Minimum Distance Estimation of Quantile Panel Data Models[*]

Blaise Melly[†]and Martina Pons[‡]

First version: November 2020
This version: February 24, 2023
Preliminary version. Link to the newest version

Abstract

We propose a minimum distance estimation approach to quantile panel data models where the unit effects may be correlated with the covariates. The estimation method is computationally straightforward to implement and fast. We first compute a quantile regression within each individual and then apply GMM to the fitted values from the first stage. The suggested estimators apply (i) to grouped data, where we observe data at the individual level, but the treatment varies at the group level, and (ii) to classical panel data, where we follow the same units over time. Depending on the variables assumed to be exogenous, this approach provides quantile analogs of the classical least squares panel data estimators such as the fixed effects, random effects, between, and Hausman-Taylor estimators. For grouped (instrumental) quantile regression, we provide a more precise estimator than the existing estimators. We establish the asymptotic properties of our estimators when both the number of units and observations per unit jointly diverge to infinity. We suggest an inference procedure that automatically adapts to the (potentially) unknown rate of convergence of the estimators. Monte Carlo simulations show that our estimator and inference procedure also perform well in finite samples when the number of observations per unit is small. In an empirical application, we find that the introduction of the food stamp program increased the birth weights only at the bottom of the distribution.

1 Introduction

Quantile regression, as introduced by Koenker and Bassett (1978), is the method of choice when we are interested in the effect of a policy on the distribution of an outcome. The quantile treatment effect function provides more information than the average treatment effect; for instance, it allows evaluating the impact of the treatment on inequality. When panel data are available, new identification and estimation strategies become feasible. The researchers can alleviate endogeneity concerns, for instance, by allowing for correlated group effects. They can obtain more precise estimates, for example, by using a random-effects estimator; or they

[*] We are grateful to Bo Honoré, Aleksey Tetenov and seminar participants at Bern University (4th of November 2020), the 2021 scientific meeting of the Italian statistical society (24th of June 2021), Geneva University (3rd of March 2022), the 27th International Panel Data Conference, and the COMPIE 2022 Conference, for useful comments.

[†] Department of Economics, University of Bern, Schanzeneckstrasse 1, 3001 Bern, Switzerland. blaise.melly@unibe.ch.

[‡] Department of Economics, University of Bern, Schanzeneckstrasse 1, 3001 Bern, Switzerland. martina.pons@unibe.ch.

can exploit individual-level variables to identify the impact of group-level variables, e.g., with the Hausman and Taylor (1981) estimator. In this paper, we propose a minimum distance estimation approach to quantile panel data models, which provides quantile analogs of the classical least-squares panel data estimators such as the fixed effects, random effects, between, and Hausman-Taylor estimators.

We use a general notation (i and j subscripts) and terminology (individuals and groups) that is more common to group data, where individual-level data is available, and the treatment varies at the group level. For instance, in Autor et al. (2013), the groups are commuting zones in the United States while they are schools in Angrist and Lang (2004). In both cases, the treatment varies only between groups, but individual data are needed to estimate the conditional distribution of the outcome within each group. Our results also apply to classical panel data, where the data follow the same individual over time. In this part of the literature, the j units are the individuals, and the i units are the time periods. We discuss the application of our results to this framework in Section 4.

In all cases, we perform the estimation in two stages. The first stage consists of group-level quantile regressions using individual-level covariates at each quantile of interest. In the second stage, the first stage fitted values are regressed on individual-level and group-level variables. If these variables are potentially endogenous, an instrumental variable regression or, more generally, the generalized method of moment (GMM) estimator can be used. Thus, including external or internal instruments in the second stage is straightforward. This estimator is simple to implement, flexible, computationally fast, and can be used in various applied fields. While this two-step procedure may sound unusual, Section 2.3 shows that it is numerically identical to the standard estimators if we use least squares in the first stage and the appropriate instruments.

As a nonlinear estimator, first-stage quantile regression is subject to a bias that decreases as the number of observations per group increases. Inference is justified in an asymptotic framework where both the number of observations per group n, and the number of groups m diverge to infinity.[1] Recently, Galvao et al. (2020) have weakened the requirements on the relative rate of divergence of m and n for asymptotic normality of fixed effects quantile estimators. Using their results, we show that our estimator is asymptotically normal under the condition that $m(\log n)^2/n \to 0$. Under this condition and other assumptions, we show that our estimators are asymptotically normally distributed and centered at zero. The requirement on the growth rate of n relative to m can be weakened if only the coefficient vector on group-level regressors is of interest. In this case, the milder condition that $\sqrt{m}(\log n)/n \to 0$ is sufficient for an unbiased asymptotic distribution.

The asymptotic distribution of the estimator is non-standard because the speed of convergence is not the same for all coefficients. The speed of convergence depends on the moment

[1] Large n asymptotic has been widely used in the quantile and nonlinear traditional panel data literature as well as in the nonstationary and dynamic panel data literature. For seminal contributions, see Phillips and Moon (1999), Hahn and Kuersteiner (2002), and Alvarez and Arellano (2003).

conditions that are used to identify a parameter. More precisely, it depends on the interaction between the instrument and the unobserved heterogeneity. We distinguish three cases. In the first case, the moments exploit variation between groups, and there is group-level heterogeneity. The coefficients of the variables that are identified by these moments converge at the $\sqrt{m}$ rate (the 'slow' coefficients). In the second case, the moments only exploit within variation. This can happen if there is no group level heterogeneity or if the instrument is set to exploit only the variation within groups as in a within regression. The coefficient of the variables identified by these moments converges at the faster $\sqrt{mn}$ rate (the 'fast' coefficients). The third case is an intermediate case that could occur if, for example, the level of group-level heterogeneity is small. In this special case, the rate of convergence is unknown and lies between $\sqrt{m}$ and $\sqrt{mn}$. The differential rate of convergence has several consequences for the first-order asymptotic distribution: (i) The fast and slow coefficients are first-order asymptotically independent. (ii) If a coefficient is identified by both within and between variations, then the between variation is first-order asymptotically useless. For instance, the random-effect estimator is asymptotically equivalent to the fixed-effects estimator.[2] (iii) The first-order asymptotic distribution of the estimator of the slow coefficients is not affected by the first-stage estimation error. Using this better approximation, we solve these three issues. This allows us to suggest a new quantile random effects estimator that is more precise than the fixed effects estimator in finite samples. We also improve the quality of the estimated standard errors by taking the first-stage estimation error into account. Quite surprisingly, we find that clustering the standard errors in the second stage automatically takes into account the first-stage error and provides an adaptive inference procedure in the sense that it is uniformly valid in the rate of convergence of the moment conditions and the estimator, including the intermediate case. Further, our inference does not require estimating the density like in traditional quantile models.

This paper contributes to the literature on quantile panel data and IV models. A large share of the literature focused on fixed effects models (see, for example, Canay, 2011; Galvao and Kato, 2016; Gu and Volgushev, 2019). Koenker (2004) introduced a penalized quantile fixed effects estimator that treated the individual heterogeneity as a pure location shift. Kato et al. (2012) allow the group effects to depend on the quantile of interest and contribute to the asymptotic theory of the estimator. Galvao and Wang (2015) suggest a two-step minimum distance (MD) estimator as a computationally fast way to estimate fixed effects quantile panel data model. Galvao and Poirier (2019) suggest using quantile regression as an estimator in the presence of random effects. Our random effects estimator is different because we focus on the conditional quantile function that also conditions on the group effect (see Remark 1 for a discussion on conditional effects). In other words, we estimate a different parameter, and quantile regression is not consistent for this parameter even if the random effects are not correlated with the covariates.

[2] See Ahn and Moon (2014) for similar results for least-squares estimators.

Our class of estimators nests the MD estimators of Chamberlain (1994) and Galvao and Wang (2015) as special cases. We generalize the results in Chamberlain (1994) by including individual-level regressors and allowing the number of groups to go to infinity.[3] Galvao and Wang (2015) are only interested in the effect of individual-level covariates and do not exploit variation between individuals.[4] In contrast, we aim to estimate the effect of both individual-level and group-level regressors. Furthermore, we allow for both internal and external instruments.

Chetverikov et al. (2016) consider a quantile extension of the Hausman-Taylor model. They focus on the effect of variables that vary only between groups and allow for instrumental variables for identification. The main difference compared to our setting is that, in the second stage, we regress the fitted values on all variables while they regress the estimated intercept on the group-level regressors. Since they use only the intercept in the second stage, their estimator is not invariant to reparametrizations of the individual-level regressors. By keeping all the variables in the second stage, we can easily impose equality of the coefficients on the individual-level regressors, which increases precision at a minimal cost from a computational perspective and make the estimator invariant to reparametrization. Clearly, our estimator accommodates modeling heterogeneous treatment effects by including interactions. Simulations using the same data generating process as Chetverikov et al. (2016) show that our MD estimator has substantially lower variance and MSE across all sample sizes considered. From a technical point of view, we are able to weaken the growth rate of the number of observations per group n relative to the number of groups m necessary to obtain unbiased asymptotic normality of the estimator of the 'slow' coefficient. We also contribute to this literature by deriving the limiting distribution of the estimator of the 'fast' coefficients, which were not studied by Chetverikov et al. (2016).

As an empirical application, we build on the work of Almond et al. (2011) and estimate the distributional effect of the food stamp program on birth weight. Following the Food Stamp Act, the number of counties that implemented a food stamp program increased substantially in the late 1960s and in the beginning of the 1970s. To apply our minimum distance estimator, we define groups as county-trimester cells. Thus, the subscript j indexes a county-trimester cell, while the subscript i defines an individual within this cell. We estimate the model separately for black and white mothers, and we find that the food stamp program has a positive impact on the lower tail of the birth weight distribution, mostly among blacks.

The remainder of the paper is structured as follows. Section 2 presents the model and the estimator and briefly discusses equivalent methods to estimate average effects with panel data models to motivate our two-step approach. Section 3 presents the asymptotic theory. Section 4 focuses more in detail on the estimation of traditional quantile panel data models, and we

[3] Chamberlain (1994) uses a different terminology because he considers cross-sectional regressions. He analyses a quantile regression model with a finite number of combinations of values of the regressors. The number of cells is thus finite, and the regressors are constant within each cell.

[4] Galvao and Wang (2015) consider a traditional panel data setting, thus, using their terminology, they focus on estimating the effect of time-varying regressors

present a generalization of the Hausman test for the random effects assumption. Section 5 discusses the grouped quantile regression model and compares our estimator to the grouped IV quantile regression of Chetverikov et al. (2016). Monte Carlo simulations to analyze the finite sample performance are included in Sections 4 and 5. In Section 6, in an empirical application, we study the effect of the food stamp program on the distribution of birth weight. Section 7 concludes.

2 Model and Minimum Distance Estimator

2.1 Quantile Model

We want to learn the effects of the individual-level variables x_{1ij} and the group-level variables x_{2j} on the distribution of an outcome y_{ij}. We observe these variables for the groups $j = 1, \ldots, m$ and individuals $i = 1, \ldots, n$.[5] For some quantile index $0 < \tau < 1$, we assume that

$$Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j) = x'_{1ij}\beta(\tau) + x'_{2j}\gamma(\tau) + \alpha(\tau, v_j), \tag{1}$$

where $Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j)$ is the τth conditional quantile function of the response variable y_{ij} for individual i belonging to group j given the K_1-vector of individual-level regressors x_{1ij}, the K_2-vector of time invariant variables x_{2j}, and an unobserved random vector v_j of unrestricted and unknown dimension. In total, there are $K_1 + K_2 = K$ parameters to estimate. The parameters $\beta(\tau)$, $\gamma(\tau)$ and the unobserved group heterogeneity $\alpha(\tau, v_j)$ can depend on the quantile index τ. Depending on the setting, $\beta(\tau)$ or $\gamma(\tau)$ (or both) might be the parameters of interest. We normalize $\mathbb{E}[\alpha(\tau, v_j)] = 0$, which is not restrictive because x_{2j} includes a constant.

Remark 1 (Conditional versus unconditional effects). In contrast to the average effect, the definition of a quantile treatment effect depends on the conditioning variables. In this paper, we model the distribution of y_{ij} conditionally on the covariates and on the group effect $\alpha(\tau, v_j)$. Thus, even if the group effects are independent of the regressors, we identify different coefficients than those identified by quantile regression as introduced by Koenker and Bassett (1978) or by instrumental variable quantile regression as introduced by Chernozhukov and Hansen (2005). The following example illustrates the difference between these parameters. Consider an application where each group j corresponds to a region and each unit i to an individual within this region. We do not have any x_{1ij} variable. We are interested in the effect of a binary treatment x_{2j}, which has been randomized and is, therefore, independent from $\alpha(\tau, v_j)$. $\gamma(\tau)$ is the effect of this treatment for individuals that rank at the τ quantile of y_{ij} in *their* region. On the other hand, the quantile regression of y_{ij} on x_{2j} identifies the effect for individuals that rank at the τ quantile in the whole country (given the treatment status). These are different parameters except if $\alpha(\tau, v_j)$ is constant over j or if the treatment effect is homogeneous such that $\gamma(\tau) = \gamma$. Whether conditional or unconditional quantile treatment effects are of interest depends on the

[5] For notational simplicity, we assume a balanced panel. However, the results generalize to unbalanced datasets.

question at hand. For example, conditional quantile treatment effects are particularly of interest to study within-group inequalities when groups might be regions or industries. For example, Autor et al. (2016), and Engbom and Moser (2022) study the effect of the minimum wage on within-state inequality, while Autor et al. (2021) study the effect of trade shock on wage inequality within local labor markets. If the unconditional effect is of interest, one can naturally obtain the unconditional distribution functions by integrating out the group effects (and possibly the other variables) and then inverting the resulting distribution functions to obtain the unconditional quantile functions, see Chernozhukov et al. 2013.[6]

When model (1) holds, the τ quantile regression of y_{ij} on x_{1ij} and a constant using only observations for group j identifies the slope $\beta(\tau)$ and the intercept $x_{2j}'\gamma(\tau)+\alpha(\tau,v_j)$. To identify the coefficient on the group-level variables, we need to consider variation across groups. Note that model (1) implies

$$\mathbb{E}\left[Q(\tau,y_{ij}|x_{1ij},x_{2j},v_j)|x_{1ij},x_{2j}\right]=x_{1ij}'\beta(\tau)+x_{2j}'\gamma(\tau)+\mathbb{E}\left[\alpha(\tau,v_j)|x_{1ij},x_{2j}\right].$$

If $\alpha(\tau,v_j)$ is exogenous with respect to x_{1ij} and x_{2j} and the linear model is correctly specified, $\mathbb{E}[\alpha(\tau,v_j)|x_{1ij},x_{2j}]=0$ and this linear regression identifies the parameters of interest.[7] The last representation suggests a two-step estimation strategy: (i) group-level quantile regression of y_{ij} on x_{1ij}, (ii) OLS regression of the fitted values from the first stage on x_{1ij} and x_{2j}.

When the group effects $\alpha(\tau,v_j)$ are endogenous (possibly correlated with x_{1ij} and x_{2j}), we assume that there is a L-dimensional vector ($L\geq K$) of valid instruments z_{ij} satisfying

$$\mathbb{E}[z_{ij}\alpha(\tau,v_j)]=\mathbb{E}\left[z_{ij}\left(Q(\tau,y_{ij}|x_{1ij},x_{2j},v_j)-x_{1ij}'\beta(\tau)-x_{2j}'\gamma(\tau)\right)\right]=0. \tag{2}$$

Note that $\beta(\tau)$ is identified in model (1) as long as there is some variation in x_{1ij} within some groups. For instance, we can include the demeaned regressors, $\dot{x}_{1ij}=x_{1ij}-\bar{x}_{1j}$ with $\bar{x}_{1j}=n^{-1}\sum_{i=1}^{n}x_{1ij}$, in the vector of instruments z_{ij} because this variable will satisfy condition (2) under strict exogeneity.[8] On the other hand, we need additional instruments to identify $\gamma(\tau)$. Equation (2) suggests a similar estimation strategy as in the exogenous case but with the instrumental variable estimator (or more generally the GMM estimator) in the second stage: (i) group-level quantile regression of y_{ij} on x_{1ij}, (ii) GMM regression of the fitted values from the first stage on x_{1ij} and x_{2j} using z_{ij} as instrument.

Remark 2 (Skorohod representation). The following Skorohod representation implies the model defined in equation (1):

$$\begin{aligned}y_{ij}&=x_{1ij}\beta(u_{ij})+x_{2j}\gamma(u_{ij})+\alpha(u_{ij},v_j)\\&=q(x_{1ij},x_{2j},u_{ij},v_j),\end{aligned}$$

[6] We refer to Frölich and Melly (2013) for a discussion about conditional and unconditional treatment effects.

[7] Uncorrelation between $\alpha(\tau,v_j)$ and x_{1ij} and x_{2j} is sufficient to identify the linear projection.

[8] In the special case of traditional panel data, the demeaned regressors correspond to the within transformation.

where $q(x_{1ij}, x_{2j}, u_{ij}, v_j)$ is strictly increasing in u (while fixing the other arguments). We normalize $u_{ij}|x_{1ij}, x_{2j}, v_j \sim U(0,1)$ such that $q(x_{1ij}, x_{2j}, u, v_j)$ is the conditional quantile function. v_j ranks the groups while u_{ij} ranks the individuals within a group. In this model, a sufficient condition for equation (2) is $(u_{ij}, v_j) \perp\!\!\!\perp z_{ij}$. If the instrument does not vary within groups, only $v_j \perp\!\!\!\perp z_j$ is sufficient.

Remark 3 (Heterogeneous coefficients). Our model allows only the intercept to differ between groups.[9] Now consider a more general model where we also allow the slopes to differ between groups:

$$y_{ij} = x'_{1ij}\beta(u_{ij}, v_j) + x'_{2j}\gamma(u_{ij}, v_j) + \alpha(u_{ij}, v_j). \tag{3}$$

If we maintain the conditional strict monotonicity assumption with respect to u_{ij}, this model implies that

$$Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j) = x'_{1ij}\beta(\tau, v_j) + x'_{2j}\gamma(\tau, v_j) + \alpha(\tau, v_j). \tag{4}$$

In the exogenous case where $(x_{1ij}, x_{2j}) \perp\!\!\!\perp v_j$, this implies

$$\begin{aligned}\mathbb{E}\left[Q\left(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j\right)|x_{1ij}, x_{2j}\right] &= x'_{1ij}\int \beta(\tau, v)dF_V(v) + x'_{2j}\int \gamma(\tau, v)dF_V(v) + \int \alpha(\tau, v)dF_V(v)\\ &= x'_{1ij}\bar{\beta}(\tau) + x'_{2j}\bar{\gamma}(\tau)\end{aligned}$$

because we have normalized $\mathbb{E}[\alpha(\tau, v_j)] = 0$. It follows that the linear projection of $Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j)$ on x_{1ij} and x_{2j} identifies the average effects when these effects are heterogeneous. Thus, the linear projection identifies the coefficients $\beta(\tau)$ and $\gamma(\tau)$ when the homogenous model (1) holds and the average effect for all groups at the τ quantile of their conditional distribution when the heterogenous model (4) holds.[10] Naturally, it is also possible to modelize the heterogeneity between groups by estimating more flexible linear projections of $Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j)$. For instance, we can interact x_{1ij} with indicator variables for the groups, which allows for unrestricted heterogeneity of $\beta(\tau, v_j)$, or by interacting x_{2j} with observable characteristics.

Remark 4 (Least-squares versus quantile regression in the second-stage). As discussed in Remark 3, the projection identifies the average coefficients across groups when those are heterogeneous. It is possible to analyze the inter-group heterogeneity if we impose model (3), restrict v_j to be a scalar, and impose the strict monotonicity of $x'_{1ij}\beta(u, v) + x'_{2j}\gamma(u, v) + \alpha(u, v)$ with respect to v.[11] When we normalize $v_j|x_{1ij}, x_{2j} \sim U(0,1)$ and $\alpha(\tau, \theta) = 0$, we obtain in the exogeneous case, for any $0 < \theta < 1$,

$$Q(\theta, Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j)|x_{1ij}, x_{2j}) = x'_{1ij}\beta(\tau, \theta) + x'_{2j}\gamma(\tau, \theta).$$

[9] This is the same model as in Chetverikov et al. (2016), where a similar Skorohod representation is derived in their footnote 6.

[10] In the endogenous case, we obtain the instrumental variable projection instead of the standard linear projection. For instance, if x_{2j} is an endogenous binary variable and z_{ij} is a binary instrument, we identify the average treatment effects for the complying individuals at the τ quantile of their conditional distribution.

[11] In the presence of multivariate heterogeneity, quantile regression identifies local average structural derivatives of nonseparable models, see Hoderlein and Mammen (2007).

All coefficients have two quantile indices: one for the heterogeneity across groups and one for the heterogeneity within groups.[12] These heterogeneous coefficients are identified by a two-step quantile regression: (i) τ quantile regression of y_{ij} on x_{1ij}, (ii) θ quantile regression of the fitted values from the first-stage on x_{1ij} and x_{2j}. This alternative strategy identifies different parameters and is outside the scope of this paper. We focus instead on the model defined by equations (1) and (2), which is the same as in Chetverikov et al. (2016) and nests the fixed effects quantile regression model (studied e.g. in Galvao et al., 2020).

2.2 Quantile Minimum Distance Estimators

Motivated by the representation in equation (2), we suggest a quantile version of the two-steps procedure. In the first step, for each group j and quantile τ, we regress y_{ij} on individual-level variables x_{1ij} and a constant using quantile regression. The intercept of the first stage regression captures both the group effect $\alpha(\tau, v_j)$ and the term $x_{2j}'\gamma(\tau)$ as these vary only between groups. In a second step, we regress the fitted values of the first stage on x_{1ij} and x_{2j}, using GMM with instruments z_{ij}.

Formally, the first stage quantile regression solves the following minimization problem for each group and quantile separately:

$$\hat{\beta}_j(\tau) \equiv \left(\hat{\beta}_{0,j}, \hat{\beta}_{1,j}'\right)' = \underset{(b_0,b_1)\in\mathbb{R}^{K_1+1}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{ij} - b_0 - x_{1ij}'b_1), \tag{5}$$

where $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function. The true vector of coefficients for group j is given by $\beta_j(\tau) = (\alpha(\tau, v_j) + x_{2j}'\gamma(\tau), \beta(\tau)')'$. In the special case that the models do not contain any x_{1ij} variables quantile regression estimates the percentile in each group.

Notation. Throughout the paper, we will use the following notation. Let $\tilde{x}_{ij} = (1, x_{1ij}')'$ and $x_{ij} = (x_{1ij}', x_{2j}')'$. For each group j we define the following matrices. The $n \times K_1$ matrix of group-level regressors $X_{1j} = (x_{1i1}, x_{1i2}, \ldots, x_{1ij})'$, the $n \times K$ matrix containing all regressors $X_j = (x_{i1}, x_{i2} \ldots, x_{1ij})'$ and the $n \times L$ matrix of instruments $Z_j = (z_{i1}, z_{i2}, \ldots, z_{ij})'$. Further, we define two matrices for all observations. The $mn \times K$ matrix of regressors for all groups $X = (X_1', \ldots, X_m')'$ and the $mn \times L$ matrix of instruments for all groups as $Z = (Z_1', \ldots, Z_m')'$. We let Y be the $mn \times 1$ vector of the response variable. The fitted value for individual i in group j at quantile τ is $\hat{y}_{ij}(\tau) = \hat{\beta}_{0,j}(\tau) + x_{1ij}'\hat{\beta}_{1,j}(\tau)$. Denote the $n \times 1$ column vector of fitted values for group j by $\hat{Y}_j(\tau) = (\hat{y}_{i1}(\tau), \ldots, y_{i,T}(\tau))'$, and the $mn \times 1$ vector of fitted values by $\hat{Y}(\tau) = (\hat{y}_1'(\tau), \ldots, \hat{y}_m'(\tau))'$.

Remark 5 (Alternative first-stage estimators). The Koenker and Bassett (1978) quantile regression estimator is not necessarily efficient. Newey and Powell (1990) suggest a semiparametrically efficient weighted estimator of $\beta_j(\tau)$. We prefer to use the unweighted quantile regression

[12] This is similar to the instrumental variable model in Chesher (2003) and Ma and Koenker (2006), which also contains two quantile indices: one for the selection equation and one for the outcome equation.

estimator due to the difficulty of estimating the weights and the complex interpretation of the estimates in case of misspecification. In our model (1), the variation within group to be exogenous. If this was not the case, it would be possible to use an instrumental variable quantile regression (see e.g., Chernozhukov and Hansen, 2006) in the first stage, followed by the second stage GMM regression described below.[13] We do not explore this (computationally expensive) extension in this paper.

The second stage consists in a GMM regression using $\mathbb{E}[g_j(\delta, \tau)] = 0$ as a moment condition, where $g_j(\delta, \tau) = Z_j'(\tilde{X}_j \hat{\beta}_j(\tau) - X_j \delta(\tau))$ and $\delta(\tau) = (\beta(\tau)', \gamma(\tau)')'$ is the K-dimensional vector of coefficients. This moment restriction depends on the first stage and is the sample counterpart of $\mathbb{E}[Z_j' \alpha_j(\tau, v_j)]$. The closed-form expression of the second stage estimator is

$$\hat{\delta}(\hat{W}, \tau) = \left(X'Z\hat{W}(\tau)Z'X \right)^{-1} X'Z\hat{W}(\tau)Z'\hat{Y}(\tau), \tag{6}$$

where $\hat{W}(\tau)$ is a $L \times L$ symmetric weighting matrix. If $L = K$, the second step estimator in (6) simplifies to the IV estimator using Z as instrument.

Our two-step estimator is extremely simple to implement; it requires only routines performing quantile regression and GMM estimation, which are already available in many softwares. In addition, we provide general-purpose packages for both R and Stata. Quantile regression, which is computationally more demanding due to the absence of a closed-form solution, is used only in the first stage, where there are fewer observations and a limited number of parameters to estimate. The first stage is also embarrassingly parallelizable, which further increases the computational speed. For this reason, our estimator remains computationally attractive in large datasets with numerous groups. The second stage is a straightforward GMM estimator, which includes OLS and two-stage least squares as special cases. Traditional panel data methods can also be used in the second stage. For instance, in our application, we observe individuals born in a given trimester in a given county. The subscript j defines a county-trimester cell, while the subscript i defines an individual within this cell. In the second stage, we include trimester, county, and state × year fixed effects to estimate the effect of food stamps on the birth weight distribution.

Remark 6 (Interpretation as a minimum distance estimator). Our estimator can be written as a MD estimator, where the second stage imposes restrictions on the first-stage coefficients. For simplicity, in this remark, we consider the case where all the regressors are exogenous and $Z = X$. The MD estimator minimizes

$$\hat{\delta}(\tau) = \arg\min_{\beta} \sum_{j=1}^{m} (\hat{\beta}_j(\tau) - R_j \delta(\tau))' \tilde{W}(\tau) (\hat{\beta}_j(\tau) - R_j \delta(\tau)), \tag{7}$$

where $\tilde{W}(\tau)$ is a $K \times K$ weighting matrix that might depend on the quantile index. The matrix of restrictions R_j is defined such that $\tilde{X}_j R_j = X_j$:

[13] An IV extension of the MD estimator of Galvao and Wang (2015) is suggested in Dai and Jin (2021).

$$\underset{(K_1+1)\times K}{R_j} = \begin{pmatrix} x'_{2j} & 0 \\ 0 & I_{K_1} \end{pmatrix}.$$

If we set $\tilde{W}(\tau) = \tilde{X}'_j \tilde{X}_j$ then the MD estimator is algebraically identical to using OLS in the second stage.

Given the representation above, our estimator is an MD estimator. However, it does not correspond to the textbook definition of a "classical minimum distance" estimator.[14] In the classical MD setup, all the sampling variance arises in the first stage: if we know the first stage coefficients, we know the final coefficients. It follows that the efficient weighting matrix $\tilde{W}(\tau)$ is the inverse of the first-stage variance. In our case, the second stage also contributes to the variance due to the presence of the group effects $\alpha(\tau, v_j)$. Even if we know $\beta_j(\tau)$ (for a finite number of groups), we cannot exactly pinpoint $\gamma(\tau)$. The group effects play a role similar to misspecification in the classical MD, but the resulting 'bias' disappears asymptotically as the number of groups increases. This is the second important difference: the dimension of our first stage estimates increases with the sample size while it is fixed for classical MD estimators.

The estimators suggested by Chamberlain (1994) and Galvao and Wang (2015) are special cases of our estimator in which only the first stage estimation error matters, and the efficient weighting matrix is the inverse of the first stage variance.[15] In section 3, we derive the asymptotic distribution when the estimation error of both stages matters.

2.3 Least Squares Minimum Distance Estimators

Since in our model we are interested in the within-group heterogeneity, we divide the process into two steps where the first step consists of the group by group regressions. It turns out that in the special case of a least squares first stage, splitting the problem into two steps, does not affect the results. This section discusses the analogy between our minimum distance estimator and traditional least squares (panel data) estimators in the special case that least squares regression is used in the first stage.

We present some equivalent ways to compute various common estimators and show that they can be estimated using our two-stage minimum distance approach. A more detailed discussion including formal statements can be found in Appendix A with proofs in Appendix A.2.

Consider first a model with group effects and group-level regressors

$$y_{ij} = x'_{1ij}\beta + \alpha_j + \varepsilon_{ij}. \tag{8}$$

The least squares fixed effects estimator can be computed by subtracting from each variable

[14] See section 14.6 in Wooldridge (2010).

[15] The efficient MD estimator of Galvao and Wang (2015) is numerically identical to our estimator using an IV second stage with instrument $(Z^*_j(\tau), \zeta_j)$, with $Z^*_j(\tau) = \tilde{X}_j(\tilde{X}'_j\tilde{X}_jV_j(\tau)\tilde{X}'_j\tilde{X}_j)^{-1}\tilde{X}'_jZ_j = (\tilde{X}_jV_j(\tau)\tilde{X}'_j)^{+}Z_j$, where Z_j contains a constant, X_{1j}, and ζ_j is a group fixed effect. The symbol $^+$ denotes the Moore-Penrose inverse.

its group average and applying the ordinary least squares estimator.[16] This within transformation eliminates the potential endogeneity coming from α_j and provides a consistent estimator without imposing assumptions on the unobserved group-level heterogeneity. This approach is not applicable in quantile models, as there is no known transformation that eliminates the group effects. In particular, in the traditional panel data context, time-demeaning or first-differencing the variables modifies the interpretation of the quantile regression coefficients because the quantiles are nonlinear operators. A second possibility to estimate fixed effects models consists in estimating the group effects by including an indicator variable for each group. It is well-known that this is algebraically identical to the within estimator. In quantile models, the dummy variables regression is computationally unattractive, as it requires estimating many parameters.[17] In addition, this approach does not provide a way to estimate the effect of the group-level variables, especially when we need to exploit an instrumental variable to identify their effect.[18]

A third numerically equivalent way to compute the least squares fixed effects estimator consists in exploiting the exogenous within variation using instrumental variable. Setting $\dot{x}_{1ij} = x_{1ij} - \bar{x}_{1j}$ where $\bar{x}_{1j} = n^{-1}\sum_{i=1}^{n} x_{1ij}$ as an instrument in an instrumental variable regression is numerically identical to the least squares fixed effects. This estimation procedure can be divided into two stages. The first stage consists of group-level regressions for each j. The intercept of each regression will absorb the unobserved heterogeneity α_j. The second stage aggregates the individual results by regressing the fitted values from the first stage on x_{1ij} using the transformed regressors, $\dot{x}_{1ij} = x_{1ij} - \bar{x}_{1j}$ where $\bar{x}_{1j} = n^{-1}\sum_{i=1}^{n} x_{1ij}$, as an instrumental variable. As shown below, using the first-stage fitted values as the dependent variable does not affect the results in the least squares model. Whereas, the instrument exploits only the variation within individuals, thus yielding a within estimator. This procedure can be easily extended to quantile models, where it substantially reduces the computational burden of quantile fixed effects estimation.[19]

The two-step procedure is not specific to fixed effects but applies to a wide range of estimators. We include the group-level regressors x_{2j} in the model

$$y_{ij} = x'_{1ij}\beta + x'_{2j}\gamma + \alpha_j + \varepsilon_{ij} \tag{9}$$

[16] In the context of traditional panel data we would call the j units "individuals" and the j units "time periods". Thus, this transformation consists in the time demeaning applied in traditional panel data literature that eliminates time-invariant individual effects.

[17] This approach is nevertheless feasible thanks to the sparsity of the design matrix, see Koenker (2004), and Koenker and Ng (2005).

[18] Koenker (2004) suggests a penalized quantile regression estimator with group effects, which can be interpreted as a random effects estimator. However, the linear dependence between the group indicator variables and the group-level variables implies that the effect of these variables is identified only from the individuals with fully shrunken group effects, see Harding et al. (2020).

[19] There is a fourth possibility to compute the least squares fixed estimator, which is similar to the third one, but, not numerically identical to the within estimator. The first stage also consists of individual-level regressions for each j. The second stage aggregates the slope coefficients directly by taking the average of the individual slopes with weights proportional to the variance of the regressors within individuals. Galvao and Wang (2015) suggests a similar estimator for the fixed effects quantile regression model. However, this averaging method does not allow for the presence of group-level regressors and, more generally, does not exploit the between-individuals variation.

and consider our minimum distance estimator with a least squares first stage instead of quantile regression. That is, the first stage consists of group-level least squares regressions, including only the individual-level variables. The second stage is a linear GMM regression of the first-stage fitted values on both individual-level and group-level variables. This two-step estimator is algebraically identical to the one-step linear GMM regression of y_{ij} on x_{ij} under the mild condition that for each group j, the matrix of instruments lies in the column space of the matrix of first stage regressors (see Proposition 3 in Appendix A). The intuition is as follows. The fitted values of the first-stage least squares regression can be written as $P_{X_j}Y_j$ where P_{X_j} is the first-stage least squares projection matrix of group j. If the instrument matrix, Z_j, is in the column space of $\tilde{X}_j$, it follows that $P_{X_j}Z_j = Z_j$. Therefore, $Z'\hat{Y} = Z'Y$ and the two GMM regressions are numerically identical. The matrix Z_j will lies in the column space of $\tilde{X}_j$ if the instruments are contained in $\tilde{X}_j$ or are a linear combination of the columns of the matrix of first-stage regressors. For example, demeaned individual-level regressors as well as group-level variables fall into the last category. Since OLS and 2SLS are special cases of GMM, the same result follows directly.

We can numerically obtain the most common least squares panel data estimators by selecting different instrumental variables for the second step GMM regression. In essence, the instrument determines which variation we exploit in the second stage.[20] For instance, we obtain the between estimator by using the group averaged variables, $\bar{x}_{1j}$ and x_{2j} as instruments. Instrumental variable approaches are available also for random effects estimation. More precisely, while FGLS is the most common estimator for the random effects model, Im et al. (1999) show that the overidentified 3SLS estimator, with instruments $\dot{x}_{1ij}$, $\bar{x}_{1j}$, and x_{2j}, is numerically identical to the random effects estimator. Since 3SLS is a special case of a GMM estimator, using the first-stage fitted values as dependent variables does not change the estimates. Alternatively, the random effects estimator can be implemented using the theory of optimal instrument with a just identified 2SLS regression (see Im et al., 1999; Hansen, 2022). Finally, the Hausman-Taylor estimator (Hausman and Taylor, 1981) can be implemented by selecting the following instruments: $\dot{x}_{1ij}$, the group average of the exogenous regressors. External instruments might also be included. Interestingly, in all cases, clustering the standard errors at the level of the group (or at a higher level) is sufficient to capture the first stage estimation error, see Proposition 5 in Appendix A.2. These clustered standard errors are numerically identical to the standard errors obtained after using the one-step GMM estimator with y_{ij} as the dependent variable.

3 Asymptotic Theory

In this section, we state the assumptions and present the asymptotic results. All the proofs are included in Appendix B. For simplicity of notation, in the following, we write $\alpha_j(\tau)$ instead

[20] The IV approach to these panel data estimators can also be implemented in one stage with y_{ij} as the dependent variable.

of $\alpha(\tau, v_j)$. We prove weak uniform consistency and weak convergence of the whole quantile regression process for $\tau \in \mathcal{T}$, where $\mathcal{T} \in (0,1)$ is a compact set of quantile indices of interest. The symbol $\ell^\infty(\mathcal{T})$ denotes the set of component-wise bounded vector values function of $\mathcal{T}$ and $\rightsquigarrow$ denotes weak convergence.

We start by writing the sampling error of $\hat{\delta}(\hat{W}, \tau)$ as a sum of a component arising from the first stage estimation error of $\beta_j(\tau)$ and a component arising from the second stage noise $\alpha_j(\tau)$:

Lemma 1 (Sampling error). *Assume that the model in equation (1) holds, then*

$$\hat{\delta}(\hat{W}, \tau) - \delta(\tau) = \left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1} S'_{ZX}\hat{W}(\tau)\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\left(\tilde{x}'_{ij}(\hat{\beta}_j(\tau) - \beta_j(\tau)) + \alpha_j(\tau)\right),$$

where $S_{ZX} = \frac{1}{mn}\sum_{i=1}^{N}\sum_{i=1}^{n} z_{ij}x'_{ij}$.

We now state assumptions that ensure that both components are well-behaved. For the analysis of the first stage estimator, we rely on results derived in Galvao et al. (2020) and make the assumptions required in their Theorem 2:

Assumption 1 (Sampling). *(i) The processes $\{(y_{ij}, x_{ij}, z_{ij}) : i \in \mathbb{Z}\}$ are independent across j. (ii) For each j, the observations $(y_{ij}, x_{1ij}, z_{1ij})_{i=1,\dots,n}$ are i.i.d. across i.*

Assumption 2 (Covariates). *(i) For all $j = 1, \dots, m$ and all $i = 1, \dots, n$, $\|x_{ij}\| \leq C$ almost surely. (ii) The eigenvalues of $\mathbb{E}[x_{1ij}x'_{1ij}]$ are bounded away from zero and infinity uniformly across j.*

Assumption 3 (Conditional distribution). *The conditional distribution $F_{y_{ij}|x_{1ij}}(y|x)$ is twice differentiable w.r.t. y, with the corresponding derivatives $f_{y_{ij}|x_{1ij}}(y|x)$ and $f'_{y_{ij}|x_{1ij}}(y|x)$. Further, assume that*

$$f_{max} := \sup_j \sup_{y\in\mathbb{R}, x\in\mathcal{X}} |f_{y_{ij}|x_{1ij}}(y|x)| < \infty$$

and

$$\bar{f}' := \sup_j \sup_{y\in\mathbb{R}, x\in\mathcal{X}} |f'_{y_{ij}|x_{1ij}}(y|x)| < \infty.$$

where $\mathcal{X}$ is the support of x_{1ij}

Assumption 4 (Bounded density). *There exists a constant $f_{min} < f_{max}$ such that*

$$0 < f_{min} \leq \inf_j \inf_{\tau\in\mathcal{T}} \inf_{x\in\mathcal{X}} f_{y_{ij}|x_{1ij}}(Q(\tau, y_{ij}|x)|x).$$

These are quite standard assumptions in the quantile regression literature. In Assumption 1, we assume that the processes are independent across j; this assumption can also be relaxed by allowing for clustering between groups. We also assume that the observations are i.i.d. within group, but this can be relaxed at the cost of a more complex notation by applying Theorem 4 in Galvao et al. (2020), which requires only stationarity and β-mixing. The estimator of the asymptotic variance that we suggest below is consistent in both cases. Assumption 2 requires that the regressors are bounded and that $\mathbb{E}\left[x_{1ij}x'_{1ij}\right]$ is invertible. Assumptions 3 and 4 impose smoothness and boundedness of the conditional distribution, the density, and its derivatives.

For the second stage GMM regression we impose the following assumptions:

Assumption 5 (**Instruments**). *(i) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, $||z_{ij}|| \leq C$ a.s. (ii) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, $\mathbb{E}[z_{ij}\alpha_j(\tau)] = 0$. (iii) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, y_{ij} is independent of z_{ij} conditional on (x_{ij}, v_j). (iv) As $m \to \infty$, $m^{-1}\sum_{j=1}^{m}\mathbb{E}_j[z_{ij}x'_{ij}] \to \Sigma_{ZX}$ where the singular values of Σ_{ZX} are bounded from below and from above.*

Assumption 6 (**Group effects**).
(i) For all $j = 1, \ldots, m$, $\mathbb{E}\left[\sup_{\tau\in\mathcal{T}}|\alpha_j(\tau)|^{4+\varepsilon_C}\right] \leq C$ for $\varepsilon_C > 0$. (ii) For some (matrix-valued) function $\Omega_2 : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{L\times L}$, $m^{-1}\sum_{j=1}^{m}\mathbb{E}[\alpha_j(\tau_1)\alpha_j(\tau_2)z_{ij}z'_{ij}] \to \Omega_2(\tau_1, \tau_2)$ uniformly over $\tau_1, \tau_2 \in \mathcal{T}$. (iii) For all $\tau_1, \tau_2 \in \mathcal{T}$, $|\alpha_j(\tau_2) - \alpha_j(\tau_1)| \leq C|\tau_2 - \tau_1|$.

Assumption 7 (**Coefficients**). *For all $\tau_1, \tau_2 \in \mathcal{T}$ and $j = 1, \ldots, m$, $||\beta_j(\tau_2) - \beta_j(\tau_1)|| \leq C|\tau_2 - \tau_1|$.*

These assumptions are the same as in Chetverikov et al. (2016). For the instrumental variables, we assume that (i) they are bounded, (ii) they are not correlated with the group effect (exclusion restriction), (iii) they do not affect the first stage estimation (this is often satisfied by construction, e.g. when the instruments do not vary within individuals or are a linear transformation of the first stage regressors), and (iv) they satisfy the relevance conditions. For the group effects we assume that they have a finite fourth moment, and the average variance of $z_{ij}\alpha_j(\tau)$ converges to a well-defined matrix. Finally, we assume that the group effects and the coefficients are continuous functions of the quantile index.

Since the unobserved heterogeneity $\alpha(\tau, v_j)$ is group-specific, we require that the number of groups m diverges to infinity. The first stage quantile regression estimator is a nonlinear estimator that has potentially a bias of order $\frac{1}{n}$. Hence, for consistency, the number of observations per group n must also diverge to infinity. For unbiased asymptotic normality, we need the bias to shrink faster than the standard deviation of the estimator. We will show that some elements of $\hat{\delta}(W, \tau)$ converge at the $\sqrt{m}$ rate such that we need that n goes to infinity more quickly than $\sqrt{m}$. On the other hand, other elements converge at the $\sqrt{mn}$ rate so that n goes to infinity more quickly than m. We state these three different relative growth rates in the following assumption:

Assumption 8 (**Growth rates**). *As $m \to \infty$, we have*

(a) $\frac{\log m}{n} \to 0$,

(b) $\frac{\sqrt{m}\log n}{n} \to 0$,

(c) $\frac{m(\log n)^2}{n} \to 0.$

In our first result, we establish uniform consistency of our estimator. In addition to the previously stated conditions, we assume that the estimated weighting matrix uniformly converges to a strictly positive definite matrix that is continuous in the quantile index.[21]

Theorem 1 (**Uniform consistency**). *Let the model in equation (1), Assumptions 1-7 as well as Assumption 8(a) hold. Uniformly in* $\tau \in \mathcal{T}$, $\hat{W}(\tau) \underset{p}{\to} W(\tau)$ *where* $W(\tau)$ *is strictly positive definite and, for all* $\tau_1, \tau_2 \in \mathcal{T}$, $||W(\tau_2) - W(\tau_1)|| \leq C|\tau_2 - \tau_1|$. *Then,*

$$\sup_{\tau \in \mathcal{T}} \|\hat{\delta}(\tau) - \delta(\tau)\| = o_p(1).$$

We now study the asymptotic distribution of our estimator. In Lemma 1 we see that the sample moment condition is made of two terms. It is useful to consider them separately:

$$\bar{g}^{(1)}_{mn}(\hat{\delta}, \tau) := \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} z_{ij} \tilde{x}'_{ij} \left(\hat{\beta}_j(\tau) - \beta_j(\tau)\right) \tag{10}$$

$$\bar{g}^{(2)}_{mn}(\hat{\delta}, \tau) := \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} z_{ij} \alpha_j(\tau) \tag{11}$$

such that total moment condition is the sum of both components: $\bar{g}_{mn}(\hat{\delta}, \tau) := \bar{g}^{(1)}_{mn}(\hat{\delta}, \tau) + \bar{g}^{(2)}_{mn}(\hat{\delta}, \tau) = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} g_{ij}(\hat{\delta}, \tau)$. Lemma 2 establishes joint asymptotic normality for the entire moment condition processes.

Lemma 2 (**Asymptotic distribution of the sample moments**). *We assume that Assumptions 1-7 hold.*

(i) Under Assumption 8(c), as $m \to \infty$,

$$\sqrt{mn}\bar{g}^{(1)}_{mn}(\hat{\delta}, \cdot) \rightsquigarrow Z_1(\cdot), \text{ in } l^\infty(\mathcal{T}), \tag{12}$$

where $Z_1(\cdot)$ *is a mean-zero Gaussian process with uniformly continuous sample paths and covariance function* $\Omega_1(\tau, \tau') = \mathbb{E}_j \left[\Sigma_{ZXj} V_j(\tau, \tau') \Sigma'_{ZXj}\right]$ *with* $\Sigma_{ZXj} = \mathbb{E}[z_{1j}\tilde{x}'_{1j}]$ *and* $V_j(\tau, \tau')$ *is the asymptotic variance-covariance matrix of* $\hat{\beta}_j(\tau)$ *and* $\hat{\beta}_j(\tau')$*:*

$$V_j(\tau, \tau') = \mathbb{E}[f_{y|x}(Q_{y|x}(\tau|\tilde{x}_{1j})|\tilde{x}_{1j})\tilde{x}_{1j}\tilde{x}'_{1j}]^{-1}(\min(\tau, \tau') - \tau\tau')\mathbb{E}[\tilde{x}_{1j}\tilde{x}'_{1j}]\mathbb{E}_t[f_{y|x}(Q_{y|x}(\tau'|\tilde{x}_{1j})\tilde{x}_{1j}\tilde{x}'_{1j}]^{-1}$$

(ii) Under Assumption 8(b), As $m \to \infty$,

$$\sqrt{m}\bar{g}^{(2)}_{mn}(\hat{\delta}, \cdot) \rightsquigarrow Z_2(\cdot), \text{ in } l^\infty(\mathcal{T}), \tag{13}$$

where $Z_2(\cdot)$ *is a mean-zero Gaussian process with uniformly continuous sample paths and covariance function* $\Omega_2(\tau, \tau')$, *which is defined in Assumption 6(ii).*

[21] The efficient weighting matrix suggested below may actually be asymptotically singular. For this reason, in Appendix B we also provide an alternative consistency result (Theorem 1′) that applies to the efficient estimator.

(iii) Under Assumption 8(c), as $m \to \infty$, $\sup_{\tau \in \mathcal{T}} \left(\bar{g}^{(1)}_{mn}(\hat{\delta}, \cdot), \bar{g}^{(2)}_{mn}(\hat{\delta}, \cdot) \right) = o_p\left(\frac{1}{mn}\right)$.

$\bar{g}^{(1)}_{mn}(\hat{\delta}, \cdot)$ reflects the estimation error that arises in the first stage quantile regression estimation. Since the first-stage regressors vary within groups, the relevant number of observations is mn and, correspondingly, the variance is proportional to $1/(mn)$. On the other hand, since the bias of the first-stage quantile regression is of order $1/n$, for asymptotic unbiasedness, we must require that n goes to infinity slightly faster than m. In the proof we build on results derived in Volgushev et al. (2019) and in Galvao et al. (2020). $\bar{g}^{(2)}_{mn}(\hat{\delta}, \cdot)$ reflects the estimation error due to the randomness in $\alpha_j(\tau)$. This moment can also be interpreted as the moment that would be relevant if we knew $\beta_j(\tau)$. Since $\alpha_j(\tau)$ varies only between groups, the relevant number of observations here is m and, accordingly, the variance of this moment converges at the slower rate of $1/m$. For asymptotic unbiasedness, we need only the weaker condition 8(b), which requires that n goes to infinity slightly faster than $\sqrt{m}$.

The moment condition $\bar{g}_{mn}(\hat{\delta}, \tau)$ is, thus, the sum of two components that converge at different rates to zero. Its first-order asymptotic distribution will be dominated by the slowest component, which is $\bar{g}^{(2)}_{mn}(\hat{\delta}, \tau)$, except if its variance is zero. Since the degree of group-level heterogeneity affects this variance, it is useful to consider three cases: strong heterogeneity, no heterogeneity, and weak heterogeneity. We first derive the asymptotic distribution of our estimator when the level of heterogeneity is known and then we suggest an estimator that is efficient in all cases and adaptive inference procedures.

We start with the case of strong heterogeneity that we define to be $\mathrm{Var}(\alpha_j(\tau)) > \varepsilon > 0$ uniformly in τ.[22] We must distinguish between two sort of instruments: L_1 instruments in z_{1ij} satisfy $\bar{z}_{1j} = 0$ for all j, while L_2 instruments in z_{2ij} satisfy $\bar{z}_{2j} \neq 0$ at least for some groups j. We order the instruments such that $z_{ij} = (z'_{1ij}, z'_{2ij})'$. Note that instruments that vary only within groups can be normalized to have mean zero. For instance, we can identify the effect of individual-level variables x_{ij} by instrumenting it with $\dot{x}_{ij} = x_{ij} - \bar{x}_j$, which has a mean of zero in all groups and corresponds to the fixed effects estimator. For instruments of the first type, the associated elements of $\bar{g}^{(2)}_{mn}(\hat{\delta}, \cdot) = 0$ such that the corresponding elements of $\bar{g}_{mn}(\hat{\delta}, \cdot)$ converge at the $\sqrt{mn}$ rate. Thus, we have L_1 'fast' moments that converge at the $\sqrt{mn}$ rate and L_2 'slow' moments that converge at the $\sqrt{m}$ rate.

The rate of convergence of each element of $\hat{\delta}(W, \cdot)$ is determined by the rate of convergence of the moments that are used asymptotically to estimate this parameter. In the exactly identified case, our estimator simplifies to the instrumental variable estimator such that the estimation error is

$$\hat{\delta}(\tau) - \delta(\tau) = S_{ZX}^{-1} \bar{g}_{mn}$$

[22] We will allow later for different levels of heterogeneity at different quantiles but keep the exposition simple at the moment.

and $S_{ZX} \underset{p}{\rightarrow} \Sigma_{ZX}$. We partition Σ_{ZX} and assume that it is block lower triangular:

$$\Sigma_{ZX} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{14}$$

where Σ_{11} is $L_1 \times K_1$, Σ_{12} is $L_1 \times K_2$, Σ_{21} is $L_2 \times K_1$ and Σ_{22} is $L_2 \times K_2$. Then it follows that the first K_1 elements of $\hat{\delta}$ are determined only by the 'fast' moments while the remaining K_2 elements are also determined by the 'slow' moments.

The coefficients on variables that vary only between groups cannot be estimated using the 'fast' instruments z_{1ij}, which do not vary between groups. Formally, the full rank condition 5(iv) requires that group-level variables are identified with instruments that vary between groups. The related coefficients will necessarily converge at the slow $\sqrt{m}$ rate. On the other hand, coefficients on individual-level variables can be estimated using only the within-group variation, that is by using $\dot{x}_{ij}$ as instrument for x_{ij}. These instruments are by construction uncorrelated with variables that vary only across group, which satisfies condition 14. For this reason and to avoid additional notation, we denote the number of individual-level variables and of 'fast' coefficients by the same symbol K_1.[23]

When the model is overidentified, condition (14) does not guarantee that the 'slow' instrument will not contaminate if the asymptotic GMM weighting matrix $W(\tau)$ has full rank, then the slow moments will asymptotically dominate. Since the rate of convergence of these processes are different, the rate of convergence and the first-order asymptotic approximation will be determined by the slower moment, which is $\bar{g}^{(2)}_{mn}(\hat{\delta}, \cdot)$.

We summarize and formalize this discussion in the following theorem.

Theorem 2 **(Asymptotic distribution when the degree of heterogeneity is known).** *Let Assumptions 1-7 hold. In addition, $\hat{W}(\tau) \underset{p}{\rightarrow} W(\tau)$ uniformly in $\tau \in \mathcal{T}$. We partition the $L \times L$ weighting matrix as follows*

$$W(\tau) = \begin{pmatrix} W_{11}(\tau) & W_{12}(\tau) \\ W_{21}(\tau) & W_{22}(\tau) \end{pmatrix}$$

where $W_{11}(\tau)$ is a $L_1 \times L_1$ matrix and $W_{22}(\tau)$ a $L_2 \times L_2$ matrix. For all $\tau_1, \tau_2 \in \mathcal{T}$, $||W(\tau_2) - W(\tau_1)|| \leq C|\tau_2 - \tau_1|$. Let Σ_{11} and Σ_{22} be the $L_1 \times K_1$ upper-left and $L_2 \times K_2$ bottom-right submatrices of Σ_{ZX}.

(i) Let Assumption 8(c) hold, $W_{11}(\tau)$ is strictly positive definite, $W_{12}(\tau) = 0$, $W_{21}(\tau) = 0$, and $W_{22}(\tau) = 0$. Then,

$$\sqrt{mn}(\hat{\delta}_1(\hat{W}(\cdot), \cdot) - \delta_1(\cdot)) \rightsquigarrow G_1(\cdot) Z_1(\cdot), \textit{ in } l^\infty(\mathcal{T}), \tag{15}$$

where $G_1(\tau) = \left(\Sigma_{11}' W_{11}(\tau) \Sigma_{11}\right)^{-1} \Sigma_{11}' W_{11}(\tau)$.

[23] The results can be trivially extended to the case where some individual-level variables are identified using only group-level variables.

(ii) Let Assumption 8(b) hold and $W(\tau)$ is strictly positive definite, then

$$\sqrt{m}(\hat{\delta}_2(\hat{W}(\cdot),\cdot) - \delta_2(\cdot)) \rightsquigarrow G_2(\cdot)Z_2(\cdot), \text{ in } l^\infty(\mathcal{T}), \tag{16}$$

where $G_2(\tau) = \left(\Sigma_{22}' W_{22}(\tau)\Sigma_{22}\right)^{-1} \Sigma_{22}' W_{22}(\tau)$.

The asymptotic distribution in Theorem 2(ii) is the same as in Chetverikov et al. (2016) but we were able to weaken the growth rate condition from $\frac{m^{2/3}\log n}{n} \to 0$ to $\frac{m^{1/2}\log n}{n} \to 0$ by exploiting new results in Galvao et al. (2020). $\hat{\delta}_1(W,\tau)$ and $\hat{\delta}_2(W,\tau)$ have both different rates of convergence, and require different growth conditions. These estimators are first-order efficient if $W_1(\tau) = \Omega_1(\tau)^{-1}$ and $W_2(\tau) = \Omega_2(\tau)^{-1}$. These results have several weaknesses. First, the asymptotic distribution is not uniform in $\text{Var}(\alpha_j(\bar{z}_j\tau))$ and the rate of convergence changes discontinuously if, for example, $\text{Var}(\bar{z}_j\alpha_j(\tau))$ converges to zero. Second, for $\hat{\delta}_1(W,\tau)$, we do not use the information contained in the slow moments. Consider, for example, the random effects estimator where all the regressors are individual-level. The vector of instruments consists of $x_{ij} - \bar{x}_j$ and $\bar{x}_j$. Considering only the first-order asymptotic distribution, we can obtain a first-order efficient estimator by giving zero weights to the slow moments. In other words, the instruments $\bar{x}_j$ are not used because their contribution is asymptotically negligible, and the random effects estimator would be equivalent to the fixed effects estimator.[24] Third, for the $\hat{\delta}_2(W,\tau)$, variance coming from the first stage does not appear in the asymptotic distribution because it converges to zero at a quicker rate. Consequently, inference may have poor properties. To solve these issues, we keep both moments together, implement adaptive inference that takes the first stage error into account, and use the slow moment with weights that decline at the n^{-1} rate.

Note that

$$\sqrt{m}\bar{g}_{mn}(\hat{\delta},\cdot) \rightsquigarrow \frac{Z_1(\cdot)}{n} + Z_2(\cdot). \tag{17}$$

Following standard GMM arguments, the efficient weighting matrix is given by

$$W(\tau)^* = \left(\Omega_1(\tau)/n + \Omega_2(\tau)\right)^{-1}. \tag{18}$$

With this weight matrix, asymptotically, the fast moments get weighted infinitely more than the slow moments, so the parameters identified by the fast moments will converge at the $\sqrt{mn}$ rate. The parameters that are identified only by the slow moments will not be affected by W_1 such that their asymptotic distribution will depend only on W_2.[25]

We estimate this weighting matrix by

$$\hat{W}(\tau)^* = \hat{\Omega}(\tau)^{-1} = \left(\frac{1}{m}\sum_{j=1}^{m} Z_j'\tilde{u}_j(\tau)\tilde{u}_j(\tau)'Z_j\right)^{-1}.$$

[24] This is not specific to quantile models and also affects least squares models with large n (see Ahn and Moon, 2014).

[25] As we can see in the proof of Proposition 2, the fact that W_1 is multiplied with n does not cause a problem because the weighting matrix matters only up to scale.

The $n \times 1$ vector $\tilde{u}_j(\tau)$ contains the residuals from a preliminary second stage regression using an inefficient weighting matrix. Consistency of $\hat{\Omega}$ follows directly by the proof of Proposition 1 below.

For asymptotic normality of the adaptive estimator we need to impose the strongest growth rate condition 8(c):

Theorem 3 (Adaptive estimator). *Let Assumptions 1-7 and 8(c) hold. Then,*

$$\sqrt{m}(\hat{\delta}(\hat{\Omega}(\cdot), \cdot) - \delta(\cdot)) \rightsquigarrow G(\cdot)\left(\frac{Z_1(\cdot)}{n} + Z_2(\cdot)\right), \text{ in } l^{\infty}(\mathcal{T}), \tag{19}$$

where $Z_1(\cdot)$ and $Z_2(\cdot)$ are defined in Lemma 2.

We have seen that the convergence rate and the asymptotic distribution of our estimator changes substantially depending on the data generating process. For this reason, we want to suggest an adaptive inference procedure that is uniformly valid over different degree of heterogeneity and convergence rates of the estimator. Surprisingly, we find that clustering the second-stage covariance matrix at the individual level yields uniformly consistent estimator of the covariance matrix over different degrees of unobserved heterogeneity and convergence rates. Thus, inference does not require estimating the density of the quantile regression in the first stage and is computationally straightforward. Clustering automatically takes the first stage variance into account also for $\sqrt{m}$-consistent parameters, thus providing a higher-order improvement. This simple procedure might work in a broader range of situations, and it is of interest on its own. A similar bootstrap-based procedure is suggested in Fernández-Val et al. (2022).

To estimate the covariance matrix, define the $n \times 1$ vector of residuals $\hat{u}_j(\tau) = \tilde{X}_j \hat{\beta}_j(\tau) - X_j \hat{\delta}(\tau)$. Then the covariance matrix of $\hat{\delta}(\tau)$ is estimated by

$$\widehat{V}_{\delta}(\tau) = \left(X'Z\hat{W}Z'X\right)^{-1} X'Z\hat{W}\left(\sum_{j=1}^{m} Z_j'\hat{u}_j(\tau)\hat{u}_j(\tau)'Z_j\right)\hat{W}Z'X\left(X'Z\hat{W}Z'X\right)^{-1}.$$

The following proposition shows that the covariance matrix consistent of uniformly in the variance of $\bar{z}_j\alpha_j$.

Proposition 1 (Consistency of the estimated covariance matrix). *Let assumptions 1-7 and 8(c) hold. Let $\eta \in \mathbb{R}^K$ with $||\eta|| > \varepsilon > 0$. Let*

$$V_{\delta}(\tau) = G(\tau)\left(\frac{\Omega_1(\tau)}{mn} + \frac{\Omega_2(\tau)}{n}\right)G(\tau)'.$$

Then,

$$\frac{\eta'\hat{V}_{\delta}(\tau)\eta}{\eta' V_{\delta}(\tau)\eta} = o_p(1).$$

Proposition 1 show that the clustered covariance matrix provides uniformly valid inference in $\text{Var}(\bar{z}_j\alpha_j)$ and therefore, valid regardless of the speed of convergence of the moment conditions. Similar results based on cross-sectional (clustered) bootstrap are suggested in Liao and Yang (2018); Lu and Su (2022); Fernández-Val et al. (2022).

If there are more moment conditions than parameters to estimate ($L > K$), it is possible to test overidentifying restrictions with an overidentification test in the second stage (see e.g. Hansen, 1982). More precisely, we want to test the hypothesis $\mathbb{H}_0 : \mathbb{E}[Z_j'\alpha_j(\tau)] = 0$. Compared to a traditional GMM, our overidentification test has to deal with the possible different convergence rates of the elements of $\hat{\delta}$. We solve this issue by rescaling the efficient weight matrix by Λ_n, where Λ_n is a $K \times K$ diagonal matrix with $\sqrt{n}$ for the first K_1 elements and 1 otherwise. Let $g_j(\delta,\tau) = Z_j'\left(\hat{Y}_j(\tau) - X_j\delta(\tau)\right)$ and $\bar{g}_m(\delta,\tau) = \frac{1}{m}\sum_{j=1}^{m} g_j(\delta,\tau)$. Define the GMM criterion function

$$J\left(\hat{\delta}(\tau)\right) = m\bar{g}_m(\hat{\delta},\tau)'\hat{S}^{-1}(\tau)\bar{g}_m(\hat{\delta},\tau), \tag{20}$$

where $\hat{S} = \Lambda_n^{-1}\hat{\Omega}\Lambda_n^{-1}$ is the inverse of the second order optimal weighting matrix. Note that this weight matrix is uniformly efficient over different degrees of unobserved heterogeneity and different convergence rates of the moment conditions.

Proposition 2. *Under the $\mathbb{H}_0$ and Assumptions 1-6 and 8(c) as n and $m \to \infty$, $J(\hat{\delta}(\tau)) \xrightarrow{d} \chi^2_{L-K}$.*

Hence, the criterion function $J\left(\hat{\delta}(\tau)\right)$ can be used to assess the validity of the instruments. In the next section, we show how this overidentification test can be used as a generalization of the Hausman Test for the random effects estimator.

4 Traditional Quantile Panel Data Estimators

4.1 Fixed Effects, Random Effects and Between Estimators

In this section, we discuss a particular case of our model in which j indexes the individuals and i indexes the time periods. This case corresponds to the traditional panel data setting. In this literature, the x_{1ij} variables are usually called time-varying variables, and the x_{2j} variables are called time-invariant or time-constant variables. Through this section, we will use this terminology (individuals and time periods) that is commonly used in traditional panel data. The MD estimator can be used for many panel data models, including the fixed effects, the random effects, the between, and the Hausman-Taylor model. The first stage estimator uses only data for one individual at the time and is unaffected. In the second stage, as for least squares estimation (see section 2.3), we compute panel data estimators by selecting different instruments. Depending on the model, the instrument z_{ij} will be defined so that the orthogonality condition in equation (2) holds. More precisely, for fixed effects estimation, the instrument z_{ij} contains the demeaned regressor $\dot{x}_{1ij}$ and varies only within j. For the between estimator, z_{ij} equals the individual mean of the regressors $\bar{x}_{ij}$. Finally, for the pooled estimator, $z_{ij} = x_{ij}$.[26] Implementing efficient estimation is one of the main challenges of quantile random effects as the model is overidentified. We suggest two different estimators. The first is an efficient GMM estimator, while the second uses optimal instruments. Given the first stage, we have the following moment restriction:

$$\mathbb{E}[Z_j'(\tilde{X}_j\hat{\beta}_j(\tau) - X_j\delta(\tau)] = 0. \tag{21}$$

If the instrument Z_j contains both the mean and the demeaned regressors, the efficient GMM will optimally weight the within and between variation. The moment condition in equation (21) contains both fast and slow moments, but the fast moments are sufficient to identify the coefficients on the individual-level regressors. The first-order efficient weighting matrix would give zero weights to the slow moment, and the random effects estimator would be identical to the fixed effects estimator. Using an efficient weighting matrix, we obtain a more efficient random effects estimator by also exploiting the between variation. The weighting matrix can be computed as described in section 3. As $m \to \infty$, the relative weights given to the slow moments converge to 0, and the random effect estimator converges to the fixed effects estimator (see Baltagi, 2021; Ahn and Moon, 2014 for a similar argument in least squares models).

If we impose the stronger assumption that the moment restriction in equation (21) holds conditional on Z_j, we can use the theory of optimal instruments to derive a random effects estimator. Optimal instruments are relevant when a researcher has a conditional moment restriction of the form $\mathbb{E}[g_j(\delta,\tau)|Z_j] = 0$. When a moment condition holds conditional on Z_j, an infinite set of valid moments exist, and one could use additional moments to increase efficiency. The goal is to select the instrument that minimizes the asymptotic variance, which takes the form $Z_j^* = \mathbb{E}[g_j(\delta,\tau)g_j(\delta,\tau)'|Z_j]^{-1}R_j(\delta,\tau)$, with $R_j(\delta,\tau) = \mathbb{E}[\frac{\partial}{\partial\delta}g_j(\delta,\tau)|Z_j]$ (see, e.g., Chamberlain, 1987 and Newey, 1993). To implement the random effect estimator with optimal instruments, we set $Z_j = X_j$. Under the additional assumption that $\mathbb{E}[\alpha_j^2(\tau)|X_j] = \sigma_\alpha^2(\tau)$, the optimal instrument simplifies to $Z_j^*(\tau) = \left(\tilde{X}_j\frac{V_j(\tau)}{n}\tilde{X}_j' + \mathbf{l}_n'\mathbf{l}_n\sigma_\alpha^2(\tau)\right)^+ X_j$, where $V_j(\tau)$ is the asymptotic variance from the first stage for a group j, $\mathbf{l}_n$ is a n-dimenstional vector of ones, and $^+$ denotes the Moore-Penrose inverse.[27] If $\mathbb{E}[\alpha_j^2(\tau)|X_j] = \sigma_\alpha^2(\tau)$, the random effect estimator based on optimal instruments is efficient.

A few remarks about the optimal instruments follow. First, under standard random effects assumptions, the optimal instrument applied to mean random effects models is numerically identical to the FGLS estimator. Second, in least squares models, using the moment restrictions

[26] The fixed effects estimator, in general, does not allow estimating γ, as the effect of time-constant variables is not identified separately from the group effects. In some situations, it is still possible to estimate γ by strengthening the assumption on the group-level regressors x_{2j} without changing the assumptions on individual-level regressors x_{1ij}. If x_{2j} is uncorrelated with α_j, it is possible to consistently estimate γ by regressing the fitted values for each quantile τ on x_{ij} using demeaned x_{1ij} and x_{2j} as instruments. Therefore, our two-step approach allows us to consistently estimate the effect of group-level regressors using the same approach as with linear regression.

[27] Since the matrix $(\tilde{X}_j\frac{V_j(\tau)}{n}\tilde{X}_j' + \mathbf{l}_n'\mathbf{l}_n\sigma_\alpha^2(\tau))$ is singular, we use the Moore-Penrose inverse.

with the true outcome or the first stage fitted values imply the same optimal instrument. To put it differently, under random effects assumptions, the matrix $\tilde{X}_j \frac{V_j}{n} \tilde{X}_j' + \mathbf{l}_n' \mathbf{l}_n \sigma_\alpha^2$ simplifies to the usual random effects structure. These results are summarized in Proposition 4 in Appendix A.2. Third, if $\sigma_\alpha = 0$, this estimator is identical to the efficient MD estimator (see Proposition 6 in Appendix C). Fourth, the optimal instrument depends on n analogously to the efficient weighting matrix of the GMM estimator. As n increases, the first stage variance converges to zero, and the generalized inverse will give infinitely more weights to the within variation and asymptotically converge to the fixed effects estimator.

To make the optimal instrument approach operational, we need a consistent estimator of Z_j^*. In the following, we assume that $\mathbb{E}[\alpha_j^2(\tau)|X_j] = \sigma_\alpha^2(\tau)$ and we suggest estimators for $V_j(\tau)$ and $\sigma_\alpha^2(\tau)$. Compared to the classical random effects structure, we use the first stage variance.[28] This formula has two main advantages. First, it is straightforward to compute $\hat{V}_j$. Second, it is possible to allow for dependence in the errors in the first stage regressions. The first stage variance can be estimated by $\hat{V}_j(\tau) = \hat{A}_j^{-1}(\tau)\hat{B}_j(\tau)\hat{A}_j^{-1}(\tau)$ where $\hat{A}_j(\tau) = \tau(1-\tau)\frac{1}{n}\sum_{i=1}^n \tilde{x}_{ij}\tilde{x}_{ij}'$ and $B_j(\tau)$ can be computed using the Kernel Density estimator of Powell (1991):

$$\hat{B}_j(\tau) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{y_{ij} - \tilde{x}_{ij}'\beta_j(\tau)}{h}\right)\tilde{x}_{ij}\tilde{x}_{ij}', \tag{22}$$

where $K(\cdot)$ is the uniform kernel $K(u) = \frac{1}{2}I(|u| \leq 1)$. Alternatively, $V_j(\tau)$ can be estimated by bootstrapping the first stage for each individual separately. We estimate $\sigma_\alpha(\tau)$ using the estimator suggested by Nerlove (1971):

$$\hat{\sigma}_\alpha^2(\tau) = \frac{m}{m-1}\sum_{j=1}^m (\hat{\alpha}_j(\tau) - \bar{\hat{\alpha}}_j(\tau))^2, \tag{23}$$

where $\bar{\hat{\alpha}}_j = \frac{1}{m}\sum_{j=1}^m \hat{\alpha}_j$ and the α_j are estimated by a preliminary least squares dummy variable regression of $\hat{y}_{ij}(\tau)$ on x_{ij}.[29]

Compared to the optimal instrument approach, efficient GMM relies on weaker conditions, and it is simpler to implement as it does not require a direct estimation of $V_j(\tau)$. Instead, it requires only the consistent estimation of the efficient weighting matrix, which is simpler to estimate.

4.2 Hausman and Taylor Model

The Hausman-Taylor model allows to find instrumental variables from inside the model. It is a middle ground between fixed effects and random effects. On the one hand, the random effects estimator relies on the orthogonality between $\alpha_j(\tau)$ and x_{ij}. On the other hand, the fixed effects estimator only identifies the effect of individual-level variables. To estimate the

[28] Using the first stage variance will not impose equality on the estimated densities of the errors $\hat{f}_{Y_j - \tilde{X}_j\beta_j}(0)$ in the second stage. Thus, observations will be weighted differently, depending on the first stage variance.

[29] The formula can be modified to accommodate unbalanced panels.

effect of group-level variables, Hausman and Taylor (1981) assume that some elements of X_j are uncorrelated with $\alpha_j(\tau)$. We consider model (1) but we partition X into four types of variables, $X = [X^x_{1j}\ X^n_{1j}\ X^x_{2j}\ X^n_{2j}]$, where the superscript x indicates that the variable is exogenous, and the superscript n indicates that it might be endogenous. Thus,

$$\mathbb{E}[X^x_{1j}\alpha_j(\tau)] = 0,$$
$$\mathbb{E}[X^x_{2j}\alpha_j(\tau)] = 0.$$

The assumptions imply that we can estimate $\delta(\tau)$ using the instrument $Z_j = (\dot{X}^x_{1j}, \dot{X}^n_{1j}, \bar{X}^x_{1j}, X^x_{2j})$. While X^n_{2j} is potentially endogenous, the within variation is uncorrelated with $\alpha_j(\tau)$ as it varies only between j. Identification requires that there are at least as many instruments as parameters to estimate. Hence, we need $dim(x^x_{1ij}) \geq dim(x^n_{2j})$. If the model is overidentified, it is possible to implement efficient GMM, and if conditional moment restrictions are available, optimal instruments can be used. Implementation of the optimal instrument is not straightforward as it requires the estimation of $\mathbb{E}[X_j|Z_j]$, usually estimated nonpametrically (see Newey, 1993). In this paper, we do not contribute in this direction. In the special case where there is no X^n_{1j}, so that all individual-level regressors are exogenous, the optimal instrument approach can more easily be implemented as the first stage includes only exogenous variables.

4.3 Hausman Test

Consistency of the random effects estimator relies on stronger orthogonality conditions compared to the fixed effects estimator. Under these stronger assumptions, both estimators are consistent, but the fixed effects is inefficient. Hausman (1978) suggested a test for the null hypothesis of random effects against the alternative of fixed effects. This subsection explains how we can use the overidentification test presented in Section 3 as a quantile version of the Hausman test for our two-step estimator. Various generalizations of the Hausman test have been suggested in the literature (see, e.g., Chamberlain, 1982; Mundlak, 1978; Wooldridge, 2019). Arellano (1993) considers an heteroskedasticity and autocorrelation robust generalization based on a Wald test. Ahn and Low (1996) propose a GMM test based on a 3SLS regression as an equivalent method for the Hausman test. In Section 4, we suggest efficient GMM as a possibility to perform random effects estimation. The assumption of correct specification of the first stage is maintained both under the null and the alternative hypotheses. Compared to the fixed effects estimator, consistency of the random effects estimator additionally requires that X_j is uncorrelated with $\alpha_j(\tau)$ so that $\mathbb{E}[\dot{X}'_{1j}\alpha_j(\tau)] = 0$ and $\mathbb{E}[\bar{X}'_j\alpha_j(\tau)] = 0$ are a valid moment conditions. By contrast, the fixed effects rely only on the moment condition $\mathbb{E}[\dot{X}'_{1j}\alpha_j(\tau)] = 0$. Consequently, the overidentification test suggested in Section 3 can be used as a test of the $\mathbb{H}_0 : \mathbb{E}[\dot{X}'_{1j}\alpha_j(\tau)] = 0$ and $\mathbb{E}[\bar{X}'_j\alpha_j(\tau)] = 0$, which is a test of the random effects orthogonality conditions. Compared to the traditional Hausman test, our test does not rely on the assumption of conditional homoskedasticity of the errors and is robust to clustering.

4.4 Simulations

This section presents simulation results for the different panel data estimators and the Hausman-type test presented in the previous subsections. These simulations focus on the estimation of $\beta(\tau)$, while the next section includes results for $\gamma(\tau)$. We consider the following data generating process where all variables are scalars:

$$y_{ij} = \beta x_{1ij} + \alpha_j + (1 + 0.1x_{1ij})\nu_{ij}. \tag{24}$$

We let $\beta = 1$ and $\nu_{ij} \sim \mathcal{N}(0,1)$. The regressor is defined by $x_{1ij} = h_j + 0.5u_{ij}$, with $u_{ij} \sim \mathcal{N}(0,1)$ and

$$\begin{pmatrix} h_j \\ \alpha_j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix}\right).$$

If $\lambda \neq 0$, x_{1ij} is correlated with α_j. For the simulation of the panel data estimators, we let $\lambda = 0$ so that all estimators are consistent. In contrast, in the Monte Carlo study of the Hausman test, we set $\lambda = \{0, 0.1, 0.2, 0.3, 0.4\}$. The true coefficient takes the values $\beta(\tau) = \beta + 0.1F^{-1}(\tau)$ where F is the standard normal CDF. We consider the samples with $n = \{10, 25, 200\}$ and $m = \{25, 200\}$ and focus on the set of quantiles $\mathcal{T} = \{0.1, 0.5, 0.9\}$. All simulation results are based on 10,000 replications. Table 1 shows the bias and the standard deviations, and Table 2 shows the coverage probability of the confidence intervals. Simulation results of the rejection probabilities of the Hausman test are in Table 3.

As shown in Table 1, the estimators perform well also when both m and n are small. The RE-GMM (random effects implemented by GMM) estimator performs similarly to the RE-OI (random effects implemented with optimal instruments) estimator and, in some cases, even better. As expected, asymptotically, the RE-GMM, the RE-OI, and the fixed effects (FE) estimators become indistinguishable as n increases. Whereas with small n, there is an apparent gain in using a random effects estimator. From the standard deviations, it is possible to see the different rates of convergence of the estimators. The precision of the fixed effects and random effects estimators increases in similar magnitude when m or n increases. In contrast, the standard deviation of the pooled and between (BE) estimator decreases only when m increases. The pooled and the between estimators have the smallest bias but, in most cases, also the largest variance.

The coverage probabilities of the 95% confidence intervals are in Table 2. The confidence intervals of the pooled and the fixed effects estimator perform well in all sample sizes considered. On the other hand, the confidence bands of the random effects estimators slightly undercover the true parameter mostly when n is small. In larger samples, all the coverage probabilities are close to the theoretical level.

Table 3 shows the rejection probabilities of the overidentification test for different values of λ. When $\lambda = 0$, the $\mathbb{H}_0$ is satisfied, so we should reject the null at a rate close to 5%. If $\lambda \neq 0$, X_j is correlated with α_j some moment conditions used by the RE-GMM estimator are not valid.

Quantile	Pooled	BE	FE	RE	GMM
(m, n) = (25, 10)					
0.1	0.009	0.002	0.037	0.044	0.014
	(0.193)	(0.235)	(0.261)	(0.177)	(0.178)
0.5	0.000	0.000	-0.001	0.000	0.000
	(0.182)	(0.224)	(0.172)	(0.168)	(0.143)
0.9	-0.010	-0.003	-0.039	-0.045	-0.015
	(0.195)	(0.235)	(0.259)	(0.181)	(0.180)
(m, n) = (200, 10)					
0.1	0.011	0.005	0.040	0.046	0.019
	(0.068)	(0.080)	(0.092)	(0.067)	(0.061)
0.5	0.001	0.001	0.001	0.001	0.001
	(0.063)	(0.076)	(0.059)	(0.063)	(0.047)
0.9	-0.010	-0.003	-0.040	-0.045	-0.018
	(0.067)	(0.080)	(0.091)	(0.068)	(0.060)
(m, n) = (25, 25)					
0.1	0.003	0.000	0.015	0.016	0.008
	(0.175)	(0.222)	(0.141)	(0.120)	(0.124)
0.5	-0.003	-0.004	0.000	-0.002	-0.002
	(0.171)	(0.218)	(0.102)	(0.106)	(0.099)
0.9	-0.009	-0.007	-0.017	-0.018	-0.013
	(0.177)	(0.223)	(0.138)	(0.120)	(0.124)
(m, n) = (200, 25)					
0.1	0.006	0.004	0.015	0.017	0.011
	(0.061)	(0.075)	(0.049)	(0.042)	(0.041)
0.5	0.000	0.000	0.000	0.000	0.000
	(0.059)	(0.073)	(0.036)	(0.036)	(0.032)
0.9	-0.006	-0.004	-0.015	-0.017	-0.012
	(0.061)	(0.075)	(0.049)	(0.042)	(0.041)
(m, n) = (25, 200)					
0.1	0.001	0.002	0.002	0.002	0.002
	(0.163)	(0.211)	(0.049)	(0.047)	(0.056)
0.5	0.001	0.001	0.000	0.000	0.001
	(0.163)	(0.210)	(0.035)	(0.035)	(0.045)
0.9	0.000	0.001	-0.002	-0.002	-0.002
	(0.163)	(0.211)	(0.049)	(0.046)	(0.056)
(m, n) = (200, 200)					
0.1	0.000	0.000	0.002	0.002	0.002
	(0.058)	(0.073)	(0.017)	(0.016)	(0.017)
0.5	0.000	0.000	0.000	0.000	0.000
	(0.058)	(0.072)	(0.013)	(0.012)	(0.012)
0.9	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.058)	(0.073)	(0.017)	(0.017)	(0.017)

Note:
The table reports bias and standard deviation (in parentheses) of the simulations for $\beta(\tau)$ from 10,000 Monte Carlo Simulations.

Table 1: Bias and Standard Deviation of $\hat{\beta}(\tau)$

Quantile	Pooled	BE	FE	RE	GMM
		(m, n) = (25, 10)			
0.1	0.948	0.922	0.946	0.914	0.918
0.5	0.946	0.920	0.948	0.914	0.919
0.9	0.950	0.924	0.950	0.912	0.916
		(m, n) = (200, 10)			
0.1	0.942	0.941	0.927	0.874	0.933
0.5	0.947	0.943	0.952	0.943	0.948
0.9	0.946	0.942	0.932	0.877	0.935
		(m, n) = (25, 25)			
0.1	0.949	0.921	0.949	0.933	0.923
0.5	0.946	0.918	0.948	0.931	0.923
0.9	0.946	0.917	0.951	0.931	0.921
		(m, n) = (200, 25)			
0.1	0.947	0.945	0.942	0.928	0.940
0.5	0.950	0.945	0.954	0.948	0.948
0.9	0.948	0.946	0.938	0.927	0.938
		(m, n) = (25, 200)			
0.1	0.950	0.923	0.950	0.947	0.920
0.5	0.949	0.926	0.952	0.948	0.916
0.9	0.947	0.925	0.950	0.948	0.918
		(m, n) = (200, 200)			
0.1	0.948	0.943	0.947	0.950	0.946
0.5	0.947	0.943	0.951	0.950	0.950
0.9	0.948	0.944	0.949	0.949	0.945

Note:
Results based on 10,000 Monte Carlo simulations. The table reports the coverage probabilities of the confidence intervals of $\beta(\tau)$.

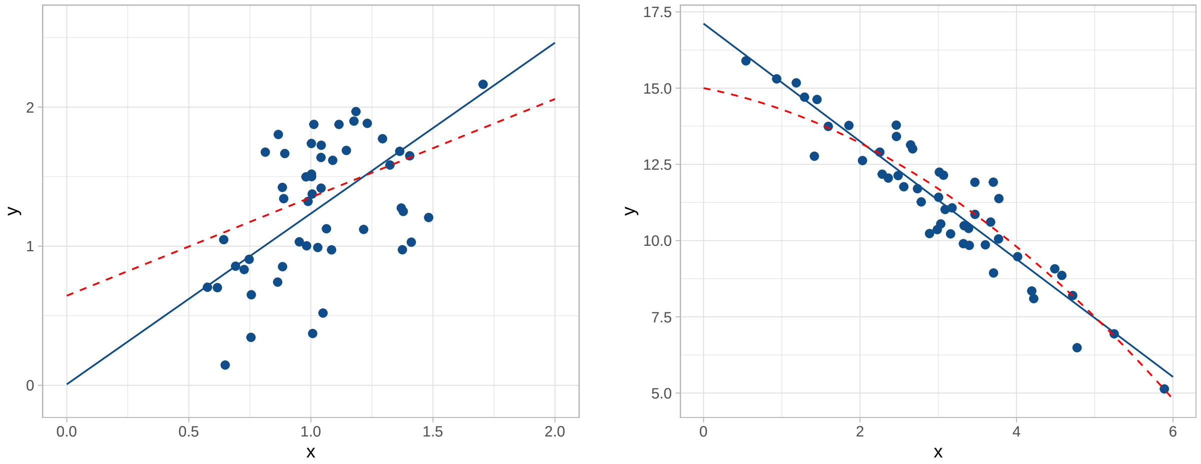
Table 2: Properties of the 95% Confidence Invervals

	λ				
Quantile	0.0	0.1	0.2	0.3	0.4
(m, n) = (25, 10)					
0.1	0.052	0.058	0.077	0.118	0.181
0.5	0.057	0.073	0.117	0.195	0.306
0.9	0.050	0.067	0.095	0.147	0.224
(m, n) = (200, 10)					
0.1	0.062	0.085	0.276	0.578	0.844
0.5	0.050	0.177	0.533	0.872	0.987
0.9	0.058	0.193	0.483	0.782	0.949
(m, n) = (25, 25)					
0.1	0.060	0.075	0.121	0.209	0.342
0.5	0.064	0.087	0.152	0.269	0.430
0.9	0.059	0.081	0.140	0.231	0.363
(m, n) = (200, 25)					
0.1	0.051	0.167	0.555	0.898	0.994
0.5	0.051	0.232	0.691	0.963	0.999
0.9	0.049	0.231	0.646	0.938	0.997
(m, n) = (25, 200)					
0.1	0.086	0.119	0.212	0.366	0.567
0.5	0.101	0.138	0.248	0.417	0.615
0.9	0.085	0.118	0.218	0.374	0.570
(m, n) = (200, 200)					
0.1	0.054	0.262	0.773	0.986	1.000
0.5	0.055	0.276	0.792	0.989	1.000
0.9	0.053	0.273	0.787	0.987	1.000

Note:
The table reports rejection probabilities of the Hausman test. The results are based on 10,000 Monte Carlo simulations. The first column, shows the empirical size, while the other columns show the power of the test.

Table 3: Hausman Test

In this case, higher rejection probabilities suggest a more powerful test. The first column shows that the empirical sizes of the test are close to the theoretical levels in most sample sizes. The power of the test is higher in large samples and increases the larger the correlation between $\bar{x}_{1j}$ and the unobserved heterogeneity α_j. An increase in m substantially improves the power of the test, while a larger number of time periods n improves the results to a lesser extent. In general, the test performs better both in terms of size and power when m is large, which is most often the case in empirical applications. Although as n increases the random effects estimator converges to the fixed effects estimator, and the random effects estimator of β will be consistent even if $\lambda \neq 0$, the size and power of the test do not deteriorate. This result is consistent with the findings in Ahn and Moon (2014).



(a) Extrapolation

(b) Misspecification

Figure 1: First Stage Regressions

The figure illustrates the venerability of the intercept to extrapolation and misspecification. Both panels show generated data for one group and a first-stage fit. Panel (a) uses the same DGP as in the simulation of CLP. The solid line shows the regression line estimated by median regression, and the dashed red line is the true regression line. Panel (b) uses a different DGP where $y = 15 - 0.5x - 0.2x^2 + u$, where $u \sim N(0,1), x \sim N(3,1)$. The solid line is estimated by least squares regression without the quadratic term (misspecified). The dashed red line is the true regression line.

5 Grouped (IV) Quantile Regression Model

In this section, we discuss the group (IV) quantile regression model and compare our estimator with the estimator suggested by Chetverikov et al. (2016).

5.1 Chetverikov et al. (2016)

Chetverikov et al. (2016) (CLP) suggest an estimator for the group (IV) quantile regression model. They consider two variations of the model. The first one is identical to model (1). The second model assumes that

$$\begin{aligned} Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j) =& \beta_{0,j}(\tau) + x_{1ij}'\beta_j(\tau) && (25)\\ \beta_{0,j}(\tau) =& x_{2j}'\gamma(\tau) + \alpha(\tau, v_j). && (26)\end{aligned}$$

Compared to model (1), the coefficient on x_{1ij} is allowed to vary over j, and the between variation in x_{2j} might be endogenous. They suggest a two-step estimator. The first stage consists of regressing y_{ij} on x_{1ij} and a constant using quantile regression separately for each group j and quantile τ. In the second stage, they regress the *intercept* from the first stage on x_{2j}. Their estimator focuses on estimating $\gamma(\tau)$ and does not directly estimate $\beta(\tau)$.

The main difference between the two estimators is that CLP use the estimated intercept of the first stage (i.e., fitted values evaluated at $x_{1ij} = 0$) as a dependent variable in the second stage. Thus, the CLP estimator is consistent for the quantile treatment effect at $x_{1ij} = 0$. We discuss this parameter in more detail below. A linear reparametrization of the individual-level variable changes the intercept so that the CLP estimator is not invariant to linear transformations of the individual-level covariates.

Further, the CLP estimator might have lower precision than the MD estimator. (i) It only includes the individual level covariates x_{1ij} in the first stage. Thus, it does not exploit (potentially exogenous) variation in the individual-level covariates between groups. (ii) It does not impose equality of the $\beta_j(\tau)$ in the first stage. Additionally, the estimator might extrapolate the intercept in the first stage, making it more vulnerable to misspecification and estimation error in finite samples.

Figure 1 illustrates these issues. The figure shows two groups from two different samples. Each Panel shows an estimated first-stage regression line (solid blue line) and the true regression line (dashed red line). Both panels show that if the support of the covariates does not cover zero in all groups, the intercept is extrapolated. As shown in Panel (a), a small estimation error in the slope parameter can lead to large estimation errors in the intercept. Further, the value of the (true) intercept and its estimation error change with the reparametrization of the individual-level covariates so that reparametrization leads to different results. If the first stage slopes are allowed to change over groups, reparametrization of the covariates changes the estimand. By imposing equality of the slopes across groups, our estimator becomes invariant to reparametrizations of the covariates. Similarly, when x_{1it} contains discrete variables, if some groups do not contain any observations in the base category or if some variables exhibit variation only within some group, the interpretation of the intercept changes over groups. Panel (b) shows how model misspecification in the first stage can lead to a large estimation error with the CLP estimator. The model is misspecified, as it fits a linear regression model instead of a quadratic one. The consequences of misspecification are substantially larger outside the support of the covariates, e.g., at $x_{1ij} = 0$. In comparison, the misspecification error in the fitted values is negligible.

Now we consider both estimators in the context of models (1) and model (25)-(26). If model (1) is correct, our estimator will, in general, outperform the CLP estimator for the reasons listed above. Further, imposing equality also makes the estimator invariant to reparametrizations of the individual-level variables. For this model, the CLP estimator is consistent for the treatment effect at $x_{1ij} = 0$, which equals the quantile treatment effects. We obtain a more precise estimator by estimating all the parameters simultaneously and imposing all the assumptions. On the other hand, if the model (25)-(26) is correct, we should not exploit the between-variation in the individual-level variables and use the demeaned individual-level regressors as instruments. If the slopes are systematically correlated with the treatment variable, the treatment effect is heterogeneous, and CLP estimates the quantile treatment effects at $x_{1ij} = 0$, which may not be particularly interesting. In such a case, one could parametrize the treatment effect on the random slope and estimate the effect on the intercept and the effect on the slope separately. In a second step, combining both estimates gives, for instance, an average (in x_{1ij}) QTE. Using our approach, we estimate the best linear approximation of the treatment effect, and we can also allow for heterogeneous effects by including interaction terms between x_{1ij} and x_{2j}.

We want to show that asymptotically our estimator has lower variance than the CLP estimator uniformly over different values of $\text{Var}(\bar{z}_j\alpha_j(\tau))$.[30] From section 3 we know that the variance comprises two terms, one that accounts for first-stage error and the other accounts for the second-stage noise. The sum of these two terms determines the asymptotic behavior of the estimator. Thus, to show that our MD estimator is more precise, it suffices to show that both components of the variance are smaller. If $\text{Var}(\bar{z}_j\alpha_j(\tau)) = 0$, there is no second stage noise, both estimators converge at the $\sqrt{mn}$ rate, and only the variance coming from the first stage matters. On the other hand, if $\text{Var}(\bar{z}_j\alpha_j(\tau)) > \varepsilon > 0$, both estimators converge at the $\sqrt{m}$ rate, and the variance coming from the first stage does not enter the first-order asymptotic distribution. Thus, in the latter case, we will consider the estimators as if we knew the true first stage.

In the following, we assume that the more widely used model (1) is correct and focus on a case with exogenous regressors. We consider our MD estimator implemented using optimal instruments, as this estimator simultaneously minimizes both components of the variance. To apply optimal instruments, we need to impose the stronger assumption that $\mathbb{E}[\alpha_j(\tau)|X_j] = 0$. Below, we provide some results without this assumption.

First, we consider the variance arising from the first stage, which shows up in the first-order asymptotic distribution of $\hat{\delta}_1(\tau)$. To study this part of the variance, we can assume, without loss of generality, that $\bar{g}^{(2)}_{mn}(\hat{\delta}, \tau) = 0$. The optimal instrument is $Z^*_j = \left(\tilde{X}_j V_j(\tau) \tilde{X}'_j\right)^+ X_j$, which yields an estimator that is algebraically identical to the efficient minimum distance estimator (see Proposition 6). Remark 6 provides the minimum distance representation of our estimator. The CLP estimator can also be written as a minimum distance estimator that minimizes

$$\frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j(\tau) - \tilde{R}_j\delta(\tau)\right)'\left(\hat{\beta}_j(\tau) - \tilde{R}_j\delta(\tau)\right), \tag{27}$$

where

$$\underset{(K_1+1)\times(K_1\cdot m+K_2)}{\tilde{R}_j} = \begin{pmatrix} x'_{2j} & 0 \\ 0 & l'_j \otimes I_{K_1} \end{pmatrix},$$

and l_j is a m-dimensional vector of zeros with a 1 in the j position. The restriction matrix $\tilde{R}_j$ is different from the restriction matrix of our estimator, as it does not impose equality of the first stage coefficients implied by the model. Further, CLP use an identity weighting matrix so that their estimator is inefficient relative to an efficient MD estimator with restriction matrix $\tilde{R}_j$. Since our estimator imposes the additional (correct) restriction, our efficient MD estimator has a smaller variance than any alternative (efficient) MD estimator with restriction matrix $\tilde{R}_j$, including the CLP estimator. In the special case of quantile independence, the weighting matrix of the efficient MD estimator reduces to $\hat{W}_j = \tilde{X}'_j\tilde{X}_j$, which corresponds to using OLS in the

[30] The asymptotic results in Chetverikov et al. (2016) implicitly assume that $\text{Var}(\bar{z}_j\alpha_j(\tau)) > \varepsilon > 0$ so that their estimator converges at the $\sqrt{m}$ rate.

second stage. In this case, our estimator with a least squares second stage is efficient and will have a lower variance than the CLP estimator.

Next, we focus on the component of the variance coming from the second-stage error. For this term, we can assume that we know the true first stage. We start by noting that we numerically obtain the CLP estimator by regressing the first stage fitted values on x_{2j}, $x_{1ij} \cdot d_1, \ldots, x_{1ij} \cdot d_m$ with instruments x_{2j}, $\dot{x}_{1ij} \cdot d_1, \ldots, \dot{x}_{1ij} \cdot d_m$ where d_j is a group indicator. In the special case where we know the true first stage, we can recover the CLP point estimates if we regress the fitted values on x_{1ij} and x_{2j} with instruments $\dot{x}_{1ij}$ and x_{2ij} without the interactions. From this representation, it is clear that the CLP estimator only exploits the within variation of x_{1ij}. Differently, our estimator uses the entire variation of x_{1ij} efficiently. If we know the true first stage, the optimal instrument implied by the conditional moment restriction is $Z_j^* = \mathbb{E}[\alpha_j(\tau)^2|X_j]^{-1}X_j$, which implies that our second stage is a GLS regression which is efficient.

One backdrop of this analysis is that it relies on the stronger conditional moment restriction $\mathbb{E}[\alpha_j(\tau)|X_j] = 0$. Nonetheless, we can show that regardless the value of $\text{Var}(\bar{z}_j\alpha_j(\tau))$, we can implement an estimator that is more precise than the CLP estimator. More precisely, if $\text{Var}(\bar{z}_j\alpha_j(\tau)) = 0$ for all j, the efficient minimum distance is optimal. Differently, if $\text{Var}(\bar{z}_j\alpha_j(\tau)) > \varepsilon > 0$ using an efficient GMM estimator with instruments $\dot{x}_{1ij}, \bar{x}_{1j}, x_{2j}$ yields more precise point estimates as it exploits all moment conditions efficiently. This GMM estimator exploits the between variation of x_{1ij} by including $\bar{x}_{ij}$ in the instrument set. By adding an instrument, asymptotically, our estimator will have a weakly lower variance (see Proposition 4.51 in White, 2001).

5.2 Simulations

This subsection presents Monte Carlo simulations comparing the MD and CLP estimators. The simulations are based on the same data generating process and sample sizes as in Chetverikov et al. (2016). That is, $(m, n) = \{(25, 25), (200, 25), (25, 200), (200, 200)\}$. For both estimator we use a OLS (or 2SLS) second stage. The generated data include one individual-level regressor, one group-level regressor, and one instrument. Heterogeneity is introduced via a rank variable u_{ij}. Since the effect of the individual-level covariates is constant across groups, $\beta_j(\tau) = (\beta_{j,0}(\tau), \beta_j(\tau)')' = (\beta_{j,0}(\tau), \beta(\tau)')'$, where $\beta_{i,0}(\tau)$ is the constant of the first stage. The data is generated as follows:

$$y_{ij} = \beta_0(u_{ij}) + x_{1ij}\beta(u_{ij}) + x_{2j}\gamma(u_{ij}) + \alpha_j(u_{ij}), \tag{28}$$

$$z_j = x_{2j} + \eta_j + \nu_j, \tag{29}$$

$$\alpha_j(u_{ij}) = u_{ij}\eta_j - \frac{u_{ij}}{2}, \tag{30}$$

where x_{1ij}, x_{2j} and ν_j are distributed $\exp(0.25 \cdot N[0,1])$ and η_j as well as the rank variable u_{ij} are $U[0,1]$ distributed. The data generating process implies that $\mathbb{E}[\alpha(u_{ij})|x_{2j}] = \mathbb{E}[u_{ij}\eta_j - \frac{u_{ij}}{2}|x_{2j}] =$

$\mathbb{E}[\frac{u_{ij}}{2} - \frac{u_{ij}}{2}|x_{2j}] = 0$. At quantiles $\tau \in (0, 1)$, the true parameters $\gamma(\tau)$ and $\beta(\tau)$ equal $\sqrt{\tau}$ and, $\alpha_1(\tau) = \frac{\tau}{2}$. Consequently, $\gamma(u_{ij}) = \beta(u_{ij}) = \sqrt{u_{ij}}$ and $\beta_0(u_{ij}) = \frac{u_{ij}}{2}$.

The simulations consider three cases. In the first one (baseline), $\alpha_j(\tau) = 0$ for all j and all τ. In this case, conditioning on the individual does not affect the quantile function and quantile regression is consistent for the same parameter. Further, as $\alpha_j(\tau) = 0$, the estimators with a least square second stage are $\sqrt{mn}$-consistent. In the second case, there are group-specific effects ($\alpha_j(\tau) \neq 0$), which are uncorrelated with the regressors. The individual heterogeneity is multiplied with the rank variable u_{ij}. Thus, in the lower tail of the distribution, we see a faster convergence rate. In the third case, $\alpha_j(\tau)$ is correlated with the regressor of interest, such that x_{2j} is endogenous. In this case, we use 2SLS in the second stage. Since the data generating process of Chetverikov et al. (2016) has a weak instrument when m is small, one should pay

	Baseline			Exogenous			Endogenous		
Quantile	MD	CLP	Rel. MSE	MD	CLP	Rel. MSE	MD	CLP	Rel. MSE
				(m, n) = (25, 25)					
0.1	0.022	-0.011	0.051	0.022	-0.010	0.052	0.049	0.001	0.404
	(0.192)	(0.858)		(0.195)	(0.860)		(3.218)	(5.062)	
0.5	-0.010	-0.001	0.061	-0.011	0.000	0.088	-0.017	0.039	0.318
	(0.166)	(0.673)		(0.204)	(0.691)		(3.098)	(5.491)	
0.9	-0.019	-0.003	0.049	-0.020	-0.004	0.216	-0.052	-0.011	0.409
	(0.094)	(0.435)		(0.227)	(0.490)		(3.239)	(5.065)	
				(m, n) = (200, 25)					
0.1	0.024	0.003	0.060	0.024	0.004	0.063	0.023	0.006	0.057
	(0.066)	(0.284)		(0.067)	(0.285)		(0.106)	(0.456)	
0.5	-0.006	-0.001	0.059	-0.006	0.000	0.086	-0.009	-0.003	0.071
	(0.056)	(0.232)		(0.069)	(0.238)		(0.097)	(0.366)	
0.9	-0.017	-0.004	0.060	-0.017	-0.003	0.223	-0.022	-0.009	0.142
	(0.031)	(0.145)		(0.075)	(0.164)		(0.086)	(0.234)	
				(m, n) = (25, 200)					
0.1	0.003	-0.002	0.059	0.003	-0.001	0.066	-0.027	-0.076	0.130
	(0.070)	(0.289)		(0.074)	(0.291)		(2.025)	(5.618)	
0.5	-0.001	-0.002	0.060	-0.001	-0.001	0.233	-0.082	-0.094	0.580
	(0.060)	(0.247)		(0.134)	(0.278)		(3.485)	(4.575)	
0.9	-0.002	0.000	0.061	-0.001	0.001	0.769	-0.118	-0.114	1.126
	(0.030)	(0.121)		(0.217)	(0.247)		(3.780)	(3.561)	
				(m, n) = (200, 200)					
0.1	0.003	-0.003	0.057	0.003	-0.003	0.062	0.002	-0.004	0.058
	(0.024)	(0.100)		(0.025)	(0.101)		(0.039)	(0.162)	
0.5	-0.001	0.000	0.059	-0.001	-0.001	0.222	-0.004	-0.004	0.141
	(0.020)	(0.084)		(0.044)	(0.093)		(0.051)	(0.136)	
0.9	-0.002	0.000	0.067	-0.003	-0.001	0.762	-0.009	-0.007	0.617
	(0.010)	(0.040)		(0.071)	(0.082)		(0.074)	(0.095)	

Note:
The table reports mean bias, standard deviation and relative MSE from the simulations for $\gamma(\tau)$ from 10000 Monte Carlo simulations using the MD estimator and the CLP estimator. The relative MSE gives the MSE of the MD estimator relative to that of the CLP estimator.

Table 4: Bias, Standard Deviation and MSE of $\hat{\gamma}(\tau)$

	Baseline			Exogenous			Endogenous		
	Rel. length	Coverage Rate		Rel. length	Coverage Rate		Rel. length	Coverage Rate	
Quantile	MD/CLP	MD	CLP	MD/CLP	MD	CLP	MD/CLP	MD	CLP
				(m, n) = (25, 25)					
0.1	0.232	0.941	0.938	0.235	0.939	0.938	0.227	0.966	0.972
0.5	0.244	0.940	0.942	0.301	0.942	0.945	0.262	0.964	0.972
0.9	0.223	0.942	0.949	0.501	0.940	0.946	0.373	0.957	0.972
				(m, n) = (200, 25)					
0.1	0.230	0.932	0.947	0.233	0.932	0.948	0.230	0.942	0.953
0.5	0.245	0.947	0.944	0.296	0.945	0.946	0.267	0.952	0.949
0.9	0.220	0.925	0.947	0.475	0.941	0.945	0.368	0.953	0.952
				(m, n) = (25, 200)					
0.1	0.241	0.943	0.940	0.256	0.943	0.943	0.240	0.968	0.974
0.5	0.242	0.937	0.944	0.496	0.938	0.944	0.370	0.949	0.971
0.9	0.248	0.948	0.941	0.884	0.934	0.945	0.771	0.939	0.955
				(m, n) = (200, 200)					
0.1	0.241	0.944	0.944	0.254	0.947	0.945	0.246	0.951	0.950
0.5	0.244	0.946	0.945	0.483	0.952	0.948	0.377	0.957	0.951
0.9	0.246	0.942	0.953	0.872	0.950	0.950	0.772	0.954	0.955

Note:
Results based on 10000 Monte Carlo simulations. The table provides the coverage rate and median length of the confidence intervals of $\gamma(\tau)$. The relative length provides the length of the confidence interval of the MD estimator relative to that of the CLP estimator. Robust standard errors are used for the CLP estimator, and clustered standard errors at the group level are used for the MD estimator.

Table 5: Properties of the 95% Confidence Intervals

attention when looking at the simulation results for the endogenous case.[31] In empirical research, it is straightforward to construct confidence intervals that remain valid even if identification is weak. We perform 10,000 Monte Carlo replications for the set of quantiles $\tau \in \{0.1, 0.5, 0.9\}$. Since the CLP estimator does not directly provide an estimate for $\beta(\tau)$, we present only results for $\gamma(\tau)$.

Table 4 reports the bias, standard deviation and relative MSE of the estimators. The relative MSE reports the MSE of the MD estimator relative to that of the CLP estimator. Thus, a number smaller than 1 indicates that the MD estimator has a lower MSE. The CLP estimator seems to have a smaller bias than the MD estimator when $n = 25$. When n increases to 200, the difference disappears. There are more remarkable differences in the variance of the estimators. The standard deviation of the MD estimator is four times smaller compared to that of the CLP estimator in the baseline case. The difference is somewhat smaller in the exogenous and endogenous cases but remains substantial.[32] This difference in precision explains the large discrepancies in MSE. The MSE of the CLP estimator is over 10 times larger than that of the MD estimator when $\alpha_j(\tau) = 0$ and remains substantially larger in all scenarios considered. If

[31] With $m = 25$ in over 40% of the draws, the F-statistics of the first stage of the 2SLS estimations is below 10. The issue disappears when $m = 200$.

[32] The standard deviations in the endogenous case with $m = 25$ should be interpreted with caution due to the weak instrument.

$\alpha_j(\tau) = 0$, quantile regression is a consistent estimator for $\beta(\tau)$. Although not shown here, simulation results comparing our estimator with traditional quantile regression show that the two estimators are indistinguishable in terms of bias and variance in large samples.

Table 5 show the performance of the 95% confidence intervals suggested with our inference procedure. The table reports the coverage rate of the confidence intervals and the median length of the confidence interval of our estimator relative to that of the CLP estimator. Our suggested inference procedure has coverage close to 95% in all cases. Compared to the CLP estimator, our confidence bands are substantially shorter. In most cases, our estimator yields confidence bands less than half the length of those for the CLP estimator.

6 Empirical Application: The Effect of the Food Stamps Program on Birth Weight

In this section, we apply our minimum distance approach to estimate the impact of the food stamp program on the birth weight distribution using grouped data. We complement the analysis of Almond et al. (2011) by providing distributional effects. Food stamps constitute an important means-tested program that gives entitled households coupons they can redeem at approved retail food stores. The Food Stamp Act (FSA) was introduced in 1964 and enabled counties to start their own federally funded food stamp program (FSP). In the subsequent years, counties increasingly adopted such programs, and in 1973, an amendment to the FSA required all counties to establish a FSP by 1975. Thus, the share of counties with an FSP increased steadily from 1964 to 1974, and identification exploits the variation in the timing of the adoption across counties. Almond et al. (2011) use data from 1968 (when about 40% of the counties had introduced the program) to 1977 (two years after the FSP was implemented everywhere) to analyze the effect of the program.

Given the negative consequences of low birth weight, besides estimating the effect of the policy on average weight, Almond et al. (2011) estimate the effect on the probability that birth weight falls below a certain threshold. As discussed in Melly and Santangelo (2015) this procedure leads to biased results unless there is no time effect or group effect, or the outcome is uniformly distributed.

In this section, we use the subscripts i, c, and t to denote the birth, the county, and the trimester of birth, respectively.[33] The variable of interest is a binary variable that is coded 1 if there was a food stamp program in place three months before birth. The treatment is assigned to county-month cells, and in around 1% of cases, it also varies within groups.

We consider the following model separately for blacks and whites:

$$Q(\tau, bw_{ict}|fsp_{ct}, x_{1ict}, x_{2ct}, v_{ct}) = fsp_{ct}\gamma_1(\tau) + x_{1ict}\beta(\tau) + x_{2ct}\gamma_2(\tau) + \alpha(\tau, v_{ct}), \qquad (31)$$

[33] Using the same notation as in the paper, the j units are county-trimester combinations, and the i units index individual births within a county in a given trimester. In this section, we use three subscripts for clarity.

where $Q(\tau, bw_{ict}|fsp_{ct}, x_{1ict}, x_{2ct}, v_{ct})$ is the conditional quantile function of the outcome given all the variables. fsp_{ct} is a variable indicating whether there is a food stamp program in place, x_{1ict} are variables related to the individual births, such as gender, mother age, and its square as well as the legitimacy status of the birth. Group level control variables x_{2ct} include annual county-level controls (real per capita income, government transfers to individuals, medical spending, and retirement and disability payments) and 1960 county-level characteristics (county population and the shares of urban population, black population, and of farmland) interacted with a linear time trend. Further, x_{2j} also includes county, state-year fixed effects, and time fixed effects.

Figure 2 illustrates the results. For the estimation, we drop groups that have less than 20 degrees of freedom.[34] The estimations are performed using a sample of 2,822,091 individual observations divided into 19,482 groups for blacks and 16,038,235 individual births divided into 80,289 groups in the sample of whites.[35] The results for black are in the left panel, while the results for whites are in the right panel. As shown, the effect is substantially larger among blacks. The results suggest a positive effect of the food stamp program on the lower tail of the conditional distribution. The estimates suggest that the food stamp program is associated with an increase in birth weight by almost 30 grams for blacks at the 5th percentile of the conditional distribution. For whites, there seems to be an effect only at the left tail of the distribution, and the effects are small. For blacks, the coefficients are large in the left tail and remain positive, albeit of small magnitude, until the 75% percentile. However, for higher quantiles, the effects are not statistically different from zero.

7 Conclusion

This paper suggests a MD estimator for quantile panel data models. The estimator is of practical relevance with classical panel data settings where the units are observed over time and with grouped data, where individuals are divided into groups, and the treatment varies at the group level. The coefficient on the individual-level and group-level variables can be estimated. The estimator is computationally fast and straightforward to compute and consists of a first stage individual level quantile regressions, followed by a GMM regression with the fitted values as the dependent variable. We show that our two-step procedure applied to linear estimators is algebraically identical to traditional one-step estimators. We suggest a quantile counterpart to traditional panel data estimators, including the pooled, the fixed effects, and the random effects estimators. In the second stage, both internal and external instruments can be

[34] Essentially, if there are K_1 individual level variables we drop all groups with less than K_1+1+20 observations. Since some variables might vary at the individual level in some groups only, this threshold is group specific.

[35] We have a different number of groups compared to Almond et al. (2011) due to multiple reasons. First, they give higher weights to births in groups where only 50% of the births are included in the natality data; thus, when they drop groups with less than 25 births, the number of births in these groups is inflated. Second, since they take the group average, they keep births with missing values for birth weight. We drop those births as we work with individual-level data.

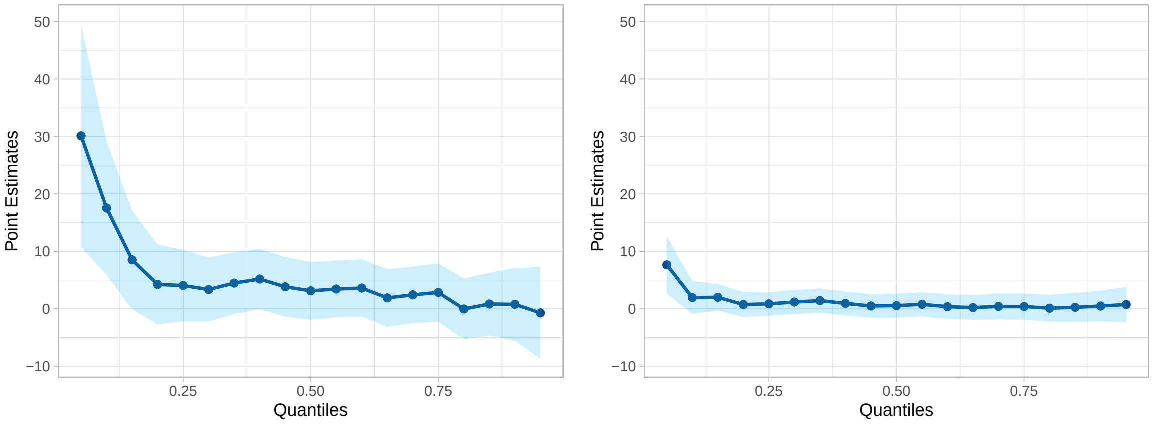


Figure 2: Impact of Food Stamp Introduction on the Distribution of Birth Weight

The figure shows the impact of the food stamp introduction on the conditional distribution of birth weight. The panels show point estimates and 95% confidence bands (shaded area) computed using standard errors clustered at the county level. The panel on the left (right) shows the effects for blacks (whites). The regressions include county, time, and state-year fixed effects.

used in a Hausman and Taylor or traditional instrumental variables framework. Further, an overidentification test can be implemented if the model is overidentified.

The asymptotic distribution of the estimator is non-standard because the speed of convergence is not the same for all coefficients. The speed of convergence depends on the moment conditions that are used to identify a parameter. For the coefficient converging at the faster ($\sqrt{mn}$) rate, only the variance coming from the first stage enters the first-order asymptotic distribution. On the other hand, since n diverges to infinity, the first stage variance does not appear in the first-order asymptotic variance of the coefficients converging at the slower ($\sqrt{m}$) rate. In other words, the first-order asymptotic distribution is the same as if we knew the true first stage. We suggest an inference procedure that is uniformly valid regardless of the convergence rate of the estimator and, importantly, takes the first stage variance into account, thus providing better inference. Monte Carlo simulations show that our estimator and the suggested standard errors perform well in finite samples. Compared to the grouped estimator of Chetverikov et al. (2016), the MD estimator has a much smaller MSE. Finally, in an empirical application, we study the effect of the food stamp program on birth weight, and we find positive effects for blacks in the lower tail of the conditional distribution.

References

Ahn, S. C. and S. Low (1996): “A reformulation of the Hausman test for regression models with pooled cross-section-time-series data,” *Journal of Econometrics*, 71, 309–319.

Ahn, S. C. and H. R. Moon (2014): “Large-N and Large-T Properties of Panel Data Estimators and the Hausman Test," in *Festschrift in Honor of Peter Schmidt*, Springer.

Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011): "Inside the war on poverty: The impact of food stamps on birth outcomes," *Review of Economics and Statistics*, 93, 387–403.

Alvarez, J. and M. Arellano (2003): "The time series and cross-section asymptotics of dynamic panel data estimators," *Econometrica*, 71, 1121–1159.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Angrist, J. D. and K. Lang (2004): "Does school integration generate peer effects? Evidence from Boston's metco program," *American Economic Review*, 94, 1613–1634.

Arellano, M. (1993): "On the testing of correlated effects with panel data," *Journal of Econometrics*, 59, 87–97.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 103, 2121–2168.

——— (2021): "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, 1, 161–178.

Autor, D. H., A. Manning, and C. L. Smith (2016): "The contribution of the minimum wage to US wage inequality over three decades: A reassessment," *American Economic Journal: Applied Economics*, 8, 58–99.

Baltagi, B. H. (2021): *Econometric Analysis of Panel Data*, Springer, 6 ed.

Brown, B. M. (1971): "Martingale central limit theorems," *The Annals of Mathematical Statistics*, 59–66.

Canay, I. A. (2011): "A Simple Approach to Quantile Regression for Panel Data," *Econometrics Journal*, 14, 368–386.

Chamberlain, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18, 5–46.

——— (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1994): "Quantile Regression, Censoring, and the Structure of Wages," *Advances in econometrics*, 1, 171–209.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405–1441.

Chetverikov, D., B. Larsen, and C. Palmer (2016): "IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade," *Econometrica*, 84, 809–833.

Dai, X. and L. Jin (2021): "Minimum distance quantile regression for spatial autoregressive panel data models with fixed effects," *PLoS ONE*, 16, 1–13.

Engbom, N. and C. Moser (2022): "Earnings Inequality and the Minimum Wage: Evidence from Brazil," *American Economic Review*, 112, 3803–3847.

Fernández-Val, I., W. Y. Gao, Y. Liao, and F. Vella (2022): "Dynamic Heterogeneous Distribution Regression Panel Models, with an Application to Labor Income Processes," 1–45.

Frölich, M. and B. Melly (2013): "Unconditional Quantile Treatment Effects Under Endogeneity," *Journal of Business and Economic Statistics*, 31, 346–357.

Galvao, A. and A. Poirier (2019): "Quantile Regression Random Effects," *Annals of Economics and Statistics*, 109–148.

Galvao, A. F., J. Gu, and S. Volgushev (2020): "On the unbiased asymptotic normality of quantile regression with fixed effects," *Journal of Econometrics*, 218, 178–215.

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

Galvao, A. F. and L. Wang (2015): "Efficient Minimum Distance Estimator for Quantile Regression Fixed Effects Panel Data," *Journal of Multivariate Analysis*, 133, 1–26.

Gu, J. and S. Volgushev (2019): "Panel Data Quantile Regression with Grouped Fixed Effects," *Journal of Econometrics*, 213, 68–91.

Hahn, J. and G. Kuersteiner (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large," *Econometrica*, 70, 1639–1657.

Hansen, B. E. (2022): *Econometrics*, Princeton University Press.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Harding, M., C. Lamarche, and M. H. Pesaran (2020): "Common correlated effects estimation of heterogeneous dynamic panel quantile regression models," *Journal of Applied Econometrics*, 35, 294–314.

Hausman, J. and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Hoderlein, S. and E. Mammen (2007): "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica*, 75, 1513–1518.

Im, K. S., S. C. Ahn, P. Schmidt, and J. M. Wooldridge (1999): "Efficient estimation of panel data models with strictly exogenous explanatory variables," *Journal of Econometrics*, 93, 177–201.

Kato, K., A. Galvao Jr., and G. V. Montes-Rojas (2012): "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics*, 170, 76–91.

Knight, K. (1998): "Limiting Distributions for L1 Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755–770.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33.

Koenker, R. and P. Ng (2005): "A frisch-newton algorithm for sparse quantile regression," *Acta Mathematicae Applicatae Sinica*, 21, 225–236.

Liao, Y. and X. Yang (2018): "Uniform Inference for Characteristic Effects of Large Continuous-Time Linear Models," 1–52.

Lu, X. and L. Su (2022): "Uniform inference in linear panel data models with two-dimensional heterogeneity," *Journal of Econometrics*.

Ma, L. and R. Koenker (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471–506.

Melly, B. and G. Santangelo (2015): "The changes-in-changes model with covariates," 1–32.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

Nerlove, M. (1971): "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, 39, 359.

Newey, W. K. (1993): "Efficient estimation of models with conditional moment restrictions," in *Handbook of Statistics*, Elsevier, vol. 11, chap. 16, 419–454.

Newey, W. K. and J. L. Powell (1990): "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

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

Powell, J. L. (1991): "Estimation of Monotonic Regression Models under Quantile Restriction," in *Nonparametric and Semiparametric Model in Economics*, Cambridge: Cambridge University Press.

Volgushev, S., S.-K. Chao, and G. Cheng (2019): "Distributed inference for quantile regression processes," *The Annals of Statistics*, 47, 1634–1662.

White, H. (2001): *Asymptotic theory for econometricians*, Academic press.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press.

——— (2019): "Correlated random effects models with unbalanced panels," *Journal of Econometrics*, 211, 137–150.

A Least Squares Panel Data Models

A.1 Formal results

This section complements subsection 2.3 by discussing more in detail the relationship between least squares estimator and the minimum distance approach. Throughout the section, we define the $n \times (K_1+1)$ matrix of first-stage regressors $\tilde{X}_{1j} = (\tilde{x}_j, \tilde{x}_{i2}, \ldots, \tilde{x}_{ij})'$, and the $mn \times K_1$ matrix of individual-level regressors $X_1 = (X'_{1j}, \ldots, X'_{1N})'$. Further, we use the matrices $P_j = l(l'l)^{-1}l'$ and $Q_j = I_j - P_j$, where l is a $n \times 1$ vector of ones. Thus, $P_j X_j = \bar{X}_j$ and $Q_j X_{1j} = \dot{X}_j$. We consider a linear version of our estimator, where OLS instead of quantile regression is used in the first stage. We consider model (9). In this section, we show that mean models can be estimated using a two-step procedure. Notation is the same as in the paper, except that the fitted values are computed using an OLS regression. More precisely, the vector of fitted values of group j is

$$\hat{Y}_j = \tilde{X}_j \hat{\beta}_j = \tilde{X}_j \left(\tilde{X}'_j \tilde{X}_j \right)^{-1} \tilde{X}'_j Y_j.$$

The following Proposition states the equivalence of the two-step procedure using the fitted values and the conventional one-step estimator in mean models.

Proposition 3. *Denote $\hat{\delta}^{MD}_{GMM}$ the coefficient vector of a linear GMM regression of $\hat{Y}$ on X with instrument Z. Let $\hat{\delta}_{GMM}$ be the coefficient vector of the same GMM regression but with regressand Y. If $\tilde{X}_j$ lies in the column space of Z_j, $\hat{\delta}^{MD}_{GMM} = \hat{\delta}_{GMM}$.*

The proof of this Proposition and all subsequent proofs are in Appendix A.2. Proposition 3 implies that any linear model can be computed by a two-step estimator as long as the matrix of instruments of group j, Z_j lies in the column space of the matrix of first-stage regressors of group j, $\tilde{X}_j$.[36] This result applies to a wide range of estimators. Since OLS is a special case of GMM, the result for pooled OLS follows directly, while the result for the within estimator is summarized in the following Corollary.

Corollary 1. *Denote $\hat{\delta}^{MD}_{FE}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on $\tilde{X}$ with instruments $\dot{X}_1$. Let $\hat{\delta}_{FE}$ be the coefficient vector of the within estimator, that is, of a regression of $\dot{Y}$ on $\dot{X}_1$. Then $\hat{\delta}^{MD}_{FE} = \hat{\delta}_{FE}$.*

The between estimator is usually computed by regressing $\bar{Y}$ on $\bar{X}$. Alternatively, it can be estimated by an IV regression of Y (or $\hat{Y}$) on X using $\bar{X}$ as an instrument, where it exploits only the variation between individuals.

Corollary 2. *Denote $\hat{\delta}^{MD}_{BE}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on X with instruments $\bar{X}$. Let $\hat{\delta}_{BE}$ be the coefficient vector of the between estimator, that is, of a regression of $\bar{Y}$ on $\bar{X}$. Then $\hat{\delta}^{MD}_{BE} = \hat{\delta}_{BE}$.*

[36] Since $\tilde{X}_j$ includes a constant, the presence of group-level variables in Z_j will not affect its column space.

It is worth noting that the IV approach to these panel data estimators also works in one stage with Y as the dependent variable. Further, it is possible to estimate between (within) models using average (demeaned) fitted values and regressors.
The pooled OLS and the between estimators can estimate both β and γ but are not efficient. The random effects estimator optimally combines between and the within variation to find a more efficient estimator. While FGLS is the most common estimator for the random effects model, Im et al. (1999) show that the overidentified 3SLS estimator, with instruments $Z_j = (\dot{X}_{1j}, \bar{X}_j)$, is identical to the random effects estimator. The 3SLS estimator is a special case of GMM with weighting matrix $W = \mathbb{E}[Z_j'\tilde{\Omega}Z_j]$ where $\tilde{\Omega}$ follows the usual random effects covariance structure. Thus, by Proposition 3, the random effects estimator can also be computed in two steps using the fitted values in the second stage.

Corollary 3. *Denote $\hat{\delta}_{RE}^{MD}$ the coefficient vector of a 3SLS regression of $\hat{Y}$ on X with instruments $(\dot{X}_{1j}, \bar{X}_j)$. Let $\hat{\delta}_{RE}$ be the coefficient vector of a random effects regression of Y on X. Then $\hat{\delta}_{RE}^{MD} = \hat{\delta}_{RE}$.*

Alternatively, the random effects estimator can be implemented using the theory of optimal instruments and a just identified 2SLS regression. Starting from a conditional moment restriction, the idea of optimal instruments is to select an instrument and weights that minimize the asymptotic variance (see, e.g. Newey, 1993). Relevant to our two-step procedure, under homoskedasticity of the errors, the conditional moments $\mathbb{E}[Y_j - X_j\delta|X_j] = 0$ and $\mathbb{E}[\hat{Y}_j - X_j\delta|X_j] = 0$ imply the same optimal instrument:

Proposition 4. *Assume $\mathbb{E}[\varepsilon_{ij}^2|X_j] = \sigma_\varepsilon^2$ and $\mathbb{E}[\alpha_j^2|X_j] = \sigma_\alpha^2$. The conditional moments $\mathbb{E}[\hat{Y}_j - X_j\delta|X_j] = 0$ and $\mathbb{E}[Y_j - X_j\delta|X_j] = 0$ imply the same optimal instrument.*

The Hausman-Taylor model (Hausman and Taylor, 1981) is a middle ground between the fixed effects and the random effects models where some regressors are assumed to be uncorrelated with α_j. In contrast, no restriction is placed on the relationship between the other regressors and the unobserved heterogeneity. The matrix of regressors X is partitioned as $X = [X_1^x \ X_1^n \ X_2^x \ X_2^n]$ where X_1^x and X_2^x are orthogonal to α_j. No assumption is placed on the relationship between α_j and X_1^n and X_2^n. The model can be estimated by IV using instruments $Z = (\dot{X}_1^x, \dot{X}_1^n, \bar{X}_1^x, X_2^x)$ (see, e.g., Hansen, 2022). Thus, it follows by Proposition 3 that the Hausman-Taylor model can be estimated in two stages.

Corollary 4. *Denote $\hat{\delta}_{HT}^{MD}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on X with instruments $(\dot{X}_1^x, \dot{X}_1^n, \bar{X}_1^x, X_2^x)$. Let $\hat{\delta}_{HT}$ be the coefficient vector of the Hausman Taylor Estimator based on a regression Y on X. Then $\hat{\delta}_{HT}^{MD} = \hat{\delta}_{HT}$.*

Finally, we show that not only the point estimates but also the standard errors can be obtained using the two-stage minimum distance approach. This requires clustering the standard errors in the second stage at a level weakly higher than the individual i. Let $g = 1, \ldots, G$ index the clusters and assume that each of the clusters has N_g observations. This nests the case where one wishes to cluster at the individual level or at a higher level. For example, if i are county-year combinations, one might cluster at the county level. For an estimator $\hat{\delta}$ the clustered covariance matrix is estimated by

$$\hat{V}_\delta = \left(\frac{1}{G} \sum_{g=1}^{G} X_g' Z_g \hat{W} \frac{1}{G} \sum_{g=1}^{G} Z_g' X_g \right)^{-1} \frac{1}{G} \sum_{g=1}^{G} X_g' Z_g \hat{W} \left(\frac{1}{G} \sum_{g=1}^{G} Z_g' \tilde{u}_g \tilde{u}_g' Z_g \right)$$
$$\cdot \hat{W} \frac{1}{G} \sum_{g=1}^{G} Z_g' X_g \left(\frac{1}{G} \sum_{g=1}^{G} X_g' Z_g \hat{W} \frac{1}{G} \sum_{g=1}^{G} Z_g' X_g \right)^{-1},$$

where $\tilde{u}_g$ is a N_g-dimensional vector of estimated errors for the observations in cluster g.

Proposition 5. *Denote $\hat{V}_\delta$ the clustered covariance matrix of $\hat{\delta}$ estimated by a GMM regression of Y on X with instrument Z. Let $\hat{V}_{\delta^{MD}}$ be the clustered covariance matrix of $\hat{\delta}^{MD}$ estimated by GMM regression of $\hat{Y}$ on X with instrument Z, where $\hat{Y}$ are estimated by an OLS first-stage. Let the clusters be at weakly higher level than i. Then, $\hat{V}_{\delta^{MD}} = \hat{V}_\delta$.*

A.2 Proofs of the least squares results

Proof of Proposition 3. Define the projection matrix $\tilde{P} = \tilde{X}_j(\tilde{X}_j'\tilde{X}_j)^{-1}\tilde{X}_j'$. Since Z_j is in the column space of $\tilde{X}_j$,

$$\tilde{P} Z_j = Z_j \tag{32}$$

The MD estimator with a GMM second stage is:

$$\hat{\delta}^{MD}_{GMM} = \left(X' Z \hat{W} Z' X \right)^{-1} X' Z \hat{W} Z' \hat{Y}.$$

For $\hat{\delta}^{MD}_{GMM}$ to be equal to $\hat{\delta}_{GMM}$, it suffices that $Z'\hat{Y} = Z'Y$. Note that

$$\begin{aligned} Z'\hat{Y} &= \sum_{j=1}^{m} Z_j' \hat{Y}_j \\ &= \sum_{j=1}^{m} Z_j' \tilde{X}_j \hat{\beta}_j \\ &= \sum_{j=1}^{m} Z_j' \tilde{X}_j (\tilde{X}_j' \tilde{X}_j)^{-1} \tilde{X}_j' Y_j \\ &= \sum_{j=1}^{m} (\tilde{P} Z_j)' Y_j \\ &= \sum_{j=1}^{m} Z_j' Y_j = Z'Y, \end{aligned}$$

where the third line uses $\hat{Y}_j = \tilde{X}_j \hat{\beta}_j$, the fourth line uses the definition of the OLS estimator in the first stage and the last line uses equation (32). Thus, it follows directly that $\hat{\delta}_{MD}$ equals $\hat{\delta}_{GMM}$. ■

Proof of Corollary 1. First, note that since $Q_jX_{1j} = \dot{X}_{1j}$, $\dot{X}_{1j}$ lies in the column space of X_{1j}. Then, we apply Proposition 3 and since $K = L$, the 2SLS estimator reduces to the IV estimator. It follows that a 2SLS (or IV) regression of $\hat{Y}$ on X_{1j} with instrument Z_j is algebraically identical to a 2SLS regression with Y_j as dependent variable. Then,

$$\begin{aligned}
\hat{\delta}^{MD}_{FE} &= \left(\sum_{j=1}^{m} Z_j' X_{1j}\right)^{-1} \sum_{j=1}^{m} Z_j' Y_j \\
&= \left(\sum_{j=1}^{m} \dot{X}_{1j}' X_{1j}\right)^{-1} \sum_{j=1}^{m} \dot{X}_{1j}' Y_j \\
&= \left(\sum_{j=1}^{m} X_{1j}' Q_j X_{1j}\right)^{-1} \sum_{j=1}^{m} X_{1j}' Q_j Y_j \\
&= \left(\sum_{j=1}^{m} \dot{X}_{1j}' \dot{X}_{1j}\right)^{-1} \sum_{j=1}^{m} \dot{X}_{1j}' \dot{Y}_j = \hat{\delta}_{FE},
\end{aligned}$$

where the second line follows since $Z_j = \dot{X}_{1j}$, the third and last line by $Q_jX_{1j} = \dot{X}_{1j}$, $Q_jY_j = \dot{Y}_j$ and since Q_j is idempotent. ■

Proof of Corollary 2. First, note that since $P_j\tilde{X}_j = \bar{X}_j$, $\bar{X}_j$ lies in the column space of $\tilde{X}_j$. Then, we apply Proposition 3 and since $K = L$, the 2SLS estimator reduces to an IV estimator. It follows that a 2SLS regression of $\hat{Y}_j$ on X_j with instrument Z_j is algebraically identical to a 2SLS regression with Y_j as dependent variable. Then,

$$\begin{aligned}
\hat{\delta}^{MD}_{BE} &= \left(\sum_{j=1}^{m} Z_j' X_j\right)^{-1} \sum_{j=1}^{m} Z_j' Y_j \\
&= \left(\sum_{j=1}^{m} \bar{X}_j' X_j\right)^{-1} \sum_{j=1}^{m} \bar{X}_j' Y_j \\
&= \left(\sum_{j=1}^{m} X_j' P_j X_j\right)^{-1} \sum_{j=1}^{m} X_j' P_j Y_j \\
&= \left(\sum_{j=1}^{m} \bar{X}_j' \bar{X}_j\right)^{-1} \sum_{j=1}^{m} \bar{X}_j' \bar{Y}_j = \hat{\delta}_{BE}
\end{aligned}$$

where the second line follows since $Z_j = \bar{X}_j$, the third and last line by $P_jX_j = \bar{X}_j$, $P_jY_j = \bar{Y}_j$ and, since P_j is idempotent. ■

Proof of Proposition 4. The optimal instrument takes the form $Z_j^* = \mathbb{E}[g_j(\delta)g_j(\delta)'|Z_j]^{-1}R_j(\delta,\tau)$, where $R_j(\delta,\tau) = \mathbb{E}[\frac{\partial}{\partial\delta}g_j(\delta,\tau)|Z_j]$. For both moment conditions, $R_j(\delta,\tau)$ is identical. Then for the first moment restriction, we have:

$$\begin{aligned}\mathbb{E}[(\hat{Y}_j - X_j\delta)(\hat{Y}_j - X_j\delta)'|X_j] &= \mathbb{E}[(\tilde{X}_j(\hat{\beta}_j - \beta) + \tilde{X}_j\beta - X_j\delta)(\tilde{X}_j(\hat{\beta}_j - \beta) + \tilde{X}_j\beta - X_j\delta)'|X_j] \qquad (33)\\ &= \mathbb{E}[(\tilde{X}_j(\hat{\beta}_j - \beta) + \alpha_j)(\tilde{X}_j(\hat{\beta}_j - \beta) + \alpha_j)'|X_j]\\ &= \tilde{X}_j\frac{V_j}{n}\tilde{X}_j' + \mathbf{l}_n\mathbf{l}_n'\sigma_\alpha^2.\end{aligned}$$

The matrix $\tilde{X}_j\frac{V_j}{n}\tilde{X}_j' + \mathbf{l}_n\mathbf{l}_n'\sigma_\alpha^2$ is singular, so that we suggest using the Moore-Penrose inverse to construct the optimal instrument.

For the second moment restriction, we have:

$$\begin{aligned}\mathbb{E}[(Y_j - X_j\delta)(Y_j - X_j\delta)'|X_j] &= \mathbb{E}[(\alpha_j + \varepsilon_{ij})(\alpha_j + \varepsilon_{ij})'|X_j]\\ &= \left(\mathbf{I}_n\sigma_\varepsilon^2 + \mathbf{l}_n\mathbf{l}_n'\sigma_\alpha^2\right).\end{aligned}$$

Then note that $\left(\mathbf{I}_n\sigma_\varepsilon^2 + \mathbf{l}_n\mathbf{l}_n'\sigma_\alpha^2\right)^{-1} = (\tilde{X}_j\tilde{X}_j^+\sigma_\varepsilon^2 + \mathbf{l}_n'\mathbf{l}_n\sigma_\alpha^2)^+ = (\tilde{X}_j(\tilde{X}_j'\tilde{X}_j)^{-1}\tilde{X}_j'\sigma_\varepsilon^2 + \mathbf{l}_n'\mathbf{l}_n\sigma_\alpha^2)^+ = (\tilde{X}_j\frac{V_j}{n}\tilde{X}_j' + \mathbf{l}_n'\mathbf{l}_n\sigma_\alpha^2)^+X_j$, where $V_j = (\frac{1}{n}\tilde{X}_j'\tilde{X}_j)^{-1}\sigma_\varepsilon^2$ and since for a full column rank matrix $\tilde{X}_j$, $\tilde{X}_j\tilde{X}_j^+ = I_n$ and $\tilde{X}_j^+ = (\tilde{X}_j'\tilde{X}_j)^{-1}\tilde{X}_j'$. ■

Proof of Proposition 5. Define $Z_g = (z_{1g}, \dots, z_{n_g g})'$, $X_g = (x_{1g}, \dots, x_{n_g g})'$, $Y_g = (y_{1g}, \dots, y_{n_g g})'$ and $\hat{Y}_g = (\hat{y}_{1g}, \dots, \hat{y}_{n_g g})'$. The first and third terms of the expression are identical for both estimators. Thus, we focus on the middle term. Let $\hat{u}_g = Y_g - X_g\hat{\delta}$ be the vector of residuals from the regression using Y as dependent variable, and let $\hat{u}_g^{MD} = \hat{Y}_g - X_g\hat{\delta}^{MD}$ be the vector of residuals of the estimator using the fitted values as regressand. We show that $Z_g'\hat{u}_g = Z_g'\hat{u}_g^{MD}$ for all g. By Proposition 3, $\hat{\delta}^{MD} = \hat{\delta}$. Thus, the fitted values of both estimators are identical. Next, define $\breve{X}_g = \text{diag}\{\tilde{x}_{1g}, \dots, \tilde{x}_{n_g g}\}$ and recall that regressing Y_g on $\breve{X}_g$ is the same as performing G separate regressions. Let $\breve{\beta}_g$ be the coefficient vector of a OLS regression of Y_g on $\breve{X}_g$. Note that Z_g is in the column space of $\tilde{X}_g$. Define the projection matrix $\breve{P} = \breve{X}_g(\breve{X}_g'\breve{X}_g)^{-1}\breve{X}_g'$. Since Z_j is in the column space of $\breve{X}_g$,

$$\breve{P}Z_g = Z_g. \qquad (34)$$

Then,

$$\begin{aligned}Z_g'\hat{u}_g^{MD} &= Z_g'\left(\hat{Y}_g - X_g\hat{\delta}^{MD}\right)\\ &= Z_g'\tilde{X}_g\breve{\beta}_g - Z_gX_g\hat{\delta}\\ &= Z_g'\tilde{X}_g(\breve{X}_g'\breve{X}_g)^{-1}\breve{X}_g'Y_g - Z_gX_g\hat{\delta}\\ &= Z_g'(Y_g - X_g\hat{\delta}) = Z_g'\hat{u}_g,\end{aligned}$$

where the fourth line follows by (34). Since this holds for all g, the desired result follows directly.

■

B Proofs of the asymptotic results

B.1 Proof of Lemma 1

Proof of lemma 1. Starting from the definition of the estimator we obtain

$$
\begin{aligned}
\hat{\delta}(\tau) &= \left(X'Z\hat{W}(\tau)Z'X\right)^{-1} X'Z\hat{W}(\tau)Z'\hat{y}(\tau) \\
&= \left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1} S'_{ZX}\hat{W}(\tau)\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\tilde{x}'_{ij}\hat{\beta}_j(\tau) \\
&= \left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1} S'_{ZX}\hat{W}(\tau)\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\left(\tilde{x}'_{ij}\left(\hat{\beta}_j(\tau)-\beta_j(\tau)\right)+\tilde{x}'_{ij}\beta_j(\tau)\right) \\
&= \left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1} S'_{ZX}\hat{W}(\tau)\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\left(\tilde{x}'_{ij}\left(\hat{\beta}_j(\tau)-\beta_j(\tau)\right)+x'_{ij}\delta(\tau)+\alpha_j(\tau)\right) \\
&= \delta(\tau)+\left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1} S'_{ZX}\hat{W}(\tau)\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\left(\tilde{x}'_{ij}\left(\hat{\beta}_j(\tau)-\beta_j(\tau)\right)+\alpha_j(\tau)\right).
\end{aligned}
$$

■

B.2 Proofs of Theorems 1 and 1′

As a preliminary step to prove the uniform consistency of our estimator, we show uniform (in τ and j) consistency of the group-level quantile regressions.

Lemma 3 (Uniform consistency of $\hat{\beta}_j(\tau)$). *Under Assumptions 1-4 and 8(a), we have*

$$\sup_{\tau\in\mathcal{T}}\max_{1\leq j\leq m}\|\hat{\beta}_j(\tau)-\beta_j\| = o_p(1).$$

Proof of Lemma 3. Angrist et al. (2006) show uniform consistency in τ but not in j of the quantile regression estimator (see their Theorem 3) while Galvao and Wang (2015) show uniform consistency in j but not in τ (see their Lemma 1). We show here uniformity in both dimensions by following the steps of the proof in Galvao and Wang (2015) and extending it. We define $\mathbb{Q}_{nj}(\tau,\beta) := \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_{ij}-\tilde{x}'_{ij}\beta)-\rho_\tau(y_{ij}-\tilde{x}'_{ij}\beta_j(\tau))$ and $Q_j(\tau,\beta) := E[\rho_\tau(y_{ij}-\tilde{x}'_{ij}\beta)-\rho_\tau(y_{ij}-\tilde{x}'_{ij}\beta_j(\tau))]$. Angrist et al. (2006) show that the empirical process for each group j is stochastically equicontinuous because $|\mathbb{Q}_{nj}(\tau',\beta')-\mathbb{Q}_{nj}(\tau'',\beta'')| \leq C_1\cdot|\tau'-\tau''|+C_2\cdot\|\beta'-\beta''\|$ where $C_1 = 2\cdot C\cdot\sup_{\beta\in\mathcal{B}}\|\beta\|$ for any compact set $\mathcal{B}$ and $C_2 = 2\cdot C$. The constant C is defined in Assumption 2. Note that C_1 and C_2 are neither functions of j nor τ.

Fix any $\delta > 0$. Let $B_j(\delta,\tau) := \{\beta : \|\beta-\beta_j(\tau)\| \leq \delta\}$, the ball with center $\beta_j(\tau)$ and radius δ. For each $\beta \notin B_j(\delta,\tau)$, define $\tilde{\beta} = r_j\beta+(1-r_j)\beta_j(\tau)$ where $r_j = \frac{\delta}{\|\beta-\beta_j(\tau)\|}$. So $\tilde{\beta}\in\partial B_j(\delta,\tau) := \{\beta : \|\beta-\beta_j(\tau)\| = \delta\}$, the boundary of $B_j(\delta,\tau)$. Since $\mathbb{Q}_{nj}(\beta,\tau)$ is convex in β for all τ, and $\mathbb{Q}_{nj}(\beta_j(\tau),\tau) = 0$, we have

$$r_j\mathbb{Q}_{nj}(\beta,\tau) \geq \mathbb{Q}_{nj}(\tilde{\beta},\tau) = Q_j(\tilde{\beta},\tau)+\mathbb{Q}_{nj}(\tilde{\beta},\tau)-Q_j(\tilde{\beta},\tau) > \epsilon_\delta+\mathbb{Q}_{nj}(\tilde{\beta},\tau)-Q_j(\tilde{\beta},\tau) \tag{35}$$

uniformly in j and τ, where

$$\epsilon_\delta := \inf_{\tau\in\mathcal{T}} \inf_{1\le j\le m} \inf_{\|\beta-\beta_j(\tau)\|=\delta} \mathbb{E}\left[\int_0^{\tilde{x}_{ij}'(\beta-\beta_j(\tau))} \left(1(y_{ij}-\tilde{x}_{ij}'\beta_j(\tau)\le s) - 1(y_{ij}-\tilde{x}_{ij}'\beta_j(\tau)\le 0)\right) ds\right]$$

by the identity of Knight (1998) and $\epsilon_\delta > 0$ by Assumptions 3 and 4.

Thus, we have the following

$$\begin{aligned}\left\{\sup_{\tau\in\mathcal{T}} \max_{1\le j\le m} \|\hat{\beta}_j(\tau)-\beta_j(\tau)\| > \delta\right\} &\overset{(a)}{\subseteq} \{\exists\tau_j\in\mathcal{T}, \exists\beta_j\notin B_j(\delta,\tau_j) : \mathbb{Q}_{nj}(\beta_j,\tau_j)\le 0\} \\ &\overset{(b)}{\subseteq} \cup_{j=1}^m \left\{\sup_{\tau\in\mathcal{T}} \sup_{\beta_j\in B_j(\delta,\tau_j)} |\mathbb{Q}_{nj}(\beta_j,\tau_j) - Q_j(\beta_j,\tau_j)| \ge \epsilon_\delta\right\}\end{aligned}$$

Relation (a) holds because, by definition, $\hat{\beta}_j(\tau)$ minimizes $\mathbb{Q}_{nj}(\beta,\tau)$, and $\mathbb{Q}_{nj}(\beta_j(\tau),\tau) = 0$. Relation (b) holds by the rightmost inequality of line (35). Then, it follows that

$$\begin{aligned}\mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \max_{1\le j\le m} \|\hat{\beta}_j(\tau)-\beta_j(\tau)\| > \delta\right\} &\le \mathrm{P}\left\{\cup_{j=1}^m \left\{\sup_{\tau\in\mathcal{T}} \sup_{\beta_j\in B_j(\delta,\tau)} |\mathbb{Q}_{nj}(\beta_j,\tau) - Q_j(\beta_j,\tau)| \ge \epsilon_\delta\right\}\right\} \\ &\le \sum_{j=1}^m \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \sup_{\beta_j\in B_j(\delta\tau)} |\mathbb{Q}_{nj}(\beta_j,\tau) - Q_j(\beta_j,\tau)| \ge \epsilon_\delta\right\} \\ &\le m \max_{1\le j\le m} \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \sup_{\beta_j\in B_j(\delta,\tau)} |\mathbb{Q}_{nj}(\beta_j,\tau) - Q_j(\beta_j,\tau)| \ge \epsilon_\delta\right\}\end{aligned}$$

Therefore, if we can show that

$$\max_{1\le j\le m} \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \sup_{\beta_j\in B_j(\delta,\tau)} |\mathbb{Q}_{nj}(\beta_j,\tau) - Q_j(\beta_j,\tau)| \ge \epsilon_\delta\right\} = o\left(\frac{1}{m}\right)$$

the proof of the lemma will be completed.

Without loss of generality, we assume $\beta_j(\tau) = 0$ for all j and $\tau\in\mathcal{T}$. Then the balls $B_j(\delta,\tau)$ for all j and $\tau\in\mathcal{T}$ are identical and we denote them by $B(\delta)$. Because the closed ball $B(\delta)$ is compact, there exist K balls with center β^k, $k = 1, ..., K$, and radius $\frac{\epsilon}{3C_2}$ such that the collection of them covers $B(\delta)$. For any $\epsilon > 0$, we can find a finite K that satisfies this condition and is independent of j and τ. Therefore, for any $\beta\in B(\delta)$, there is some $k\in\{1, ..., K\}$ such that

$$\begin{aligned}|\mathbb{Q}_{nj}(\beta,\tau) - Q_j(\beta,\tau)| - |\mathbb{Q}_{nj}(\beta^k,\tau) - Q_j(\beta^k,\tau)| &\le |\mathbb{Q}_{nj}(\beta,\tau) - Q_j(\beta,\tau) - \mathbb{Q}_{nj}(\beta^k,\tau) + Q_j(\beta^k,\tau)| \\ &\le |\mathbb{Q}_{nj}(\beta,\tau) - \mathbb{Q}_{nj}(\beta^k,\tau)| + |Q_j(\beta,\tau) - Q_j(\beta^k,\tau)| \\ &\le C_2\frac{\epsilon}{3C_2} + C_2\frac{\epsilon}{3C_2} = \frac{2\epsilon}{3}\end{aligned}$$

uniformly in j and $\tau\in\mathcal{T}$. The third line is justified by the stochastic equicontinuity of $\mathbb{Q}_{nj}(\beta,\tau)$.

It then follows that, for any $\epsilon > 0$,

$$\sup_{\tau\in\mathcal{T}} \sup_{\beta\in B(\delta)} |\mathbb{Q}_{nj}(\beta,\tau) - Q_j(\beta,\tau)| \le \sup_{\tau\in\mathcal{T}} \max_{1\le k\le K} |\mathbb{Q}_{nj}(\beta^k,\tau) - Q_j(\beta^k,\tau)| + \frac{2\epsilon}{3}$$

and

$$\begin{aligned}\mathrm{P}\left\{\sup_{\tau\in\mathcal{T}}\sup_{\beta\in B(\delta)}|\mathbb{Q}_{nj}(\beta,\tau)-Q_j(\beta,\tau)>\epsilon\right\} &\leq \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}}\max_{1\leq k\leq K}|\mathbb{Q}_{nj}(\beta^k,\tau)-Q_j(\beta^k,\tau)|+\frac{2\epsilon}{3}>\epsilon\right\}\\ &= \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}}\max_{1\leq k\leq K}|\mathbb{Q}_{nj}(\beta^k,\tau)-Q_j(\beta^k,\tau)|>\frac{\epsilon}{3}\right\}\\ &\leq \sup_{\tau\in\mathcal{T}}\sum_{k=1}^{K}\mathrm{P}\left\{|\mathbb{Q}_{nj}(\beta^k,\tau)-Q_j(\beta^k,\tau)|>\frac{\epsilon}{3}\right\}\end{aligned}$$

For each $\tau\in\mathcal{T}$, $\mathbb{Q}_{nj}(\beta^k,\tau)$ is the sample mean of n i.i.d. terms bounded in absolute values by $2\cdot C\cdot\delta$. By Hoeffding's inequality, it follows that

$$\begin{aligned}\sum_{k=1}^{K}\mathrm{P}\left\{|\mathbb{Q}_{nj}(\beta^k,\tau)-Q_j(\beta^k,\tau)|>\frac{\epsilon}{3}\right\} &\leq 2K\exp\left\{-\frac{2n\epsilon^2}{3^2 2^2 C^2\delta^2}\right\}\\ &= 2K\exp\left\{-\frac{n\epsilon^2}{18C^2\delta^2}\right\}\\ &= O(\exp(-n))\end{aligned}$$

This upper bound is deterministic and not a function of τ such that it also applies to the supremum over τ. Since $\frac{\log m}{n}\to 0$ by Assumption 8(a), it follows that $O(\exp(-n))=o(1/m)$. ∎

Proof of Theorem 1. We start from Lemma 1 and show that the last factor converges to zero while the other factors converge to finite values.

First, it follows from Assumptions 1(ii), 2(i) and 5(i) that $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}x'_{ij}\right)=o_p\left(\frac{1}{n}\right)$ and $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}z_{ij}x'_{ij}\right]=\mathbb{E}[z_{ij}x'_{ij}]$. Hence, by Assumption 1(i), $\mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}x_{ij}z'_{ij}\right)=o_p\left(\frac{1}{mn}\right)$. By Chebyshev's inequality,

$$\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}x'_{ij}-\mathbb{E}[z_{ij}x'_{ij}]\right)\underset{p}{\to}0.$$

In addition, by Assumption 5(iii), $m^{-1}\sum_{j=1}^{m}\mathbb{E}[z_{ij}x'_{ij}]\to\Sigma_{ZX}$. It follows that

$$S_{ZX}\underset{p}{\to}\Sigma_{ZX}$$

Uniformly in $\tau\in\mathcal{T}$, $\hat{W}(\tau)\underset{p}{\to}W(\tau)$ where $W(\tau)$ is uniformly continuous. Together with the boundedness of Σ_{ZX} and the invertibility of $\Sigma'_{ZX}W(\tau)\Sigma_{ZX}$, it follows that

$$\sup_{\tau\in\mathcal{T}}\left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1}S'_{ZX}\hat{W}(\tau)\underset{p}{\to}\left(\Sigma'_{ZX}W(\tau)\Sigma'_{ZX}\right)^{-1}\Sigma'_{ZX}W(\tau)\tag{36}$$

By lemma 3, $\hat{\beta}_j(\tau)$ is consistent for $\beta_j(\tau)$ uniformly in j and τ. Together with the boundedness of x_{ij} in Assumption 2(i) and of z_{ij} in Assumption 5(i), it follows that

$$\sup_{\tau\in\mathcal{T}}\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}z_{ij}\tilde{x}'_{ij}(\hat{\beta}_j(\tau)-\beta_j(\tau))\underset{p}{\to}0.\tag{37}$$

By Assumption 5(ii), $\mathbb{E}[z_{ij}\alpha_j(\tau)] = 0$ uniformly in τ. By Assumption 5, $\mathrm{Var}(z_{ij}\alpha_j(\tau))$ is uniformly bounded. In addition, z_{ij} is bounded and $\alpha_j(\tau)$ is uniformly continuous in τ. Hence,

$$\sup_{\tau\in\mathcal{T}} \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\alpha_j(\tau) \underset{p}{\to} 0. \tag{38}$$

The result of the proposition follows from equations 36, 37, and 38. ■

When the rates of convergence of the moments are heterogeneous, i.e. when $L_1 > 0$ instruments satisfy $\mathrm{Var}(\bar{z}_{1j}\alpha_j(\tau) = 0$ and $L_2 > 0$ instruments satisfy $\mathrm{Var}(\bar{z}_{1j}\alpha_j(\tau) > 0$, then the rates of convergence of the different elements of the efficient weighting matrix will also be heterogeneous. In such a case, depending on the scaling of W (the estimator is invariant to the scaling of W), either some elements converge to 0 or other elements diverge to infinity such that Theorem 1 does not apply. However, Theorem 1′ shows that this does not preclude the consistency of the estimator.

Theorem 1′ (Uniform consistency when the weighting matrix is asymptotically singular). *Let the model in equation (1), Assumptions 1-7, Assumption 8(a) hold.* $\hat{W}(\tau) \underset{p}{\to} W(\tau)$ *uniformly in* $\tau \in \mathcal{T}$, *where* $W(\tau)$ *is symmetric. For all* $\tau_1, \tau_2 \in \mathcal{T}$, $||W(\tau_2) - W(\tau_1)|| \leq C|\tau_2 - \tau_1|$. *For all* $\tau \in \mathcal{T}$, *we partition*

$$W(\tau) = \begin{pmatrix} W_{11}(\tau) & W_{12}(\tau) \\ W_{21}(\tau) & W_{22}(\tau) \end{pmatrix},$$

where the $L_1 \times L_1$ *matrix* $W_{11}(\tau)$ *and the* $L_2 \times L_2$ *matrix* $W_{22}(\tau)$ *have eigenvalues bounded away from zero and infinity uniformly across* τ. *For any element* w_1 *of* $W_{11}(\tau)$ *and any element* w_2 *of* $W_{12}(\tau)$, $W_{21}(\tau)$ *or* $W_{22}(\tau)$, *we have* $w_2/w_1 = o_p(1)$.

Then,

$$\sup_{\tau\in\mathcal{T}} ||\hat{\delta}(\tau) - \delta(\tau)|| = o_p(1)$$

Proof of Theorem 1′. The proof is similar to the proof of Theorem 1 except that $\Sigma'_{ZX} W \Sigma_{ZX}$ is not necessarily invertible. We partition the matrix Σ_{ZX} as follows

$$\Sigma_{ZX} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where Σ_{11} is $L_1 \times K_1$, Σ_{12} is $L_1 \times K_2$, Σ_{21} is $L_2 \times K_1$ and Σ_{22} is $L_2 \times K_2$. Note that $\Sigma_{12} = 0$:

$$\Sigma_{12} = \mathbb{E}\left[z_{1ij}x'_{2ij}\right] = \mathbb{E}_j\left[\mathbb{E}_i\left[z_{1ij}x'_{2ij}\right]\right] = \mathbb{E}_j\left[\mathbb{E}_i\left[z_{1ij}x_{2j}\right]\right] = \mathbb{E}_j\left[\mathbb{E}_i\left[z_{1ij}\right]x_{2j}\right] = 0$$

To simplify the notation we suppress the dependency of W on τ. Note that

$$\begin{aligned} \Sigma'_{ZX} W \Sigma_{ZX} &= \begin{pmatrix} \Sigma'_{11} & \Sigma'_{21} \\ 0 & \Sigma'_{22} \end{pmatrix} \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \\ &= \begin{pmatrix} A_{11} + B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \end{aligned}$$

where $A_{11} = \Sigma'_{11} W_{11} \Sigma_{11}$, $B_{11} = \Sigma'_{11} W_{12} \Sigma_{21} + \Sigma'_{21} W_{21} \Sigma_{11} + \Sigma'_{21} W_{22} \Sigma_{21}$, $B_{12} = \Sigma'_{11} W_{12} \Sigma 22 + \Sigma'_{21} W_{22} \Sigma_{22}$, $B_{21} = \Sigma_{22} W_{21} \Sigma_{11} + \Sigma_{22} W_{22} \Sigma_{21}$, and $B_{22} = \Sigma'_{22} W_{22} \Sigma_{22}$. Note that A_{11} and B_{22} are invertible by Assumptions 5 and the strict positive definiteness of W_{11} and W_{22}. Note also that B_{11} is relatively negligible compared to A_{11} in the sense that $A_{11}^{-1} B_{11} = o_p(1)$ By the inverse of a partitioned matrix, we obtain

$$
\begin{aligned}
&(\Sigma'_{ZX} W \Sigma_{ZX}{}^{-1} \\
&\quad = \begin{pmatrix} (A_{11} + B_{11})^{-1} & -\left(A_{11} + B_{11}\right)^{-1} B_{12} \left(B_{22} - B_{21}(A_{11} + B_{11})^{-1} B_{12}\right)^{-1} \\ -B_{22}^{-1} B_{21} (A_{11} + B_{11})^{-1} & \left(B_{22} - B_{21}(A_{11} + B_{11})^{-1} B_{12}\right)^{-1} \end{pmatrix}
\end{aligned}
$$

where Σ Similarly,

$$
\Sigma'_{ZX} W = \begin{pmatrix} \Sigma'_{11} W_{11} + \Sigma'_{21} W_{21} & \Sigma'_{11} W_{12} + \Sigma'_{21} W_{22} \\ \Sigma'_{22} W_{21} & \Sigma'_{22} W_{22} \end{pmatrix}
$$

where $\Sigma'_{11} W_{11} + \Sigma'_{21} W_{21} = \Sigma'_{11} W_{11} + o_p(1)$. Combining these results, we obtain

$$
\left(\Sigma'_{ZX} W \Sigma_{ZX}\right)^{-1} \Sigma'_{ZX} W = \begin{pmatrix} A_{11}^{-1} \Sigma'_{11} W_{11} & 0 \\ -B_{22}^{-1} B_{21} A_{11}^{-1} \Sigma'_{11} W_{11} + B_{22}^{-1} \Sigma'_{22} W_{21} & B_{22}^{-1} \Sigma'_{22} W_{22} \end{pmatrix} + o_p(1)
$$

All the terms in this matrix are finite. The rest of the proofs follows as in the proof of Theorem 1. ∎

Proof of Lemma 2. **Part (i)**

Lemma 3 in Galvao et al. (2020) provides the uniform Bahadur representation for the individual-level quantile regression coefficient under our assumptions:

$$
\hat{\beta}_j(\tau) - \beta_j(\tau) = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,\tau}(\tilde{x}_{ij}, y_{ij}) + R_{nj}^{(1)}(\tau) + R_{nj}^{(2)}(\tau), \tag{39}
$$

where

$$
\phi_{j,\tau}(\tilde{x}_{ij}, y_{ij}) = -B_{j,\tau}^{-1} \tilde{x}_{ij} (1(y_{ij} \leq \tilde{x}_{ij} \beta_j(\tau)) - \tau) \tag{40}
$$

with $B_{j,\tau} = \mathbb{E}[f_{y|x}(Q_{y|x}(\tau|\tilde{x}_{ij})|\tilde{x}_{ij}) \tilde{x}_{ij} \tilde{x}'_{ij}]$ and

$$
\sup_{j} \sup_{\tau \in \mathcal{T}} \left\| R_{nj}^{(2)}(\tau) \right\| = O_p\left(\frac{\log n}{n}\right) \tag{41}
$$

$$
\sup_{j} \sup_{\tau \in \mathcal{T}} \left\| \mathbb{E}\left[R_{nj}^{(1)}(\tau)\right] \right\| = O\left(\frac{\log n}{n}\right) \tag{42}
$$

$$
\sup_{j} \sup_{\tau \in \mathcal{T}} \left\| \mathbb{E}\left[\left(R_{nj}^{(1)}(\tau) - \mathbb{E}[R_{nj}^{(1)}(\tau)]\right) \left(R_{nj}^{(1)}(\tau) - \mathbb{E}[R_{nj}^{(1)}(\tau)]\right)'\right] \right\| = O\left(\left(\frac{\log n}{n}\right)^{3/2}\right) \tag{43}
$$

It follows that

$$
\begin{aligned}
\frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} z_{ij} \tilde{x}'_{ij} \left(\hat{\beta}_j(\tau) - \beta_j(\tau)\right) = & \frac{1}{m} \sum_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n} z_{ij} \tilde{x}'_{ij}\right) \left(\frac{1}{n} \sum_{i=1}^{n} \phi_{j,\tau}(\tilde{x}_{ij}, y_{ij})\right) && (44) \\
& + \frac{1}{m} \sum_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n} z_{ij} \tilde{x}'_{ij}\right) R_{nj}^{(1)}(\tau) && (45) \\
& + \frac{1}{m} \sum_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n} z_{ij} \tilde{x}'_{ij}\right) R_{nj}^{(2)}(\tau) && (46)
\end{aligned}
$$

Consider first the third term (46). By assumptions 2(i) and 5(i), x_{ij} and z_{ij} are bounded by C such that the sample mean of their product is also bounded. Therefore, (41) implies that

$$\sup_{\tau\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'\right)R_{nj}^{(2)}(\tau)=O_p\left(\frac{\log n}{n}\right) \tag{47}$$

Consider now the second term of (44). Since $\text{Var}\left(R_{nj}^{(1)}(\tau)\right)=o\left(\frac{1}{n}\right)$ by (43), x_{ij} and z_{ij} are bounded by assumptions 2(i) and 5(i), and observations are independent across individuals, it follows that $\text{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'\right)R_{nj}^{(1)}(\tau)\right)=o_p\left(\frac{1}{mn}\right)$. In addition, by (42), $\sup_{j}\sup_{\tau\in\mathcal{T}}\mathbb{E}\left[R_{nj}^{(1)}\right]=O\left(\frac{\log n}{n}\right)$ such that $\sup_{\tau\in\mathcal{T}}\mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'\right)R_{nj}^{(1)}(\tau)\right]=O\left(\frac{\log n}{n}\right)$. Putting this together, by the Chebyshev inequality and under Assumption 8(c),

$$\sup_{\tau\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'\right)R_{nj}^{(1)}(\tau)=o_p\left(\frac{1}{\sqrt{mn}}\right) \tag{48}$$

It follows that both remainder terms are $o_p\left(\frac{1}{\sqrt{mn}}\right)$ uniformly over τ.

Consider now the term (44). Let $\Sigma_{ZXj}=\mathbb{E}[z_{ij}\tilde{x}_{ij}'|v_j]$, i.e. Σ_{ZXj} is the expected value of $z_{ij}\tilde{x}_{ij}'$ for group j.

$$\begin{aligned}&\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)=\\&\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'-\Sigma_{ZXj}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)+\frac{1}{m}\sum_{j=1}^{m}\Sigma_{ZXj}\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)\end{aligned} \tag{49}$$

By the boundedness of z_{ij} and x_{ij} and the independence of the observations over time, it follows that $\left\|\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'-\Sigma_{ZXj}\right\|=o(1)$ uniformly in j. In addition, $\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)=O\left(\frac{1}{n}\right)$. Hence,

$$\text{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'-\Sigma_{ZXj}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{i,\tau}(\tilde{x}_{ij},y_{ij})\right)\right)=o\left(\frac{1}{mn}\right)$$

The model in equation (1) and Assumption 5(iv) imply that $\mathbb{E}\left[1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))|\tilde{x}_{ij},z_{ij},v_j\right]=\tau$, which implies that

$$\mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'-\Sigma_{ZXj}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)\right]=0$$

uniformly in τ. Therefore, by Chebyshev inequality,

$$\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'-\Sigma_{ZXj}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)=o_p\left(\frac{1}{\sqrt{mn}}\right) \tag{50}$$

uniformly in τ.

Since all other terms are $o_p\left(\frac{1}{\sqrt{mn}}\right)$ uniformly over τ, the limiting distribution of the process $\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\tilde{x}_{ij}'\left(\hat{\beta}_j(\tau)-\beta_j(\tau)\right)$ is the same as the limiting distribution of

$$\frac{1}{m}\sum_{j=1}^{m}\Sigma_{ZXj}\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau}(\tilde{x}_{ij},y_{ij})\right)$$
$$=\frac{1}{m}\sum_{j=1}^{m}\Sigma_{ZXj}\left(\frac{-B_{j,\tau}^{-1}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))-\tau)\right):=\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}s_{ij}(\tau) \quad (51)$$

This is a sample mean over mn independent (but not necessarily identically distributed) observations denoted by $s_{ij}(\tau)$. The model in equation (1) and Assumption 5(iv) imply that $\mathbb{E}\left[1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))|\tilde{x}_{ij},z_{ij},v_j\right]=\tau$, which implies that $\mathbb{E}[s_{ij}(\tau)]=0$. In addition,

$$\text{Var}(s_{ij}(\tau))=\mathbb{E}[\Sigma_{ZXj}\,\text{Var}(\phi_{i,\tau})\Sigma_{ZXj}']=\mathbb{E}[\Sigma_{ZXj}B_{j,\tau}^{-1}\tau(1-\tau)\mathbb{E}[x_{ij}x_{ij}'|v_j]B_{j,\tau}^{-1}\Sigma_{ZXj}'] \quad (52)$$

Pointwise asymptotic normality follows by an application of the Lindeberg CLT.

Next we note that $\left\{\Sigma_{ZXj}\left(\frac{-B_{j,\tau}^{-1}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\leq\tilde{x}_{ij}\beta)-\tau)\right),\tau\in\mathcal{T},\beta\in\mathcal{B}\right\}$ is a Donsker class for any compact set $\mathcal{B}$. This follows by noting that $\{1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau)),\tau\in\mathcal{T},\beta\in\mathcal{B}\}$ is a VC subgraph class and hence a bounded Donsker class. Hence,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\leq\tilde{x}_{ij}\beta)-\tau),\tau\in\mathcal{T},\beta\in\mathcal{B}\right\}$$

is also bounded Donsker with a square-integrable envelope $2\cdot\max_{t\in 1,\ldots,T}|\tilde{x}_{ij}|\leq 2\cdot C$. The whole function is then Donsker by the boundedness of Σ_{ZXj} and $B_{j,\tau}^{-1}$. The weak convergence result follows by application of the functional central limit theorem for independent but not identically distributed random variables, see for instance Theorem 3 in Brown (1971).

Part (ii) follows directly by Lemma 3 in Chetverikov et al. (2016).

Part (iii) The first moment is asymptotically equivalent to (up to a term, which is uniformly $o_p\left(\frac{1}{mn}\right)$)

$$\frac{1}{m}\sum_{j=1}^{m}\Sigma_{ZXj}\left(\frac{-B_{j,\tau}^{-1}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))-\tau)\right).$$

We have already shown that both moments have mean zero. By Assumption 1, the observations are independent across i and j such that we only need to consider the correlation between both moments for the same individual and time period.

$$\begin{aligned}&\text{Cov}(\tilde{x}_{ij}(1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))-\tau),z_{ij}\alpha_j(\tau'))\\&\quad=\mathbb{E}[\tilde{x}_{ij}(1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))-\tau)z_{ij}'\alpha_j(\tau')]\\&\quad=\mathbb{E}[\tilde{x}_{ij}\mathbb{E}[(1(y_{ij}\leq\tilde{x}_{ij}\beta_j(\tau))-\tau)|x_{ij},z_{ij},v_j]z_{ij}'\alpha_j(\tau')]=0\end{aligned}$$

It follows that

$$\sup_{\tau,\tau'\in\mathcal{T}}\|\text{Cov}(\bar{g}_{mn}^{(1))}(\delta,\tau),\bar{g}_{mn}^{(2))}(\delta,\tau'))\|=o_p\left(\frac{1}{mn}\right)$$

■

Proof of Theorem 2. From the definition of the estimator,

$$\hat{\delta}(\tau) - \delta(\tau) = \left(S'_{ZX}\hat{W}S_{ZX}\right)^{-1} S'_{ZX}\hat{W}\bar{g}_{mn}(\delta, \tau)$$

As shown in the proof of Theorem 1, $S_{ZX} \rightarrow \Sigma_{ZX}$. We can partition the matrix Σ_{ZX} as follows

$$\Sigma_{ZX} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where Σ_{11} is $L_1 \times K_1$, Σ_{12} is $L_1 \times K_2$, Σ_{21} is $L_2 \times K_1$ and Σ_{22} is $L_2 \times K_2$. Note that $\Sigma_{12} = 0$:

$$\Sigma_{12} = \mathbb{E}\left[z_{1ij}x'_{2ij}\right] = \mathbb{E}_j\left[\mathbb{E}_t\left[z_{1ij}x'_{2ij}\right]\right] = \mathbb{E}_j\left[\mathbb{E}_t\left[z_{1ij}x_{2j}\right]\right] = \mathbb{E}_j\left[\mathbb{E}_t\left[z_{1ij}\right]x_{2j}\right] = 0$$

To simplify the notation we suppress the dependency of W on τ in the rest of the proof. Note that

$$\begin{aligned} \Sigma'_{ZX}W\Sigma_{ZX} &= \begin{pmatrix} \Sigma'_{11} & \Sigma'_{21} \\ 0 & \Sigma'_{22} \end{pmatrix} \begin{pmatrix} W_1 & 0 \\ 0 & W_{22} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \\ &:= \begin{pmatrix} \Sigma'_{11}W_1\Sigma_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \end{aligned}$$

where $A_{11} = \Sigma'_{11}W_1\Sigma_{11}$, $B_{11} = \Sigma'_{11}W_{11}\Sigma_{11} + \Sigma'_{11}W_{12}\Sigma_{21} + \Sigma'_{21}W_{21}\Sigma_{11} + \Sigma'_{21}W_{22}\Sigma_{21}$, $B_{12} = \Sigma'_{11}W_{12}\Sigma 22 + \Sigma'_{21}W_{22}\Sigma_{22}$, $B_{21} = \Sigma_{22}W_{21}\Sigma_{11} + \Sigma_{22}W_{22}\Sigma_{21}$, and $B_{22} = \Sigma'_{22}W_{22}\Sigma_{22}$. Note that A_{11} and B_{22} are invertible by Assumptions 5 and the assumption on the weighting matrix. By the inverse of a partitioned matrix, we obtain

$$\begin{aligned} &\left(\Sigma'_{ZX}W\Sigma_{ZX}\right)^{-1} = \\ &\begin{pmatrix} \left(A_{11}T + B_{11} - B_{12}B_{22}^{-1}B_{21}\right)^{-1} & -\left(A_{11}T + B_{11}\right)^{-1}B_{12}\left(B_{22} - B_{21}(A_{11}T + B_{11})^{-1}B_{12}\right)^{-1} \\ -B_{22}^{-1}B_{21}(A_{11}n + B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1} & \left(B_{22} - B_{21}(A_{11}T + B_{11})^{-1}B_{12}\right)^{-1} \end{pmatrix} \end{aligned}$$

Similarly,

$$\Sigma'_{ZX}W = \begin{pmatrix} \Sigma'_{11}W_1T + \Sigma'_{11}W_{11} + \Sigma'_{21}W_{21} & \Sigma'_{11}W_{12} + \Sigma'_{21}W_{22} \\ \Sigma'_{22}W_{21} & \Sigma'_{22}W_{22} \end{pmatrix}$$

As $T \rightarrow \infty$, applying L'Hospital's rule for the first column, we obtain

$$\left(\Sigma'_{ZX}W\Sigma_{ZX}\right)^{-1}\Sigma'_{ZX}W \rightarrow \begin{pmatrix} A_{11}^{-1}\Sigma'_{11}W_1 & 0 \\ -B_{22}^{-1}B_{21}A_{11}^{-1}\Sigma'_{11}W_1 + B_{22}^{-1}\Sigma'_{22}W_{21} & B_{22}^{-1}\Sigma'_{22}W_{22} \end{pmatrix}$$

All the terms in this matrix are finite by the invertibility of A_{11} and B_{22}.

Next, we partition the matrix S_{ZX} to separate the z_{1ij} from the z_{2ij} components as well as the x_{1ij} from the x_{2ij} components:

$$S_{ZX} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} S_{11} & 0 \\ S_{21} & S_{22}, \end{pmatrix}$$

where S_{11} is $L_1 \times K_1$, S_{12} is $L_1 \times K_2$, S_{21} is $L_2 \times K_1$ and S_{22} is $L_2 \times K_2$. Note that $S_{12} = 0$:

$$S_{12} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} x_{2ij}z_{1ij} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} x_{2j}z_{1ij} = \frac{1}{mn}\sum_{j=1}^{m} x_{2j}\sum_{i=1}^{n} z_{1ij} = \frac{1}{m}\sum_{j=1}^{m} x_{2j}\bar{z}_{1j} = 0.$$

This means that the fast moments (individual-level-instruments) cannot identify the coefficients on the group-level covariates.

It follows that

$$\begin{aligned}
\Lambda_n^{-1} S'_{ZX} \hat{W} S_{ZX} \Lambda_n^{-1} &= \begin{pmatrix} S'_{11}/\sqrt{T} & S'_{21}/\sqrt{T} \\ 0 & S'_{22} \end{pmatrix} \left(\begin{pmatrix} \hat{W}_1 T & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \hat{W}_{11} & \hat{W}_{12} \\ \hat{W}_{21} & \hat{W}_{22} \end{pmatrix} \right) \begin{pmatrix} S_{11}/\sqrt{T} & 0 \\ S_{21}/\sqrt{T} & S_{22} \end{pmatrix} \\
&= \begin{pmatrix} S'_{11} \hat{W}_1 S_{11} & 0 \\ 0 & S'_{22} \hat{W}_{22} S_{22} \end{pmatrix} + o_p(1).
\end{aligned}$$

and

$$\begin{aligned}
\Lambda_n^{-1} S'_{ZX} \hat{W} \Lambda_n^{-1} &= \begin{pmatrix} S'_{11}/\sqrt{T} & S'_{21}/\sqrt{T} \\ 0 & S'_{22} \end{pmatrix} \left(\begin{pmatrix} \hat{W}_1 T & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \hat{W}_{11} & \hat{W}_{12} \\ \hat{W}_{21} & \hat{W}_{22} \end{pmatrix} \right) \begin{pmatrix} I_{L_1}/\sqrt{T} & 0 \\ 0 & I_{L_2} \end{pmatrix} \\
&= \begin{pmatrix} S'_{11} \hat{W}_1 & 0 \\ 0 & S'_{22} \hat{W}_{22} \end{pmatrix} + o_p(1).
\end{aligned}$$

Using the results derived in the proof of Proposition 1, we obtain

$$\begin{aligned}
&\left(\Lambda_n^{-1} S'_{ZX} \hat{W} S_{ZX} \Lambda_n^{-1} \right)^{-1} \Lambda_n^{-1} S'_{ZX} \hat{W} \Lambda_n^{-1} \\
&\qquad \underset{p}{\rightarrow} \begin{pmatrix} \left(\Sigma'_{11} W_1 \Sigma_{11} \right)^{-1} \Sigma'_{11} W_1 & 0 \\ 0 & \left(\Sigma'_{22} W_{22} \Sigma_{22} \right)^{-1} \Sigma'_{22} W_{22} \end{pmatrix} = G. \qquad (53)
\end{aligned}$$

Combining this result and Lemma 2 with Slutzky's lemma, we obtain

$$\Lambda_{mn} (\hat{\delta}(\tau) - \delta(\tau)) \underset{d}{\rightarrow} N(0, G \Omega G').$$

■

B.2.1 Covariance Matrix

Proof of Proposition 1.

$$\widehat{\text{Var}}(\sqrt{m}(\hat{\delta} - \delta)) = \left(S'_{ZX} \hat{W} S_{ZX} \right)^{-1} S'_{ZX} \hat{W} \left(\frac{1}{mn^2} \sum_{j=1}^{m} Z'_j \hat{u}_j \hat{u}'_j Z_j \right) \hat{W} S_{ZX} \left(S'_{ZX} \hat{W} S_{ZX} \right)^{-1}$$

By the proof of Theorem 3 it follows that

$$\left(S'_{ZX} \hat{W}(\tau) S_{ZX} \right)^{-1} S'_{ZX} \hat{W}(\tau) = G(\tau) + o_p(1).$$

Consider now the term in the middle and insert $\hat{u}_j = \tilde{X}_j \hat{\beta}_j - X_j \hat{\delta} = \tilde{X}_j (\hat{\beta}_j - \beta_j) + X_j (\delta - \hat{\delta}) + \alpha_j$, to obtain,

$$\begin{aligned}
\frac{1}{mn^2} \sum_{j=1}^{m} Z'_j \hat{u}_j \hat{u}'_j Z_j =& \frac{1}{mn^2} \sum_{j=1}^{m} \left(Z'_j \left(\tilde{X}_j (\hat{\beta}_j - \beta_j) + X_j (\delta - \hat{\delta}) + \alpha_j \right) \cdot \left(\tilde{X}_j (\hat{\beta}_j - \beta_j) + X_j (\delta - \hat{\delta}) + \alpha_j \right)' Z_j \right) \\
=& \frac{1}{mn^2} \sum_{j=1}^{m} \Big(Z'_j \Big(\tilde{X}_j (\hat{\beta}_j - \beta_j)(\hat{\beta}_j - \beta_j)' \tilde{X}'_j + \alpha_j \alpha'_j + X_j (\hat{\delta} - \delta)(\hat{\delta} - \delta)' X'_j \\
&+ X_j (\delta - \hat{\delta})(\hat{\beta}_j - \beta_j)' \tilde{X}'_j + \tilde{X}_j (\hat{\beta}_j - \beta_j)(\delta - \hat{\delta})' X'_j \\
&+ \alpha_j (\hat{\beta}_j - \beta_j)' \tilde{X}'_j + \tilde{X}_j (\hat{\beta}_j - \beta_j) \alpha'_j + \alpha_j (\delta - \hat{\delta})' X'_j + X_j (\delta - \hat{\delta}) \alpha'_j \Big) Z_j \Big).
\end{aligned}$$

Next, we want to show that all but the first two terms converge to zero quickly. We consider each term separately. Let $\zeta_{mn}(\tau) = \frac{1}{\sqrt{mn}} + \frac{1}{\sqrt{m}} \cdot ||V_\alpha(\tau)||^{1/2}$, where $V_\alpha(\tau) = \mathrm{Var}(\bar{z}_j \alpha_j(\tau))$. Note that $(\delta - \hat{\delta}) = O_p\left(\zeta_{mn}(\tau)\right)$ and $(\hat{\beta}_j - \beta_j) = O_P\left(\frac{1}{n^{1/2}}\right)$.

Consider the first term. By the proof of Lemma 2(i), it follows that

$$\begin{aligned}\frac{1}{mn^2}\sum_{j=1}^{m} Z_j'\tilde{X}_j(\hat{\beta}_j - \beta_j)(\hat{\beta}_j - \beta_j)'\tilde{X}_j' =& \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} z_{ij}\tilde{x}_{ij}'(\hat{\beta}_j - \beta_j)\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_{ij}\tilde{x}_{ij}'(\hat{\beta}_j - \beta_j)\right)'\\ =&\mathbb{E}\left[\left(\Sigma_{ZXj}\frac{1}{n}\sum_{i=1}^{n}\phi_{i,\tau}(\tilde{x}_{ij}, z_{ij})\right)\left(\Sigma_{ZXj}\frac{1}{n}\sum_{i=1}^{n}\phi_{i,\tau}(\tilde{x}_{ij}, z_{ij})\right)'\right]\\ &+ o_p\left((mn)^{-1}\right)\\ =&\frac{\Omega_1}{n} + o_p\left((mn)^{-1}\right).\end{aligned}$$

For the second term, we have

$$\frac{1}{m}\sum_{j=1}^{m}\bar{z}_j\bar{z}_j'\alpha_j^2 = \mathrm{Var}(\bar{z}_j\alpha_j) + O_p\left(\frac{\mathrm{Var}(\bar{z}_j\alpha_j)}{\sqrt{m}}\right) = \Omega_2 + O_p\left(\frac{1}{\sqrt{m}}\right)\cdot||V_\alpha||,$$

where the first equality is follows by the central limit theorem,

$$\sqrt{m}\frac{1}{m}\sum_{j=1}^{m}\left[\mathrm{Var}(\bar{z}_j\alpha_j)^{-1}\bar{z}_j\bar{z}_j'\alpha_j - 1\right] = O_p(1),$$

which implies that $\frac{1}{m}\sum_{j=1}^{m}\bar{z}_j\bar{z}_j'\alpha_j - \mathrm{Var}(\bar{z}_j\alpha_j) = O_p\left(\frac{\mathrm{Var}(\bar{z}_j\alpha_j)}{\sqrt{m}}\right)$.

Consider now the third term. We have that

$$\begin{aligned}\frac{1}{mn^2}\sum_{j=1}^{m} Z_j'X_j(\hat{\delta} - \delta)(\hat{\delta} - \delta)'X_j'Z_j =& \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} z_{ij}x_{ij}'\right)(\hat{\delta} - \delta)(\hat{\delta} - \delta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_{ij}z_{ij}'\right)\\ =&O_p(\zeta_{mn}^2).\end{aligned}$$

The fifth term is just the transpose of the fourth term. Thus we will consider only the fourth one.

For the sixth (and seventh) term(s), we have that

$$\frac{1}{mn^2}\sum_{j=1}^{m} Z_j'\alpha_j(\hat{\beta}_j - \beta_j)'\tilde{X}_j'Z_j = \frac{1}{m}\sum_{j=1}^{m}\bar{z}_j\alpha_j(\hat{\beta}_j - \beta_j)'\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{ij}z_{ij}'\right) = O_p\left(\zeta_{mn}\cdot n^{-1/2}\right).$$

Finally, for the eighth (and ninth) term(s), it follows that

$$\frac{1}{mn^2}\sum_{j=1}^{m} Z_j'\alpha_j(\delta - \hat{\delta})'X_j'Z_j = \frac{1}{mn^2}\sum_{j=1}^{m}\bar{z}_j\alpha_j(\delta - \hat{\delta})'\left(\frac{1}{n}\sum_{i=1}^{n} x_{ij}z_{ij}'\right) = O_p\left(\zeta_{mn}^2\right).$$

Hence,

$$\left(\frac{1}{mn^2}\sum_{j=1}^{m} Z_j'\hat{u}_j\hat{u}_j'Z_j\right) = \frac{\Omega_1}{n} + \Omega_2 + O_p(\zeta_{mn}^2 + \zeta_{mn}n^{-1/2}),$$

and the final result follows directly by the continuous mapping theorem. ■

B.2.2 Overidentification Test

Proof of Proposition 2. First, we want to rewrite the J-statistics, in a way that accounts for the different convergence rates of the moments conditions:

$$\begin{aligned}
J(\hat{\delta}) &= m\bar{g}_{mn}(\hat{\delta})'\hat{S}^{-1}\bar{g}_{mn}(\hat{\delta})\\
&= m\left(\Lambda_{mn}\bar{g}_{mn}(\hat{\delta})\right)'\Lambda_{mn}^{-1}\hat{S}^{-1}\Lambda_{mn}^{-1}\Lambda_{mn}\bar{g}_{m}(\hat{\delta})\\
&= \left(\Lambda_{mn}\bar{g}_{mn}(\hat{\delta})\right)'\Lambda_{n}^{-1}\hat{S}^{-1}\Lambda_{n}^{-1}\Lambda_{mn}\bar{g}_{mn}(\hat{\delta}).\\
&= \left(\Lambda_{mn}\bar{g}_{mn}(\hat{\delta})\right)'\hat{\Omega}^{-1}\Lambda_{mn}\bar{g}_{mn}(\hat{\delta}).
\end{aligned}$$

where $\sqrt{m}\Lambda_n = \Lambda_{mn}$ and $\hat{\Omega}^{-1} = \left(\Lambda_n\hat{S}\Lambda_n\right)^{-1}$.
Second, we want to show that $\bar{g}_{mn}(\hat{\delta}) = \hat{B}\left(\bar{g}_{mn}(\delta) + \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n} z_{ij}\tilde{x}_{ij}'(\hat{\beta}_j - \beta_j)\right)$. Recall that $\hat{Y}_j = X_j\delta + \alpha_j + \tilde{X}_j(\hat{\beta}_j - \beta_j)$. Hence, we can write

$$\begin{aligned}
Z_j'\hat{Y}_j &= Z_j'X_j\delta + Z_j'\alpha_j + Z_j'\tilde{X}_j(\hat{\beta}_j - \beta)\\
S_{Z\hat{Y}} &= S_{ZX}\delta + \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n} z_{ij}\alpha_j + \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n} z_{ij}\tilde{x}_{ij}'(\hat{\beta}_j - \beta_j)\\
S_{Z\hat{Y}} &= S_{ZX}\delta + \bar{g}_{mn}(\delta).
\end{aligned}$$

Then, note that

$$\begin{aligned}
\bar{g}_{mn}(\hat{\delta}) &= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n} z_{ij}(\hat{y}_{ij} - x_{ij}'\hat{\delta})\\
&= S_{Z\hat{Y}} - S_{ZX}\hat{\delta}\\
&= S_{Z\hat{Y}} - S_{ZX}\left(S_{ZX}\hat{S}^{-1}S_{Z}X\right)^{-1}S_{ZX}\hat{S}^{-1}S_{Z\hat{Y}} = \hat{B}S_{Z\hat{Y}},
\end{aligned}$$

where $\hat{B} = \left(I_L - S_{ZX}\left(S_{ZX}'\hat{S}^{-1}S_{ZX}\right)^{-1}S_{ZX}'\hat{S}^{-1}\right)$.
Thus,

$$\begin{aligned}
\bar{g}_{mn}(\hat{\delta}) &= \hat{B}S_{Z\hat{Y}}\\
&= \left(I_L - S_{ZX}\left(S_{ZX}'\hat{S}^{-1}S_{ZX}\right)^{-1}S_{ZX}'\hat{S}^{-1}\right)(S_{ZX}\delta + \bar{g}_{mn}(\delta))\\
&= \hat{B}\bar{g}_{mn}(\delta).
\end{aligned}$$

Since Ω is positive definite there exist a matrix C such that $\hat{\Omega}^{-1} = C'C$.
We define $A \equiv C\Lambda_n S'_{ZX}$ and $M \equiv I_L - A(A'A)^{-1}A'$.
In this third part, we show that

$$\hat{B}'\Lambda_{mn}\hat{\Omega}^{-1}\Lambda_{mn}\hat{B} = \Lambda_{mn}C'MC\Lambda_{mn}.$$

Note that

$$\begin{aligned}
C\Lambda_{mn}\hat{B} &= C\Lambda_{mn}\left(I_L - S_{ZX}\left(S'_{ZX}\hat{S}^{-1}S_{ZX}\right)^{-1}S'_{ZX}\hat{S}^{-1}\right) \\
&= \left(C\Lambda_{mn} - C\Lambda_{mn}S_{ZX}\left(S'_{ZX}\hat{S}^{-1}S_{ZX}\right)^{-1}S'_{ZX}\hat{S}^{-1}\right) \\
&= \left(C\Lambda_{mn} - C\Lambda_{mn}S_{ZX}\left(S'_{ZX}\Lambda_n C'C\Lambda_n S_{ZX}\right)^{-1}S'_{ZX}\Lambda_n C'C\Lambda_n\right) \\
&= \left(C\Lambda_{mn} - C\Lambda_{n}S_{ZX}\left(S'_{ZX}\Lambda_n C'C\Lambda_n S_{ZX}\right)^{-1}S'_{ZX}\Lambda_n C'C\Lambda_{mn}\right) \\
&= \left(I_L - A\left(A'A\right)^{-1}A'\right)C\Lambda_{mn} \\
&= MC\Lambda_{mn}.
\end{aligned}$$

Where the third line uses $\hat{\Omega}^{-1} = \Lambda_n^{-1}\hat{S}^{-1}\Lambda_n^{-1} = C'C$. The fourth line follows because $\Lambda_{mn} = \sqrt{m}\Lambda_n$. In the last two lines, we use the definitions of A and M.
M is symmetric and idempotent. Thus

$$\begin{aligned}
\hat{B}'\Lambda_{mn}\hat{\Omega}^{-1}\Lambda_{mn}\hat{B} &= \hat{B}'\Lambda_{mn}C'C\Lambda_{mn}\hat{B} \\
&= (C\Lambda_{mn}\hat{B})'C\Lambda_{mn}\hat{B} \\
&= (MC\Lambda_{mn})'MC\Lambda_{mn} \\
&= \Lambda_{mn}C'MC\Lambda_{mn}.
\end{aligned}$$

The rank of M is the trace of M, which is $L - K$.
Since $\Omega = \Lambda_n S\Lambda_n$ is positive definite, there exist a matrix Q such that

$$Q'Q = \Omega^{-1}$$

and the probability limit of C is Q. We define $v \equiv C\Lambda_{mn}\bar{g}_{mn}(\delta)$.
It follows that

$$\Lambda_{mn}\bar{g}_{mn}(\delta) \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Omega_{11} & 0\\ 0 & \Omega_{22}\end{pmatrix}\right) \sim N(0,\Omega).$$

Thus it follows directly that

$$v \xrightarrow{d} N(0, Q\Omega Q') = N(0, Q(Q'Q)^{-1}Q') = N(0, I_L).$$

Now we can come back to our test statistic:

$$\begin{aligned}
J(\hat{\delta}) &= N\bar{g}_{mn}(\hat{\delta})'\hat{S}^{-1}\bar{g}_{mn}(\hat{\delta}) \\
&= \left(\Lambda_{mn}\bar{g}_{mn}(\hat{\delta})\right)' \hat{\Omega}^{-1}\Lambda_n^{-1}\bar{g}_{mn}(\hat{\delta}) \\
&= \left(\Lambda_{mn}\hat{B}\bar{g}_m(\delta)\right)' \hat{\Omega}^{-1}\Lambda_{mn}\left(\hat{B}\bar{g}_{mn}(\delta)\right) \\
&= \bar{g}_{mn}(\delta)'\hat{B}'\Lambda_{mn}\hat{\Omega}^{-1}\Lambda_{mn}\hat{B}\bar{g}_{mn}(\delta) \\
&= \bar{g}_{mn}(\delta)'\Lambda_{mn}C'MC\Lambda_{mn}\bar{g}_{mn}(\delta) \\
&= \left[C\Lambda_{mn}\bar{g}_{mn}(\delta)\right]' M \left[C\Lambda_{mn}\bar{g}_{mn}(\delta)\right].
\end{aligned}$$

Since M is idempotent with rank L, it follows that

$$J(\hat{\delta}) \xrightarrow{d} \chi^2_{L-K}.$$

■

C Optimal Instruments and Minimum Distance

In this section, we show that if $\alpha_j(\tau) = 0$ for all j and τ, efficient minimum distance can be implemented by optimal instruments. From equation (33) we have that if $\alpha_j(\tau) = 0$ for all j and all τ, $\mathbb{E}[(\tilde{X}_j\hat{\beta}_j(\tau) - X_j\delta(\tau))(\tilde{X}_j\hat{\beta}_j(\tau) - X_j\delta(\tau))'|X_j] = \tilde{X}_j\frac{V_j(\tau)}{n}\tilde{X}_j'$. This implies the optimal instrument $Z_j^* = (\tilde{X}_j\frac{V_j(\tau)}{n}\tilde{X}_j')^+X_j$. Since n is a scalar, using $Z_j^*(\tau) = (\tilde{X}_jV_j(\tau)\tilde{X}_j')^+X_j$ leads to the same results.

Proposition 6. *The IV regression with instrument $Z_j^*(\tau) = (\tilde{X}_jV_j(\tau)\tilde{X}_j')^+X_j$ is the efficient MD estimator.*

Proof.

$$\begin{aligned}
\hat{\delta}_{EMD}(\tau) &= \left(\sum_{j=1}^{m} R_j'\hat{V}_j^{-1}(\tau)R_j\right)^{-1}\left(\sum_{j=1}^{m} R_j'\hat{V}_j^{-1}(\tau)\hat{\beta}_j(\tau)\right) \\
&= \left(\sum_{j=1}^{m} X_j'\tilde{X}_j\left(\tilde{X}_j'\tilde{X}_j\hat{V}_j(\tau)\tilde{X}_j'\tilde{X}_j\right)^{-1}\tilde{X}_j'X_j\right)^{-1}\left(X_j'\tilde{X}_j\left(\tilde{X}_j'\tilde{X}_j\hat{V}_j(\tau)\tilde{X}_j'\tilde{X}_j\right)^{-1}\tilde{X}_j'\hat{Y}_j(\tau)\right) \\
&= \left(\sum_{j=1}^{m} X_j'\left(\tilde{X}_j\hat{V}_j(\tau)\tilde{X}_j'\right)^{+}X_j\right)^{-1}\left(X_j'\left(\tilde{X}_j\hat{V}_j(\tau)\tilde{X}_j'\right)^{+}\hat{Y}_j(\tau)\right) = \hat{\delta}_{Oj}(\tau).
\end{aligned}$$

The second line follows by the relationship between $\tilde{X}_j$ and X_j, that is $\tilde{X}_jR_j = X_j$ and the third line follows since for a full column rank matrix $\tilde{X}_j$, $\tilde{X}_j^+ = (\tilde{X}_j'\tilde{X}_j)^{-1}\tilde{X}_j'$. ■