{A}}_{y,1i}\right]. \\
\widehat{V}_\gamma(y) &= \frac{1}{N}\sum_{i=1}^{N}(S'_{wz}\boldsymbol{w}_i\boldsymbol{w}'_i S_{wz}) \otimes (\widehat{\gamma}_i(y)\widehat{\gamma}_i(y)')
\end{aligned}$$

where computing the estimators $\widehat{\mathbb{A}}_{y,1i}$ and $\widehat{\gamma}_i(y)$ are straightforward. Meanwhile, we apply the Newey-West type estimator $\Xi(y)$ to estimate $\mathbb{E}(\frac{1}{T}\sum_{s,t\leq T}\psi_{it}(y_k)\psi_{it}(y_l)' \mid W)$, which is given by, for the bandwidth L,

$$\Xi(y) := \frac{1}{T}\sum_{t=1}^{T}\widehat{\psi}_{it}(y)\widehat{\psi}_{it}(y)' + \frac{1}{T}\sum_{h=1}^{L}(1-\frac{h}{L})\sum_{t>h}[\widehat{\psi}_{it}(y)\widehat{\psi}_{i(t-h)}(y)' + \widehat{\psi}_{i(t-h)}(y)\widehat{\psi}_{it}(y)'].$$

Table 6.1 summarizes the coverage probabilities of $\{\theta(y) : y \in \mathcal{Y}\}$ where $\mathcal{Y} = \{1.7, 1.8, ..., 2.3\}$ out of 1,000 replications. The results are genearlly as expected although the conservative bootstrap does not appear conservative.

TABLE 6.1. Coverage Probabilities of $\{\theta(y) : y \in \mathcal{Y}\}$

		Methods				
T	N	Proposed	No-debias	Conser-boot	Plugin-over	Plugin-under
50	300	0.942	0.576	0.944	0.994	0.894
	400	0.945	0.440	0.952	0.998	0.899
100	300	0.946	0.813	0.957	0.995	0.854
	400	0.947	0.740	0.943	0.996	0.852
200	300	0.951	0.914	0.954	0.975	0.632
	400	0.958	0.883	0.947	0.970	0.621

6.2. **Coverage probabilities for quantile treatment effects.** We now investigate the performance of our proposed estimators of counterfactual QEs arising from the change of w_i, and the corresponding inferential methods. Specifically, we investigate how the analytical debiasing and the jack-knife debiasing help reduce the MSEs of the estimators and improve the coverage probabilities of the confidence intervals relative those without debiasing. We consider the QE_t at $t = 1$ of a counterfactual increase in w_i by the amont of 0.5 for all i and the consequent changes in $\beta_i(y)$, while keeping $y_{i,t-1}$ of each i unchanged. We set $N = 200$ and $T = 50$.

TABLE 6.2. QE_t at $t = 1$

	Estimator MSE $\times 10^{-6}$					95% CI Coverage					
quantiles	15%	25%	50%	75%	85%	15%	25%	50%	75%	85%	joint
No-debias	0.47	0.32	0.27	0.39	0.63	0.95	0.94	0.90	0.89	0.93	0.89
Analytical	0.44	0.32	0.24	0.29	0.47	0.92	0.93	0.89	0.95	0.92	0.91
Jackknife	0.45	0.31	0.24	0.28	0.47	0.94	0.94	0.91	0.96	0.94	0.94

The left panel of Table 6.2 reports the MSE of the three versions of our estimators for the QE at the 15%, 25%, 50%, 70% and 85% quantiles. The right panel reports the coverage rates of the 95% confidence intervals, first separately for each of the five quantiles separately, and then uniformly for the five quantiles together ("joint"). The CIs are constructed based on our cross-sectional bootstrap procedures in Algorithm A.3. The results illustrate that both the analytical debiasing and the jackknife debiasing improve the finite-sample performances of our QE estimators and CIs. The estimator MSEs under the analytical debiasing and the jackknife debiasing are uniformly lower than those without debiasing across all five quantiles. There is also a noticeable improvement in the coverage rates of the uniform CIs with debiasing.

7. Conclusion

We develop estimation and inference methods for dynamic distribution regression panel models that incorporate heterogeneity both within and between units. An empirical investigation of labor income processes illustrates some economic insights our approach can provide. We find that accounting for individual heterogeneity is important in studying the potential impact of taxes on future income and evaluating how the income distribution responds to increases in the education levels of subpopulations of the data. Individual heterogeneity is also important in understanding income mobility and poverty persistence. Our model can be employed in a large number of empirical settings. Our focus here is a panel comprising repeated time series observations on the same unit. However, our approach may be applicable to a network setting in which there is contemporaneous dependence across units. We leave this extension to future work.

In the econometric analysis, the unknown degree of heterogeneity affects both the rate of convergence and the asymptotic distribution, making them unknown and *continuously varying* across different assumptions on the heterogeneity. While analytical plug-in methods in performing inference break down when degrees of heterogeneity vary, we prove that a simple cross-sectional bootstrap method is uniformly valid for a large class of data generating processes including the case of homogeneous coefficients.

Appendix A. The bootstrap algorithms

In this section we introduce the bootstrap algorithm for confidence bands.

Algorithm A.1 (Confidence Band for Projections of Coefficients)**.**

Step 0: Pick the confidence level p, number of bootstrap repetitions B, region $\mathcal{Y}$ and a component of the linear projection. This amounts to selecting a vector $\boldsymbol{\eta}$ such that $\boldsymbol{\eta}'\mathsf{vec}(\boldsymbol{\theta}(y))$ over $y \in \mathcal{Y}$ is the function of interest.

Step 1: For any $y \in \mathcal{Y}$, obtain the debiased DR coefficient estimates

$$\widehat{\boldsymbol{\beta}}(y) := \{\widehat{\boldsymbol{\beta}}_i(y) : i = 1, ..., N_{01}(y)\}$$

using (3.1), and the estimates of the linear projection, $\widehat{\boldsymbol{\theta}}(y)$, using (3.2).

Step 2: For any $y \in \mathcal{Y}$, let $\{(\widehat{\boldsymbol{\beta}}^*_i(y), \boldsymbol{w}^*_i, \boldsymbol{z}^*_i) : i = 1, ..., N_{01}(y)\}$ be a random sample with replacement from $\{(\widehat{\boldsymbol{\beta}}_i(y), \boldsymbol{w}_i, \boldsymbol{z}_i) : i = 1, ..., N_{01}(y)\}$. Compute

$$\widehat{\boldsymbol{\theta}}^*(y) = \sum_{i=1}^{N_{01}(y)} \widehat{\boldsymbol{\beta}}^*_i(y)\widehat{\boldsymbol{z}}^*_i(y)' \left(\sum_{i=1}^{N_{01}(y)} \widehat{\boldsymbol{z}}^*_i(y)\widehat{\boldsymbol{z}}^*_i(y)'\right)^{-1},$$

$$\widehat{\boldsymbol{z}}^*_i(y) := \sum_{j=1}^{N_{01}(y)} \boldsymbol{z}^*_j \boldsymbol{w}^{*\prime}_j \left(\sum_{j=1}^{N_{01}(y)} \boldsymbol{w}^*_j \boldsymbol{w}^{*\prime}_j\right)^{-1} \boldsymbol{w}^*_i.$$

Step 3: Repeat Step 2 for B times to obtain $\{\widehat{\boldsymbol{\theta}}^*_b(y)\}_{b=1}^B$ for each $y \in \mathcal{Y}$.

Step 4: Let q_τ be the booststrap τ-quantile of

$$\left\{\sup_{y\in\mathcal{Y}} \left|\frac{\boldsymbol{\eta}'\mathsf{vec}(\widehat{\theta}^*_b(y) - \widehat{\boldsymbol{\theta}}(y))}{s^*(y)}\right|\right\}_{b=1}^B$$

where $s^*(y)$ could be either the bootstrap standard deviation or rescaled interquartile range of $\{\boldsymbol{\eta}'\mathsf{vec}(\widehat{\boldsymbol{\theta}}^*_b(y)\}_{b=1}^B$. See remark A.1 below.

Step 5: Compute the asymptotic p-confidence band

$$\mathsf{CI}_p(\boldsymbol{\eta}'\mathsf{vec}(\boldsymbol{\theta}(y))) := [\boldsymbol{\eta}'\mathsf{vec}(\widehat{\boldsymbol{\theta}}(y)) - q_\tau s^*(y), \boldsymbol{\eta}'\mathsf{vec}(\widehat{\boldsymbol{\theta}}(y)) + q_\tau s^*(y)].$$

Remark A.1 (Standard Errors)**.** We show in the appendix that the bootstrap standard deviation $s^*(y)$ is consistent, $(s^*(y) - \sigma(y))/\sigma(y) = o_P(1)$, uniformly in y, where $\sigma(y) = \sqrt{\boldsymbol{\eta}'\Sigma_{NT}(y)\boldsymbol{\eta}}$. The bootstrap interquartile range rescaled with the standard normal distribution is an alternative: $s^*(y) = (q^*_{.75}(y) - q^*_{.25}(y))/(z_{.75} - z_{.25})$, where q^*_p is the bootstrap p-quantile of $\boldsymbol{\eta}'\mathsf{vec}(\widehat{\boldsymbol{\theta}}^*_b(y) - \widehat{\boldsymbol{\theta}}(y))$ and z_p is the p-quantile of the standard normal. Our theory covers both cases.

For the actual and counterfactual distributions, it is convenient to express the estimator in (3.3) as

$$\widehat{G}_t(y) = \frac{1}{N}\sum_{i=1}^N \Psi_i(y; h(\boldsymbol{x}_{it}), \widehat{\boldsymbol{\beta}}^g_i(y))$$

with

$$\begin{aligned}\Psi_i(y;\boldsymbol{x},\boldsymbol{b}) &= 1\{i \le N_{01}(y)\}\Lambda(-\boldsymbol{x}'\boldsymbol{b}) + \frac{N_1(y)}{N} \\ &\quad -1\{i \le N_{01}(y)\}\frac{1}{2T}\mathsf{tr}\left(\ddot{\Lambda}(-\boldsymbol{x}'\boldsymbol{b})\boldsymbol{x}\boldsymbol{x}'\widehat{\Sigma}_i(y)^{-1}\right),\end{aligned}$$

to simplify the notation.

Algorithm A.2 (Confidence Band for Actual and Counterfactual Distribution)**.**

Step 0: Pick the confidence level p, number of bootstrap repetitions B, and region $\mathcal{Y}$.

Step 1: For each $y \in \mathcal{Y}$, obtain the debised estimate $\widehat{G}_t$ from (3.3).

Step 2: Let $\{(\boldsymbol{x}_{it}^*, \widehat{\boldsymbol{\beta}}_i^*(y), \boldsymbol{w}_i^*, \boldsymbol{z}_i^*) : i = 1, ..., N_{01}(y)\}$ be a random sample with replacement from $\{(\boldsymbol{x}_{it}, \widehat{\boldsymbol{\beta}}_i(y), \boldsymbol{w}_i, \boldsymbol{z}_i) : i = 1, ..., N_{01}(y)\}$. Compute

$$\widehat{G}_t^*(y) = \frac{1}{N}\sum_{i=1}^{N}\Psi_i(y; h_{it}(\boldsymbol{x}_{it}^*), \widehat{\boldsymbol{\beta}}_i^{g*}(y), \quad \widehat{\boldsymbol{\beta}}_i^{g*}(y) = \widehat{\boldsymbol{\beta}}_i^*(y) + \widehat{\boldsymbol{\theta}}^*(y)[g(\boldsymbol{z}_i^*) - \boldsymbol{z}_i^*],$$

where $\widehat{\boldsymbol{\theta}}^*(y)$ is defined as in Step 2 of Algorithm A.1

Steps 3-5: The same as Steps 3-5 of Algorithm A.1, with $(\widehat{G}^*, \widehat{G})$ in place of $(\boldsymbol{\eta}'\mathsf{vec}(\widehat{\boldsymbol{\theta}}^*), \boldsymbol{\eta}'\mathsf{vec}(\widehat{\boldsymbol{\theta}}))$.

The bootstrap inference for the actual distribution $F_t(y)$ is a special case with $h(\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}$ and $g(\boldsymbol{z}_i) = \boldsymbol{z}_i$. Finally, the algorithm below computes the confidence band for the quantile effects.

Algorithm A.3 (Confidence Bands for Quantile Effect)**.**

Step 0: Pick the confidence level p, number of bootstrap repetitions B, and region of quantile indexes $\mathcal{T}$.

Step 1: For any $\tau \in \mathcal{T}$, obtain the estimate $\widehat{\mathsf{QE}}_t(\tau)$ using (3.5).

Step 2: Compute the bootstrap draws of $\widehat{\mathsf{QE}}_t(\tau)$:

(1) Obtain $\widehat{F}_t^*$ and $\widehat{G}_t^*$ as in step 2 of Algorithm A.2. For $\widehat{F}_t^*$, set $h(\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}$ and $g(\boldsymbol{z}_i) = \boldsymbol{z}_i$.

(2) For any $\tau \in \mathcal{T}$, calculate

$$\widehat{\mathsf{QE}}_t^*(\tau) = \widetilde{\phi}(\widehat{G}_t^*, \tau) - \widetilde{\phi}(\widehat{F}_t^*, \tau).$$

Steps 3-5: The same as Steps 3-5 of Algorithm A.1, with $(\widehat{\mathsf{QE}}_t^*, \widehat{\mathsf{QE}}_t)$ in place of $(\boldsymbol{\eta}'\mathsf{vec}(\widehat{\theta}^*), \boldsymbol{\eta}'\mathsf{vec}(\widehat{\theta}))$.

In step 2 of Algorithm A.3, it is important to use the same bootstrap sample to obtain $\widehat{F}_t^*$ and $\widehat{G}_t^*$ in order to mimic the dependence of $\widehat{F}_t$ and $\widehat{G}_t$ in the original sample.

Remark A.2 (Computation)**.** The most computationally expensive task is the computation of coefficient estimates, which is conducted only in Step 1 of the algorithms.

Remark A.3 (Stationary Distributions and Effects)**.** The bootstrap algorithms for stationary distributions and quantile effects are omitted because their steps are similar to the corresponding steps in Algorithms A.2 and A.3.

References

Arellano, M., Blundell, R. and Bonhomme, S. (2017). Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica* **85** 693–734.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies* **58** 277–297.

Arellano, M. and Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies* **79** 987–1020.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs* **43** 381.

Arellano, M. and Weidner, M. (2017). Instrumental variable quantile regressions in large panels with fixed effects. *Unpublished manuscript* .

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal* **63** 318–351.

Chen, S. (2021). Quantile regression with group-level treatments. Working Paper.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125.

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica* **81** 2205–2268.

Chernozhukov, V., Fernandez-Val, I. and Weidner, M. (2018). Network and panel quantile effects via distribution regression. *arXiv preprint:1803.08154* .

Chetverikov, D., Larsen, B. and Palmer, C. (2016). Iv quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica* **84** 809–833.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies* **82** 991–1030.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* **150** 71–85.

Fernández-Val, I. and Lee, J. (2013). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics* **4** 453–481.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics* **192** 291–312.

Fernández-Val, I. and Weidner, M. (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics* **10** 109–138.

Foresi, S. and Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association* **90** 451–466.

Galvao, A. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics* **164** 142–157.

Galvao, A. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics* **193** 92–112.

Gonçalves, S. and Kaffo, M. (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics* **186** 407–426.

Gu, J. and Koenker, R. (2017). Unobserved heterogeneity in income dynamics: An empirical bayes perspective. *Journal of Business & Economic Statistics* **35** 1–16.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* **27** 1152–1191.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* **72** 1295–1319.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Hart, P. E. (1976). The dynamics of earnings, 1963-1973. *The Economic Journal* **86** 551–565.

Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric society* 1377–1398.

Hirano, K. (2002). Semiparametric bayesian inference in autoregressive panel data models. *Econometrica* **70** 781–799.

Hsiao, C. and Pesaran, M. H. (2008). Random coefficient models. In *The econometrics of panel data*. Springer, 185–213.

Hu, Y., Moffitt, R. and Sasaki, Y. (2019). Semiparametric estimation of the canonical permanent-transitory model of earnings dynamics. *Quantitative Economics* **10** 1495–1536.

Kaffo, M. (2014). Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects. Tech. rep., mimeo.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal* **11** 377–395.

Kato, K., Galvao, A. F. and Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics* **170** 76–91.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91** 74–89.

Lillard, L. A. and Willis, R. J. (1978). Dynamic aspects of earning mobility. *Econometrica* **46** 985–1012.

Moffitt, R. and Zhang, S. (2018). Income volatility and the psid: Past research and new results. In *AEA Papers and Proceedings*, vol. 108.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society* 69–85.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.

Nickell, S. J. (1981). Biases in dynamic models with fixed effects. *Econometrica* **49** 1417–26.

Okui, R. and Yanagi, T. (2019). Panel data analysis with heterogeneous dynamics. *Journal of Econometrics* **212** 451–475.

Phillips, P. C. B. and Moon, H. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* **67** 1057–1111.

PSID, I. f. S. R., Survey Research Center (2020). Panel study of income dynamics, public use dataset .

Shorrocks, A. F. (1976). Income mobility and the markov assumption. *The Economic Journal* **86** 566–578.

Su, L., Shi, Z. and Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica* **84** 2215–2264.

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica: Journal of the Econometric Society* 311–323.

Zhang, Y., Wang, H. J. and Zhu, Z. (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics* **213** 54–67.

Department of Economics, Boston University, Boston, MA
Email address: ivanf@bu.edu

Department of Economics, University of Pennsylvania, Philadelphia, PA
Email address: waynegao@upenn.edu

Department of Economics, Rutgers University, New Brunswick, NJ 08901
Email address: yuan.liao@rutgers.edu

Department of Economics, Georgetown University,
Email address: fgv@georgetown.edu