NBER WORKING PAPER SERIES

QUANTILE REGRESSION WITH PANEL DATA

Bryan S. Graham
Jinyong Hahn
Alexandre Poirier
James L. Powell

Working Paper 21034
http://www.nber.org/papers/w21034

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2015

Earlier versions of this paper, with an initial draft date of March 2008, were presented under a variety of titles. We would like to thank seminar participants at Berkeley, CEMFI, Duke, UIUC, University of Michigan, Université de Montréal, NYU, Northwestern and at the 2009 North American Winter Meetings of the Econometric Society, the 2009 All-California Econometrics Conference at UC - Riverside, the 2014 Midwest Econometrics Group at the University of Iowa, and the 2nd annual IAAE Conference. We also thank the co-editors and two anonymous referees for their comments and insights. Financial support from the National Science Foundation (SES #0921928) is gratefully acknowledged. All the usual disclaimers apply. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Quantile Regression with Panel Data
Bryan S. Graham, Jinyong Hahn, Alexandre Poirier, and James L. Powell
NBER Working Paper No. 21034
March 2015, Revised August 2016
JEL No. C23,C31,J31

ABSTRACT

We propose a generalization of the linear quantile regression model to accommodate possibilities afforded by panel data. Specifically, we extend the correlated random coefficients representation of linear quantile regression (e.g., Koenker, 2005; Section 2.6). We show that panel data allows the econometrician to (i) introduce dependence between the regressors and the random coefficients and (ii) weaken the assumption of comonotonicity across them (i.e., to enrich the structure of allowable dependence between different coefficients). We adopt a "fixed effects" approach, leaving any dependence between the regressors and the random coefficients unmodelled. We motivate different notions of quantile partial effects in our model and study their identification. For the case of discretely-valued covariates we present analog estimators and characterize their large sample properties. When the number of time periods (T) exceeds the number of random coefficients (P), identification is regular, and our estimates are $\sqrt{N}$-consistent. When $T=P$, our identification results make special use of the subpopulation of stayers – units whose regressor values change little over time – in a way which builds on the approach of Graham and Powell (2012). In this just-identified case we study asymptotic sequences which allow the frequency of stayers in the population to shrink with the sample size. One purpose of these "discrete bandwidth asymptotics" is to approximate settings where covariates are continuously-valued and, as such, there is only an infinitesimal fraction of exact stayers, while keeping the convenience of an analysis based on discrete covariates. When the mass of stayers shrinks with N, identification is irregular and our estimates converge at a slower than $\sqrt{N}$ rate, but continue to have limiting normal distributions. We apply our methods to study the effects of collective bargaining coverage on earnings using the National Longitudinal Survey of Youth 1979 (NLSY79). Consistent with prior work (e.g., Chamberlain, 1982; Vella and Verbeek, 1998), we find that using panel data to control for unobserved worker heteroegeneity results in sharply lower estimates of union wage premia. We estimate a median union wage premium of about 9 percent, but with, in a more novel finding, substantial heterogeneity across workers. The 0.1 quantile of union effects is insignificantly different from zero, whereas the 0.9 quantile effect is of over 30 percent. Our empirical analysis further suggests that, on net, unions have an equalizing effect on the distribution of wages.

Bryan S. Graham
University of California - Berkeley 530
Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
bgraham@econ.berkeley.edu

Alexandre Poirier
Department of Economics
University of Iowa
W210 John Pappajohn Business Building
Iowa City, IA 52242
alexandre-poirier@uiowa.edu

Jinyong Hahn
University of California at Los Angeles
Box 951477
Los Angeles, CA 90095-1477.
hahn@econ.ucla.edu

James L. Powell
University of California at Berkeley
Department of Economics
508-1 Evans Hall #3880
Berkeley, CA 94720-3880
powell@econ.berkeley.edu

A online appendix is available at http://www.nber.orgappendix/w21034

Linear quantile regression analysis is a proven complement to least squares methods. Chamberlain (1994) and Buchinsky (1994) represent important applications of these methods to the analysis of earnings distributions, an area where continued application has proved especially fruitful (e.g., Angrist, Chernozhukov and Fernández-Val, 2006; Kline and Santos, 2013). Recent work has applied quantile regression methods to counterfactual and decomposition analysis (e.g., Machado and Mata, 2005; Firpo, Fortin and Lemieux, 2009; Chernozhukov, Fernández-Val and Melly, 2013), program evaluation (Athey and Imbens, 2006; Firpo, 2007) and triangular systems with endogenous regressors (e.g., Ma and Koenker, 2006; Chernozhukov and Hansen, 2007; Imbens and Newey, 2009).

The application of quantile regression methods to panel data analysis has proven to be especially challenging (e.g., Koenker, 2004 and Koenker, 2005, Section 8.7). The non-linearity and non-smoothness of the quantile regression criterion function in its parameters is a key obstacle. In an important paper, Kato, Galvao and Montes-Rojas (2012) show that a linear quantile regression model with individual and quantile-specific intercepts is consistent and asymptotically normal in an asymptotic sequence where both N and T grow. Unfortunately T must grow quickly relative to rates required in other large-N, large-T panel data analyses (e.g., Hahn and Newey, 2004). In a recent working paper, Arellano and Bonhomme (2016), develop correlated random effects estimators for panel data quantile regression. They extend a method of Wei and Carroll (2009), developed for mismeasured regressors, to operationalize their identification results. Other recent attempts to integrate quantile regression and panel data include Abrevaya and Dahl (2008), Rosen (2012), Chernozhukov, Fernández-Val, Hahn and Newey (2013), Harding and Lamarche (2014) and Chernozhukov, Fernández-Val, Hoderlein, Holzmann and Newey (2015). We return to the relationship between our own and prior work in the supplemental appendix to our paper: see Graham, Hahn, Poirier and Powell (2016).

Our contribution is a quantile regression method that accommodates some of the possibilities afforded by panel data. A key attraction of panel data for empirical researchers is its ability to control for unobserved correlated heteroegeneity (e.g., Chamberlain, 1984). A key attraction of quantile regression, in turn, is its ability to accommodate heterogeneous effects (e.g., Abrevaya, 2001). Our method incorporates both of these attractions. Our approach is a "fixed effects" one: it leaves the structure of dependence between the regressors and unobserved heterogeneity unrestricted. We further study identification and estimation in settings where T is small and N is large.

The starting point of our analysis is the textbook linear quantile regression model of Koenker and Bassett (1978). This model admits a (one-factor) random coefficients representation (e.g., Koenker, 2005, Section 2.6). While this representation provides a structural interpre-

tation for the slope coefficients associated with different regression quantiles, it also requires strong maintained assumptions. We show how panel data may be used to substantially weaken these assumptions in ways likely to be attractive to empirical researchers. In evaluating the strengths and weakness of our approach, we emphasize that our model is a strict generalization of the textbook quantile regression model.

In the next section we introduce our notation and model. Section 2 motivates several quantile partial effects associated with our model and discusses their identification. Section 3 presents our estimation results. Our formal results are confined to the case of discretely-valued regressors. This is an important special case, accommodating our empirical application, as well as applications in, for example, program evaluation as we describe below. The assumption of discrete regressors simplifies our asymptotic analysis, allowing us to present rigorous results in a relatively direct way.[1] Each of our estimators begins by estimating the conditional quantiles of the dependent variable in each period given all leads and lags of the regressors. This is a high-dimensional regression function and our asymptotic analysis needs to properly account for sampling error in our estimate of it. With discretely-valued regressors, we do not need to worry about the effects of bias in this first stage of estimation. This is convenient and substantially simplifies what nevertheless remains a complicated analysis of the asymptotic properties of our estimators.

While our theorems only apply to the discrete regressor case, we conjecture that our rates-of-convergence calculations and asymptotic variance expressions, would continue to hold in the continuous regressor case. This would, of course, require additional regularity conditions and assumptions on the first stage estimator. We elaborate on this argument in Section 5 below.

We present large sample results for two key cases. First, the regular case, where the number of time periods (T) exceeds the number of regressors (P). Our analysis in this case parallels that given by Chamberlain (1992) for average effects with panel data. Second, the irregular case, where $T = P$. This is an important special case, arising, for example, in a two period analysis with a single policy variable. Our analysis in this case makes use of so-called 'stayers', units whose regressor values do not change over time. Stayer units serve as a type of control group, allowing the econometrician to identify aggregate time trends (as in the textbook difference-in-differences research design).

With continuously-valued regressors there will generally be only an infinitesimal fraction of stayers in the population. Graham and Powell (2012) show that this results in slower than $\sqrt{N}$ rates of convergence for average effects. We mimic this continuous case in our quantile

[1] Chernozhukov, Fernández-Val, Hahn and Newey (2013) study identification in discrete choice panel data models with discrete regressors.

effects context by considering asymptotic sequences which place a shrinking mass on stayer regressor realizations as the sample size grows. We argue that these "discrete bandwidth asymptotics" approximate settings where covariates are continuously-valued and, as such, there is only an infinitesimal fraction of exact stayers, while keeping the convenience of an analysis based on discrete covariates. This tool may be of independent interest to researchers interested in studying identification and estimation in irregularly identified semiparametric models.[2] Our approach is similar in spirit to Chamberlain's (1987, 1992) use of multinomial approximations in the context of semiparametric efficiency bound analysis.

Section 4 illustrates our methods in a study of the effect of collective bargaining coverage on the distribution of wages using an extract from the National Longitudinal Survey of Youth 1979 (NLSY79). The relationship between unions and wage inequality is a long-standing area of analysis in labor economics. Card, Lemieux and Riddell (2004) provide a recent survey of research. Like prior researchers we find that allowing a worker's unobserved characteristics to be correlated with their union status sharply reduces the estimated union wage premium (e.g., Chamberlain, 1982; Jakubson, 1991; Card, 1995; Vella and Verbeek, 1998). This work has focused on models admitting intercept heterogeneity in earnings functions. Our model incorporates slope heterogeneity as well. It further allows for the recovery of quantiles of these slope coefficients. We find a median union wage effect of 9 percent, close to the mean effect found by, for example, Chamberlain (1982). In a more novel finding, however, we find substantial heterogeneity in this effect across workers. For many workers the returns to collective bargaining coverage are close to, and insignificantly different from, zero. While, for a smaller proportion of workers, the returns to coverage are quite high, in excess of 20 percent.

We are only able to identify quantile effects for the subpopulation of workers that move in and/or out of the union sector during our sample period (i.e., "mover" units). Movers constitute just over 25 percent of our sample. For this group we can study inequality in a world of universal collective bargaining coverage versus one with no such coverage. We find that the average conditional 90-10 log earnings gap would be over 20 percent lower in the universal coverage counterfactual. Our results are consistent with unions having a substantial compressing effect on the distribution of wages (at least within the subpopulation of movers).

While the asymptotic analysis of our estimators is non-trivial, their computation is straightforward.[3] The first two steps of our procedure are similar to those outlined in Chamberlain

[2]Examples include sample selection models with "identification at infinity", (smoothed) maximum score and regression discontinuity models.

[3]A short STATA script which replicates our empirical application is available for download from the first authors' website.

(1994), consisting of sorting and weighted least squares operations. The final step of our procedures consist of either averaging, or a second sorting step, depending on the target estimand. While we do not provide a formal justification for doing so, we recommend the use of the bootstrap as a convenient tool for inference (the results of, for example, Chernozhukov, Fernández-Val and Melly (2013), suggest that the use of the bootstrap is valid in our setting).

Section 5 outlines a few simple extensions of our basic approach. Section 6 concludes with some suggestions for further research and application. All proofs are relegated to the appendix.

1 Setup and model

The econometrician observes N independently and identically distributed random draws of the $T \times 1$ outcome vector $\mathbf{Y} = (Y_1, \ldots, Y_T)'$ and $T \times P$ regressor matrix $\mathbf{X} = (X_1, \ldots, X_T)'$. Here Y_t corresponds to a random unit's period t outcome and $X_t \in \mathbb{X}_{tN} \subset \mathbb{R}^P$ to a corresponding vector of period t regressors.[4] The outcome is continuously-valued with a conditional cumulative distribution function (CDF), given the entire regressor sequence $\mathbf{X} = \mathbf{x}$, of $F_{Y_t|\mathbf{X}}(y_t|\mathbf{x})$. This CDF is invertible in y_t, yielding the conditional quantile function

$$Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x}) = F^{-1}_{Y_t|\mathbf{X}}(y|\mathbf{x}).$$

Let $Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}) = \left(Q_{Y_1|\mathbf{X}}(\tau|\mathbf{x}), ..., Q_{Y_T|\mathbf{X}}(\tau|\mathbf{x})\right)'$ be the $T \times 1$ stacked vector of period-specific conditional quantile functions. Let $\mathbf{W} = \mathbf{w}(\mathbf{X})$ denote a $T \times R$ matrix of deterministic functions of $\mathbf{X}$ (and $\mathbf{w} = \mathbf{w}(\mathbf{x})$). We assume that $Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x})$ takes the semiparametric form

$$Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}) = \mathbf{x}'\beta(\tau;\mathbf{x}) + \mathbf{w}'\delta(\tau) \tag{1}$$

for all $\mathbf{x} \in \mathbb{X}_N^T = \times_{t\in\{1,\ldots,T\}}\mathbb{X}_{tN}$ and all $\tau \in (0,1)$. While a subset of our estimands only require (1) to hold for a single (known) τ, for convenience, we maintain the stronger requirement that (1) holds for all $\tau \in (0,1)$.

A key feature of (1) is that the coefficients multiplying the elements of X_t – $\beta(\tau;\mathbf{x})$ – are nonparametric functions of $\mathbf{x}$, while those multiplying the elements of W_t – $\delta(\tau)$ – are constant in $\mathbf{x}$ (W_t corresponds to the transpose of the t^{th} row of $\mathbf{W}$). In what follows we will refer to $\delta(\tau)$ as the common coefficients and $\beta(\tau;\mathbf{x})$ as, depending on the context, the

[4]The first element of this vector is a constant unless noted otherwise. The notation $\mathbb{X}_{tN}$ reflects the fact that we allow the support of $\mathbf{X}$ to vary with the sample size N in a way that is specified later on.

correlated, heterogenous or individual-specific coefficients.[5]

The model of equation (1) is closely related to the class of varying coefficient (or functional) quantile regression models, studied in Honda (2004) and Kim (2007). In particular, the fact that the coefficient on $\mathbf{w}$ does not depend on $\mathbf{x}$ implies that our model is a partially varying coefficient quantile regression model: see Wang et al. (2009) and Cai and Xiao (2012). Letting $\mathbf{V}$ be an additional observed covariate, we can write that model as

$$Q_{\mathbf{Y}|\mathbf{X},\mathbf{V}}(\tau|\mathbf{x},\mathbf{v}) = \mathbf{x}'\beta(\tau;\mathbf{v}) + \mathbf{w}'\delta(\tau),$$

and letting $\mathbf{V} = \mathbf{X}$ yields our model as a special case. Despite this connection, the identification of our model cannot be established using results from this literature since they require non-degeneracy of the conditional distribution of $\mathbf{V}|\mathbf{X}$: see for example assumptions (C2) and (C3) in Cai and Xiao (2012) or condition 2 in Kim (2007). This implies the necessary exclusion restriction that $\mathbf{V}$ cannot be a subset of the matrix $\mathbf{X}$. We will use the panel structure of our model will allow us to achieve identification of the distribution of the varying coefficients.

Model (1), with conditional expectations replacing conditional quantiles, was introduced by Chamberlain (1992) and further analyzed by Graham and Powell (2012) and Arellano and Bonhomme (2012). The quantile formulation is new.

A direct justification for (1) is provided by the one-factor random coefficients model

$$Y_t = X_t'\beta\left(U_t;\mathbf{X}\right) + W_t'\delta(U_t),\ \ U_t|\,\mathbf{X} \sim \mathcal{U}\left[0,1\right]. \tag{2}$$

Validity of the resulting linear quantile representation (1) – which must be nondecreasing in the argument τ almost surely in $\mathbf{X}$ – requires further restrictions on the functions $\beta\left(\tau;\mathbf{x}\right)$ and $\delta\left(\tau\right)$ and the regressors X_t and $W_t = w_t(\mathbf{X})$ (cf., Koenker (2005)), which we implicitly assume throughout.

We provide two, more primitive, derivations of (1) immediately below. The first follows from a generalization of the linear quantile regression model for cross sectional data (e.g., Koenker and Bassett, 1978; Koenker, 2005). The second follows from a generalization of the textbook linear panel data model (e.g., Chamberlain, 1984).

[5]We will be interested in identifying and estimating functionals of $\beta(\tau;\mathbf{x})$, the correlated random coefficients, and therefore we do not consider the object of interest to be the nonparametric function $\mathbf{x}'\beta(\tau;\mathbf{x})$, as it would be in a partially linear quantile regression model, e.g. Lee (2003).

Generalizing the linear quantile regression model

The strongest interpretation of the estimands we introduce below occurs when we can characterize the relationship between the quantile regression coefficients in (1) and quantiles of the individual components of B_t in the random coefficients model:

$$Y_t = X_t' B_t. \tag{3}$$

The τ^{th} quantile of B_{pt} – $F^{-1}_{B_{pt}}(\tau)$ – has a simple economic interpretation: the "return" to a unit increase in the p^{th} component of X_t is smaller for 100τ percent of units, and greater for $100(1-\tau)$ percent of units. In what follows we call $F^{-1}_{B_{pt}}(\tau)$ the τ^{th} *unconditional quantile effect* (UQE) of a (period t) unit change in X_{pt}.

In the cross-section setting ($T = 1$) we can construct a mapping between quantiles of the individual elements of B_1 in (3) and their corresponding quantile regression coefficients in the linear quantile regression of Y_1 onto X_1 if (i) X_1 is independent of B_1, (ii) there exists a non-singular rotation $B_1^* = A^{-1}B_1$ such that the elements of B_1^* are comonotonic (i.e., perfectly concordant) and (iii) the elements of $x_1'A$ are non-negative for all $x_1 \in \mathbb{X}_1$.

Under (i) through (iii) we have

$$Q_{Y_1|X_1}(\tau|\, x_1) = x_1' b(\tau)$$

for all $x_1 \in \mathbb{X}_1$ and $\tau \in (0,1)$ and, critically, that

$$b_p(U) \sim B_{p1},\ U \sim \mathcal{U}[0,1]. \tag{4}$$

Under (4) quantiles of B_{p1} (i.e, the UQE of a unit change in X_{p1}) are identified by the *rearranged* quantile regression coefficients on X_{p1}:

$$\begin{aligned} \beta_p(\tau) &= \inf\{c \in \mathbb{R} : \Pr(B_{p1} \leq c) \geq \tau\} \\ &= \inf\{c \in \mathbb{R} : \Pr(b_p(U) \leq c) \geq \tau\}, \end{aligned}$$

where $\beta_p(\tau)$ equals the τ^{th} unconditional quantile effect (UQE) of a unit change in X_{p1}.

Requirement (iii) is related to the quality of the linear approximation of the quantile regression process. Requirements (i) and (ii) are economic in nature and restrictive.[6] Assuming independence of X_1 and B_1 is very strong outside of particular settings (e.g., randomized

[6]The requirement that comonotonicity of the random coefficients needs to hold for only a *single* rotation is an implication of equivariance of quantile regression to reparametrization of design (e.g., Koenker and Bassett, 1978).

control trials), but the issues involved, and how to reason about them, are familiar. The requirement of comonotonicity of the random coefficients, possibly after rotation, is more subtle and less familiar. It too has strong economic content.

To illustrate some of the issues associated with the comonotonicity requirement, as well as how panel data may be used to weaken it (as well as the assumption of independence), it is helpful to consider, as we do in the empirical application below, the relationship between the distribution of wages and collective bargaining coverage.

If we let Y_t equal the logarithm of period t wages, and UNION_t be a binary variable indicating whether a worker's wages are covered by a collective bargaining agreement in period t or not, we can write, without loss of generality,

$$Y_t = B_{1t} + B_{2t}\text{UNION}_t,\ t = 1, \ldots, T. \tag{5}$$

The the τ^{th} quantile of B_{2t} – $F^{-1}_{B_{2t}}(\tau)$ – has a simple economic interpretation: the "return" to collective bargaining coverage is smaller for 100τ percent of workers, and greater for $100(1-\tau)$ percent of workers.

Now consider the coefficient on UNION_1 in the τ^{th} linear quantile regression of log wages in period 1 onto a constant and UNION_1. This coefficient, $b_2(\tau)$, equals

$$b_2(\tau) = F^{-1}_{B_{11}+B_{21}|X_1}(\tau|\,\text{UNION}_1 = 1) - F^{-1}_{B_{11}|X_1}(\tau|\,\text{UNION}_1 = 0),$$

which, without further assumptions, is not a quantile effect.

Requirement (i) – independence – yields the simplification

$$b_2(\tau) \;=\; F^{-1}_{B_{11}+B_{21}}(\tau) - F^{-1}_{B_{11}}(\tau).$$

Requirement (ii) – comonotonicity – implies that there exists at least one rotation $B_1^* = A^{-1}B_1$ such that B_{11}^* and B_{21}^* are comonotonic. Different rotations have different economic content. For example if B_{11} and $B_{11} + B_{21}$ are comonotonic, then the workers with the highest potential earnings in the union sector coincide with those with the highest potential earnings in the non-union sector and vice versa. This rules out comparative advantage. If, instead, B_{11} and B_{21} are comonotonic, then those workers which benefit the most from collective bargaining coverage are also those who earn the most in its absence. Both of these comonotonicity assumptions imply (4). As a final example, if B_{1t} and -B_{2t} are comonotonic,

such that low earners in the absence of coverage gain the most from acquiring it, then

$$\begin{aligned} b_2(\tau) &= F^{-1}_{B_{11}|X_t}(\tau) + F^{-1}_{B_{21}|X_t}(1-\tau) - F^{-1}_{B_{11}|X_t}(\tau) \\ &= F^{-1}_{B_{21}|X_t}(1-\tau), \end{aligned}$$

which also implies (4).

These examples illustrate both the flexibility and restrictiveness of the comonotonicity requirement. Depending on the setting, it may be reasonable to assume comonotonicity of $B_t^* = A^{-1}B_t$ for some non-singular rotation A. Certain rotations may be more plausible than others. Nevertheless the assumption is often difficult to justify. Even in the program evaluation context, where independence of X_1 and B_1 may hold by design, researchers are often reluctant to interpret quantile treatment effects as anything more than the difference in two marginal survival functions (e.g., Koenker, 2005, pp. 30-31; Firpo, 2007).

At the same time, it is worth noting that textbook linear models with additive heterogeneity imply stronger rank invariance properties. For example, the basic models fitted by Chamberlain (1982), Jakubson (1991) and Card (1995) all have the implication that those workers with the highest potential earnings in the union sector coincide with those with the highest potential earnings in the non-union sector (cf., Vella and Verbeek (1998) for discussion).

The availability of panel data may be used to substantially weaken the assumptions of both comonotonicity and independence of the random coefficients. In particular we can replace (i) and (ii) above, with the requirement that the elements of $B_t^* = A(\mathbf{x})^{-1} B_t$ are comonotonic within the subpopulation of workers with common history $\mathbf{X} = \mathbf{x}$:[7]

$$A(\mathbf{x})^{-1} B_t \big| \mathbf{X} = \mathbf{x} \overset{D}{=} \left(F^{-1}_{B_{1t}|\mathbf{X}}(U_t|\mathbf{x}), \ldots, F^{-1}_{B_{Pt}|\mathbf{X}}(U_t|\mathbf{x}) \right), \; U \sim \mathcal{U}[0,1], \tag{6}$$

for some non-singular $A(\mathbf{x})$. Under (5) and (6) we have

$$Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x}) = x_t'\beta_t(\tau;\mathbf{x})$$

and, critically, also that

$$\beta_{pt}(U;\mathbf{x}) \sim B_{pt}|\mathbf{X} = \mathbf{x}, \; U \sim \mathcal{U}[0,1].$$

Note that the rotation of B_t that ensures conditional comonotonicity can vary with $\mathbf{X} = \mathbf{x}$.[8]

In addition to conditional comonotonicity, we also, as is typical in panel data models, need

[7] We also require that $x_t' A(\mathbf{x})$ is non-negative for all $x_t \in \mathbb{X}_t$.

[8] Clotilde and Napp (2004) present basic results on conditionally comonotonic random variables.

to impose some form of stationarity in the distribution of B_t over time. A convenient, but flexible, assumption is to require that the distribution of B_{p1} and B_{pt}, for $t > 1$, are related according to

$$\beta_{pt}(\tau; \mathbf{x}) - \beta_{p1}(\tau; \mathbf{x}) \equiv \delta_{pt}(\tau), \quad t = 2, \ldots, T, \quad p = 1, \ldots, P. \tag{7}$$

Restriction (7) corresponds to a "common trends" assumption. Under assumption (7) it is convenient to define, in a small abuse of notation, $\beta_p(\tau; \mathbf{x}) = \beta_{p1}(\tau; \mathbf{x})$.

Under restriction (7) differences in the conditional quantile functions of B_{pt} and B_{ps} for $t \neq s$ do not depend on $\mathbf{X}$. Under (3), (6) and (7) the conditional quantiles for $\mathbf{Y}$ given $\mathbf{X}$ satisfy (1) with

$$\mathbf{W} = \begin{pmatrix} \underline{0}_P' & \cdots & \underline{0}_P' \\ \mathbf{X}_2' & \cdots & \underline{0}_P' \\ \vdots & \ddots & \vdots \\ \underline{0}_P' & \cdots & \mathbf{X}_T' \end{pmatrix}, \quad \delta(\tau) = \begin{pmatrix} \delta_2(\tau) \\ \vdots \\ \delta_T(\tau) \end{pmatrix},$$

where $\underline{0}_P$ denotes a $P \times 1$ vector of zeroes. Here $\dim(\delta(\tau)) = R = (T-1)P$, since we allow the entire coefficient vector multiplying X_t to vary across periods. In practice, additional exclusion restrictions might be imposed or tested. For example one could impose the restriction that all components of $\delta_t(\tau)$ corresponding to the non-constant components of $\mathbf{X}_t$ are zero. In that case we could set $\mathbf{W} = \begin{pmatrix} \underline{0}_{T-1}, & I_{T-1} \end{pmatrix}'$.

To understand the generality embodied in (5), (6) and (7) relative to the cross-section case, it is again helpful to return to our empirical example. Suppose that $X_t = (1, \text{UNION}_t)'$ with $T = 2$, so that there are just four possible sequences of collective bargaining coverage:

$$(\text{UNION}_1, \text{UNION}_2) \in \{(0,0), (0,1), (1,0), (1,1)\}.$$

With panel data we can assume, for example, that B_{1t} and $B_{1t} + B_{2t}$ are comonotonic within the subpopulation of union joiners (i.e., $(\text{UNION}_1, \text{UNION}_2) = (0,1)$), while B_{1t} and $-B_{2t}$ are comonotonic within the subpopulation of union leavers (i.e., $(\text{UNION}_1, \text{UNION}_2) = (1,0)$). There may be no rotation of B_1 in which comonotonicity holds unconditionally on $\mathbf{X} = \mathbf{x}$. Other than the assumption of conditional comonotonicity, all other features of the joint distribution of B_t and $\mathbf{X}$ are unrestricted. This allows for dependence between X_t and B_t. For example it may be that the distribution of B_{2t}, the returns to collective bargaining coverage, across workers in the union sector both periods, stochastically dominates that across workers not in the union sector both periods (cf., Card, Lemieux and Riddell, 2004).

Equations (5), (6) and (7) show how our semiparametric model arises as a strict general-

ization of the textbook linear quantile regression model. Here, relative to the cross section case, the presence of panel data allows for (i) a relaxation of comonotonicity of the random coefficients, (ii) the introduction of correlated heterogeneity and (iii) a structured form of non-stationarity over time.

Generalizing the linear panel data model

In our exposition, for reasons of clarity, we emphasize an interpretation of (1) based on the data generating process defined by (5), (6) and (7). However it is also straightforward to derive variants of (1) from a generalization of the textbook linear panel data model (e.g., Chamberlain, 1984):

$$Y_{it} = X_{it}'\beta + A_i + V_{it},\ \mathbb{E}\left[V_{it}|\, X_{i1}, \ldots, X_{iT}, A_i\right] = 0,\ t = 1, \ldots, T. \tag{8}$$

In this model, the "fixed effects" A_i are treated as an incidental, individual-specific parameter that can be estimated or differenced out. The strict exogeneity condition allows us to identify the common coefficients β.

We modify this model with respect to multiple dimensions. First, we allow the common coefficients β to vary with the time index t, and we also assume that A_i follows a distribution that is identical for each individual.[9] Omitting the i subscript, the resulting model is

$$Y_t = X_t'\beta_t + A + V_t,\ \mathbb{E}\left[V_t|\, X_1, \ldots, X_T, A\right] = 0,\ t = 1, \ldots, T. \tag{9}$$

Second, we further generalize this model by considering a location-scale version of (9) (cf., Arellano and Bonhomme, 2011)

$$Y_t = X_t'\beta_t + X_t' g\left(A + V_t\right), \tag{10}$$

with $x_t' g\left(a + v_t\right)$ strictly increasing in $a + v_t$ for all $a + v_t \in \mathbb{A} + \mathbb{V}_t$ and all $x_t \in \mathbb{X}_{tN}$ and V_t obeying the marginal stationarity restriction of Manski (1987):

$$V_1|\, \mathbf{X} \overset{d}{=} V_t|\, \mathbf{X},\ t = 2, \ldots, T. \tag{11}$$

Relative to the textbook model, (10) allows for the marginal effect of a unit change in X_{tp} to be heterogenous across units and correlated with $\mathbf{X}$ since the individual effects' dependence on $\mathbf{X}$ is left unrestricted. The textbook model imposes homogeneity of marginal effects, a

[9]The formulation of the model in equation (8) allows for an i.n.i.d. distribution of the individual effects.

strong restriction which is useful to relax. Equations (10) and (11) generate the period t conditional quantile function

$$Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x}) = x_t'(\beta(\tau;\mathbf{x}) + \delta_t)$$

for $\beta(\tau;\mathbf{x}) = \beta_1 + g\left(Q_{A+V_1|\mathbf{X}}(\tau|\mathbf{x})\right)$ and $\delta_t = \beta_t - \beta_1$. This model implies that the time effects take a pure location-shift form, which is not a implication of (1).

Our semiparametric model (1) therefore nests both the textbook quantile regression and linear panel data models as special cases. It also strictly generalizes those models, introducing heterogenous effects and/or the dependence of these effects on the regressors.

2 Estimands and identification

In this section we introduce three estimands based on (1). We motivate these estimands vis-a-vis the correlated random coefficients model defined by (5), (6) and (7) above, although this is not essential to our formal results. Indeed a subset of our estimands only require that (1) hold for a single known τ.

Our first estimand is the $R \times 1$ vector of common coefficients $\delta(\tau)$. Recall that in our motivating data generating process the elements of $\delta(\tau)$ coincide with time effects.

Our second estimand is the $P \times 1$ vector of average conditional quantile effects (ACQEs):

$$\bar{\beta}(\tau) = \mathbb{E}[\beta(\tau;\mathbf{X})]. \tag{12}$$

Equation (12) coincides with an average of the conditional quantiles of B_1 in (5) over $\mathbf{X}$. It is similar to the average derivative quantile regression coefficients studied in Chaudhuri, Doksum and Samarov (1997).

The ACQE is also closely related to a measure of conditional inequality used by labor economists. Angrist, Chernozhukov and Fernández-Val (2006; Table 1) report estimates of the average $\mathbb{E}[X_1]'(\beta(0.9) - \beta(0.1))$, with $\beta(\tau)$ the coefficient on X_1 in the τ^{th} linear quantile regression of log earnings Y_1 on worker characteristics X_1. They interpret this as a measure of average conditional earnings inequality or 'residual' wage inequality.[10] In our panel data set-up the analogous measure of period t conditional earnings inequality would

[10]This measure captures a notion of 'residual' wage inequality in that it measures the average amount of inequality in earnings that is left-over after first conditioning on covariates (cf., Autor, Katz and Kearney, 2008).

be

$$\mathbb{E}\left[X_t'\left(\beta\left(0.9;\mathbf{X}\right)-\beta\left(0.1;\mathbf{X}\right)\right)\right]+\mathbb{E}\left[W_t\right]'\left(\delta\left(0.9\right)-\delta\left(0.1\right)\right). \tag{13}$$

Equation (13) measures the average period t conditional 90-10 earnings gap across all sub-populations of workers defined in terms of their covariate histories $\mathbf{X}$. It is a "residual" inequality measure because it is an average of earnings dispersion measures which condition on observed covariates. Under our assumptions (13) has counterfactual content. To see this consider the average conditional period t 90-10 earnings gap that we would observe if, contrary to fact, worker characteristics remained fixed at their base year values:

$$\mathbb{E}\left[X_1'\left(\beta\left(0.9;\mathbf{X}\right)-\beta\left(0.1;\mathbf{X}\right)\right)\right]+\mathbb{E}\left[W_1\right]'\left(\delta\left(0.9\right)-\delta\left(0.1\right)\right). \tag{14}$$

The difference between (13) and (14) is a measure of the increase in 'residual' earnings inequality due to changes in worker characteristics between periods 1 and t. Similar reasoning leads to more complicated decomposition estimands.

Our final estimand, which makes full use of our set-up, is the unconditional quantile effect (UQE), defined implicitly by, for $p=1,\ldots,P$,

$$\begin{aligned}\beta_p\left(\tau\right) &= Q_{B_{1p}}\left(\tau\right)\\ &= \inf\left\{b\in\mathbb{R}\,:\,\Pr\left(B_{1p}\leq b\right)\geq\tau\right\}\\ &= \inf\left\{b\in\mathbb{R}\,:\,\Pr\left(\beta_p\left(U_1,\mathbf{X}\right)\leq b\right)\geq\tau\right\}\end{aligned}$$

where $U_1\sim\mathcal{U}\left[0,1\right]$, independent of $\mathbf{X}$. The UQE $\beta_p\left(\tau\right)$ corresponds to the τ^{th} quantile of the p^{th} component of the random coefficient vector B_1. If we took a random draw from the population and increased her p^{th} regressor value by one unit, then with probability τ the effect on Y_1 would be less than or equal to $\beta_p\left(\tau\right)$, while with probability $1-\tau$ it would be greater. To get the total effect for a t^{th} period intervention, we would need to take into account the effect on $\mathbf{W}_t'\delta(\tau)$ of the change in regressor $\mathbf{W}_t$ (as a function of $\mathbf{X}_t$).

The UQE is the quantile analog of an average partial effect (APE).

Identification

We present two sets of identification results. The first requires that the time dimension of the panel (T) strictly exceed the number of regressors (P). We refer to this as the "regular" case. Chamberlain (1992) studied identification of average partial effects in this setting. Second we study identification when $T=P$. This is the case studied in Graham and Powell (2012). We refer to this case as "irregular". Both are empirically relevant (cf., Graham and Powell,

2012). Throughout this section we assume that the joint distribution of the observable data matrix $(\mathbf{Y}, \mathbf{X})$ is known – in particular, that the $T \times 1$ vector $Q_{\mathbf{Y}|\mathbf{X}}(\tau|\,\mathbf{X})$ is known for all $\tau \in (0,1)$ and $\mathbf{X} \in \mathbb{X}_N^T$.

Regular case $(T > P)$

Let $A(\mathbf{X})$ be any $T \times T$ positive definite matrix, possibly a function of $\mathbf{X}$, and define the residual-maker matrix

$$M_A(\mathbf{X}) = I_T - \mathbf{X}\left(\mathbf{X}'A(\mathbf{X})\,\mathbf{X}\right)^{-1}\mathbf{X}'A(\mathbf{X}). \tag{15}$$

If $\mathbb{E}\left[\|\mathbf{X}\|^2 + \|\mathbf{W}\|^2\right] < \infty$ and $\mathbb{E}\left[\mathbf{W}'M_A(\mathbf{X})\mathbf{W}\right]$ is invertible we can recover $\delta(\tau)$ by

$$\delta(\tau) = \mathbb{E}\left[\mathbf{W}'M_A(\mathbf{X})\,\mathbf{W}\right]^{-1} \times \mathbb{E}\left[\mathbf{W}'M_A(\mathbf{X})\,Q_{\mathbf{Y}|\mathbf{X}}(\tau|\,\mathbf{X})\right]. \tag{16}$$

Equation (16) shows that $\delta(\tau)$ indexes a (generalized) within-unit, double residual, linear regression function. The dependent variable associated with this regression is $Q_{Y_t|\mathbf{X}}(\tau|\,\mathbf{X})$ deviated from its *unit-specific* "mean" and the independent variable the deviation of $\mathbf{W}_t$ about its corresponding "mean" .

Once $\delta(\tau)$ is identified, we can also identify the τ^{th} conditional quantile of the random coefficient, for $\mathbf{X}$ realizations with full rank, through the relation

$$\beta(\tau;\mathbf{X}) = \left(\mathbf{X}'A(\mathbf{X})\,\mathbf{X}\right)^{-1}\mathbf{X}'A(\mathbf{X})\left(Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\delta(\tau)\right). \tag{17}$$

If all realizations of $\mathbf{X}$ are of full rank, we can directly recover the average conditional quantile effect (ACQE) from (17) by

$$\begin{aligned}\bar{\beta}(\tau) &= \mathbb{E}\left[\beta(\tau;\mathbf{X})\right] \\ &= \mathbb{E}\left[\left(\mathbf{X}'A(\mathbf{X})\,\mathbf{X}\right)^{-1}\mathbf{X}'A(\mathbf{X})\left(Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\delta(\tau)\right)\right],\end{aligned} \tag{18}$$

while the unconditional quantile $\beta_p(\tau)$ of the p^{th} component of B_1 is identified as the solution to the equation

$$\mathbb{E}\left[\mathbf{1}(\beta_p(U;\mathbf{X}) \leq \beta_p(\tau)) - \tau\right] = 0, \tag{19}$$

with U uniformly distributed on $(0,1)$, independently from $\mathbf{X}$. The UQE $\beta_p(\tau)$ will be uniquely identified if the distribution of $\beta(U;\mathbf{X})$ is continuous around its τ^{th} quantile with positive density.

Equations (18) and (19) do not follow if the probability π_0 that $\mathbf{X}$ is rank deficient is non-

zero. Denote by $\mathbb{X}^M$ the region of the support of $\mathbf{X}$ where its rank is full, and denote by π_0 the probability that the rank of $\mathbf{X}$ is less than P. Similarly, denote by $\mathbb{X}^S$ the region of the support where $\mathbf{X}$ is rank deficient. When $\pi_0 > 0$ it is still possible to identify $\delta(\tau)$, by using the observations where $\mathbf{X} \in \mathbb{X}^M$, via

$$\delta(\tau) = \mathbb{E}\left[\mathbf{W}' M_A(\mathbf{X})\mathbf{W} | \mathbf{X} \in \mathbb{X}^M\right]^{-1} \times \mathbb{E}\left[\mathbf{W}' M_A(\mathbf{X}) Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) | \mathbf{X} \in \mathbb{X}^M\right],$$

if we now assume that $\mathbb{E}\left[\mathbf{W}' M_A(\mathbf{X})\mathbf{W} | \mathbf{X} \in \mathbb{X}^M\right]$ is invertible and under the same moments existence requirements. It is also possible to identify $\beta(\tau; \mathbf{X})$ through the same argument, but only for the subpopulation of units where $\mathbf{X}$ has full rank. These full rank units represent fraction $1 - \pi_0$ of the population, as opposed to its entirety. Despite the non-identification of $\beta(\tau; \mathbf{X})$ for units with non-full rank, it is clearly possible to point identify the "movers' ACQE" and the "movers' UQE", defined as

$$\begin{aligned} \bar{\beta}^M(\tau) &= \mathbb{E}\left[\beta(\tau; \mathbf{X}) | \mathbf{X} \in \mathbb{X}^M\right] \\ &= \mathbb{E}\left[\left(\mathbf{X}' A\left(\mathbf{X}\right)\mathbf{X}\right)^{-1} \mathbf{X}' A\left(\mathbf{X}\right)\left(Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\delta(\tau)\right) | \mathbf{X} \in \mathbb{X}^M\right] \end{aligned} \tag{20}$$

and the solution $\beta_p^M(\tau)$ to the equation

$$\mathbb{E}\left[\mathbf{1}(\beta_p(U; \mathbf{X}) \leq \beta_p^M(\tau)) - \tau | \mathbf{X} \in \mathbb{X}^M\right] = 0. \tag{21}$$

Although $\bar{\beta}(\tau)$ and $\beta_p(\tau)$, the full population average and unconditional quantile effects, are not point identified when $\pi_0 > 0$, it is possible to construct bounds for them. The Law of Total Probability gives

$$\bar{\beta}(\tau) = \bar{\beta}^M(\tau) \Pr(\mathbf{X} \in \mathbb{X}^M) + \mathbb{E}[\beta(\tau; \mathbf{X}) | \mathbf{X} \in \mathbb{X}^S] \Pr(\mathbf{X} \in \mathbb{X}^S).$$

Let $[\underline{b}_p, \bar{b}_p]$ denote bounds on the support of $\beta_p(\tau; \mathbf{X})$. The existence of such bounds, although not their magnitude, is implied by Assumption 5 below. The identified set for $\bar{\beta}_p(\tau)$ is then

$$\left[\bar{\beta}_p^M(\tau) \Pr(\mathbf{X} \in \mathbb{X}^M) + \underline{b}_p \Pr(\mathbf{X} \in \mathbb{X}^S), \bar{\beta}_p^M(\tau) \Pr(\mathbf{X} \in \mathbb{X}^M) + \bar{b}_p \Pr(\mathbf{X} \in \mathbb{X}^S)\right]$$

for any $p = 1, \ldots, P$. This result requires us to assume that $\bar{b}_p$ and $\underline{b}_p$ are known.

A somewhat more satisfying result is available for $\beta_p(\tau)$. We give this result as a Theorem, although the required assumptions are not stated until the next section.

Theorem 1. (Partial Identification of $\beta_p(\tau)$) *Under Assumptions 1 through 5 stated below and* $\mathbb{E}\left[\mathbf{W}' M_{\mathbf{A}}(\mathbf{X})\mathbf{W} | \mathbf{X} \in \mathbb{X}^M\right]$ *invertible, the UQE for the* p^{th} *coefficient is partially*

identified with identification region:

$$\beta_p(\tau) \in \left[\beta_p^M\left(\frac{\tau - \Pr(\mathbf{X} \in \mathbb{X}^S)}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\right), \beta_p^M\left(\frac{\tau}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\right)\right]. \tag{22}$$

where $\beta_p^M(\tau) \equiv \underline{b}_p$ *for* $\tau < 0$ *and* $\beta_p^M(\tau) \equiv \bar{b}_p$ *for* $\tau > 1$.

Since the movers' UQE is identified, as well as $\Pr(\mathbf{X} \in \mathbb{X}^M)$ and $\Pr(\mathbf{X} \in \mathbb{X}^S) = 1 - \Pr(\mathbf{X} \in \mathbb{X}^M)$, the analog estimators for the lower and upper bounds of the identified set given in (22) are easy to compute.[11] If prior bounds on the random coefficient are unknown, these bounds are only meaningful for τ in a subset of $(0, 1)$. The width of this subset depends on the fraction of stayers. When τ is close to either 0 or 1, we must rely on prior bounds $\bar{b}_p$ and $\underline{b}_p$ to set identify the UQE, as is the case for the ACQE.

Irregular case $(T = P)$

We now consider the $T = P$ case. Our approach builds on that of Graham and Powell (2012) for average effects in a conditional mean variant of (1). While identification in the regular case is based solely on the subpopulation of movers, the irregular case utilizes both movers and stayers. The role of stayers is to identify the common parameter $\delta(\tau)$, which in our motivating data generating process, captures aggregate time effects. Stayers, as we detail below, serve as a type of control group, allowing the econometrician to identify "common trends" affecting all units.

Let $D = \det(\mathbf{X})$, $\mathbf{X}^*$ denote the adjugate (or adjoint) matrix of $\mathbf{X}$ (i.e., the matrix such that $\mathbf{X}^{-1} = \frac{1}{D}\mathbf{X}^*$ when $D \neq 0$), and $\mathbf{W}^* = \mathbf{X}^*\mathbf{W}$. Premultiplying equation (1) by $\mathbf{X}^*$ gives

$$\mathbf{X}^* Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) = \mathbf{W}^*\delta(\tau) + D\beta(\tau; \mathbf{X}). \tag{23}$$

Assuming that zero is in the support of the determinant, D, $\mathbb{E}\left[\|\mathbf{W}^*\|^2\right] < \infty$ and that $\mathbb{E}[\mathbf{W}^{*\prime}\mathbf{W}^*|D = 0]$ is of full rank, we can identify $\delta(\tau)$, using only stayer (i.e., $D = 0$) observations, by:

$$\delta(\tau) = \mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^* \middle| D = 0\right]^{-1} \times \mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^* Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X})|D = 0\right]. \tag{24}$$

Given identification of $\delta(\tau)$, we can then recover $\beta(\tau; \mathbf{X})$ by

$$\beta(\tau; \mathbf{X}) = \mathbf{X}^{-1}(Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\delta(\tau)) \tag{25}$$

[11] The endpoints' joint asymptotic distribution can be readily inferred from the process convergence of the UQE process established below.

for all $\mathbf{X}$ where $\mathbf{X}^{-1} = \frac{1}{D}\mathbf{X}^*$ is well-defined (i.e., for "mover" realizations of $\mathbf{X}$).

As long as $\Pr(D = 0) = 0$ it follows that the conditional effect $\beta(\tau; \mathbf{X})$ will be identified with probability one. However, the identification of the ACQE and UQE estimands is more delicate than in the regular case, due to the fact that if the density of D is positive in a neighborhood of 0 (which we require for identification of $\delta(\tau)$), expectations involving $\mathbf{X}^{-1} = \frac{1}{D}\mathbf{X}^*$ will not exist in general (e.g., Khan and Tamer (2010) and Graham and Powell (2012)). In order to identify the ACQE, $\bar{\beta}(\tau)$, we write it as the limit of a sequence of "trimmed" expectations

$$\begin{aligned}\bar{\beta}(\tau) &= \mathbb{E}[\beta(\tau; \mathbf{X})] \\ &= \lim_{h \downarrow 0} \mathbb{E}[\beta(\tau; \mathbf{X})\mathbf{1}(|D| > h)] \\ &= \lim_{h \downarrow 0} \mathbb{E}[\mathbf{X}^{-1}(Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\delta(\tau))\mathbf{1}(|D| > h)], \end{aligned} \tag{26}$$

where the second equality holds because

$$\bar{\beta}(\tau) - \mathbb{E}[\beta(\tau; \mathbf{X})\mathbf{1}(|D| > h)] = \mathbb{E}[\beta(\tau; \mathbf{X})\mathbf{1}(|D| \leq h)] = O(h)$$

under sufficient smoothness conditions. This trimming is not strictly necessary at the identification stage, under the maintained assumption that $\bar{\beta}(\tau)$ exists, but is introduced in anticipation of its estimation. In particular, replacing $Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X})$ with a nonparametric estimate introduces noise into the numerator of the sample analog of (26). This sampling error may cause the expectation of the estimated conditional effect

$$\hat{\beta}(\tau; \mathbf{X}) = \mathbf{X}^{-1}(\hat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\hat{\delta}(\tau))$$

to be undefined due to a lack of moments of the remainder term

$$\hat{\beta}(\tau; \mathbf{X}) - \beta(\tau; \mathbf{X}) = \mathbf{X}^{-1}\left(\hat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}(\hat{\delta}(\tau) - \delta(\tau))\right).$$

We can also characterize the identification of $\beta_p(\tau)$, the UQE associated with the p^{th} regressor, in terms of a sequence of trimmed means. Assuming that the distribution of $\beta(U; \mathbf{X})$ given $D = t$ is continuously differentiable in a neighborhood of $t = 0$, we can write $\beta_p(\tau)$ as the solution to

$$\begin{aligned} 0 &= \mathbb{E}\left[\mathbf{1}(\beta_p(U; \mathbf{X}) \leq \beta_p(\tau)) - \tau\right] \\ &= \lim_{h \downarrow 0} \mathbb{E}\left[(\mathbf{1}(\beta_p(U; \mathbf{X}) \leq \beta_p(\tau)) - \tau)\mathbf{1}(|D| > h)\right]. \end{aligned}$$

Our approach to estimation exploits this characterization.

If there is a point mass of stayer units with $D = 0$ (i.e. $\pi_0 > 0$), the same identification issues arise here as in the regular ($T > P$) case. In this case we can continue to identify $\delta(\tau)$ as before, but $\beta(\tau; \mathbf{X})$ will be unidentified for a set of $\mathbf{X}$ values with positive probability. It is still possible identify the movers' ACQE and UQE using straightforward modifications of the arguments given for the regular case in the previous subsection.

As a simple example of irregular identification consider a two period version of (1) with a single time-varying regressor (X_{2t}) and an intercept time shift.[12] This setup yields conditional quantiles for each period of

$$\begin{aligned} Q_{Y_1|\mathbf{X}}(\tau|\mathbf{x}) &= \beta_1(\tau; \mathbf{x}) + \beta_2(\tau; \mathbf{x}) X_{21} \\ Q_{Y_2|\mathbf{X}}(\tau|\mathbf{x}) &= \delta(\tau) + \beta_1(\tau; \mathbf{x}) + \beta_2(\tau; \mathbf{x}) X_{22}. \end{aligned}$$

Here X_{2t} might be a policy variable, such as an individuals' workers compensation benefit level, which depends on own earnings as well as state-specific benefit schedules, and Y_t an outcome of interest to a policymaker, such as time out of work following an injury.

Evaluating (24) we get

$$\delta(\tau) = \mathbb{E}\left[\omega(X_{21})\left\{Q_{Y_2|\mathbf{X}}(\tau|\mathbf{X}) - Q_{Y_1|\mathbf{X}}(\tau|\mathbf{X})\right\} \middle| D = 0\right], \quad \omega(X_1) = \frac{1 + X_{21}^2}{\mathbb{E}\left[1 + X_{21}^2 \middle| D = 0\right]},$$

so that $\delta(\tau)$ is identified by a weighted average of changes in the τ^{th} quantile of Y_t between periods 1 and 2 *across the subpopulation of stayers*. Stayer units, who in this case correspond to units where $X_{21} = X_{22}$ (i.e., the nonconstant regressor stays fixed over time), serve as a type of "control group", identifying aggregate time effects or "common trends".

The conditional quantile effect of a unit change in X_{2t} is given by the second element of (25), which evaluates to

$$\beta_p(\tau; \mathbf{x}) = \frac{Q_{Y_2|\mathbf{X}}(\tau|\mathbf{x}) - Q_{Y_1|\mathbf{X}}(\tau|\mathbf{x}) - \delta(\tau)}{\triangle x_2},$$

for all $\mathbf{x}$ with $x_{22} - x_{21} = \triangle x_2 \neq 0$. Hence $\beta_p(\tau; \mathbf{x})$ is identified by a "difference-in-differences".

[12] This corresponds to (1) with $T = P = 2$ and $\mathbf{W}$ and $\mathbf{X}$ equal to

$$\mathbf{W} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{21} \\ 1 & X_{22} \end{pmatrix}$$

so that $D = X_{22} - X_{21} = \triangle X_2$.

3 Estimation

In this section we present analog estimators, based upon the identification results presented above. Our estimators utilize preliminary nonparametric estimates of the conditional quantiles of Y_t given $\mathbf{X}$ for $t = 1, \ldots, T$. Our formal results cover the case where $\mathbf{X}$ is discretely valued with M points of support: $\mathbf{X} \in \mathbb{X}_N^T = \{\mathbf{x}_{1N}, \ldots, \mathbf{x}_{MN}\}$. This case covers many empirical applications of interest, including the one developed below. It is also, as described in the introduction, technically simpler, allowing analysis to proceed conditional on discrete, non-overlapping, cells. However, by considering asymptotic sequences where the location and probability mass attached to the different support points of X changes with N, we show how our results would extend to the case of continuously-valued regressors (albeit under additional regularity conditions).

After stating our main assumption we discuss estimation in the regular case ($T > P$) and irregular case ($T = P$).

Assumptions

Assumption 1. (Data Generating Process) *The conditional quantiles of $Y_1, \ldots, Y_T$ given $\mathbf{X}$ are of the form (1) for all $\mathbf{X} \in \mathbb{X}_N^T$ and $\tau \in (0, 1)$.*

For estimation of the common parameter, $\delta(\tau)$, and the ACQE, $\bar{\beta}_p(\tau)$, we only require that (1) hold at τ. The stronger implications of Assumption 1 are required for estimation of the UQE, $\beta_p(\tau)$, $p = 1, \ldots, P$.

Assumption 2. (Support and Support Convergence) *(i) $\mathbf{X} \in \mathbb{X}_N^T = \{\mathbf{x}_{1N}, \ldots, \mathbf{x}_{MN}\}$ with $p_{lN} = \Pr(\mathbf{X} = \mathbf{x}_{lN})$ for $l = 1, \ldots, N$; (ii) as $N \to \infty$, we have $\mathbf{x}_{lN} \to \mathbf{x}_l$, $p_{lN} \to p_l$ and $Np_{lN} \to \infty$ for any $l = 1, \ldots, M$ for some well defined $\mathbf{x}_l$ and p_l; (iii) the elements of $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ are bounded.*

Assumption 2 has two non-standard features. First, while it restricts the number of support points of $\mathbf{X}$, the location of these points is allowed to vary with N. Second, the probability mass attached to each support point may also vary with N. Both of these sequences have well-defined limits. An important feature of Assumption 2 is that it allows the probability mass attached to some points of support to shrink to zero. The rate at which this occurs is limited by the requirement that $Np_{lN} \to \infty$. This assumption ensures that the conditional quantile of Y_t given $\mathbf{X} = \mathbf{x}_l$ is consistently estimable for all $l = 1, \ldots, M$. However the rate of

convergence of these estimates will be slower for points of support with shrinking probability mass as N grows large.[13]

For the analysis which follows it is convenient to partition the support of $\mathbf{X}$ as follows.

1. Units with $\mathbf{X} = \mathbf{x}_{mN}$ for $m = 1, \ldots, L_1 < M$ correspond to **movers**. Movers correspond to units where, recalling that $\mathbf{x}_{mN} \to \mathbf{x}_m$, $\mathbf{x}_m$ and $\mathbf{x}_{mN}$ are of full rank. Intuitively movers are units whose covariate values "vary a lot" over time.

2. Units with $\mathbf{X} = \mathbf{x}_{mN}$ for $m = L_1 + 1, \ldots, L < M$ correspond to **near stayers**. Near stayers correspond to units where $\mathbf{x}_{mN}$ is of full rank, but its limit $\mathbf{x}_m$ is not. We will be more precise about the behavior of these units' design matrices along the path to the limit below. Intuitively near stayers are units who covariate values change "very little" over time.

3. Units with $\mathbf{X} = \mathbf{x}_{mN}$ for $m = L + 1, \ldots, M$ correspond to **stayers**. Stayers are units where $\mathbf{x}_{mN}$ is neither of full rank along the path nor in the limit. Stayers correspond to units where the number of distinct rows of $\mathbf{X}$ is less than P (i.e., whose regressor sequences display substantial persistence).

We let $\mathbb{X}_N^M = \{\mathbf{x}_{1N}, \ldots \mathbf{x}_{LN}\}$ denote the set of mover support points (including near stayers). The set of stayer support points is denoted by $\mathbb{X}_N^S = \{\mathbf{x}_{L+1N}, \ldots \mathbf{x}_{MN}\}$. We introduce more structure to this basic set-up (as needed) below.

Assumption 3. (Random Sampling) *$\{\mathbf{Y}_i, \mathbf{X}_i\}_{i=1}^N$ is a random (i.i.d.) sample from the population of interest.*

Assumption 4. (Bounded and Continuous Densities) *The conditional distribution $F_{Y_t|X}(y_t|\mathbf{x})$ has density $f_{Y_t|X}(y_t|\mathbf{x})$ such that $\psi(\tau;\mathbf{x}) = f_{Y_t|X}\left(F_{Y_t|X}^{-1}(\tau|\mathbf{x})\middle|\mathbf{x}\right)$ and $\frac{\partial\psi(\tau;\mathbf{x})}{\partial\tau}$ are uniformly bounded and bounded away from zero for all $\tau \in (0,1)$, all $\mathbf{x} \in \mathbb{X}_N^T$, and all $t = 1, \ldots, T$. Also, this conditional distribution does not vary with the sample size N. Finally, $f_{Y_t|X}(y_t|\mathbf{x})$, $F_{Y_t|X}(y_t|\mathbf{x})$ and $F_{Y_t|X}^{-1}(y_t|\mathbf{x})$ are all continuous in $\mathbf{x}$.*

Assumption 5. (Bounded Coefficients) *The support of B_{tp} is compact and its density is bounded away from 0 for any $p = 1, \ldots, P$ and $t = 1, \ldots, T$.*

[13] Also note that, since $\mathbf{W}$ is a function of $\mathbf{X}$ alone, it too has M points of support: $\mathbf{W} \in \mathbb{W}_N^T = \{\mathbf{w}_{1N}, \ldots, \mathbf{w}_{MN}\}$.

Assumptions 3 is standard, as is Assumption 4 is the quantile regression context. Assumption 5, in conjunction with Assumption 2 ensures that Y has bounded support.

The first step of our estimation procedure involves computing estimates of the conditional quantiles of Y_t given $\mathbf{X}$ for all $\mathbf{X} \in \mathbb{X}_N^T$ and all $t = 1, \ldots, T$. This must be done for a single τ in the case of the common coefficients, $\delta(\tau)$ and the movers' Average Conditional Quantile Effect (ACQE), $\bar{\beta}^M(\tau)$, and for a uniform grid of $\tau \in (0,1)$ in the case of the movers' Unconditional Quantile Effect (UQE), $\beta_p^M(\tau)$. When $\mathbf{X}$ is discretely valued this first step of estimation is very simple (cf., Chamberlain, 1994).

Under Assumption 2 preliminary estimation of the conditional quantiles of Y_t given $\mathbf{X}$ is straightforward. Let

$$\widehat{F}_{Y_t|\mathbf{X}}(y_t|\mathbf{x}_{mN}) = \left[\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})\right]^{-1} \times \left[\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})\mathbf{1}(Y_{it} \leq y_t)\right],$$

be the empirical cumulative distribution function of Y_t for the subsample of units with $\mathbf{X} = \mathbf{x}_{mN}$. Our estimate of the τ^{th} conditional quantile of Y_t equals

$$\widehat{Q}_{Y_t|\mathbf{X}}(\tau|\mathbf{x}_{mN}) = \widehat{F}^{-1}_{Y_t|\mathbf{X}}(y_t|\mathbf{x}_{mN}) = \inf\left\{y_t : \widehat{F}_{Y_t|\mathbf{X}}(y_t|\mathbf{x}_{mN}) \geq \tau\right\}.$$

Note that $\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau;\mathbf{x}_{lN})$ and $\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{mN})$ for $l \neq m$ are conditionally uncorrelated given $\{\mathbf{X}_i\}_{i=1}^N$.

In practice estimation of $\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau;\mathbf{x}_{mN})$ is very simple (cf., Chamberlain, 1994). Let $N_m = \sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})$ equal the number of units in cell $\mathbf{X} = \mathbf{x}_{mN}$. Let $Y_t^{(j,m)}$ denote the j^{th} order statistic of Y_t in the $\mathbf{X} = \mathbf{x}_{mN}$ subsample. We estimate $Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x}_{mN})$ by $Y_t^{(j,m)}$ where j satisfies

$$\frac{j}{N_m + 1} < \tau \leq \frac{j+1}{N_m + 1}.$$

Alternatively we could use $(Y_t^{(j,m)} + Y_t^{(j+1,m)})/2$ as our estimate.

To characterize the large sample properties of $\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau;\mathbf{x}_{mN})$ we require some additional notation. Let

$$\rho_{st}(\tau,\tau';\mathbf{X}) = \frac{\Pr\left(Y_s \leq Q_{Y_s|\mathbf{X}}(\tau|\mathbf{X}), Y_t \leq Q_{Y_t|\mathbf{X}}(\tau|\mathbf{X})\right) - \tau\tau'}{\min(\tau,\tau') - \tau\tau'}, \quad s,t = 1,\ldots,T \tag{27}$$

and $\Lambda(\tau, \tau'; \mathbf{X})$ equal

$$\begin{pmatrix} \frac{1}{f_{Y_1|\mathbf{X}}\left(Q_{Y_1|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right) f_{Y_1|\mathbf{X}}\left(Q_{Y_1|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right)} & \cdots & \frac{\rho_{1T}(\tau,\tau';\mathbf{X})}{f_{Y_1|\mathbf{X}}\left(Q_{Y_1|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right) f_{Y_T|\mathbf{X}}\left(Q_{Y_T|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right)} \\ \vdots & \ddots & \vdots \\ \frac{\rho_{1T}(\tau,\tau';\mathbf{X})}{f_{Y_1|\mathbf{X}}\left(Q_{Y_1|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right) f_{Y_T|\mathbf{X}}\left(Q_{Y_T|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right)} & \cdots & \frac{1}{f_{Y_T|\mathbf{X}}\left(Q_{Y_T|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right) f_{Y_T|\mathbf{X}}\left(Q_{Y_T|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X}\right)} \end{pmatrix}. \quad (28)$$

Using this notation, an adaptation of standard results on quantile processes in the cross sectional context, to allow for both moving support points and probabilities, gives our first result:

Lemma 1. *Suppose that Assumptions 1 through 5 are satisfied, then*

$$\sqrt{Np_{mN}}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{mN}) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{mN})\right)$$

converges in distribution to a mean zero Gaussian process $\mathbf{Z}_Q(\cdot,\cdot)$ *on* $\tau \in (0,1)$ *and* $\mathbf{x}_m \in \mathbb{X}$*, where* $\mathbf{Z}_Q(\cdot,\cdot)$ *is defined by its covariance function* $\Sigma_Q(\tau, \mathbf{x}_l, \tau', \mathbf{x}_m) = \mathbb{E}\left[\mathbf{Z}_Q(\tau, \mathbf{x}_l)\mathbf{Z}_Q(\tau', \mathbf{x}_m)'\right]$ *with*

$$\Sigma_Q(\tau, \mathbf{x}_l, \tau', \mathbf{x}_m) = (\min(\tau, \tau') - \tau\tau')\Lambda(\tau, \tau'; \mathbf{x}_l) \cdot \mathbf{1}(l = m) \quad (29)$$

for $l, m = 1, \ldots, M$.

Lemma 1 generalizes textbook process convergence results for unconditional quantiles. Consider a support point with $p_{mN} = h_N$ with $h_N \to 0$ and $Nh_N \to \infty$. The rate-of-convergence for such support points will be order $\sqrt{Nh_N}$, since the effective sample size used for estimation is proportional to Nh_N rather than N. Convergence here is on $\tau \in (0,1)$ since Y_t has compact support (by Assumptions 2 and 5 and that Y_t is the product of X_t and B_t). If Y_t's support was unbounded, Lemma 1 would instead hold uniformly for $\tau \in [\epsilon, 1-\epsilon]$ with arbitrary ϵ satisfying $0 < \epsilon < 1/2$.

Regular case ($T > P$)

We initially develop results for the fixed support case with $x_{mN} = x_m$ for all $m = 1, \ldots, L$. In this setting there are no near stayers so that $L_1 = L$. There may or may not be pure stayers in this case. This changes the identified effect, but not our approach to estimation – which utilizes movers alone – as explained below.

Let $\mathbf{\Pi}(\tau) = (Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_1)', \ldots, Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_L)')'$ be a $TL \times 1$ vector with movers' conditional

quantiles and notice that, under (1),

$$\mathbf{\Pi}(\tau) = G\gamma(\tau)$$

for $\tau \in (0,1)$ with

$$\underset{(R+PL)\times 1}{\gamma(\tau)} = \begin{pmatrix} \delta(\tau) \\ \beta(\tau;\mathbf{x}_1) \\ \vdots \\ \beta(\tau;\mathbf{x}_L) \end{pmatrix}, \quad \underset{TL\times(R+PL)}{G} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{x}_1 & \cdots & \underline{0}_T\underline{0}_P' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}_L & \underline{0}_T\underline{0}_P' & \cdots & \mathbf{x}_L \end{pmatrix}. \tag{30}$$

Since $\text{rank}(G) = \dim(\gamma(\tau))$ we have

$$\gamma(\tau) = (G'AG)^{-1} G'A\mathbf{\Pi}(\tau), \tag{31}$$

for any $TL \times TL$ positive-definite weight matrix A.

When A is block diagonal with m^{th} $T \times T$ block $p_m A(\mathbf{x}_m)$, it is straightforward to demonstrate that the first R elements of $\gamma(\tau)$ in (31) can be expressed as

$$\delta(\tau) = \mathbb{E}\left[\mathbf{W}'M_A(\mathbf{X})\mathbf{W}|\mathbf{X} \in \mathbb{X}^M\right]^{-1} \times \mathbb{E}\left[\mathbf{W}'M_A(\mathbf{X})Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X})|\mathbf{X} \in \mathbb{X}^M\right],$$

which coincides with (16) above. Manipulation of (31) also yields, for all $\mathbf{X} \in \mathbb{X}^M$,

$$\beta(\tau;\mathbf{X}) = \left[\mathbf{X}'A(\mathbf{X})\mathbf{X}\right]^{-1}\mathbf{X}'A(\mathbf{X})\left(Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}) - \mathbf{W}\delta(\tau)\right),$$

which coincides with (17) above.

Our analog estimator is

$$\hat{\gamma}(\tau) = (G'\hat{A}G)^{-1}G'\hat{A}\widehat{\mathbf{\Pi}}(\tau) \tag{32}$$

where $\hat{A}$ is a consistent estimator of a positive definite weight matrix and $\widehat{\mathbf{\Pi}}(\tau)$ is as defined above. To get precise results we make the following assumption on the weight matrix.

Assumption 6. (WEIGHT MATRIX) $\hat{A} = diag\{\hat{p}_{1N}, \ldots, \hat{p}_{LN}\} \otimes I_T$ *where* $\hat{p}_l = \frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_{lN})$.

This assumption is made to simplify the analysis and because weighting each support point by its relative frequency is often a reasonable choice. Although we do not develop this point here, it is straightforward to show, by adapting the argument given by Chamberlain (1994), that this choice of weight matrix also allows for easy characterization of the large sample properties of $\hat{\gamma}(\tau)$ under misspecification (i.e., when (1) does not hold).

Define

$$\overline{\mathbf{W}} = M_I(\mathbf{X})\,\mathbf{W}, \qquad M_I(\mathbf{X}) = I_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$
$$K(\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}, \qquad \Gamma = \mathbb{E}\left[\overline{\mathbf{W}}'\overline{\mathbf{W}}\middle|\mathbf{X}\in\mathbb{X}^M\right].$$

Theorem 2. *Suppose that Assumptions 1 through 6 are satisfied, the distribution of* $\mathbf{X}$ *is fixed, and* $\mathbb{E}\left[\overline{\mathbf{W}}'\overline{\mathbf{W}}|\mathbf{X}\in\mathbb{X}^M\right]$ *is invertible, then (i)* $\sqrt{N}\left(\widehat{\delta}(\cdot)-\delta(\cdot)\right)$ *converges in distribution to a mean zero Gaussian process* $\mathbf{Z}_\delta(\cdot)$*, where* $\mathbf{Z}_\delta(\cdot)$ *is defined by its covariance function*

$$\Sigma_\delta(\tau,\tau') = \mathbb{E}[\mathbf{Z}_\delta(\tau)\mathbf{Z}_\delta(\tau')'] = (\min(\tau,\tau')-\tau\tau')\frac{\Gamma^{-1}\mathbb{E}\left[\overline{\mathbf{W}}'\Lambda(\tau,\tau';\mathbf{X})\overline{\mathbf{W}}\middle|\mathbf{X}\in\mathbb{X}^M\right]\Gamma^{-1}}{\Pr(\mathbf{X}\in\mathbb{X}^M)}, \tag{33}$$

and (ii) $\sqrt{N}\left(\widehat{\beta}(\cdot;\cdot)-\beta(\cdot;\cdot)\right)$ *also converges in distribution to a mean zero Gaussian process* $\mathbf{Z}(\cdot,\cdot)$*, where* $\mathbf{Z}(\cdot,\cdot)$ *is defined by its covariance function*

$$\begin{aligned}\Sigma(\tau,\mathbf{x}_l,\tau',\mathbf{x}_m) &= \mathbb{E}\left[\mathbf{Z}(\tau,\mathbf{x}_l)\mathbf{Z}(\tau',\mathbf{x}_m)'\right]\\ &= (\min(\tau,\tau')-\tau\tau')\frac{(\mathbf{x}_l'\mathbf{x}_l)^{-1}\mathbf{x}_l'\Lambda(\tau,\tau';\mathbf{x}_l)\mathbf{x}_l(\mathbf{x}_l'\mathbf{x}_l)^{-1}}{p_l}\cdot\mathbf{1}(l=m)\\ &\quad + K(\mathbf{x}_l)\Sigma_\delta(\tau,\tau')K(\mathbf{x}_m)',\end{aligned} \tag{34}$$

for $l,m=1,\ldots,L$.

When $\mathbf{X}$ has finite support and, as maintained here, the location and probability mass attached to this support does not change with N, the rate of convergence of $\widehat{\beta}(\tau;\mathbf{x}_m)$ and $\widehat{\delta}(\tau)$ is $\sqrt{N}$. At the same time it is not possible to identify $\beta(\tau;\mathbf{x}_m)$ for the stayer realizations $m=L+1,\ldots,M$. Hence under the fixed support assumption $\bar{\beta}(\tau)$, the average conditional quantile effect, is not point identified. However, the movers' ACQE, defined as $\bar{\beta}^M(\tau) = \mathbb{E}[\beta(\tau;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M]$, *is* consistently estimable by

$$\widehat{\overline{\beta}}^M(\tau) = \frac{\sum_{i=1}^N \mathbf{1}\left(\mathbf{X}_i\in\mathbb{X}^M\right)\widehat{\beta}(\tau;\mathbf{X}_i)}{\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i\in\mathbb{X}^M)}. \tag{35}$$

Theorem 3. *Under the assumptions maintained in Theorem 2 above,* $\sqrt{N}(\widehat{\overline{\beta}}^M(\cdot)-\overline{\beta}^M(\cdot))$ *converges in distribution to a mean zero Gaussian process* $\mathbf{Z}_{\bar{\beta}}(\cdot)$*, where* $\mathbf{Z}_{\bar{\beta}}(\cdot)$ *is defined by*

its covariance function

$$\begin{aligned}\Sigma_{\bar{\beta}}(\tau,\tau') &= \mathbb{E}\left[\mathbf{Z}_{\bar{\beta}}(\tau)\,\mathbf{Z}_{\bar{\beta}}(\tau')'\right] \\ &= \frac{\mathbb{C}\left(\beta(\tau;\mathbf{X}),\beta(\tau';\mathbf{X})'\middle|\,\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)} + \Upsilon_1(\tau,\tau') + K^M\Sigma_\delta(\tau,\tau')K^{M\prime} \qquad (36)\end{aligned}$$

where

$$\begin{aligned}\Upsilon_1(\tau,\tau') &= \frac{(\min(\tau,\tau')-\tau\tau')}{\Pr(\mathbf{X}\in\mathbb{X}^M)}\mathbb{E}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Lambda(\tau,\tau';\mathbf{X})\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\,|\mathbf{X}\in\mathbb{X}^M\right], \\ K^M &= \mathbb{E}[K(\mathbf{X})|\mathbf{X}\in\mathbb{X}^M].\end{aligned}$$

The first term in the asymptotic distribution of $\widehat{\bar{\beta}}^M(\tau)$ arises from variation in the random coefficients across the subpopulation of movers. It would be zero if $F_{\mathbf{X}}$ were known. The second and third terms reflect sampling uncertainty in $\hat{\beta}(\tau;\mathbf{x})$ and $\hat{\delta}(\tau)$ respectively (which arises because the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is unknown). The form of $\Sigma_{\bar{\beta}}(\tau,\tau')$ mirrors that derived by Chamberlain (1992) for averages of random coefficients.

As with the ACQE, unconditional quantile effects are only identified across the subpopulation of movers. Our estimate of $\beta_p^M(\tau)$ is given by the solution to the empirical counterpart of (21) above

$$\sum_{i=1}^{N}\mathbf{1}\left(\mathbf{X}_i\in\mathbb{X}^M\right)\left\{\int_{u=0}^{u=1}\left[\mathbf{1}(\hat{\beta}_p(u;\mathbf{X}_i)\le\hat{\beta}_p^M(\tau))-\tau\right]\mathrm{d}u\right\}=0. \qquad (37)$$

The integral in (37) can be calculated exactly since $\hat{\beta}_p(u;\mathbf{x}_l)$ is piecewise linear for each x_l with finitely many pieces. Alternatively it may be approximated by a finite sum of the integrand evaluated at H evenly spaced points $u_1,\ldots,u_H$ between zero and one. In that case $\hat{\beta}_p^M(\tau)$ has a simple order statistic representation. Let $N_{\text{MOVER}}=\sum_{i=1}^N\mathbf{1}\left(\mathbf{X}_i\in\mathbb{X}^M\right)$ equal the number of movers in the sample and construct the list $\left\{\left\{\hat{\beta}_p(u_h;\mathbf{X}_i)\right\}_{h=1}^H\right\}_{i=1}^{N_{\text{MOVER}}}$.[14] The j^{th} order statistic of this list is our estimate of $\beta_p^M(\tau)$ where j satisfies

$$\frac{j}{HN_{\text{MOVER}}+1}<\tau\le\frac{j+1}{HN_{\text{MOVER}}+1}.$$

Theorem 4. *Under the assumptions maintained in Theorem 2 above* $\sqrt{N}\left(\widehat{\beta}_p^M(\cdot)-\beta_p^M(\cdot)\right)$

[14]We assume, without loss of generality, that the sample is ordered such that mover realizations appear first with indices $i=1,\ldots,N_{\text{MOVER}}$.

converges in distribution to a mean zero Gaussian process $\mathbf{Z}_{\beta_p}(\cdot)$*, where* $\mathbf{Z}_{\beta_p}(\cdot)$ *is defined by its covariance function,* $\Sigma_{\beta_p}(\tau,\tau')$

$$\Sigma_{\beta_p}(\tau,\tau') = \frac{1}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}\frac{\Upsilon_2(\tau,\tau')+\Upsilon_3(\tau,\tau')+\Upsilon_4(\tau,\tau')}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}\left(\beta_p^M(\tau)\middle|\,\mathbf{X}\in\mathbb{X}^M\right)f_{B_p|\mathbf{X}\in\mathbb{X}^M}\left(\beta_p^M(\tau')\middle|\,\mathbf{X}\in\mathbb{X}^M\right)}, \quad (38)$$

where

$$\Upsilon_2(\tau,\tau') = \frac{\mathbb{C}\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}),F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}, \quad (39)$$

$$\begin{aligned}\Upsilon_3(\tau,\tau') = \Pr\left(\mathbf{X}\in\mathbb{X}^M\right)^{-1}\mathbb{E}\big[&\big(\min\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}),F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\right)\\ &-F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X})F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\big)\\ &\times e_p'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\Lambda\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}),F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X});\mathbf{X}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}e_p\\ &\times f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X})f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\big|\,\mathbf{X}\in\mathbb{X}^M\big], \end{aligned} \quad (40)$$

and

$$\begin{aligned}\Upsilon_4(\tau,\tau') = e_p'\Bigg(\int\int\big[&f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x})f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{x}})\\ &\times K(\mathbf{x})\,\Sigma_\delta\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}),F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{x}})\right)K(\tilde{\mathbf{x}})'\big]\\ &\times f_{\mathbf{X}|\mathbf{X}\in\mathbb{X}^M}\left(\mathbf{x}|\,\mathbf{x}\in\mathbb{X}^M\right)f_{\mathbf{X}|\mathbf{X}\in\mathbb{X}^M}\left(\tilde{\mathbf{x}}|\,\tilde{\mathbf{x}}\in\mathbb{X}^M\right)\mathrm{d}\mathbf{x}\mathrm{d}\tilde{\mathbf{x}}\Bigg)e_p \end{aligned} \quad (41)$$

with e_p *a* $P\times 1$ *vector with a 1 in its* p^{th} *row and zeros elsewhere.*

While the form of the covariance function $\Sigma_{\beta_p}(\tau,\tau')$ is complicated, each term in it has a straightforward interpretation. The $\Upsilon_2(\cdot)$ term reflects sampling variability arising from the econometrician's lack of knowledge of the marginal distribution of $\mathbf{X}$. It would also be zero if the distribution of $B_p|\,\mathbf{X}=\mathbf{x}$ were constant across all $\mathbf{x}\in\mathbb{X}^M$ (i.e., no correlated heterogeneity). This term is analogous to the first term appearing the covariance expression for the ACQE in Theorem 3 above. The $\Upsilon_3(\cdot)$ term measures estimation error associated with a lack of knowledge of the distribution of $\mathbf{Y}$ given $\mathbf{X}$; specifically it captures the influence of sampling error in conditional quantiles of Y_t on the sampling variability of the estimated UQE. The $\Upsilon_4(\cdot)$ term is due to the estimation of the common coefficient $\delta(\cdot)$.

Irregular case ($T = P$)

Our estimation procedure for the regular $T > P$ case only utilizes information on movers. Our analysis of the irregular $T = P$ case additionally utilizes information on stayers and near stayers, making full use of the possibilities implied by Assumption 2. In the irregular case, similar to Graham and Powell (2012), estimation of the common coefficient, $\delta(\tau)$, requires the availability of stayers.

We introduce the presence of near stayers to illustrate how the population-wide ACQE and UQE, not just their movers' counterparts, may be consistently estimated.

Our analysis relies on "discrete bandwidth asymptotics", we argue that our approach, in addition to being of value on its own terms, approximates many features of an analysis with continuously-valued covariates. To motivate this claim we begin by reproducing the results of Graham and Powell (2012).

Discrete bandwidth framework

With $T = P$, the $\mathbf{X}$ matrix is square, with full rank if and only if $\det \mathbf{X} \neq 0$. Let $D = \det \mathbf{X}$ with $D \in \mathbb{D}_N = \{d_1, \ldots, d_K, -h_N, h_N, 0\}$, the support of the determinant of $\mathbf{X}$. The first K elements of $\mathbb{D}_N$ correspond to the L_1 mover support points of $\mathbf{X}$. The next two elements of $\mathbb{D}_N$ correspond to the $L - L_1$ near stayer support points of $\mathbf{X}$. The final element of $\mathbb{D}_N$ corresponds to the $M - L$ stayer support points of $\mathbf{X}$.

We let $\Pr(D = h_N) = \Pr(D = -h_N) = \phi_0 h_N$ for some $\phi_0 \geq 0$, defining $d_{K+1,N} = -h_N$, $d_{K+2,N} = h_N$ and $d_{K+3} = 0$. The mover support points d_k for $k = 1, \ldots, K$ are bounded away from 0 for all N.

We also let the probability of observing a singular $\mathbf{X}$ be $\Pr(D = 0) = 2\phi_0 h_N$. Finally, $\Pr(D = d_k) = \pi_k^N$ for $k = 1, \ldots, K$ with $\sum_{k=1}^K \pi_k^N = 1 - 4\phi_0 h_N$, with $4\phi_0 h_N < 1$ for all N. We also let $\pi_k = \lim_{N\to\infty} \pi_k^N$, so that $\sum_{k=1}^K \pi_k = 1$.[15]

In this setup, observations with $D = 0$ are stayers, $D = \pm h_N$ are near-stayers, while those with $D = d_k$ for $k = 1, \ldots, K$ correspond to movers. The inclusion of near-stayers is a way to approximate a continuous distribution of D, letting near stayers (those with $D = \pm h_N$) have characteristics very similar to those of stayers ($D = 0$).

Let $q_{mN|k} = \Pr(\mathbf{X} = \mathbf{x}_{mN} | D = d_k)$, $q_{mN|-h} = \Pr(\mathbf{X} = \mathbf{x}_{mN} | D = -h_N)$, $q_{mN|h} = \Pr(\mathbf{X} = \mathbf{x}_{mN} | D = h_N)$ and $q_{mN|0} = \Pr(\mathbf{X} = \mathbf{x}_{mN} | D = 0)$. For simplicity, we assume that $q_{mN|\cdot}$ does not vary with N, so that conditional on the value of the determinant, which has varying

[15] The π_k are well defined limits by Assumption 2.

support, the distribution of $\mathbf{X}$ does not depend on the sample size. We also assume that $\lim_{N\to\infty} q_{m|h} = \lim_{N\to\infty} q_{m|-h} = q_{m|0}$ for all $m = 1, \ldots, M$. This is a smoothness assumption.
Recall that $\mathbf{X}^* = \text{adj}(\mathbf{X})$ denote the adjoint of $\mathbf{X}$ such that $\mathbf{X}^{-1} = \frac{1}{D}\mathbf{X}^*$,when $\mathbf{X}^{-1}$ exists. We also let $\mathbf{Y}^* = \mathbf{X}^*\mathbf{Y}$ and $\mathbf{W}^* = \mathbf{X}^*\mathbf{W}$.

Average partial effects under discrete bandwidth asymptotics

To illustrate the operation of our discrete bandwidth framework in a familiar setting we revisit the conditional mean model studied by Graham and Powell (2012):

$$\mathbb{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{W}\delta_0 + \mathbf{X}\beta_0(\mathbf{X}).$$

For the case where X_t is continuously-valued, $T = P$, and other maintained assumptions, Graham and Powell (2012) estimate δ_0 and the average $\beta_0 = \mathbb{E}[\beta(\mathbf{X})]$ by (cf., equations (24) and (25) in their paper).

$$\hat{\delta} = \left(\frac{1}{Nh_N}\sum_{i=1}^{N} \mathbf{W}_i^{*\prime}\mathbf{W}_i^* \mathbf{1}(|D_i| < h_N)\right)^{-1} \left(\frac{1}{Nh_N}\sum_{i=1}^{N} \mathbf{W}_i^{*\prime}\mathbf{Y}_i^* \mathbf{1}(|D_i| < h_N)\right) \tag{42}$$

$$\hat{\beta} = \frac{\frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i^{-1}(\mathbf{Y}_i - \mathbf{W}_i\hat{\delta})\mathbf{1}(|D_i| \geq h_N)}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(|D_i| \geq h_N)} \tag{43}$$

where $\mathbf{Y}_i^* = \mathbf{X}_i^*\mathbf{Y}_i$.[16]
We now compute the asymptotic distribution of $\hat{\delta}$ and $\hat{\beta}$ under discrete bandwidth asymptotics. First, the numerator of term (42) is equal to $\frac{1}{h_N}\sum_{l=L+1}^{M} \mathbf{w}_{lN}^{*\prime}\mathbf{w}_{lN}^*\hat{p}_{lN} \xrightarrow{p} \sum_{l=L+1}^{M} \mathbf{w}_l^{*\prime}\mathbf{w}_l^* q_{l|0} 2\phi_0 = 2\mathbb{E}[\mathbf{W}^{*\prime}\mathbf{W}^*|D = 0]\phi_0$.
Let $\mathbf{U}_i^* = \mathbf{X}_i^*(\mathbf{Y}_i - \mathbf{W}_i\delta_0 - \mathbf{X}_i\beta_0(\mathbf{X}_i))$. As in equation (46) of Graham and Powell (2012), the numerator of $\hat{\delta} - \delta_0$ is equal to

$$\frac{1}{Nh_N}\sum_{i=1}^{N} \mathbf{W}_i^{*\prime}(D_i\beta_0(\mathbf{X}_i) + \mathbf{U}_i^*)\mathbf{1}(|D_i| < h_N) = \frac{1}{Nh_N}\sum_{i=1}^{N} \mathbf{W}_i^{*\prime}\mathbf{U}_i^*\mathbf{1}(D_i = 0).$$

This expression has mean zero since $\mathbb{E}[\mathbf{U}^*|\mathbf{X}] = 0$, and, letting $\Sigma(\mathbf{X})$ denote $\mathbb{V}(\mathbf{U}|\mathbf{X})$, we can verify that its asymptotic variance when premultiplied by $\sqrt{Nh_N}$ is $2\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^*\Sigma(\mathbf{X})\mathbf{X}^{*\prime}\mathbf{W}^*|D = 0\right]\phi_0$

[16]Note that, relative to their expressions, we have changed the definition of stayers from units with $|D_i| \leq h_N$ to units with $|D_i| < h_N$ and conversely for movers. This change has no impact when D is continuously distributed, and is made here to allow the near-stayers in our framework to be categorized as movers rather than stayers.

through a simple analysis. Therefore, we have

$$\sqrt{Nh_N}(\hat{\delta} - \delta_0) \xrightarrow{d} N\left(0, \frac{\Lambda_0}{2\phi_0}\right)$$

where $\Lambda_0 = \mathbb{E}[\mathbf{W}^{*\prime}\mathbf{W}^*|D=0]^{-1} \times \mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^*\Sigma(\mathbf{X})\mathbf{X}^{*\prime}\mathbf{W}^*|D=0\right] \times \mathbb{E}[\mathbf{W}^{*\prime}\mathbf{W}^*|D=0]^{-1}$.

In an analogy to Graham and Powell (2012), we see that ϕ_0 in this setup plays the exact same role as the density function of the determinant evaluated at 0. When a larger fraction of the sample is concentrated near or at $D = 0$, we can obtain more precision in our estimate of the common coefficient δ_0. We now decompose $\hat{\beta}$ into an infeasible version

$$\hat{\beta}^I = \frac{\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i^{-1}(\mathbf{Y}_i - \mathbf{W}_i\delta_0)\mathbf{1}(|D_i| \geq h_N)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(|D_i| \geq h_N)}$$

and a second term that contains the estimate of the common coefficient δ_0:

$$\hat{\beta} = \hat{\beta}^I + \hat{\Xi}_N(\hat{\delta} - \delta_0)$$

where $\hat{\Xi}_N = \frac{\frac{1}{N}\sum_{i=1}^N D_i^{-1}\mathbf{X}_i^*\mathbf{W}_i\mathbf{1}(|D_i|\geq h)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(|D_i|\geq h)}$. Since they are computed with different subsamples, $\hat{\beta}^I$ and $\hat{\delta}$ are independent.

The denominator of $\hat{\Xi}_N$ converges in probability to 1 since $h_N \to 0$. The numerator can be decomposed in two separate terms: $\sum_{l=1}^{L_1} \mathbf{x}_{lN}^{-1}\mathbf{w}_{lN}\hat{p}_{lN}$, which converges to $\sum_{l=1}^{L_1} \mathbf{x}_l^{-1}\mathbf{w}_l p_l$ and $\sum_{l=L_1+1}^{L} D_{lN}^{-1}\mathbf{w}_{lN}^*\hat{p}_{lN}$ which converges to a finite limit since D_{lN}^{-1} is either $\pm h_N^{-1}$ and $\hat{p}_{lN}$ is of order $O_p(h_N)$, and these two orders will cancel out. Therefore, as in Graham and Powell (2012), $\hat{\Xi}_N$ converges to well defined probability limit we denote by Ξ_0.

Finally, we see that

$$\begin{aligned}\hat{\beta}^I - \beta_0 = &\frac{\frac{1}{N}\sum_{i=1}^N(\beta_0(\mathbf{X}_i) - \beta_0)\mathbf{1}(|D_i| \geq h_N)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(|D_i| \geq h_N)} \\ &+ \frac{\frac{1}{N}\sum_{i=1}^N D_i^{-1}\mathbf{U}_i^*\mathbf{1}(|D_i| > h_N)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(|D_i| \geq h_N)} + \frac{\frac{1}{N}\sum_{i=1}^N D_i^{-1}\mathbf{U}_i^*\mathbf{1}(|D_i| = h_N)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(|D_i| \geq h_N)}.\end{aligned}$$

The denominators of all terms above, $\frac{1}{N}\sum_{i=1}^N \mathbf{1}(|D_i| \geq h)$, converges to 1 since the fraction of movers converges to 1. The numerator of the first term is equal to $\sum_{l=1}^{L} \beta_0(\mathbf{x}_{lN})(\hat{p}_{lN} - p_{lN})$. This term will be of order $\sqrt{N}$ since $\sqrt{N}(\hat{p}_{lN} - p_{lN}) = O_p(1)$ for $l = 1, \ldots, L_1$ and $O_p(\sqrt{h_N})$ for $l = L_1 + 1, \ldots, L$ by equation (61) in the appendix.

The numerator of the second term will be of order $\sqrt{N}$ since it concerns strict movers only, which have non-shrinking probabilities and D_i bounded away from 0. For these reasons, the

usual limit theorem can be applied to show this term exhibits a standard rate of convergence. The numerator of the third term's convergence is more delicate. Premultiplying this term by $\sqrt{Nh_N}$, its variance is equal to

$$\begin{aligned}
\mathbb{E}_N\left[\mathbf{X}^*\Sigma(\mathbf{X})\mathbf{X}^{*\prime}\frac{h\mathbf{1}(|D|=h_N)}{D^2}\right] &= \mathbb{E}_N\left[\mathbf{X}^*\Sigma(\mathbf{X})\mathbf{X}^{*\prime}\frac{\mathbf{1}(|D|=h_N)}{h_N}\right] \\
&= \mathbb{E}_N\left[\mathbf{X}^*\Sigma(\mathbf{X})\mathbf{X}^{*\prime}||D|=h_N\right]\frac{\Pr(|D|=h_N)}{h_N} \\
&\to 2\mathbb{E}\left[\mathbf{X}^*\Sigma(\mathbf{X})\mathbf{X}^{*\prime}|D=0\right]\phi_0 \\
&= 2\Upsilon_0\phi_0
\end{aligned}$$

since $\Pr(|D|=h_N) = 2\phi_0 h_N$ and by the continuity of the conditional distribution of $\mathbf{X}|D$ in D near 0. Combining results for these terms, we get that

$$\sqrt{Nh_N}(\hat{\beta}^I - \beta_0) \xrightarrow{d} N\left(0, 2\Upsilon_0\phi_0\right),$$

and using the independence of $\hat{\beta}^I$ and $\hat{\delta}$, we see that

$$\sqrt{Nh_N}(\hat{\beta} - \beta_0) \xrightarrow{d} \left(0, 2\Upsilon_0\phi_0 + \frac{\Xi_0\Lambda_0\Xi_0'}{2\phi_0}\right),$$

exactly as in Graham and Powell (2012, Theorem 2.1).

To make these results coincide, it is important to let $\Pr(|D| = h_N) = \Pr(D = 0)$. We make this assumption for the following reason: in the continuous setup, the fraction of the sample considered as stayers is approximately $2\phi_0 h_N$, and these stayers solely determine the asymptotic distribution of $\hat{\delta} - \delta_0$. For the estimation of β_0, we consider individuals with $|D| \geq h_N$, but the asymptotic behavior of $\hat{\beta} - \beta_0$ is solely driven by individuals with $|D|$ arbitrarily close to h_N. This is due to the D^{-1} term which diverges for individuals where $|D| = h_N$. In both cases, the set of individuals considered converges to the infinitesimal set of individuals with $D = 0$, since $h_N \to 0$ as $N \to \infty$, therefore, in a sense, the subsamples that generate the asymptotic variation in $\hat{\delta} - \delta_0$ and $\hat{\beta} - \beta_0$ are the same. This is why we place the same discrete probabilities on $|D| = h_N$ and on $D = 0$.

Quantile effects under discrete bandwidth asymptotics

We now study the estimation of the various quantile estimands introduced in Section 2 in the irregular $T = P$ case. To begin, we estimate $\delta(\tau)$, proceeding in analogy to the identification

analysis given above, by

$$\widehat{\delta}(\tau) = \left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{W}^{*\prime}_i \mathbf{W}^*{}_i \mathbf{1}(D_i < h_N)\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{W}^{*\prime}_i \mathbf{X}^*_i \widehat{Q}_{\mathbf{Y}|\mathbf{X}}\left(\tau|\,\mathbf{X}_i\right) \mathbf{1}(D_i < h_N)\right]. \quad (44)$$

With $\widehat{\delta}(\tau)$ in hand, we estimate the conditional quantiles of the random coefficients for all mover and near stayer support points by[17]

$$\hat{\beta}(\tau; \mathbf{x}_{lN}) = \mathbf{x}_{lN}^{-1}\left(\hat{Q}_{\mathbf{Y}|\mathbf{X}}\left(\tau|\,\mathbf{x}_{lN}\right) - \mathbf{w}_{lN}\hat{\delta}\left(\tau\right)\right) \text{ for } l = 1, \ldots, L.$$

To develop a formal result on the sampling properties of these two estimates we add the following assumption.

Assumption 7. (Irregular Case) *(i)* $\mathbb{E}[\mathbf{W}^{*\prime}\mathbf{W}^*|D=0]$ *is invertible, (ii)* $\|\mathbf{E}_N[\beta(\tau;\mathbf{X})|D = h_N] - \mathbf{E}_N[\beta(\tau;\mathbf{X})|D=0]\|$ *converges to 0 as* $N \to \infty$*, (iii)* $Nh_N \to \infty$ *and* $h_N \to 0$ *as* $N \to \infty$.

The second part of Assumption 7 implies that we can learn about the conditional distribution of random coefficients across stayers by studying that observed across near stayers; a smoothness condition.

Our first result for the irregular case is:

Theorem 5. *Suppose that Assumptions 1 through 5 and 7 are satisfied, then*
(i) $\sqrt{Nh_N}\left(\widehat{\delta}\left(\cdot\right) - \delta\left(\cdot\right)\right)$ *converges in distribution to a mean zero Gaussian process* $\mathbf{Z}_\delta\left(\cdot\right)$*, where* $\mathbf{Z}_\delta\left(\cdot\right)$ *is defined by its covariance function*

$$\begin{aligned}\Sigma_\delta(\tau,\tau') &= \mathbb{E}[\mathbf{Z}_\delta\left(\tau\right)\mathbf{Z}_\delta\left(\tau'\right)'] \\ &= \frac{(\min(\tau,\tau') - \tau\tau')}{2\phi_0}\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^*|D=0\right]^{-1} \times \\ &\quad \mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^*\Lambda(\tau,\tau';\mathbf{X})\mathbf{X}^{*\prime}\mathbf{W}^*|D=0\right]\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^*|D=0\right]^{-1}, \end{aligned} \quad (45)$$

(ii) $\sqrt{Nh_N}\left(\widehat{\beta}\left(\cdot;\mathbf{x}_{lN}\right) - \beta\left(\cdot;\mathbf{x}_{lN}\right)\right)$ *also converges in distribution for each* $l = 1, \ldots, L_1$ *to a mean zero Gaussian process* $\mathbf{Z}\left(\cdot,\mathbf{x}_l\right)$*, where* $\mathbf{Z}\left(\cdot,\mathbf{x}_l\right)$ *is defined by its covariance function*

$$\Sigma(\tau,\mathbf{x}_l,\tau',\mathbf{x}_m) = \mathbb{E}\left[\mathbf{Z}\left(\tau,\mathbf{x}_l\right)\mathbf{Z}\left(\tau',\mathbf{x}_m\right)'\right] = \mathbf{x}_l^{-1}\mathbf{w}_l\Sigma_\delta(\tau,\tau')\mathbf{w}'_m\mathbf{x}_m^{-1\prime} \quad (46)$$

[17]Note that, in our set-up, $\mathbf{1}(|D_i| < h_N) = \mathbf{1}(D_i = 0)$. We use the former representation to highlight how our results would extend to settings with continuously-valued covariates. Since (44) conditions on a subpopulation with mass shrinking to zero, estimation of $\delta(\tau)$ will not be possible at the regular rate of $\sqrt{N}$.

for $l, m = 1, \ldots, L_1$ *and*
(iii) $\sqrt{Nh_N^3}\left(\widehat{\beta}\left(\cdot; \mathbf{x}_{lN}\right) - \beta\left(\cdot; \mathbf{x}_{lN}\right)\right)$ *also converges in distribution for each* $l = L_1 + 1, \ldots, L$ *to a mean zero Gaussian process* $\mathbf{Z}\left(\cdot, \mathbf{x}_l\right)$*, where* $\mathbf{Z}\left(\cdot, \mathbf{x}_l\right)$ *is defined by its covariance function*

$$\begin{aligned}\Sigma(\tau, \mathbf{x}_l, \tau', \mathbf{x}_m) &= \mathbb{E}\left[\mathbf{Z}\left(\tau, \mathbf{x}_l\right) \mathbf{Z}\left(\tau', \mathbf{x}_m\right)'\right] \\ &= \left(\min\left(\tau, \tau'\right) - \tau\tau'\right) \frac{\mathbf{x}_l^* \Lambda\left(\tau, \tau'; \mathbf{x}_l\right) \mathbf{x}_l^{*\prime}}{q_{l|0} 2\phi_0} \cdot \mathbf{1}\left(l = m\right) \\ &+ \mathbf{w}_l^* \Sigma_\delta(\tau, \tau') \mathbf{w}_m^{*\prime}\end{aligned} \tag{47}$$

for $l, m = L_1 + 1, \ldots, L$.

The rate of convergence for $\hat{\delta}(\tau)$ coincide with that which would be expected when $\mathbf{X}$ has a continuous distribution, as in Graham and Powell (2012). The $\delta(\tau)$ estimator relies on the sample with $D = 0$, which has fraction equal to $2\phi_0 h_N$ giving an effective sample size of approximately $2N\phi_0 h_N$ for estimation. As ϕ_0 increases, more effective observations are available for estimation, and therefore the asymptotic precision increases. The influence of the preliminary quantile estimator appears through the $\Lambda(\tau, \tau'; \mathbf{X})$ matrix.

The conditional coefficient estimates, $\widehat{\beta}\left(\tau; \mathbf{x}_{lN}\right)$, converge at different rates depending on whether $\mathbf{x}_{lN}$ has shrinking mass or not. Since these estimates depend linearly on $\hat{\delta}(\tau)$, their fastest possible rate of convergence is $\sqrt{Nh_N}$, the rate of convergence of $\hat{\delta}(\tau)$. This rate is achieved for (strict) movers, whose population frequencies are bounded away from zero. In fact, for movers, the only component of the asymptotic variance of $\widehat{\beta}\left(\tau; \mathbf{x}_{lN}\right)$ is due to sampling variability in $\hat{\delta}(\tau)$, since the other ingredient to the estimator, the conditional quantiles of Y_t, are estimated at rate $\sqrt{N}$.

For units whose covariate sequences have shrinking mass, that is, for near-stayers, the rate of convergence of $\widehat{\beta}\left(\tau; \mathbf{x}_{lN}\right)$ is $\sqrt{Nh_N^3}$. For near stayers, $\mathbf{X}^{-1} = \mathbf{X}^* D^{-1}$, which diverges since $D = h_N \to 0$ as $N \to \infty$. To account for, and cancel, this shrinking denominator term, the extra h_N term is present. Note that we do not require that $Nh_N^3 \to \infty$ as $N \to \infty$, and in fact these conditional betas will not be consistently estimated if $Nh_N^3 \to 0$. This is not a problem, since their consistent estimation is not the goal. Rather, we will show that the ACQE and UQE estimators can incorporate these inconsistent estimates and still deliver a consistent and asymptotically normal estimator for these functionals.

We now turn to the estimation of the average conditional quantile effect (ACQE). The ACQE is consistently estimable under our discrete bandwidth setup because the mass of stayers shrinks to zero as $N \to \infty$. Specifically, the ACQE is identified by the limit $\bar{\beta}(\tau) = \lim_{N\to\infty} \mathbb{E}_N\left[\beta(\tau; \mathbf{X}) | \mathbf{X} \in \mathbb{X}_N^M\right]$ since $\beta(\tau; \mathbf{X})$ is identified on $\mathbb{X}_N^M$ and the probability mass of

stayers vanishes as N goes to infinity. Our estimate of the ACQE in the $T = P$ case is

$$\widehat{\bar{\beta}}_N(\tau) = \frac{\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i^{-1}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}\left(\tau|\,\mathbf{X}_i\right) - \mathbf{W}_i\widehat{\delta}(\tau)\right)\mathbf{1}(\mathbf{X}_i \in \mathbb{X}_N^M)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}_N^M)}.$$

Theorem 6. *Under Assumptions 1 through 5, Assumption 7, and $Nh_N^3 \to 0$, we have that:*

$$\sqrt{Nh_N}\left(\widehat{\bar{\beta}}_N(\tau) - \bar{\beta}_N(\tau)\right) \xrightarrow{d} \mathbf{Z}_{\bar{\beta}}(\tau),$$

a zero mean Gaussian process, on $\tau \in (0,1)$. The variance of the Gaussian process $\mathbf{Z}_{\bar{\beta}}(\cdot)$ is defined as

$$\mathbb{E}\left[\mathbf{Z}_{\bar{\beta}}(\tau)\mathbf{Z}_{\bar{\beta}}(\tau')'\right] = \Upsilon_1(\tau,\tau') + \Xi_0\Sigma_\delta(\tau,\tau')\Xi_0' \tag{48}$$

with

$$\begin{aligned}\Upsilon_1(\tau,\tau') &= 2\phi_0\left(\min\left(\tau,\tau'\right) - \tau\tau'\right)\mathbb{E}\left[\mathbf{X}^*\Lambda(\tau,\tau',\mathbf{X})\mathbf{X}^{*\prime}|D=0\right]\\ \Xi_0 &= \lim_{N\to\infty}\mathbb{E}_N\left[\mathbf{X}^{-1}\mathbf{W}||D| \geq h_N\right].\end{aligned}$$

The rate of convergence of $\widehat{\bar{\beta}}_N(\tau)$ is $\sqrt{Nh_N}$, as is the case for the average effect studied by Graham and Powell (2012). The asymptotic variance depends only on terms with $D = 0$, since only stayers and near stayers contribute to the asymptotic distribution of the estimator. If ϕ_0 increases, it is possible to estimate the term $\Xi_0\Sigma_\delta(\tau,\tau')\Xi_0'$ with more precision since $\hat{\delta}(\tau)$ is more precisely determined when there are many units with $D = 0$. On the other hand, term $\Upsilon_1(\tau,\tau')$ increases with ϕ_0. The intuition behind this increase is that there are more near-stayers when ϕ_0 is large, and their contributions to $\widehat{\bar{\beta}}_N(\tau)$ are estimated at a slower rate than those of movers.

Finally we turn to the unconditional quantile effect, $\beta_p(\tau)$, the τ^{th} quantile of B_p. As in the regular case our estimate is the solution to (37). The only difference between the regular and irregular case is the method used to estimate the conditional quantile effects $\beta_p(\tau,\mathbf{x})$.

Theorem 7. *Fix $p \in \{1,\ldots,P\}$. Under the assumptions maintained in Theorem 6 we have that*

$$\sqrt{Nh_N}\left(\widehat{\beta}_p(\tau) - \beta_p(\tau)\right) \xrightarrow{d} \mathbf{Z}_{\beta_p}(\tau)$$

on $\tau \in (0,1)$ with $\mathbf{Z}_{\beta_p}(\cdot)$ being a zero mean Gaussian process. The covariance of this Gaussian

process is equal to:

$$\mathbb{E}\left[\mathbf{Z}_{\beta_p}(\tau)\mathbf{Z}_{\beta_p}(\tau')'\right] = \frac{\Upsilon_3(\tau,\tau') + \Upsilon_4(\tau,\tau')}{f_{B_p}(\beta_p(\tau))f_{B_p}(\beta_p(\tau'))}$$

where

$$\begin{aligned}
\Upsilon_3(\tau,\tau') &= 2\phi_0\mathbb{E}\left[e_p'\mathbf{X}^{-1}\Lambda(F_{B_p|\mathbf{X}}(\beta_p(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p(\tau')|\mathbf{X}), \mathbf{X})\mathbf{X}^{-1\prime}e_p\right. \\
&\quad \times (\min(F_{B_p|\mathbf{X}}(\beta_p(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p(\tau')|\mathbf{X})) - F_{B_p|\mathbf{X}}(\beta_p(\tau)|\mathbf{X})F_{B_p|\mathbf{X}}(\beta_p(\tau')|\mathbf{X})) \\
&\quad \left.\times f_{B_p|\mathbf{X}}(\beta_p(\tau)|\mathbf{X})f_{B_p|\mathbf{X}}(\beta_p(\tau')|\mathbf{X})|D=0\right] \\
\Upsilon_4(\tau,\tau') &= \sum_{l=1}^{L}\sum_{l'=1}^{L} e_p'(\mathbf{x}_l^{-1}\mathbf{w}_l p_l \mathbf{1}(l \le L_1) \\
&\quad + \mathbf{w}_l^* q_{l|0} 2\phi_0 \mathbf{1}(l > L_1))\Sigma_\delta(F_{B_p|\mathbf{X}}(\beta_p(\tau)|\mathbf{x}_l), F_{B_p|\mathbf{X}}(\beta_p(\tau')|\mathbf{X}_{l'})) \\
&\quad \times (\mathbf{x}_{l'}^{-1}\mathbf{w}_{l'} p_{l'} \mathbf{1}(l' \le L_1) + \mathbf{w}_{l'}^* q_{l'|0} 2\phi_0 \mathbf{1}(l' > L_1))' e_p f_{B_p|\mathbf{X}}(\beta_p(\tau)|\mathbf{x}_l) f_{B_p|\mathbf{X}}(\beta_p(\tau')|\mathbf{x}_{l'}).
\end{aligned}$$

The asymptotic distribution of the UQE depends on the conditional density of $B_p|X$ evaluated at $\beta_p(\tau)$. The term $\Upsilon_3(\cdot)$ reflects the estimation error for the near-stayers' conditional quantile effects. The overall rate of convergence is $(Nh_N)^{-1/2}$. Although the conditional quantile effects of near stayers converge at rate $(Nh_N^3)^{-1/2}$, they enter the UQE with a weight which is of order $O(h_N)$, leading to the $(Nh_N)^{-1/2}$ rate. The $\Upsilon_4(\cdot)$ term reflects the influence of estimation error in $\hat{\delta}(\tau)$. Both these terms are divided by the density of B_p evaluated at $\beta_p(\tau)$, meaning that a larger density of the random coefficient around the estimated quantile will lead to a smaller asymptotic variance. How the constant ϕ_0 enters these equation tells us that a smaller density of stayers and near-stayers will lead to smaller asymptotic contribution of term $\Upsilon_3(\cdot)$ since there are less stayers excluded from the UQE estimator. On the other hand, a lower ϕ_0 can increase $\Upsilon_4(\cdot)$ since it reduces the precision of the estimator of $\delta(\tau)$, due to a lower relative sample size.

4 Union wage premium

The effect of collective bargaining coverage on the distribution of earnings is a question of longstanding interest to labor economists (e.g., Card, Lemieux and Riddell, 2004). This is also an area where both panel data and quantile regression methods have played important roles in empirical work (e.g., Chamberlain, 1982; Jakubson, 1991; Card, 1995; Chamberlain, 1994), making an analysis which combines both approaches of particular interest.

We begin with a target sample consisting of the 4,837 male NLSY79 respondents in the cross-

Table 1: Summary statistics

	Full Sample	**Stayers**		**Movers**
		Never	Always	
Entire Sample (N=2,444)	·	0.6579	0.1100	0.2322
Black (N=2,444)	0.1168	0.0864	0.1510	0.1868
Hispanic (N=2,444)	0.0602	0.0568	0.0616	0.0692
Years of Schooling (N=2,437)	12.99	13.24	12.59	12.50
	(2.17)	(2.31)	(1.50)	(1.92)
AFQT percentile (N=2,351)	52.00	56.57	47.72	40.87
	(29.88)	(29.91)	(25.86)	(28.34)
1988 Hourly Wage (N=2,444)	19.48	19.79	22.47	17.15
	(25.29)	(30.40)	(7.06)	(10.15)

Source: National Longitudinal Survey of Youth 1979 and authors' calculations.
Notes: Analysis based of the balanced panel of NLSY79 2,444 male respondents (in 2,104 households) described in the main text. AFQT corresponds to Armed Force Qualification Test. Stayers consist of workers who are never covered by a collective bargaining agreement as well as those who are always covered. Movers consist of individuals who move in and/or out of coverage during the sample period. The first row calculates the fraction of individuals corresponding to each of the three subgroups. Sample sizes are smaller for some covariates due to item non-response.

sectional and supplemental Black and Hispanic subsamples. Our frame excludes respondents in the supplementary samples of poor whites and military personnel (cf., MaCurdy, Mroz and Gritz, 1998). We constructed a balanced panel of respondents who were (i) engaged in paid private sector or government employment in each of the years 1988 to 1992 and (ii) had complete wage and union coverage information. Exclusion from the estimation sample occurred for several reasons. We excluded all self-employed individuals, individuals with stated hourly wages less than $1, or greater than $1,000, in 2010 prices, and individuals who were not surveyed in all five calendar years. We use the hourly wage measure associated with each respondent's "CPS" job. Our measure of collective bargaining coverage is also defined vis-a-vis the CPS job.[18] Respondents were between the ages of 24 and 33 in 1988 and hence past the normal school-leaving age.

Our estimation sample is similar to that used by Chernozhukov, Fernández-Val, Hahn and Newey (2013), who also study the union wage premia using the NLSY79. Our subsample includes slightly more individuals, primarily by virtue of the fact that we follow respondents for five instead of eight years, reducing attrition.

Table 1 reports a selection of worker attributes known to be predictive of wages by collec-

[18]The "CPS" job coincides with a respondents primary employment as determined by the same criteria used in the Current Population Survey (CPS).

tive bargaining coverage status. Column 1 reports the mean of these characteristics across all individuals in our sample (standard deviations are in parentheses for non-binary-valued variables). Column 2 reports the corresponding statistics for workers who are never covered by a collective bargaining agreement during the sample period, column 3 for those who are always covered, and column 4 for those who move in and/or out of coverage during the sample period. Movers are more likely to be minority and have lower years of completed school, AFQT scores and hourly wages. Workers who are never covered, have the lowest minority share, the greatest years of completed schooling, and highest AFQT scores.

Table 2 reports out main results. All specifications allow for shifts in the intercept over time, but maintain homogeneity of slope coefficients across time. Column 1 reports the coefficient on the union dummy in a simple pooled least squares fit of log wages onto the union dummy and a vector of year dummies. Column 2 reports the union coefficient in a specification that additionally adds a vector of covariates for race, education and AFQT (see the notes to Table 2 for details). Column 3 reports the union wage premium in a specification which includes worker-specific intercepts. The estimator is as described by Arellano and Bover (1995), which is a GMM variant of Chamberlain's (1984) minimum distance estimator for linear panel data models. Column 4 reports an estimate of the movers' *average* union wage premium using the variant of Chamberlain's (1992) correlated random coefficients estimator described in Graham and Powell (2012, Section 3.3). The movers average union effect is between one-half and two-thirds of the OLS estimates of Columns 1 and 2. It is also very close to the Column 3 effect which allows for intercept heterogeneity in the earnings function, but assumes a homogenous union effect.

A researcher studying Columns 1 through 4 might conclude that, while the incorporation of correlated intercept heterogeneity into earnings functions is important, allowing for slope heterogeneity is less so. We report movers' unconditional quantile partial effects of collective bargaining coverage for $\tau = 0.25, 0.5, 0.75$ in Column 5. Here we find evidence of substantial heterogeneity in the effect of collective bargaining coverage on wages. For over 25 percent of workers, the effect of coverage is estimated to be less than 5 percent, whereas it is in excess of 15 percent for a similar proportion of workers. Our movers' UQE are relatively precisely determined, with estimated standard errors only modestly larger than the Column 3 model which assumes a homogenous effect.

Figure 1 plots our estimated movers' unconditional quantile effects as well as 95 percent point-wise confidence bands. The figure also includes quantile effects associated with a model which does not incorporate correlated heteroegeneity. These effects are estimated by a linear quantile regression of wages onto a constant, the union dummy and four time dummies. The coefficients on the union dummy, rearranged to be monotonic, are plotted as

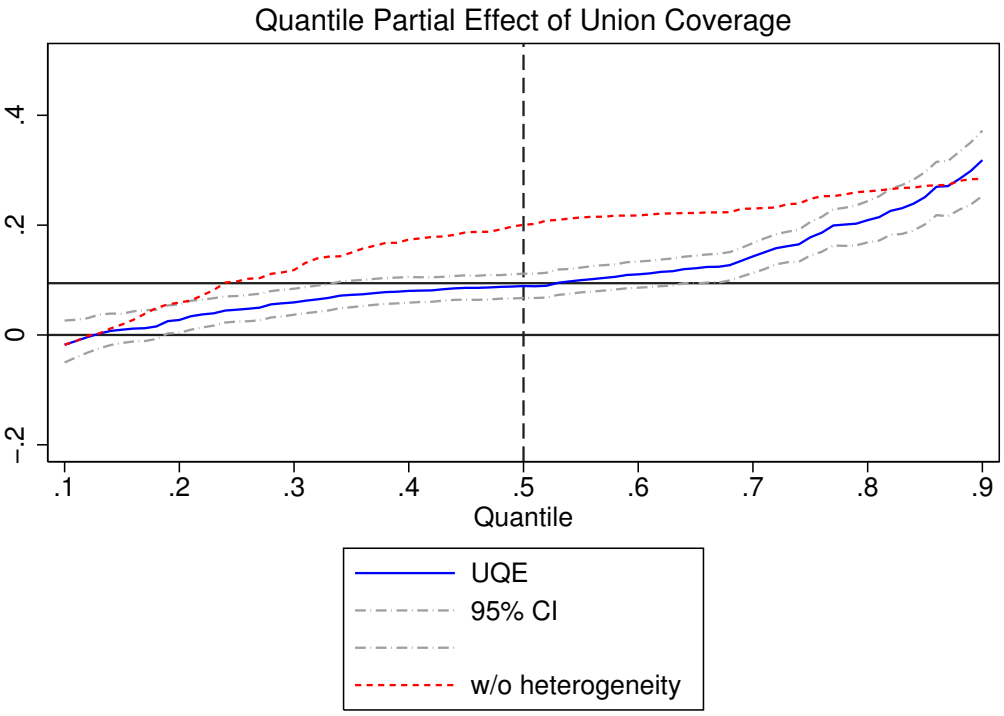
Table 2: Union wage premium

	(1) Pooled OLS	(2) Pooled OLS	(3) GMM Ch/AB	(4) CRC Avg.	(5) CRC $\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Union	0.1566	0.2225	0.0982	0.0936	0.0460	0.0891	0.1778
	(0.0186)	(0.0180)	(0.0134)	(0.0169)	(0.0141)	(0.0135)	(0.0186)
Covariates?	No	Yes	No	No		No	
$J_{(df)}$			$22.31_{(19)}$ (0.2691)				

Source: National Longitudinal Survey of Youth 1979 and authors' calculations.

Notes: All specifications include four time dummies capturing intercept shifts across periods. Column 2 additionally conditions on respondent's race (Black, Hispanic or non-Black, non-Hispanic), years of completed schooling at age 24, and AFQT percentile. Due to item non-response, this specification uses 2,348 respondents (in 2,023 households). Column 3 reports the union coefficient from a two-step GMM "fixed effects" specification where each respondent's individual-specific intercept is projected onto their entire union history and this history (plus a constant) are used as instruments for each time period. This generates $T(T+1) = 30$ moment restrictions for $2T+1 = 11$ parameters (and hence $T(T-1)-1 = 19$ over-identifying restrictions). See Arellano and Bover (1995) for estimation details. The Sargan-Hansen test statistic (and its p-value) for this specification is reported in the last row of the table. Columns 4 and 5 report correlated random coefficients specifications. Column 4 reports the movers' average union wage premium using Chamberlain's (1992) estimator following the specific implementation described in Graham and Powell (2012). Column 5 reports the movers unconditional quantile effect (UQE) using the estimator introduced here for $\tau = 0.25, 0.5, 0.75$. Standard errors reported in parentheses. For Columns 1 - 3 standard errors were analytically computed. For Columns 4 and 5 they were computed using the Bayesian Bootstrap. To be specific let $\hat{\beta}$ be the parameter estimate and $\hat{\beta}^{(b)}$ its b^{th} bootstrap value. Let $T_N^{(b)} = \hat{\beta}^{(b)} - \hat{\beta}$. A $1-\alpha$ bootstrap confidence interval is $\left[\hat{\beta} - F^{-1}_{T_N^{(b)}}(1-\alpha/2), \hat{\beta} - F^{-1}_{T_N^{(b)}}(\alpha/2)\right]$ (e.g., Hansen, 2014). The length of this interval divided by $2\Phi(1-\alpha/2)$ is the reported standard error estimate. Reported Column 4 and 5 point estimates were also biased corrected using the bootstrap.

Figure 1:



Source: National Longitudinal Survey of Youth 1979 and authors' calculations.
Notes: Blue line corresponds to the movers' unconditional quantile effect for $\tau \in (0.1, 0.9)$. Dashed grey lines are 95 percent point wise confidence intervals based on the Bayesian Bootstrap as described in the notes to Table 2 above. The dashed red line corresponds to the UQE associated with a simple pooled linear quantile regression of log wages onto a constant, the union dummy and four time dummies.

the dashed red line. As is the case for mean effects, quantile partial effects of union coverage are severely overstated in models which do not allow unobserved worker attributes to covary with union status.

Our empirical analysis also allows for assessment of the impact of collective bargaining coverage on inequality, at least within the subpopulation of movers. The counterfactual 1992 average 90-10 log wage gap in a world with no collective bargaining coverage is given by

$$\mathbb{E}\left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}' \left([\beta(0.9;\mathbf{X}) + \delta_{92}(0.9)] - [\beta(0.1;\mathbf{X}) + \delta_{92}(0.9)]\right) \middle| \mathbf{X} \in \mathbb{X}^M\right].$$

The corresponding gap in a world of universal coverage is given by

$$\mathbb{E}\left[\begin{pmatrix} 1 \\ 1 \end{pmatrix}' \left([\beta(0.9;\mathbf{X}) + \delta_{92}(0.9)] - [\beta(0.1;\mathbf{X}) + \delta_{92}(0.9)]\right) \middle| \mathbf{X} \in \mathbb{X}^M\right].$$

We estimate an average 90-10 gap in the no coverage counterfactual of 1.28. The corresponding gap in the universal coverage case is 1.07. The difference is 0.22 with a standard error of 0.08. Our analysis implies that unions have a substantially compressing effect on the distribution of wages, at least within the movers subpopulation.

5 Extensions

We briefly discuss a few extensions to our work. A longer discussion of each of these points can be found in the supplemental appendix.

Using stayers to estimate time effects when $T > P$

When identification is regular, as outlined above, $\delta(\tau)$ is estimable using mover units alone. However, it may nevertheless be advantageous to incorporate stayer units into the estimation procedure. This can improve the precision of $\hat{\delta}(\tau)$. It can also increase the precision with which the movers' ACQE and UQE are estimated, through the influence of sampling error in $\hat{\delta}(\tau)$ on the asymptotic variance of both of these objects (see Theorems 3 and 4 above). A description of the implementation procedure is described in the supplemental appendix.

Non-shrinking mass of stayers in the irregular case

In some applications, it is common to observe a positive mass of stayers at $D = 0$ along with a small number of near-stayers. We can model this in our discrete bandwidth framework by

letting $\Pr(D = 0) = \pi_0 + 2\phi_0 h_N$ where $\pi_0 > 0$ and keeping $\Pr(D = h_N) = \Pr(D = -h_N) = \phi_0 h_N$. We show how this mass of stayers requires us to modify the estimators for the ACQE and UQE and induces a slower rate of convergence of $\sqrt{Nh_N^3}$.

Continuous regressors and bandwidth selection

While not developed in this paper, the application of these models to continuously distributed $\mathbf{X}$ is an interesting research question. In the irregular case ($T = P$) we argue that the discrete bandwidth asymptotics represent a good approximation to the asymptotics when $\mathbf{X}$ is continuous. The main difficulty in extending our work to continuous $\mathbf{X}$ is the nonparametric estimation of the first-stage conditional quantiles, uniformly over τ and X. Although not shown in this paper, prior preliminary work using the calorie and expenditure data from Graham and Powell (2012) showed that results for the continuous case were similar to results where the continuous variable was discretized.[19] We discuss the potential application of existing nonparametric conditional quantile estimators (Qu and Yoon, 2015, Yu and Jones 1998, and Lee, 2013) to our framework in the supplemental appendix, and further develop the relationship between this case and the discrete bandwidth case covered earlier. We also establish a MSE-minimizing bandwidth selection rule for the tuning parameter h_N.

Trimming extremal quantiles

Estimation of quantiles can become problematic when the quantile considered is close to 0 or 1. As an example, the estimation of unconditional quantiles of a scalar random variable does not converge in process over $\tau \in (0, 1)$ when the support of the random variable is unbounded: see for example Lemma 21.4 in van der Vaart (2000). On the other hand, it does converge when considering $\tau \in [\epsilon, 1 - \epsilon]$ for any $0 < \epsilon < 1/2$. Whether the support is bounded or not, trimming extremal quantiles is common practice since estimators for them may be poorly approximated by the same asymptotic distribution as non-extremal quantiles. This could lead to a problem for the identification of $\beta(\tau)$ since it requires the use of $\beta(\tau; \mathbf{X})$ for all $\tau \in (0, 1)$. Trimming any fixed amount of quantiles will result in a loss of point identification, but we show in the supplemental appendix that partial identification of the UQE can be shown with tractable bounds which depend on the trimming parameter ϵ.

[19] A locally linear quantile regression estimator was used in the non-parametric first stage.

6 Conclusion

The extension of quantile regression methods to panel data has proved to be especially challenging. Our approach to this challenge generalizes both the textbook linear quantile regression and linear panel data models. Relative to these benchmarks our set-up allows for richer types of correlated (unobserved) heterogeneity, while still offering positive identification results. While the technical analysis of our estimators is non-trivial, their computation is not, requiring only sorting and weighted least squares operations. Our empirical analysis illustrates some of the possibilities of our approach.

One area of application where our methods may be especially attractive to researchers is for program and policy evaluation. As a concrete example consider a researcher who wishes to study the effect of minimum wage laws on the distribution of earnings using several waves of the Current Population Survey (CPS). Here $\mathbf{X}$ would encode the minimum wage level over time within a state. Since we observe many workers per state in each period, and hence per realization of $\mathbf{X}$, our discrete covariate results apply. Specifically $F_{Y_t|\mathbf{X}}(y|\mathbf{x})$ may be estimated by the empirical distribution function of period t wages in states with minimum wage sequence $\mathbf{X} = \mathbf{x}$. Applications in educational policy, where the entire distribution of test scores may be of interest, are also natural.

References

[1] Abrevaya, Jason. (2001). "The effects of demographics and maternal behavior on the distribution of birth outcomes," *Empirical Economics* 26 (1): 247 - 257.

[2] Abrevaya, Jason and Christian M. Dahl. (2008). "The effects of birth inputs on birthweight," *Journal of Business & Economic Statistics* 26 (4): 379 - 397.

[3] Angrist, Joshua, Victor Chernozhukov, and Iván Fernández-Val. (2006). "Quantile regression under misspecification, with an application to the U.S. wage structure," *Econometrica* 74 (2): 539 - 563.

[4] Arellano, Manuel and Stéphane Bonhomme. (2011). "Nonlinear panel data analysis," *Annual Review of Economics* 3: 395 - 424.

[5] Arellano, Manuel and Stéphane Bonhomme. (2012). "Identifying distributional characteristics in random coefficients panel data models," *Review of Economic Studies* 79 (3): 987 - 1020.

[6] Arellano, Manuel and Stéphane Bonhomme. (2016). "Nonlinear panel data estimation via quantile regressions," *Econometrics Journal*, forthcoming.

[7] Arellano, Manuel and Olympia Bover. (1995). "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics* 68 (1): 29 - 51.

[8] Athey, Susan and Guido W. Imbens. (2006). "Identification and inference in nonlinear difference-in-differences models," *Econometrica* 74 (2): 431 - 497.

[9] Autor David H., Lawrence F. Katz and Melissa S. Kearney. (2008). "Trends in U.S. wage inequality: revising the revisionists," *Review of Economics and Statistics* 90 (2): 300 - 323.

[10] Buchinsky, Moshe. (1994). "Changes in the U.S. wage structure 1963-1987: application of quantile regression," *Econometrica* 62 (2): 405 - 458.

[11] Cai, Zongwu, and Zhijie Xiao. (2012) "Semiparametric quantile regression estimation in dynamic models with partially varying coefficients," *Journal of Econometrics* 167 (2): 413-425.

[12] Card, David. (1995). "The effects of unions on the structure of wages: a longitudinal analysis," *Econometrica* 64 (4): 957 - 979.

[13] Card, David, Thomas Lemieux and W. Craig Riddell. (2004). "Unions and wage inequality," *Journal of Labor Research* 25 (4): 519 - 559.

[14] Chamberlain, Gary. (1982). "Multivariate regression models for panel data," *Journal of Econometrics* 18 (1): 5 - 46.

[15] Chamberlain, Gary. (1984). "Panel data," *Handbook of Economics* 2: 1247 - 1318 (Z. Griliches & M. Intriligator, Eds.). Amsterdam: North-Holland.

[16] Chamberlain, Gary. (1987). "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics* 34 (3): 305 - 334.

[17] Chamberlain, Gary. (1992). "Efficiency bounds for semiparametric regression," *Econometrica* 60 (3): 567 - 596.

[18] Chamberlain, Gary. (1994). "Quantile regression, censoring, and the structure of wages," *Advances in Econometrics: Sixth World Congress* 2: 171 - 209.

[19] Chaudhuri, Probal, Kjell Doksum and Alexander Samarov. (1997). "On average derivative quantile regression," *Annals of Statistics* 25 (2): 715 - 744.

[20] Chernozhukov, Victor, Iván Fernández-Val and Blaise Melly. (2013). "Inference on Counterfactual Distributions," *Econometrica* 81 (6): 2205 - 2268.

[21] Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn and Whitney Newey. (2013). "Average and quantile effects in nonseparable panel models," *Econometrica* 81 (2): 535 - 580.

[22] Chernozhukov, Victor, Iván Fernández-Val, Stefan Hoderlein, Hajo Holzmann and Whitney K. Newey. (2015). "Nonparametric identification in panels using quantiles," *Journal of Econometrics* 188 (2): 378 - 392.

[23] Chernozhukov, Victor and Christian Hansen. (2007). "Instrumental variable quantile regression: a robust inference approach," *Journal of Econometrics* 142 (1): 379 - 398.

[24] Clotilde, Elyès and Jouini Napp. (2004). "Conditional comonotonicity," *Decisions in Economics and Finance* 27 (2): 153 - 166.

[25] Firpo, Sergio. (2007). "Efficient semiparametric estimation of quantile treatment effects," *Econometrica* 75 (1): 259 – 276.

[26] Firpo, Sergio, Nicole M. Fortin and Thomas Lemieux. (2009). "Unconditional quantile regressions," *Econometrica* 77 (3): 953 – 973.

[27] Graham, Bryan S. and James L. Powell. (2012). "Identification and estimation of average partial effects in 'irregular' correlated random coefficient panel data models," *Econometrica* 80 (5): 2105 - 2152.

[28] Graham, Bryan S., Jinyong Hahn, Alexandre Poirier and James L. Powell. (2016). "Supplemental appendix to A quantile correlated random coefficients panel data model," *Mimeo.*

[29] Hahn, Jingyong and Whitney K. Newey. (2004). "Jackknife and analytical bias reduction for nonlinear panel models," *Econometrica* 72 (4): 1295 – 1319.

[30] Hansen, Bruce E. (2014). *Econometrics.* `http://www.ssc.wisc.edu/bhansen/econometrics/Econometrics.pdf`.

[31] Harding, Matthew and Carlos Lamarche. (2014). "Estimating and testing a quantile regression model with interactive effects," *Journal of Econometrics* 178: 101 - 113.

[32] Honda, Toshio. (2004). "Quantile regression in varying coefficient models," *Journal of statistical planning and inference* 121 (1): 113-125.

[33] Imbens, Guido. W. and Whitney K. Newey. (2009). "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica* 77 (5): 1481 – 1512.

[34] Jakubson, George. (1991). "Estimation and testing of the union wage effect using panel data," *Review of Economic Studies* 58 (5): 971 - 991.

[35] Kato, Kengo, Antonio F. Galvao Jr., Gabriel V. Montes-Rojas. (2012). "Asymptotics for panel quantile regression models with individual effects," *Journal of Econometrics* 170 (1): 76 - 91.

[36] Khan, Shakeeb and Elie Tamer. (2010). "Irregular identification, support conditions, and inverse weight estimation," *Econometrica* 78 (6): 2021 – 2042.

[37] Kim, Mi-Ok. (2007). "Quantile regression with varying coefficients," *The Annals of Statistics* 35 (1): 92-108.

[38] Kline, Patrick and Andres Santos. (2013). "Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure," *Quantitative Economics* 4 (2): 231 – 267.

[39] Koenker, Roger. (2004). "Quantile regression for longitudinal data," *Journal of Multivariate Analysis* 91 (1): 74 - 89.

[40] Koenker, Roger. (2005). *Quantile Regression.* Cambridge: Cambridge University Press.

[41] Koenker, Roger and Gilbert Bassett, Jr. (1978). "Regression quantiles," *Econometrica* 46 (1): 33 - 50.

[42] Lee, Sokbae (2003). "Efficient semiparametric estimation of a partially linear quantile regression model," *Econometric Theory*, 19 (1): 1-31.

[43] Lee, Ying-Ying. (2013). "Nonparametric weighted average quantile derivative," *Mimeo*, Nuffield College, Oxford University.

[44] Lemieux, Thomas. (2006). "Postsecondary education and increasing wage inequality," *American Economic Review* 96 (2): 195 - 199.

[45] Ma, Lingjie and Roger Koenker. (2006). "Quantile regression methods for recursive structural equation models," *Journal of Econometrics* 134 (2): 471 - 506.

[46] MaCurdy, Thomas, Thomas Mroz and R. Mark Gritz. (1998). "An evaluation of the National Longitudinal Survey on Youth," *Journal of Human Resources* 33 (2): 345 - 436.

[47] Machado, Jose. A. F. and Jose Mata. (2005). "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics* 20 (4): 445 – 465.

[48] Manski, Charles F. (1987). "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica* 55 (2): 357 - 362.

[49] Qu, Zhongjun and Jungmo Yoon. (2015) "Nonparametric estimation and inference on conditional quantile processes," *Journal of Econometrics* 185 (1): 1 - 19.

[50] Rosen, Adam. (2012). "Set identification via quantile restrictions in short panels," *Journal of Econometrics* 166 (1):127 - 137.

[51] van der Vaart, Aad W. *Asymptotic Statistics.* Cambridge University Press, 2000.

[52] Vella, Francis and Marno Verbeek. (1998). "Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men?" *Journal of Applied Econometrics* 13 (2): 163 - 183.

[53] Wang, Huixia Judy, Zhongyi Zhu, and Jianhui Zhou. (2009) "Quantile regression in partially linear varying coefficient models," *The Annals of Statistics* 37 (6B): 3841-3866.

[54] Wei, Ying and Raymond J. Carroll. (2009). "Quantile regression with measurement error," *Journal of the American Statistical Association* 104 (487): 1129 - 1143.

[55] Yu, Keming amd Jones, M. C. (1998). "Local linear quantile regression." *Journal of the American statistical Association* 93 (441): 228-237.

Proof of Lemmas and Theorems

Proof of Lemma 1

First consider the asymptotic distribution of the conditional CDF estimate

$$\widehat{F}_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN}) = \frac{\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(Y_{it}\leq c, \mathbf{X}_i = \mathbf{x}_{mN})}{\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})}. \tag{49}$$

We have that $\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN}) - p_{mN} = O_p\left(\frac{1}{\sqrt{Np_{mN}}}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(Y_{it}\leq c, \mathbf{X}_i = \mathbf{x}_{mN}) - \Pr(Y_{it}\leq c, \mathbf{X} = \mathbf{x}_{mN}) = O_p\left(\frac{1}{\sqrt{Np_{mN}}}\right)$. Using the delta method we get

$$\widehat{F}_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN}) - F_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{1}(Y_{it}\leq c, \mathbf{X}_i = \mathbf{x}_{mN})}{p_{mN}} - \frac{\Pr_N(Y_t\leq c|\mathbf{X} = \mathbf{x}_{mN})}{p_{mN}}\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})\right) + O_p\left(\frac{1}{Np_{mN}}\right). \tag{50}$$

By Lyapunov's Central Limit Theorem, and Assumptions 2 to 5, we have that, for fixed c, the normalized difference $\sqrt{Np_{mN}}\left(\widehat{F}_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN}) - F_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN})\right)$ is asymptotically normal with limiting variance equal to

$$\lim_{N\to\infty}\Pr(Y_t\leq c|\mathbf{X} = \mathbf{x}_{mN})(1-\Pr(Y_t\leq c|\mathbf{X} = \mathbf{x}_{mN})) = \Pr(Y_t\leq c|\mathbf{X} = \mathbf{x}_m)(1-\Pr(Y_t\leq c|\mathbf{X} = \mathbf{x}_m)).$$

Note that continuity of the conditional CDF of Y_{it} given $\mathbf{X}$, a component of Assumption 4, is important for this result.

The next step is to show that the convergence of the normalized difference is uniform in $c \in \mathbb{R}$ and that the limiting process is Gaussian. The normalized summand $\frac{\mathbf{1}(Y_{it} \leq c, \mathbf{X}_i = \mathbf{x}_{mN})}{\sqrt{p_{mN}}} - \frac{\Pr(Y_t \leq c|\mathbf{X}=\mathbf{x}_{mN})}{\sqrt{p_{mN}}}\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})$ can be shown to have a finite bracketing integral for any N (since indicator functions have finite bracketing integrals) and $\Pr(Y_t \leq c|\mathbf{X} = \mathbf{x}_{mN})$ has bounded derivatives in c since the conditional density of Y_t given $\mathbf{X}$ is uniformly bounded in all arguments by Assumption 4. Consider the function $G_N = \frac{\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})}{\sqrt{p_{mN}}}$ and notice it is an envelope function for

$$\frac{\mathbf{1}(Y_{it} \leq c, \mathbf{X}_i = \mathbf{x}_{mN})}{\sqrt{p_{mN}}} - \frac{\Pr(Y_t \leq c|\mathbf{X} = \mathbf{x}_{mN})}{\sqrt{p_{mN}}}\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN}). \tag{51}$$

This envelope function has the following properties: $\mathbb{E}[G_N^2] = 1$ and for $\varepsilon > 0$, $\mathbb{E}[G_N^2 \mathbf{1}(G_N > \varepsilon\sqrt{N})] \to 0$ as long as $\sqrt{Np_{mN}} \to \infty$, which is assumed (Assumption 2). Therefore, this estimator satisfies the conditions of Theorem 19.28 in van der Vaart (2000) and $\sqrt{Np_{mN}}\left(\widehat{F}_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN}) - F_{Y_t|\mathbf{X}}(c|\mathbf{x}_{mN})\right)$ is P-Donsker and therefore converges in process over $c \in \mathbb{R}$ for any $t = 1, \ldots, T$ and any $m = 1, \ldots, M$.

Next we use the fact that Y_t is bounded, its positive density, and Corollary 21.5 in van der Vaart (2000) (or Lemma 12.8 (ii) in Kosorok (2007)) to show that the inverse of the conditional CDF process, i.e. the conditional quantile process, converges over $\tau \in (0, 1)$ to a mean zero Gaussian process with the asymptotically linear representation:

$$\begin{aligned}
&\sqrt{Np_{mN}}(\widehat{Q}_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN}) - Q_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN})) \\
&= \frac{-1}{\sqrt{Np_{mN}}} \sum_{i=1}^{N} \frac{\left(\mathbf{1}(Y_{it} \leq Q_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN})) - \tau\right)\mathbf{1}(\mathbf{X}_i = \mathbf{x}_{mN})}{f_{Y_t|\mathbf{X}}(Q_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN})|\mathbf{x}_{mN})} + O_p\left(\frac{1}{\sqrt{Np_{mN}}}\right).
\end{aligned}$$

Since both t and m have finite range, the convergence of this process is uniform over all values of t and m (as well as on $\tau \in (0, 1)$). Using the continuity of $f_{Y_t|\mathbf{X}}(Q_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN})|\mathbf{x}_{mN})$ and the boundedness of indicator functions, we apply Lyapunov's Central Limit Theorem to show that the covariance kernel of the limiting process of $\sqrt{Np_{mN}}(\widehat{Q}_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN}) - Q_{Y_t|\mathbf{X}}(\tau|\,\mathbf{x}_{mN}))$ is equal to $\Sigma_Q(\tau, \mathbf{x}_l, \tau', \mathbf{x}_m) = (\min(\tau, \tau') - \tau\tau')\Lambda(\tau, \tau'; \mathbf{x}_l)\mathbf{1}(l = m)$ as claimed.

Proof of Theorem 1

By the Law of Total Probability

$$\tau = \Pr(B_p \leq \beta_p(\tau)) = \Pr(B_p \leq \beta_p(\tau)|\mathbf{X} \in \mathbb{X}^M)\Pr(\mathbf{X} \in \mathbb{X}^M) + \Pr(B_p \leq \beta_p(\tau)|\mathbf{X} \in \mathbb{X}^S)\Pr(\mathbf{X} \in \mathbb{X}^S).$$

The quantities $\Pr(\mathbf{X} \in \mathbb{X}^M)$, $\Pr(\mathbf{X} \in \mathbb{X}^S)$ and the conditional distribution of B_p given $\mathbf{X} \in \mathbf{X}^M$ are identified by arguments detailed above. The quantity $\Pr(B_p \leq \beta_p(\tau)|\mathbf{X} \in \mathbb{X}^S)$ can take arbitrary values in the $[0, 1]$ interval. Therefore, the identified set of $\beta_p(\tau)$ is defined by:

$$\left\{b_p(\tau) : \Pr(B_p \leq b_p(\tau)|\mathbf{X} \in \mathbb{X}^M)\Pr(\mathbf{X} \in \mathbb{X}^M) + q\Pr(\mathbf{X} \in \mathbb{X}^S) = \tau, q \in [0, 1]\right\}.$$

Since $\Pr(B_p \leq b_p(\tau)|\mathbf{X} \in \mathbb{X}^M)$ is monotone in $b_p(\tau)$, we can get bounds on $\beta_p(\tau)$ by inverting this region at $q = 0$ and $q = 1$:

$$\begin{aligned}\Pr(B_p \leq \beta_p(\tau)|\mathbf{X} \in \mathbb{X}^M) &\in \left[\frac{\tau - 1 \times \Pr(\mathbf{X} \in \mathbb{X}^S)}{\Pr(\mathbf{X} \in \mathbb{X}^M)}, \frac{\tau - 0 \times \Pr(\mathbf{X} \in \mathbb{X}^S)}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\right] \\ &\Rightarrow \beta_p(\tau) \in \left[\beta_p^M\left(\frac{\tau - \Pr(\mathbf{X} \in \mathbb{X}^S)}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\right), \beta_p^M\left(\frac{\tau}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\right)\right].\end{aligned}$$

Finally observe that if $\frac{\tau - \Pr(\mathbf{X} \in \mathbb{X}^S)}{\Pr(\mathbf{X} \in \mathbb{X}^M)} < 0$ or if, $\frac{\tau}{\Pr(\mathbf{X} \in \mathbb{X}^M)} > 1$, the quantiles of $B_p|\mathbf{X} \in \mathbb{X}^M$ are not defined. For these cases, the inversion leads to bounds of $\underline{b}_p$ and $\bar{b}_p$ respectively.

Proof of Theorem 2

Recall that $\gamma(\tau) = (\delta(\tau)', \beta(\tau; \mathbf{x}_1)', \ldots, \beta(\tau; \mathbf{x}_L)')'$, manipulating (32) of the main text yields

$$\sqrt{N}(\hat{\gamma}(\tau) - \gamma(\tau)) = (G'\widehat{A}G)^{-1}G'\widehat{A}\sqrt{N}(\widehat{\mathbf{\Pi}}(\tau) - \mathbf{\Pi}(\tau)). \tag{52}$$

By a Law of Large Numbers, $\widehat{A} \xrightarrow{p} A = \text{diag}\{p_1, \ldots, p_L\} \otimes I_T$ and by the Continuous Mapping Theorem $(G'\widehat{A}G)^{-1} \xrightarrow{p} (G'AG)^{-1}$ which can be shown to be equal to

$$(G'AG)^{-1} = \begin{pmatrix} \frac{\Gamma^{-1}}{\Pr(\mathbf{X} \in \mathbb{X}^M)} & \frac{-\Gamma^{-1}}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\underline{K}' \\ -\underline{K}\frac{\Gamma^{-1}}{\Pr(\mathbf{X} \in \mathbb{X}^M)} & \text{diag}\left\{\frac{(\mathbf{x}_1'\mathbf{x}_1)^{-1}}{p_1}, \ldots, \frac{(\mathbf{x}_L'\mathbf{x}_L)^{-1}}{p_L}\right\} + \underline{K}\frac{\Gamma^{-1}}{\Pr(\mathbf{X} \in \mathbb{X}^M)}\underline{K}' \end{pmatrix} \tag{53}$$

where $\underline{\mathrm{K}} = (K(\mathbf{x}_1)', \ldots, K(\mathbf{x}_L)')'$.

The numerator, $G'\widehat{A}\sqrt{N}(\widehat{\mathbf{\Pi}}(\cdot) - \mathbf{\Pi}(\cdot))$, converges in distribution, by Slutsky's theorem, to $G'A\tilde{Z}_Q(\cdot)$ where $\tilde{Z}_Q(\cdot) = \left(\frac{Z_Q(\cdot,\mathbf{x}_1)}{\sqrt{p_1}}', \ldots, \frac{Z_Q(\cdot,\mathbf{x}_L)}{\sqrt{p_L}}'\right)'$ with $Z_Q(\cdot, \mathbf{x}_l)$ as defined in the statement of Lemma 1. The variance-covariance matrix of $\tilde{Z}_Q(\cdot)$ is

$$\tilde{\Sigma}_Q(\tau, \tau') = (\min(\tau, \tau') - \tau\tau')\text{diag}\left\{\frac{\Lambda(\tau, \tau'; \mathbf{x}_1)}{p_1}, \ldots, \frac{\Lambda(\tau, \tau'; \mathbf{x}_L)}{p_L}\right\}. \tag{54}$$

This gives an asymptotic covariance of $\sqrt{N}(\hat{\gamma}(\tau) - \gamma(\tau))$ and $\sqrt{N}(\hat{\gamma}(\tau') - \gamma(\tau'))$ equal to

$$(G'AG)^{-1}G'A\tilde{\Sigma}_Q(\tau, \tau')AG(G'AG)^{-1}. \tag{55}$$

Partitioning this matrix yields an asymptotic covariance of $\sqrt{N}(\hat{\delta}(\tau) - \delta(\tau))$ and $\sqrt{N}(\hat{\delta}(\tau') - \delta(\tau'))$ equal

$$\Sigma_\delta(\tau, \tau') = (\min(\tau, \tau') - \tau\tau')\frac{\Gamma^{-1}\mathbb{E}\left[\overline{\mathbf{W}}'\Lambda(\tau, \tau'; \mathbf{X})\overline{\mathbf{W}}\middle|\mathbf{X} \in \mathbb{X}^M\right]\Gamma^{-1}}{\Pr(\mathbf{X} \in \mathbb{X}^M)}, \tag{56}$$

as claimed in the statement of the theorem.

The asymptotic covariance of $\sqrt{N}(\hat{\beta}(\tau, \mathbf{x}_l) - \beta(\tau, \mathbf{x}_l))$ and $\sqrt{N}(\hat{\beta}(\tau', \mathbf{x}_m) - \beta(\tau', \mathbf{x}_m))$ is

$$\begin{aligned}\mathbb{E}\left[\mathbf{Z}(\tau, \mathbf{x}_l)\mathbf{Z}(\tau', \mathbf{x}_m)'\right] &= (\min(\tau, \tau') - \tau\tau')\frac{(\mathbf{x}_l'\mathbf{x}_l)^{-1}\mathbf{x}_l'\Lambda(\tau, \tau'; \mathbf{x}_l)\mathbf{x}_l(\mathbf{x}_l'\mathbf{x}_l)^{-1}}{p_l} \cdot \mathbf{1}(l = m) \\ &\quad + K(\mathbf{x}_l)\Sigma_\delta(\tau, \tau')K(\mathbf{x}_m)',\end{aligned}$$

as claimed.

Proof of Theorem 3

Manipulating (35) of the main text yields

$$\widehat{\bar{\beta}}^M(\tau) - \bar{\beta}^M(\tau) = \frac{\frac{1}{N}\sum_{i=1}^N (\hat{\beta}(\tau;\mathbf{X}_i) - \beta(\tau;\mathbf{X}_i))\mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)} \tag{57}$$

$$+ \frac{\frac{1}{N}\sum_{i=1}^N \beta(\tau;\mathbf{X}_i)\mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)} - \mathbb{E}[\beta(\tau;\mathbf{X})|\mathbf{X} \in \mathbb{X}^M]. \tag{58}$$

Consider term (57) in this expansion. Its denominator $\frac{1}{N}\sum_{i=1}^N (\mathbf{X}_i \in \mathbb{X}^M)$ converges, by a Law of Large Numbers to $\Pr(\mathbf{X} \in \mathbb{X}^M)$. The numerator will converge when normalized by $\sqrt{N}$ to the following Gaussian process:

$$\frac{1}{N}\sum_{i=1}^N \sqrt{N}(\hat{\beta}(\tau;\mathbf{X}_i) - \beta(\tau;\mathbf{X}_i))\mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M) = \sum_{l=1}^L \sqrt{N}(\hat{\beta}(\tau;\mathbf{x}_l) - \beta(\tau;\mathbf{X}_l))\hat{p}_l \xrightarrow{d} \sum_{l=1}^L \mathbf{Z}(\tau,\mathbf{x}_l)p_l. \tag{59}$$

The asymptotic covariance of (59) equals

$$\begin{aligned}
&\sum_{l=1}^L\sum_{l'=1}^L \mathbb{E}\left[\mathbf{Z}(\tau,\mathbf{x}_l)\mathbf{Z}(\tau',\mathbf{x}_l')\right] p_l p_{l'} \\
&= \sum_{l=1}^L\sum_{l'=1}^L \left(\min(\tau,\tau') - \tau\tau'\right) \frac{\left(\mathbf{x}_l'\mathbf{x}_l\right)^{-1}\mathbf{x}_l'\Lambda\left(\tau,\tau';\mathbf{x}_l\right)\mathbf{x}_l\left(\mathbf{x}_l'\mathbf{x}_l\right)^{-1}}{p_l} \cdot \mathbf{1}\left(l = l'\right) p_l p_{l'} \\
&+ \sum_{l=1}^L\sum_{l'=1}^L K\left(\mathbf{x}_l\right)\Sigma_\delta(\tau,\tau')K\left(\mathbf{x}_{l'}\right)' p_l p_{l'} \\
&= \left(\min(\tau,\tau') - \tau\tau'\right)\mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\Lambda\left(\tau,\tau';\mathbf{X}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}|\mathbf{X} \in \mathbb{X}^M\right]\Pr(\mathbf{X} \in \mathbb{X}^M) \\
&+ K^M\Sigma_\delta(\tau,\tau')K^{M\prime}\Pr(\mathbf{X} \in \mathbb{X}^M)^2.
\end{aligned}$$

Now consider term (58). Replacing the sample average by empirical probabilities yields

$$\begin{aligned}
\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^N \beta(\tau;\mathbf{X}_i)\mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)} - \mathbb{E}[\beta(\tau;\mathbf{X})|\mathbf{X} \in \mathbb{X}^M]\right) &= \sqrt{N}\left(\sum_{l=1}^L \beta(\tau;\mathbf{x}_l)\hat{q}_l^M - \sum_{l=1}^L \beta(\tau;\mathbf{x}_l)q_l^M\right) \\
&= \sum_{l=1}^L \beta(\tau;\mathbf{x}_l)\sqrt{N}\left(\hat{q}_l^M - q_l^M\right)
\end{aligned} \tag{60}$$

where q_l^M denotes the conditional probability $\Pr(\mathbf{X} = \mathbf{x}_l|\mathbf{X} \in \mathbb{X}^M)$ and $\hat{q}_l^M = \frac{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_l)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}$ its estimate. To derive the asymptotic distribution of this term, we must first derive the asymptotic distribution of

$\hat{q}_l^M - q_l^M$. Begin with the fact that

$$\sqrt{N}\begin{pmatrix} \frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_1) - p_1 \\ \vdots \\ \frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x}_L) - p_L \\ \frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M) - \Pr(\mathbf{X}_i \in \mathbb{X}^M) \end{pmatrix}$$
$$\xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} p_1(1-p_1) & \cdots & -p_1 p_L & p_1(1-\Pr(\mathbf{X}\in\mathbb{X}^M)) \\ \vdots & \ddots & \vdots & \vdots \\ -p_1 p_L & \cdots & p_L(1-p_L) & p_L(1-\Pr(\mathbf{X}\in\mathbb{X}^M)) \\ p_1(1-\Pr(\mathbf{X}\in\mathbb{X}^M)) & \cdots & p_L(1-\Pr(\mathbf{X}\in\mathbb{X}^M)) & \Pr(\mathbf{X}\in\mathbb{X}^M)(1-\Pr(\mathbf{X}\in\mathbb{X}^M)) \end{pmatrix}\right) \tag{61}$$

and use the delta method to show

$$\sqrt{N}\begin{pmatrix} \hat{q}_1^M - q_1^M \\ \vdots \\ \hat{q}_L^M - q_L^M \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\underline{0}_L, \Sigma_{q^M}\right) \tag{62}$$

where

$$\Sigma_{q^M} = \frac{1}{\Pr(\mathbf{X}\in\mathbb{X}^M)}\begin{pmatrix} q_1^M(1-q_1^M) & \cdots & -q_1^M q_L^M \\ \vdots & \ddots & \vdots \\ -q_1^M q_L^M & \cdots & q_L^M(1-q_L^M) \end{pmatrix}. \tag{63}$$

Note that some of these limiting variance matrices are singular. Combining these results yields an asymptotic variance-covariance for (60) of

$$\sum_{l=1}^L \sum_{l'=1}^L \beta(\tau;\mathbf{x}_l)\Sigma_{q^M}(l,l')\beta(\tau';\mathbf{x}_{l'})' \tag{64}$$
$$= \frac{1}{\Pr(\mathbf{X}\in\mathbb{X}^M)}\sum_{l=1}^L\sum_{l'=1}^L \beta(\tau;\mathbf{x}_l)(q_l^M\mathbf{1}(l=l') - q_l^M q_{l'}^M)\beta(\tau';\mathbf{x}_{l'})'$$
$$= \frac{1}{\Pr(\mathbf{X}\in\mathbb{X}^M)}\left(\mathbb{E}\left[\beta(\tau;\mathbf{X})\beta(\tau';\mathbf{X})'|\mathbf{X}\in\mathbb{X}^M\right] - \mathbb{E}\left[\beta(\tau;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right]\mathbb{E}\left[\beta(\tau';\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right]'\right)$$
$$= \frac{\mathbb{C}(\beta(\tau;\mathbf{X}),\beta(\tau';\mathbf{X})|\mathbf{X}\in\mathbb{X}^M)}{\Pr(\mathbf{X}\in\mathbb{X}^M)}. \tag{65}$$

Terms (59) and (60) are asymptotically independent, since the variation of (59) is conditional on $\mathbf{X}$, while the variation of (60) depends on $\mathbf{X}$ alone. Therefore, the limiting covariance of $\widehat{\widehat{\beta}}^M(\tau)$ is the sum of the covariances of its two components, which yields the claimed result.

Proof of Theorem 4

There are two main steps to the proof. The first is to recover an estimate of the distribution function of the random coefficient B_p (within the subpopulation of movers). The second is to invert this distribution

function to recover the quantile function of the movers' random coefficients.

For the first step, let $c \in \mathbb{R}$ and consider the asymptotic distribution of the estimated distribution function evaluated at c, denoted by $\hat{F}_{\hat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M}(c) = \sum_{l=1}^{L}\left[\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u\right]\hat{q}_l^M$:

$$\begin{aligned}
&\sqrt{N}\left(\hat{F}_{\hat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M}(c) - F_{B_p|\mathbf{X}\in\mathbb{X}^M}(c)\right) \\
&= \sum_{l=1}^{L}\sqrt{N}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u\right)\hat{q}_l^M \qquad (66)\\
&+ \sum_{l=1}^{L}\left[\int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u\right]\sqrt{N}\left(\hat{q}_l^M - q_l^M\right). \qquad (67)
\end{aligned}$$

Both of these two terms converge uniformly over $c \in \mathbb{R}$. For the first term, we have that

$$\sqrt{N}\left(\hat{\beta}_p(\tau;\mathbf{x}_l) - \hat{\beta}_p(\tau;\mathbf{x}_l)\right) \xrightarrow{d} \left(\mathbf{Z}(\tau,\mathbf{x}_l)\right)_p = \mathbf{Z}_p(\tau,\mathbf{x}_l)$$

over $\tau \in (0,1)$ and all $l = 1,\ldots,L$ (here $\mathbf{Z}_p(\tau,\mathbf{x}_l)$ is as defined in the statement of Theorem 4 and $(\cdot)_p$ denotes the p^{th} element of the vector). Let $F_p^c(\cdot) : C[0,1] \to \mathbb{R}$ be a functional such that $F_c(\beta(\cdot,\mathbf{x})) = \int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}) \leq c)\mathrm{d}u$. By Lemma 8 of Chernozhukov, Fernández-Val and Melly (2012), this functional is Hadamard differentiable and we can apply the functional delta method, yielding

$$\begin{aligned}
&\sqrt{N}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u\right) \\
&= \sqrt{N}\left(\hat{\beta}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l);\mathbf{x}_l) - \hat{\beta}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l);\mathbf{x}_l)\right)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l) + o_p(1) \xrightarrow{d} \mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l).
\end{aligned} \qquad (68)$$

This convergence is uniform in $c \in \mathbb{R}$ since $F_{B_p|\mathbf{X}}(c|\mathbf{x}_l)$ ranges between 0 and 1, and uniformly in $\mathbf{x}_l$ (since there are finitely many possible values for $\mathbf{x}_l$). Therefore,

$$\sum_{l=1}^{L}\sqrt{N}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u\right)\hat{q}_l^M \xrightarrow{d} \sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)q_l^M$$

for $c \in \mathbb{R}$. Also, similar to (65) above, $\sum_{l=1}^{L}\int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l) \leq c)\mathrm{d}u\sqrt{N}\left(\hat{q}_l^M - q_l^M\right)$ will converge over $c \in \mathbb{R}$ to a mean zero Gaussian process $\mathbf{Z}_{2p}(c)$ with asymptotic covariance

$$\mathbb{E}\left[\mathbf{Z}_{2p}(c)\mathbf{Z}_{2p}(c')'\right] = \frac{\mathbb{C}\left(F_{B_p|\mathbf{X}}(c|\mathbf{X}), F_{B_p|\mathbf{X}}(c'|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}. \qquad (69)$$

Note that $\mathbf{Z}_{2p}(c)$ and $\sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)q_l^M$ are uncorrelated since the variation in the latter is conditional on $\mathbf{X}$ while that in the former depends on $\mathbf{X}$ only. Therefore,

$$\sqrt{N}\left(\hat{F}_{\hat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M}(c) - F_{B_p|\mathbf{X}\in\mathbb{X}^M}(c)\right) \xrightarrow{d} \sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)q_l^M + \mathbf{Z}_{2p}(c) \qquad (70)$$

over $c \in \mathbb{R}$.

We now turn to step two of the proof: inverting the distribution of the random coefficient to recover the distribution of its quantiles. By assumption 5, B_p is assumed to have bounded support and $f_{B_p|\mathbf{X}\in\mathbb{X}^M}(c)$

is strictly positive for all $c \in \mathbb{R}$. Therefore by Kosorok (2007), Lemma 12.8 part (ii) the inverse functional is Hadamard differentiable into $\tilde{D}(0,1)$, the space of left-continuous functions with right-hand limits. Evaluating this inverse at $c = \beta_p^M(\tau)$ yields

$$\sqrt{N}\left(\hat{\beta}_p^M(\tau) - \beta_p^M(\tau)\right) \xrightarrow{d} \frac{\sum_{l=1}^L \mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l), \mathbf{x}_l) f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l) q_l^M + \mathbf{Z}_{2p}(\beta_p^M(\tau))}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau))} \tag{71}$$

$$= \mathbf{Z}_{\beta_p}(\tau) \tag{72}$$

uniformly over $\tau \in (0,1)$. To conclude the proof, we evaluate $\mathbb{E}\left[\mathbf{Z}_{\beta_p}(\tau)\mathbf{Z}_{\beta_p}(\tau')'\right]$, the asymptotic covariance of this Gaussian process:

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{Z}_{\beta_p}(\tau)\mathbf{Z}_{\beta_p}(\tau')'\right] \\
&= \frac{\sum_{l=1}^L \sum_{l'=1}^L f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l) e_p' \Sigma(\tau, F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l), \mathbf{x}_l), \tau', F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}), \mathbf{x}_{l'})) e_p f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}) q_l^M q_{l'}^M}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau)) f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau'))} \\
&+ \frac{\mathbb{E}\left[\mathbf{Z}_{2p}(\beta_p^M(\tau))\mathbf{Z}_{2p}(\beta_p^M(\tau'))\right]}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau)) f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau'))}
\end{aligned}$$

where

$$\begin{aligned}
&\frac{\sum_{l=1}^L \sum_{l'=1}^L f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l) q_l^M e_p' \Sigma(\tau, F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l), \mathbf{x}_l), \tau', F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}), \mathbf{x}_{l'}) e_p f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}) q_{l'}^M}{f_{\mathbf{B}_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau)) f_{\mathbf{B}_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau'))} \\
&= \frac{1}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)} \mathbb{E}\left[\left(\min\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\right) - F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}) F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\right)\right. \\
&\times e_p'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\Lambda\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X}); \mathbf{X}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} e_p \\
&\left. f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}) f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X}) \middle| \mathbf{X}\in\mathbb{X}^M\right] \qquad (73) \\
&+ \mathbb{E}\left[f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}) f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{X}})\right. \\
&\left. e_p' K(\mathbf{X})\Sigma_\delta\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{X}})\right) K\left(\tilde{\mathbf{X}}\right)' e_p \middle| \mathbf{X}\in\mathbb{X}^M, \tilde{\mathbf{X}}\in\mathbb{X}^M\right] \qquad (74) \\
&= \Upsilon_3(\tau,\tau') + \Upsilon_4(\tau,\tau'),
\end{aligned}$$

where $\tilde{\mathbf{X}}$ is an independent copy of $\mathbf{X}$. Also,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{Z}_{2p}(\beta_p^M(\tau))\mathbf{Z}_{2p}(\beta_p^M(\tau'))\right] &= \frac{\mathbb{C}\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)} \qquad (75) \\
&= \Upsilon_2(\tau,\tau'),
\end{aligned}$$

which agrees with the expressions in the statement of the Theorem.

Proof of Theorem 5

From (42) we get the following asymptotically linear representation

$$\begin{aligned}
&\sqrt{Nh_N}\left(\hat{\delta}(\tau)-\delta(\tau)\right) \qquad (76)\\
&=\left(\frac{1}{Nh_N}\sum_{i=1}^{N}\mathbf{W}_i^{*\prime}\mathbf{W}_i^{*}\mathbf{1}(D_i=0)\right)^{-1}\frac{1}{Nh_N}\sum_{i=1}^{N}\mathbf{W}_i^{*\prime}\mathbf{X}_i^{*}\sqrt{Nh_N}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}_i)-Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}_i)\right)\mathbf{1}(D_i=0)\\
&=\left(\sum_{l=L+1}^{M}\mathbf{w}_{lN}^{*\prime}\mathbf{w}_{lN}^{*}\frac{\hat{p}_{lN}}{h_N}\right)^{-1}\sum_{l=L+1}^{M}\mathbf{w}_{lN}^{*\prime}\mathbf{x}_{lN}^{*}\frac{\hat{p}_{lN}}{\sqrt{h_N p_{lN}}}\sqrt{Np_{lN}}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})-Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})\right).
\end{aligned}$$

with

$$\sum_{l=L+1}^{M}\mathbf{w}_{lN}^{*\prime}\mathbf{w}_{lN}^{*}\frac{\hat{p}_{lN}}{h_N}\xrightarrow{p}\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^{*}|D=0\right]2\phi_0$$

since $\mathbf{w}_{lN}^{*}\to\mathbf{w}_l^{*}$ and $\frac{\hat{p}_{lN}}{h_N}\xrightarrow{p}q_{l|0}2\phi_0$ as $N\to\infty$. Similarly, we get

$$\sum_{l=L+1}^{M}\mathbf{w}_{lN}^{*\prime}\mathbf{x}_{lN}^{*}\frac{\hat{p}_{lN}}{\sqrt{h_N p_{lN}}}\sqrt{Np_{lN}}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})-Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})\right)\xrightarrow{p}\sqrt{2\phi_0}\sum_{l=L+1}^{M}\mathbf{w}_l^{*\prime}\mathbf{x}_l^{*}\sqrt{q_{l|0}}\mathbf{Z}_Q(\tau,\mathbf{x}_l) \qquad (77)$$

by Slutsky's theorem since $\frac{\hat{p}_{lN}}{\sqrt{h_N p_{lN}}}\xrightarrow{p}\sqrt{2\phi_0 q_{l|0}}$. The limiting distribution (77) has asymptotic covariance equal to

$$\begin{aligned}
&\mathbb{E}\left[\sqrt{2\phi_0}\sum_{l=L+1}^{M}\mathbf{w}_l^{*\prime}\mathbf{x}_l^{*}\sqrt{q_{l|0}}\mathbf{Z}_Q(\tau,\mathbf{x}_l)\left(\sqrt{2\phi_0}\sum_{l'=L+1}^{M}\mathbf{w}_{l'}^{*\prime}\mathbf{x}_{l'}^{*}\sqrt{q_{l'|0}}p\mathbf{Z}_Q(\tau',\mathbf{x}_{l'})\right)'\right]\\
&=2\phi_0\sum_{l=L+1}^{M}\sum_{l'=L+1}^{M}\mathbf{w}_l^{*\prime}\mathbf{x}_l^{*}(\min(\tau,\tau')-\tau\tau')\Lambda(\tau,\tau';\mathbf{x}_l)\cdot\mathbf{1}(l=l')\,\mathbf{x}_l^{*\prime}\mathbf{w}_l^{*}q_{l|0}\\
&=2\phi_0(\min(\tau,\tau')-\tau\tau')\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^{*}\Lambda(\tau,\tau';\mathbf{X})\mathbf{X}^{*\prime}\mathbf{W}^{*}|D=0\right]. \qquad (78)
\end{aligned}$$

To derive the asymptotic distribution of $\sqrt{Nh_N}\left(\widehat{\beta}(\cdot;\cdot)-\beta(\cdot;\cdot)\right)$ for strict movers ($l=1,\ldots,L_1$), we note that

$$\begin{aligned}
&\sqrt{Nh_N}\left(\widehat{\beta}(\tau;\mathbf{x}_{lN})-\beta(\tau;\mathbf{x}_{lN})\right) \qquad (79)\\
&=\mathbf{x}_{lN}^{-1}\sqrt{Nh_N}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})-Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})\right)+\mathbf{x}_{lN}^{-1}\mathbf{w}_{lN}\sqrt{Nh_N}\left(\hat{\delta}(\tau)-\delta(\tau)\right)\\
&=\mathbf{x}_{lN}^{-1}\sqrt{\frac{h_N}{p_{lN}}}\sqrt{Np_{lN}}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})-Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})\right)+\mathbf{x}_{lN}^{-1}\mathbf{w}_{lN}\sqrt{Nh_N}\left(\hat{\delta}(\tau)-\delta(\tau)\right)\\
&\xrightarrow{d}\mathbf{x}_l^{-1}\mathbf{w}_l\mathbf{Z}_\delta(\tau) \qquad (80)
\end{aligned}$$

since $\frac{h_N}{p_{lN}} \to 0$ for strict mover realizations. Finally, for near-stayers ($l = L_1 + 1, \ldots, L$) with $D_{lN} = h_N$:

$$\begin{aligned}
&\sqrt{Nh_N^3}\left(\widehat{\beta}\left(\tau;\mathbf{x}_{lN}\right) - \beta\left(\tau;\mathbf{x}_{lN}\right)\right) \qquad (81)\\
&= \mathbf{x}_{lN}^{-1}\sqrt{\frac{h_N^3}{p_{lN}}}\sqrt{Np_{lN}}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN}) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})\right) + \mathbf{x}_{lN}^{-1}\mathbf{w}_{lN}h_N\sqrt{Nh_N}\left(\hat{\delta}(\tau) - \delta(\tau)\right)\\
&= \mathbf{x}_{lN}^{*}\frac{h_N}{D_{lN}}\sqrt{\frac{h_N}{p_{lN}}}\sqrt{Np_{lN}}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN}) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_{lN})\right) + \mathbf{w}_{lN}^{*}\frac{h_N}{D_{lN}}\sqrt{Nh_N}\left(\hat{\delta}(\tau) - \delta(\tau)\right)\\
&\xrightarrow{d} \frac{\mathbf{x}_l^{*}\mathbf{Z}_Q(\tau,\mathbf{x}_l)}{\sqrt{q_{l|0}2\phi_0}} + \mathbf{w}_l^{*}\mathbf{Z}_\delta(\tau) \qquad (82)
\end{aligned}$$

since $h_N/D_{lN} = 1$ and by Slutsky's Theorem. For near-stayers with $D_{lN} = -h_N$,

$$\sqrt{Nh_N^3}\left(\widehat{\beta}\left(\tau;\mathbf{x}_{lN}\right) - \beta\left(\tau;\mathbf{x}_{lN}\right)\right) \xrightarrow{d} -\frac{\mathbf{x}_l^{*}\mathbf{Z}_Q(\tau,\mathbf{x}_l)}{\sqrt{q_{l|0}2\phi_0}} - \mathbf{w}_l^{*}\mathbf{Z}_\delta(\tau). \qquad (83)$$

Since $\mathbf{Z}_Q(\cdot,\mathbf{x}_l)$ and $\mathbf{Z}_\delta(\cdot)$ are independent for $l = 1,\ldots,L$, we can add the individual covariances of these processes and get the desired result.

Proof of Theorem 6

We begin with the decomposition

$$\begin{aligned}
\sqrt{Nh_N}\left(\widehat{\bar{\beta}}_N(\tau) - \bar{\beta}_N(\tau)\right) &= \sqrt{Nh_N}\left(\widehat{\bar{\beta}}_N(\tau) - \mathbb{E}\left[\beta(\tau;\mathbf{X})|\mathbf{X}\in\mathbb{X}_N^M\right]\right) + \sqrt{Nh_N}\left(\mathbb{E}\left[\beta(\tau;\mathbf{X})|\mathbf{X}\in\mathbb{X}_N^M\right] - \mathbb{E}\left[\beta(\tau;\mathbf{X})\right]\right)\\
&= \sum_{l=1}^{L}\beta(\tau;\mathbf{x}_{lN})\sqrt{Nh_N}\left(\hat{q}_{lN}^M - q_{lN}^M\right) \qquad (84)\\
&+ \sum_{l=1}^{L}\sqrt{Nh_N}\left(\hat{\beta}(\tau;\mathbf{x}_{lN}) - \beta(\tau;\mathbf{x}_{lN})\right)\hat{q}_{lN}^M \qquad (85)\\
&+ \sqrt{Nh_N}\left(\mathbb{E}\left[\beta(\tau;\mathbf{X})|\mathbf{X}\in\mathbb{X}_N^M\right] - \mathbb{E}\left[\beta(\tau;\mathbf{X})\right]\right). \qquad (86)
\end{aligned}$$

We first consider the joint asymptotic distribution of $\sqrt{Nh_N}\left(\hat{q}_{lN}^M - q_{lN}^M\right)$ for all $l = 1,\ldots,L$. We start by considering the asymptotic distribution of unconditional probabilities, which are normalized differently for near-stayers and strict-movers:

$$\begin{pmatrix} \sqrt{N}I_{L_1} & \underline{0}_{L_1}\underline{0}'_{L-L_1+1} \\ \underline{0}_{L-L_1+1}\underline{0}'_{L_1} & \sqrt{\frac{N}{h_N}}I_{L-L_1+1} \end{pmatrix} \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(X_i = \mathbf{x}_{1N}) - p_{1N} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(X_i = \mathbf{x}_{L_1N}) - p_{L_1N} \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(X_i = \mathbf{x}_{(L_1+1)N}) - p_{(L_1+1)N} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(X_i = \mathbf{x}_{LN}) - p_{LN} \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\mathbf{X}_i \in \mathbb{X}_N^M) - \Pr(\mathbf{X}\in\mathbb{X}_N^M) \end{pmatrix}$$

$$
\begin{aligned}
&= \sqrt{N} \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(X_i = \mathbf{x}_{1N}) - p_{1N} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(X_i = \mathbf{x}_{L_1 N}) - p_{L_1 N} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbf{1}(X_i = \mathbf{x}_{(L_1+1)N}) - p_{(L_1+1)N}}{\sqrt{h_N}} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbf{1}(X_i = \mathbf{x}_{LN}) - p_{LN}}{\sqrt{h_N}} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbf{1}(\mathbf{X}_i \in \mathbb{X}_N^M) - \Pr(\mathbf{X} \in \mathbb{X}_N^M)}{\sqrt{h_N}} \end{pmatrix} \\
&\xrightarrow{d} \mathcal{N}\left(\underline{0}_{L+1}, \begin{pmatrix} p_1(1-p_1) & \cdots & -p_1 p_{L_1} & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -p_1 p_{L_1} & \cdots & p_{L_1}(1-p_{L_1}) & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & q_{L_1+1|0} 2\phi_0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & q_{L|0} 2\phi_0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 2\phi_0 \end{pmatrix} \right)
\end{aligned} \tag{87}
$$

by the Lyapunov's Multivariate Central Limit Theorem. The conditions of the theorem are trivially satisfied since indicator functions have bounded moments. Note that the top left $L_1 \times L_1$ submatrix is singular since it contains probabilities that sum to 1. This is not a problem here since we do not invert that matrix later on. By the Delta method, we get that $\hat{q}_{lN}^M - q_{lN}^M = O_p\left(\frac{1}{\sqrt{N}}\right)$ for any $l = 1, \ldots, L_1$ and $O_p\left(\sqrt{\frac{h_N}{N}}\right)$ for $l = L_1 + 1, \ldots, L$. Therefore, the term $\sum_{l=1}^{L} \beta(\tau; \mathbf{x}_{lN}) \sqrt{N h_N} \left(\hat{q}_{lN}^M - q_{lN}^M\right)$ is of order $O_p(\sqrt{h_N})$ and therefore converges to 0.

Term 85 can be decomposed into its strict-movers and near-stayers components:

$$
\sum_{l=1}^{L} \sqrt{N h_N} \left(\hat{\beta}(\tau; \mathbf{x}_{lN}) - \beta(\tau; \mathbf{x}_{lN})\right) \hat{q}_{lN}^M = \sum_{l=1}^{L_1} \sqrt{N h_N} \left(\hat{\beta}(\tau; \mathbf{x}_{lN}) - \beta(\tau; \mathbf{x}_{lN})\right) \hat{q}_{lN}^M \tag{88}
$$

$$
+ \sum_{l=L_1+1}^{L} \sqrt{N h_N^3} \left(\hat{\beta}(\tau; \mathbf{x}_{lN}) - \beta(\tau; \mathbf{x}_{lN})\right) \frac{\hat{q}_{lN}^M}{h_N}. \tag{89}
$$

For term (88), we can see that $\sum_{l=1}^{L_1} \sqrt{N h_N} \left(\hat{\beta}(\tau; \mathbf{x}_{lN}) - \beta(\tau; \mathbf{x}_{lN})\right) \hat{q}_{lN}^M \xrightarrow{d} \sum_{l=1}^{L_1} \mathbf{x}_l^{-1} \mathbf{w}_l \mathbf{Z}_\delta(\tau) p_l$. This asymptotic distribution is due to the presence of $\hat{\delta}(\tau)$. Term (89) will have a non-degenerate limiting distribution:

$$
\sum_{l=L_1+1}^{L} \sqrt{N h_N^3} \left(\hat{\beta}(\tau; \mathbf{x}_{lN}) - \beta(\tau; \mathbf{x}_{lN})\right) \frac{\hat{q}_{lN}^M}{h_N} \xrightarrow{d} \sum_{l=L_1+1}^{L} \mathbf{Z}(\tau, \mathbf{x}_l) q_{l|0} 2\phi_0 \tag{90}
$$

since $\frac{\hat{q}_{lN}^M}{h} \xrightarrow{p} 2\phi_0 q_{l|0}$. The asymptotic covariance of (85) will then be

$$
\begin{aligned}
&= \left(\min(\tau, \tau') - \tau\tau'\right) \sum_{l=L_1+1}^{L} \sum_{l'=L_1+1}^{L} \mathbf{x}_l^* \Lambda(\tau, \tau'; \mathbf{x}_l) \mathbf{x}_l^{*\prime} \cdot \mathbf{1}(l = l') q_{l|0} 2\phi_0 \\
&+ \left(\sum_{l=L_1+1}^{L} \mathbf{w}_l^* q_{l|0} 2\phi_0 + \sum_{l=1}^{L_1} \mathbf{x}_l^{-1} \mathbf{w}_l p_l \right) \Sigma_\delta(\tau, \tau') \left(\sum_{l'=L_1+1}^{L} \mathbf{w}_{l'}^* q_{l'|0} 2\phi_0 + \sum_{l'=1}^{L_1} \mathbf{x}_{l'}^{-1} \mathbf{w}_{l'} p_{l'} \right).
\end{aligned}
$$

We now show that $\Xi_0 = \lim_{N\to\infty} \mathbb{E}_N\left[\mathbf{X}^{-1}\mathbf{W}||D| \geq h_N\right] = \sum_{l=L_1+1}^{L} \mathbf{w}_l^* q_{l|0} 2\phi_0 + \sum_{l=1}^{L_1} \mathbf{x}_l^{-1}\mathbf{w}_l p_l$:

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{E}_N\left[\mathbf{X}^{-1}\mathbf{W}||D| \geq h_N\right] &= \lim_{N\to\infty} \sum_{l=1}^{L} \mathbf{x}_{lN}^{-1}\mathbf{w}_{lN} q_{lN}^M \\
&= \lim_{N\to\infty} \sum_{l=L_1+1}^{L} \mathbf{w}_{lN}^* \frac{\Pr(\mathbf{X} = \mathbf{x}_{lN}|D = |h_N|)}{1 - 2\phi_0 h_N} 2\phi_0 + \lim_{N\to\infty} \sum_{l=1}^{L_1} \mathbf{x}_{lN}^{-1}\mathbf{w}_{lN} \frac{p_{lN}}{1 - 2bh_N} \\
&= \sum_{l=L_1+1}^{L} \mathbf{w}_l^* q_{l|0} 2\phi_0 + \sum_{l=1}^{L_1} \mathbf{x}_l^{-1}\mathbf{w}_l p_l
\end{aligned}$$

by continuity in N. Finally, we consider the bias term (86). We see that

$$\begin{aligned}
\sqrt{Nh_N}\left(\mathbb{E}_N\left[\beta(\tau;\mathbf{X})|\mathbf{X} \in \mathbb{X}_N^M\right] - \mathbb{E}_N\left[\beta(\tau;\mathbf{X})\right]\right) &= \sqrt{Nh_N}\left(\sum_{l=1}^{L} \beta(\tau;\mathbf{x}_{lN}) q_{lN}^M - \sum_{l=1}^{M} \beta(\tau;\mathbf{x}_{lN}) p_{lN}\right) \\
&= \sum_{l=1}^{L} \beta(\tau;\mathbf{x}_{lN})\sqrt{Nh_N}\left(q_{lN}^M - p_{lN}\right) - \sum_{l=L+1}^{M} \beta(\tau;\mathbf{x}_{lN})\sqrt{Nh_N} p_{lN}.
\end{aligned}$$

We see that for $l = 1, \ldots, L$, $\sqrt{Nh}\left(q_{lN}^M - p_{lN}\right) = p_{lN}\frac{2\phi_0\sqrt{Nh_N^3}}{1-2\phi_0 h_N} \to 0$ since $p_{lN} = O(1)$ and $Nh_N^3 \to 0$. For stayer realizations $l = L+1, \ldots, M$, we have $\sqrt{Nh_N} p_{lN} \to 0$ since $\sum_{l=L+1}^{M} p_{lN} = 2\phi_0 h_N$ and $Nh_N^3 \to 0$. Therefore, term (86) converges to 0.

Proof of Theorem 7

We start by deriving the asymptotic distribution of the empirical cumulative distribution function of $\widehat{\beta}_p(U;\mathbf{X})$ with U distributed uniformly on $[0,1]$ independently of $\mathbf{X}$, while conditioning on $\mathbf{X} \in \mathbb{X}_N^M$. The CDF estimand at $c \in \mathbb{R}$ is denoted as $F_{\mathbf{B}_p}(c)$ and the estimator is

$$\begin{aligned}
\widehat{F}_{\widehat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}_N^M}(c) &= \frac{\frac{1}{N}\sum_{i=1}^{N}\int_0^1 \mathbf{1}(\widehat{\beta}_p(u,\mathbf{X}_i) \leq c)\mathrm{d}u\mathbf{1}(\mathbf{X}_i \in \mathbb{X}_N^M)}{\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\mathbf{X}_i \in \mathbb{X}_N^M)} \\
&= \sum_{l=1}^{L}\left(\int_0^1 \mathbf{1}(\widehat{\beta}_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u\right)\widehat{q}_{lN}^M.
\end{aligned}$$

The integration over $u \in (0,1)$ can be done exactly since $\widehat{\beta}_p(u,\mathbf{x}_{lN})$ is piecewise linear for each $l \in \{1,\ldots,L\}$ with finitely many pieces. This asymptotic distribution can be written as the sum of four terms:

$$\widehat{F}_{\widehat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}_N^M}(c) - F_{B_p|\mathbf{X}\in\mathbb{X}_N}(c) = \sum_{l=1}^{L_1}\left(\int_0^1 \mathbf{1}(\widehat{\beta}_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u\right)\widehat{q}_{lN}^M \tag{91}$$

$$+ \sum_{l=L_1+1}^{L}\left(\int_0^1 \mathbf{1}(\widehat{\beta}_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u\right)\widehat{q}_{lN}^M \tag{92}$$

$$+ \sum_{l=1}^{L}\int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_l) \leq c)\mathrm{d}u\left(\widehat{q}_{lN}^M - q_{lN}^M\right) \tag{93}$$

$$+ F_{B_p|\mathbf{X}\in\mathbb{X}_N^M}(c) - F_{B_p}(c). \tag{94}$$

The first two terms represent the contribution of the estimation of the conditional coefficients for strict-movers and near-stayers respectively, while term three is due to the randomness of the coefficient across subpopulations defined in terms of $\mathbf{X}$. Term four is the bias term, which we will show vanishes asymptotically. Term (93) will be of order $O_p\left(\frac{1}{\sqrt{N}}\right)$ and therefore vanishes when premultiplied by $\sqrt{Nh_N}$.

Using Hadamard differentiability and the functional delta method discussed more fully in the proof of Theorem 4, we see that:

$$\sqrt{Nh_N^3}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_{lN}) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_{lN}) \leq c)\mathrm{d}u\right) \xrightarrow{d} \mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l) f_{B_p|\mathbf{X}}(c|\mathbf{x}_l). \tag{95}$$

for $l = L_1 + 1, \ldots, L$.

This convergence is uniform in $c \in \mathbb{R}$ since $F_{B_p|\mathbf{X}}(c|\mathbf{x}_l)$ ranges between 0 and 1, and also uniform in $\mathbf{x}_l$ since there are finitely many possible values for $\mathbf{x}_l$. Therefore, term (92) will converge in process to the following limit when normalized by $\sqrt{Nh_N}$:

$$\sum_{l=L_1+1}^{L} \sqrt{Nh_N^3}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_{lN}) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_{lN}) \leq c)\mathrm{d}u\right)\frac{\hat{q}_{lN}^M}{h_N}$$
$$\xrightarrow{d} \sum_{l=L_1+1}^{L} \mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l) f_{B_p|\mathbf{X}}(c|\mathbf{x}_l) q_{l|0} 2\phi_0$$

for $c \in \mathbb{R}$ since $\frac{\hat{q}_{lN}^M}{h_N} \xrightarrow{p} q_{l|0}2\phi_0$ as $N \to \infty$. Term (91) will converge in distribution to:

$$\sum_{l=1}^{L_1} \sqrt{Nh_N}\left(\int_0^1 \mathbf{1}(\widehat{\beta}_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u - \int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u\right)\widehat{q}_{lN}^M \xrightarrow{d} \sum_{l=1}^{L_1} \mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l) f_{B_p|\mathbf{X}}(c|\mathbf{x}_l) p_l.$$

We also verify that $\sqrt{Nh_N}$ times term (94), the bias term, asymptotically vanishes. We see that

$$\begin{aligned}
\left|F_{B_p|\mathbf{X}\in\mathbb{X}_N^M}(c) - F_{B_p}(c)\right| &= \left|\sum_{l=1}^{L}\int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u q_{lN}^M - \sum_{l=1}^{M}\int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u p_{lN}\right| \\
&\leq \sum_{l=1}^{L}\int_0^1 \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u \left|q_{lN}^M - p_{lN}\right| + \left|\sum_{l=L+1}^{M} \mathbf{1}(\beta_p(u,\mathbf{x}_{lN}) \leq c)\mathrm{d}u p_{lN}\right| \\
&\leq \sum_{l=1}^{L} p_{lN}\frac{2\phi_0 h_N}{1-2\phi_0 h_N} + \sum_{l=L+1}^{M} p_{lN} \\
&= 2\phi_0 h_N + 2\phi_0 h_N = o_p\left(\frac{1}{\sqrt{Nh_N}}\right)
\end{aligned}$$

since $Nh_N^3 \to 0$. We can conclude this proof in a manner similar to that of Theorem 4.