Minimum Distance Estimation of Quantile Panel Data Models*

Blaise Melly[†]and Martina Pons[‡]

First version: November 2020
This version: June 16, 2022

Preliminary version. Link to the newest version

Abstract

We propose a minimum distance estimation approach to quantile panel data models where the individual 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 classical panel data, where we follow the same units over time, and (ii) to grouped data, where we observe data at the individual level but the treatment varies at the group level. Depending on the variables that are 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. Monte Carlo simulations show that our estimator performs well in finite samples also when the time-series dimension is small.

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 are available. The researchers can alleviate endogeneity concerns, for instance, by allowing for correlated individual effects; or they can obtain more precise estimates, for example, by using a random-effects estimator; or they can exploit time-varying variables to identify the impact of time-invariant variables, e.g. with the Hausman and Taylor (1981) estimator. In this paper, we propose a minimum distance

*We are grateful to Bo Honore, 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), and Geneva University (3rd of March 2022) for useful comments.

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

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

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 the notation (i and t subscripts) and terminology (individuals and time periods) commonly used in the panel data literature. However, our results also apply to grouped data, where we observe data at the individual level but the treatment varies at the group level. In this part of literature, the i units are often called groups and the t units are individuals within these groups. 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. We discuss the application of our results to this framework on Section 5.

In all cases, we perform estimation in two stages. The first stage consists of individual-level quantile regressions using time-varying covariates at each quantile of interest. In the second stage, the first-stage fitted values are regressed on time-invariant and time-varying variables. If these variables are potentially endogenous, an instrumental variable regression can be used or, more generally, the generalized method of moment (GMM) estimator. 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, we show in Section **??** that it is numerically identical to the standard estimators (fixed effects, random effects, Hausman-Taylor) if we use least-squares in the first stage and the appropriate instruments.

As a non-linear estimator, quantile regression with fixed effects is subject to the incidental parameter problem with a bias that is of order T^{-1}. Inference is justified using a 'large N, large T' asymptotic framework.[1] Recently, Galvao et al. (2020) have weakened the requirements on the relative rate of divergence of N and T for asymptotic normality of fixed effects quantile estimators. Using their results, we show that our estimator is asymptotically normal under the condition that $N(\log T)^2/T \to 0$. 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 T relative to N can be weakened if only the coefficient vector on the time constant regressor is of interest. In this case, the milder condition that $\sqrt{N}(\log T)/T \to 0$ 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 coefficients of the variables that are identified by time-varying instruments converge at the $\sqrt{NT}$ rate (the 'fast' coefficients) while the other converge only at the $\sqrt{N}$ rate (the 'slow' coefficients). This has several consequences on the first-

[1] Large T asymptotic has been widely used in the quantile and nonlinear 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).

order asymptotic distribution: (i) The fast and slow coefficients are first-order asymptotically independent. (ii) If a coefficient is identified by time-varying and time-constant instruments then the time-constant instruments are 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. In simulations, we find that these asymptotic results do not provide an accurate approximation of the finite-sample behavior of the estimator. For this reason, we derive asymptotic results where we keep track of higher-order terms. 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 sample. 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 takes automatically into account the first-stage error. Thus, 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. Koenker (2004) introduced a penalized quantile fixed effects estimator that treated the individual heterogeneity as a pure location shift. A large share of the literature focused on fixed effects models (see, for example, Canay, 2011; Galvao and Kato, 2016; Gu and Volgushev, 2019). Kato et al. (2012) allow the individual effects to depend on the quantile of interest and contributed to the asymptotic theory of the estimator. Galvao and Wang (2015) suggest a two-step MD estimator as a computationally fast way to estimate fixed effects quantile panel data model. Galvao and Poirier (2019) suggest to use 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 individual 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.

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 time-varying regressors and allowing the number of individuals to go to infinity.[3] Galvao and Wang (2015) focus on estimating the effect of time-varying regressors. In contrast, we aim at estimating the effect of both time-varying and time-invariant 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 individuals (between groups using their terminology) and allow for instrumental variable for identification. We can also apply our ap-

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

[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.

proach to this setup. The main difference is that, in the second stage, we regress the fitted values on all variables while they regress the estimated intercept on the time-invariant regressors. Since they use only the intercept in the second stage, their estimator is not invariant to reparametrization of the time-varying regressors while our estimator possesses this property. In addition, by keeping all the variables in the second stage, we can easily impose equality of the coefficients on the time-varying regressors and, therefore, increase precision at a minimal cost from a computational perspective. Simulations using exactly 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 T relative to N 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).

The remainder of the paper is structured as follows. Section 2.2 presents the model, 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 quantile panel data models, and we 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. Section 6 concludes.

2 Model and Minimum Distance Estimator

2.1 Least Squares Estimators

To motivate the minimum distance approach, we begin by providing an instructive characterization of the traditional least squares panel data estimators. We discuss here some equivalent way to compute various common estimators and show that they can be estimated using a two-stage approach. A more detailed discussion including formal statements can be found in Appendix A with proofs in Appendix B.1.

We want to learn the effects of the time-varying variables x_{1it} on an outcome y_{it}. We observe these variables for the individuals $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$.[4] We use the traditional terminology of panel data, but the i index can define groups of any sort and the t index can represent any order within the groups. This will be the case in particular in section 5 when we will consider group-level treatments. Consider first a traditional panel data model with individual effects

$$y_{it} = x'_{1it}\beta + \alpha_i + \varepsilon_{it} \tag{1}$$

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

The least squares fixed effects estimator can be computed by subtracting from each variable its individual time average and applying the ordinary least squares estimator. This within transformation eliminates the potential endogeneity coming from α_i and provides a consistent estimator without any assumption on unobserved time-invariant heterogeneity. However, this approach is not applicable in quantile models, as there is no known transformation that eliminates the individual effects. In particular, 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 individual effects by including an indicator variable for each individual. 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.[5] In addition, we do not see a way to extend this approach to estimate the effect of the time-constant variables, especially when we need to exploit an instrumental variable to identify their effect.[6] A third numerically equivalent way to compute the least squares fixed effects estimator consists in dividing the problem in two steps. The first stage consists in individual-level regressions for each i. The unobserved heterogeneity α_i will be absorbed by the intercept of each regression. The second stage aggregates the individual results by regressing the fitted values from the first stage on x_{1it} using the time-demeaned regressor, $\dot{x}_{1it} = x_{1it} - \bar{x}_{1i}$ where $\bar{x}_{1i} = T^{-1}\sum_{t=1}^{T} x_{1it}$, as an instrumental variable. This instrument exploits only the variation within individuals. This procedure can be easily extended to quantile models, where it substantially reduces the computational burden of quantile fixed effects estimation.[7]

This two-step procedure is not specific to the fixed effects case, but it applies to a wide range of estimators. We include the time constant regressors x_{2i} in the model

$$y_{it} = x_{1it}'\beta + x_{2i}'\gamma + \alpha_i + \varepsilon_{it} \tag{2}$$

and define the generalized two-step procedure as follows. The first stage consists of individual-level least squares regressions including only the time-varying variables. The second stage is a linear GMM regression of the first-stage fitted values on both time-varying and time invariant variables. This two-step estimator is algebraically identical to the one-step linear GMM estimator (see Proposition 3 in Appendix A). We can numerically obtain the most common least

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

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

[7]There is actually a fourth possibility to compute the least squares fixed estimator, which is similar to the third one. The first stage consists also in individual-level regressions for each i. 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) suggest a similar estimator for the fixed effects quantile regression model. However, this averaging method does not allow for the presence of time-constant regressors and, more generally, does not exploit the between-individuals variation.

squares estimators by selecting the instrumental variables. For instance, we obtain the between estimator by using the individual time averaged variables, $\bar{x}_{1i}$ and x_{2i}, as instrumental variables. While FGLS is the most common estimator for the random effects model, Im et al. (1999) show that the overidentified 3SLS estimator, with instruments $\dot{x}_{1it}$, $\bar{x}_{1i}$, and x_{2i}, is numerically identical to the random effects estimator. Since 3SLS is a special case of a GMM estimator, it follows that 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 on 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: $\dot{x}_{1it}$, the individual time average of the exogenous regressors and the potential external instruments. Interestingly, in all cases, clustering the standard errors at the level of the individuals (or at a higher level) is sufficient to capture the first stage estimation error, see Proposition 6 in Appendix B.1. Actually, these clustered standard errors are numerically identical to the clustered standard errors obtain after using the one-step GMM estimator with y_{it} as the dependent variable.

2.2 Quantile Model

We now want to learn the effects of variables on the distribution of the outcome y_{it}. We assume that

$$Q(\tau, y_{it}|x_{1it}, x_{2i}) = x'_{1it}\beta(\tau) + x'_{2i}\gamma(\tau) + \alpha_i(\tau), \tag{3}$$

where $Q(\tau, y_{it}|x_{1it}, x_{2i})$ is the τth conditional quantile function of the response variable y_{it} for individual i in period t given the K_1-vector of time-varying regressors $x_{1,it}$ and the K_2-vector of time invariant variables $x_{2,it} = x_{2,i}$ for all t. In total, there are $K_1 + K_2 = K$ parameters to estimate. The parameters $\beta(\tau)$, $\gamma(\tau)$ and the unobserved individual heterogeneity $\alpha_i(\tau)$ can depend on the quantile index $0 < \tau < 1$. Depending on the setting, both the parameters $\beta(\tau)$ and $\gamma(\tau)$ might be of interest. We normalize $\mathbb{E}[\alpha_i(\tau)] = 0$, which is not restrictive because x_{2i} 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_{it} conditionally on the covariates and on the individual effect $\alpha_i(\tau)$. Thus, even if the individual effects are independent of the regressors, we identify different coefficients that 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 unit i corresponds to a region and each unit t to an individual within this region. We do not have any x_{1it} variable. We are interested in the effect of a binary treatment x_{2i}, which has been randomized and is therefore independent from $\alpha_i(\tau)$. $\gamma(\tau)$ is the effect of this treatment for individuals that rank at the τ quantile of y_{it} in their region. On the other hand, the quantile regression of y_{it} on x_{2i} identifies the effect for individuals that rank at the τ quantile in the whole country (given the treatment status). x_{1it} and the same individual effect $\alpha_i(\tau)$. These are different parameters except if $\alpha_i(\tau) = 0$ for all i or if the treatment effect is homogeneous such that $\gamma(\tau) = \gamma$. If the unconditional treatment effect is of interest, one can naturally obtain the unconditional distribution function by integrating out the individual effects (and possibly the other variables) and then invert the resulting distribution function, to obtain the unconditional quantile function, see Chernozhukov et al. 2013.[8]

We allow the individual effects $\alpha_i(\tau)$ to be possibly correlated with x_{1it} and x_{2i}. To deal with the potential endogeneity of $\alpha_i(\tau)$, we assume that there is a L-dimensional vector $(L \geq K)$ of valid instruments z_{it} satisfying $\mathbb{E}[z_{it}\alpha_i(\tau)] = 0$. Note that $\beta(\tau)$ is identified in model (3) as long as there is some variation in x_{1it} over time for some individuals i. For instance, we can can include $\dot{x}_{1it}$ in the vector of instruments z_{it} because this variable will satisfy the requirements under strict exogeneity. On the other hand, we do need to find additional instruments to identify $\gamma(\tau)$.

Remark 2 (Skorohod representation). The following Skorohod representation

$$\begin{aligned} y_{it} &= x_{1it}\beta(u_{it}) + x_{2i}\gamma(u_{it}) + \alpha_i(u_{it}) \\ &= x_{1it}\beta(u_{it}) + x_{2i}\gamma(u_{it}) + \alpha(u_{it}, v_i) \\ &= q(x_{1it}, x_{2i}, u_{it}, v_i) \end{aligned}$$

where $\alpha(u, v)$ and $q(x_{1it}, x_{2i}, u, v)$ are strictly increasing in u and v (while fixing the other arguments). We normalize $v_i \sim U(0, 1)$ and $u_{it}|x_{1it}, x_{2i}, v_i \sim U(0, 1)$ such that $\alpha(u, v)$ and $q(x_{1it}, x_{2i}, u, v)$ can be interpreted as quantile functions. v_i ranks the individuals while u_{it} ranks observations over time for the same individual. For the instrumental variable, we assume that

$$E[\alpha_i(u_{it})z_{it}] = E[\alpha(u_{it}, v_i)z_{it}] = 0 \tag{4}$$

. A sufficient condition for this assumption is $(u_{it}, v_i) \perp\!\!\!\perp z_{it}$. If the instrument does not vary over time, only $v_i \perp\!\!\!\perp z_i$ is sufficient.

Our model allows only the intercept to differ between individuals.[9] We now consider a more general model where we also allow the slopes to differ between individuals but heterogeneity is still restricted to be :

$$y_{it} = x'_{1it}\beta(u_{it}, v_i) + x'_{2i}\gamma(u_{it}, v_i) + \alpha(u_{it}, v_i)$$

If we maintain the strict monotonicity assumption, this model implies that

$$Q(\tau, y_{it}|x_{1it}, x_{2i}, v_i) = x'_{1it}\beta(\tau, v_i) + x'_{2i}\gamma(\tau, v_i) + \alpha(\tau, v_i)$$

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

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

In the exogenous case where $(x_{1it}, x_{2i}) \perp\!\!\!\perp v_i$, this implies

$$
\begin{aligned}
E\left[Q(\tau, y_{it}|x_{1it}, x_{2i}, v_i)|x_{1it}, x_{2i}\right] &= x_{1it}'\int_0^1 \beta(\tau, V_i)dV_i + x_{2i}'\int_0^1 \gamma(\tau, V_i)dV_i + \int_0^1 \alpha(\tau, V_i)dV_i \\
&= x_{1it}'\bar{\beta}(\tau) + x_{2i}'\bar{\gamma}(\tau)
\end{aligned}
$$

because we have normalized $E[\alpha_i(\tau)] = 0$. It follows that the linear projection of $Q(\tau, y_{it}|x_{1it}, x_{2i}, v_i)$ on x_{1it} and x_{2i} identifies the average effects when these effects are heterogeneous. Thus, our model identifies the average effect for all individual at the τ quantile of their conditional distribution.[10]

Note that it would be possible to analyze both dimensions of heterogeneity. For any $0 < \theta < 1$, if we normalize $\alpha(\tau, \theta) = 0$,

$$
Q(\theta, Q(\tau, y_{it}|x_{1it}, x_{2i}, v_i)|x_{1it}, x_{2i}) = x_{1it}'\beta(\tau, \theta) + x_{2i}'\gamma(\tau, v_i)
$$

All coefficients have two quantile indices: one for the heterogeneity across individuals and one for the heterogeneity within individuals.[11] These heterogeneous coefficients can be estimated with quantile regression in the first and second stage. Finally, it is also possible to identify the effect during the average period of individuals at the θ quantile of the distribution. This parameter requires a least squares regression in the first stage and a quantile regression in the second stage. These alternative strategies identify different parameters and are outside the scope of this paper. We focus instead on the model defined by equations (3) and (4), which is the same as in Chetverikov et al. (2016) and nests the fixed effects quantile regression model (e.g. in Galvao et al. (2020)).

Throughout the paper, we will use the following notation. Let $\tilde{x}_{it} = (1, x_{1it}')'$ and $x_{it} = (x_{1it}', x_{2i}')'$. For each individual i we define the following matrices. The $T \times K_1$ matrix of time-varying regressors $X_{1i} = (x_{1i1}, x_{1i2}, \ldots, x_{1iT})'$, the $T \times (K_1 + 1)$ matrix of first-stage regressors $\tilde{X}_{1i} = (\tilde{x}_i, \tilde{x}_{i2}, \ldots, \tilde{x}_{iT})'$, the $T \times K$ matrix containing all regressors $X_i = (x_{i1}, x_{i2} \ldots, x_{1iT})'$ and the $T \times L$ matrix of instruments $Z_i = (z_{i1}, z_{i2}, \ldots, z_{iT})'$. Further, we define three matrices for all observations. The $NT \times K_1$ matrix of time-varying regressors $X_1 = (X_{11}', \ldots, X_{1T}')'$, the $NT \times K$ matrix of regressors for all individuals $X = (X_1', \ldots, X_N')'$ and the $NT \times L$ matrix of instruments for all individuals as $Z = (Z_1', \ldots, Z_N')'$. We let Y be the $NT \times 1$ vector of the response variable. We use the matrices $P_i = \mathbf{l}_i(\mathbf{l}_i'\mathbf{l}_i)^{-1}\mathbf{l}_i'$ and $Q_i = I_i - P_i$ where $\mathbf{l}_i$ be a $T \times 1$ vector of 1. $P_i\tilde{X}_i$ gives the individual specific means, and $Q_i X_{1i}$ yields a matrix of deviation for individual means. Finally, we define the transformed matrix for all individuals as $P\tilde{X}$ and QX_1 where $P = \text{diag}\{P_1, P_2, \ldots, P_N\}$ and $Q = \text{diag}\{Q_1, Q_2, \ldots, Q_N\}$.

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

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

Before introducing the estimator, we motivate our approach for average effects in standard panel data settings. We discuss some equivalence results of various linear estimator methods and show that common panel data estimators can be estimated using a two-stage approach. A more detailed discussion is included in Appendix A, and the proofs are in Appendix B.1.

2.3 Quantile Estimator

In the last subsection, we have seen that in a linear model, most common panel data estimators can be computed using a two-step approach. In this subsection, we suggest a quantile version of the two-steps procedure to estimate model 3. While for linear models using the two-step procedure is algebraically identical to the one-step estimators, with quantile regression, the two-step procedure changes the conditioning set of the quantile function and therefore affects the estimand. In the first step, for each individual i and quantile τ, regress y_{it} on the time varying variables x_{1it} and a constant using quantile regression. The intercept of the first stage regression, captures both the individual effect $\alpha_i(\tau)$ and the term $x'_{2i}\gamma(\tau)$ as these vary only between individuals. Then, in a second step, we regress the fitted values of the first stage on x_{1it} and x_{2i} using GMM with instruments z_{it}.

The first stage regression solves the following minimization problem for each individual and quantile separately:

$$\hat{\beta}_i(\tau) \equiv \left(\hat{\beta}_{0,i}, \hat{\beta}'_{1,i}\right)' = \underset{(b_0,b_1)\in\mathbb{R}^{d_x+1}}{\arg\min} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_{it} - b_0 - x'_{1it}b_1), \tag{5}$$

where $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function. The true vector of coefficients for individual i is given by $\beta_i(\tau) = (\alpha_i(\tau) + x'_{2i}\gamma(\tau), \beta(\tau)')'$. The fitted value for individual i in period t is $\hat{y}_{it}(\tau) = \hat{\beta}_{0,i}(\tau) + x'_{1it}\hat{\beta}_{1,i}(\tau)$.

The second stage consists in a GMM regression using $\mathbb{E}[g_i(\delta,\tau)] = 0$ as a moment condition, with $g_i(\delta,\tau) = Z'_i(\tilde{X}_i\hat{\beta}(\tau) - X_i\delta(\tau))$. Thus, the moment restriction depends on the first stage and is the sample counterpart of $\mathbb{E}[Z'_i\alpha_i(\tau)]$. Denote the $T\times 1$ column vector of fitted values for individual i by $\hat{Y}_i = (\hat{y}_{i1}, \hat{y}_{i,2}, \ldots, y_{i,T})'$, and the $NT \times 1$ vector of fitted values by $\hat{Y} = (\hat{y}'_1, \ldots, \hat{y}'_n)'$. The second stage estimator can be written as[12]

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

where $\hat{W}(\tau)$ is a $L\times L$ symmetric weighting matrix and $\delta = (\beta', \gamma')'$ is the K-dimensional vector of coefficients. If $L = K$, the second step estimator in 6 simplifies to the IV estimator using Z as instrument. With an over-identified model ($L > K$), we can increase efficiency by using the asymptotically optimal weight matrix $\hat{W}(\tau) = \hat{S}(\tau)^{-1}$, where $\hat{S}(\tau)$ is a consistent estimator of $S(\tau) = \mathbb{E}[g_i(\delta,\tau)g_i(\delta,\tau)']$. This procedure, yields an efficient estimator given the first stage.

[12]Throughout the paper, we consider different second stage estimators. For example, in a variation of the second stage that we consider in section 4, we use optimal instruments. Here, we present the second stage as a GMM estimator, since most estimators are special cases of GMM.

Efficient estimation of β_i is studied in Newey and Powell (1990). Further, it would be possible to consider a Chernozhukov and Hansen (2006) IV quantile regression first stage, followed by a GMM second stage.[13] However, this paper does not explore these possibilities.

The two-step procedure that we consider in this paper has several advantages. First, the linear second stage allows for a large degree of flexibility. For example, instrumental variables, and panel data methods are straightforward to implement. The second advantage is computational. Quantile regression, which is computationally demanding due to the non-differentiable objective function, is used only in the first stage, where there are fewer observations and a limited number of parameters to estimate. Parallelization of the first stage regressions enables to further increase computational speed. For this reason, our estimator remains computationally attractive in large datasets with numerous individuals. Third, as shown before, if our two-step estimator is applied with OLS in the first stage, then our estimator is algebraically equivalent to the corresponding one-step estimators. Therefore, providing an intuitive justification for our approach.

The estimator can be written as a MD estimator, where the second stage imposes restrictions on the first stage coefficients. For simplicity, we consider the case where all regressors are exogenous and $Z = X$. The MD estimator minimizes

$$\hat{\delta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{N} (\hat{\beta}_i(\tau) - R_i\delta(\tau))'\tilde{W}(\tau)(\hat{\beta}_i(\tau) - R_i\delta(\tau)) \tag{7}$$

where $\tilde{W}(\tau)$ is a $K \times K$ weighting matrix that might depend on the quantile index. R_i is defined such that $\tilde{X}_i R_i = X_i$, that is

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

Similarly to Galvao and Wang (2015), the efficient MD estimator could be implemented by substituting $\tilde{W}$ with the estimated first-stage covariance matrix denoted $\hat{V}_i$. The solution to the minimization problem in equation 7 is

$$\hat{\delta}_{EMD}(\tau) = \left(\sum_{i=1}^{N} R'_i \hat{V}_i(\tau)^{-1} R_i \right)^{-1} \sum_{i=1}^{N} R'_i \hat{V}_i(\tau)^{-1} \hat{\beta}_i(\tau). \tag{8}$$

If instead $\tilde{W} = \tilde{X}'_i\tilde{X}_i$ then the estimator in equation 8 is algebraically identical to using OLS in the second stage.

Two remarks about the efficient MD estimator follow. First, except in the fixed effects case, we are treating $\alpha_i(\tau)$ as part of the error term. For this reason, the efficient MD estimator using the inverse of the first stage variance as a weighting matrix ignores $\alpha_i(\tau)$ and is, thus, inefficient. This is not the case in Galvao and Wang (2015) as they estimate $\alpha_i(\tau)$.[14] Second,

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

[14] The efficient MD estimator of Galvao and Wang (2015) is numerically identical to our estimator using an IV second stage with instrument $Z_i^*(\tau) = \tilde{X}_i(\tilde{X}'_i\tilde{X}_iV_i(\tau)\tilde{X}'_i\tilde{X}_i)^{-1}\tilde{X}'_iZ_i = (\tilde{X}_iV_i(\tau)\tilde{X}'_i)^{+}Z_i$, where Z_i contains a constant, x_{1it}, and individual dummies. The symbol $^{+}$ denotes the Moore-Penrose inverse.

when $\alpha_i(\tau) = 0$ for all i and τ, efficient MD can alternatively be implemented using optimal instruments for the moment equation $\mathbb{E}[\tilde{X}_i\hat{\beta}(\tau) - X_i\delta(\tau)|X_i] = 0$ (see Appendix B.2).[15] In this paper, to improve efficiency of our estimator, the GMM approach will exploit the covariance structure implied by $\alpha_i(\tau)$.

3 Asymptotic Theory

This section states the assumptions and presents the asymptotic results. Let D_C be a strictly positive and finite constant. All the proofs are included in Appendix B.3.
To understand the behavior of our estimator, it is useful to start writing the sampling error of $\hat{\delta}(\tau)$ as a sum of a component arising from the first stage estimation error of $\beta_i(\tau)$ and a component arising from the second stage noise α_i:

Lemma 1 **(Sampling error).** *Assume that the model in equation (3) holds, then*

$$\hat{\delta}(\tau) - \delta(\tau) = \left(S'_{ZX}\hat{W}S_{ZX}\right)^{-1} S'_{ZX}\hat{W}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}\left(\tilde{x}'_{it}(\hat{\beta}_i(\tau) - \beta_i(\tau)) + \alpha_i(\tau)\right)$$

where $S_{ZX} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}x'_{it}$

We now make assumptions to 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 assumptions that are equivalent to the assumptions required in their Theorem 2:

Assumption 1 **(Sampling).** *(i) The processes $\{(y_{it}, x_{it}, z_{it}) : t = 1, ..., T\}$ are independent across i. (ii) For each i, the observations $\{(y_{it}, x_{1it}, z_{1it})\}$ are i.i.d. across t.*

Assumption 2 **(Covariates).** *(i) For all $i = 1, ..., N$ and all $t = 1, ..., T$ $||x_{it}|| \leq D_C$ a.s. (ii) The eigenvalues of $\mathbb{E}[x_{1it}x'_{1it}]$ are bounded away from zero and infinity uniformly across i.*

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

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

and

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

Assumption 4 **(Bounded density).** *Let $\mathcal{T}$ be an open neighborhood of τ. Assume that uniformly across i, there exists a constant $f_{min} < f_{max}$ such that*

$$0 < f_{min} \leq \inf_{i} \inf_{\eta\in\mathcal{T}} \inf_{x\in\mathcal{X}} f_{y_{it}|x_{1it}}(q_{i,\eta}(x)|x)$$

[15] Section 4 provides a more detailed discussion about optimal instruments. If α_i varies across individuals the optimal instrument would it into account.

These are quite standard assumptions in the quantile regression literature. In Assumption 1, we assume that the observations are i.i.d. over time 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. We also assume that the processes are independent across i; this assumption can also be relaxed by allowing for clustering between individuals. Assumption 2 requires that the regressors are bounded and that $\mathbb{E}\left[x_{1it}x'_{1it}\right]$ 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:

Assumption 5 (Instruments). *(i) For all $i = 1, \ldots, N$ and all $t = 1, \ldots, T$, $||z_{it}|| \leq D_C$ a.s. (ii) For all $i = 1, \ldots, N$ and all $t = 1, \ldots, T$, $\mathbb{E}[z_{it}\alpha_i(\tau)] = 0$. (iii) As $N \to \infty$, $N^{-1}\sum_{i=1}^{N}\mathbb{E}[x'_{it}z_{it}] \to \Sigma'_{ZX}$ where the singular values of Σ'_{ZX} are bounded from below and from above. (iv) For all $i = 1, \ldots, N$ and all $t = 1, \ldots, T$, y_{it} is independent of z_{it} conditional on $(x_{it}, \alpha_i(\tau))$.*

Assumption 6 (Individuals effects). *(i) For all $i = 1, \ldots, N$, $\mathbb{E}[\alpha_i(\tau)^4]$ exists and is finite. (ii) As $N \to \infty$, $N^{-1}\sum_{i=1}^{N}\mathbb{E}[\alpha_i(\tau)^2 z_{it}z'_{it}] \to \Omega_2$.*

For the instrumental variables, we assume that (i) they are bounded, (ii) they are not correlated with the individual effect (exclusion restriction), (iii) they do not affect the first stage estimation (this is often satisfied by construction, e.g. when the instrument does not vary within individuals or is included in the first stage regressors), and (iv) they satisfy the relevance conditions. We also assume that the individual effects have finite fourth moment and the average variance of $z_{it}\alpha_i(\tau)$ converges to a well-define matrix.

Quantile regression is a nonlinear estimator that is potentially biased in finite samples. Thus, for consistency T must increase to infinity. For unbiased asymptotic normality, we need that the bias decreases more quickly than the standard error of the estimator. The bias of the first stage quantile regression estimator is of order $\frac{1}{T}$. We will see that $\hat{\gamma}(\tau)$ converges at the $\sqrt{N}$ rate such that we need that T goes to infinity more quickly than $\sqrt{N}$. On the other hand, $\hat{\beta}(\tau)$ converges at the $\sqrt{NT}$ rate such that we need that T goes to infinity more quickly than N. We state these three different relative growth rate in the following assumption:

Assumption 7 (Growth rates).

(a) $\frac{\log N}{T} \to 0$

(b) $\frac{\sqrt{N}\log T}{T} \to 0$

(c) $\frac{N(\log T)^2}{T} \to 0$

For consistency, T must increase to infinity but a very slow rate of increase is sufficient.

Theorem 1 **(Consistency).** *Let the model in equation (3) and Assumptions 1-6 as well as Assumption 7(a) hold. Assume that $\hat{W} \to W$ such that W is invertible. Then,*

$$\delta(\tau) \underset{p}{\to} \delta \text{ as } (T, N) \to \infty \text{ and } \frac{\log N}{T} \to 0$$

The proof of this proposition follows quite straightforwardly from the results in Galvao and Wang (2015).

We distinguish between two sorts of instruments: L_1 instruments in z_{1it} vary only within individuals, while L_2 instruments in z_{2it} vary also between groups. If an instrument varies both between and within groups, it must be classified in z_{2it}. Since z_{1it} vary only within group, without lack of generality we can normalize $\bar{z}_{1i} = T^{-1}\sum_{t=1}^{T} z_{1it} = 0$.[16] The moments associated with z_{1it}, for instance $x_{it} - \bar{x}_i$ in a fixed effects model, converge to zero at the $\sqrt{NT}$ rate because there are exploiting variations in all observations. On the other hand, the moments associated with z_{2it}, for instance $\bar{x}_i$ in a between model, converge to zero only at the $\sqrt{N}$ rate. Adding additional periods does not help if the instruments varies only between individuals. $\gamma(\tau)$ is identified only by between individuals variation. Thus, it converges at the $\sqrt{N}$ rate. The following theorem provides its first-order asymptotic distribution.

Theorem 2 **(Asymptotic distribution of $\hat{\gamma}$).** *Assume that conditions 1-6 and 7(b) hold. Assume also that $\hat{W} \underset{p}{\to} W$ such that W_{22}, the $L_2 \times L_2$ submatrix of W, is symmetric and positive definite. Then*

$$\sqrt{N}(\hat{\gamma}(\tau) - \gamma(\tau)) \underset{d}{\to} N(0, \left(\Sigma_{22}' W_{22} \Sigma_{22}\right)^{-1} \Sigma_{22}' W_{22} \Omega_{22} W_{22} \Sigma_{22} \left(\Sigma_{22}' W_{22} \Sigma_{22}\right)^{-1})$$

where Σ_{22} is the $L_2 \times K_1$ bottom-right submatrix of Σ_{ZX} and Ω_{22} is the $L_2 \times L_2$ bottom-right submatrix of Ω_2 defined in Assumption 6.

Note that the dependence of the weighting matrix W on τ is suppressed for brevity of notation. The first-stage estimation error does not appear in this first-order distribution of $\hat{\gamma}$ because it converges to zero at a quicker rate. To improve inference we will later keep this term and provide standard errors that take it into account. The asymptotic distribution in Theorem 2 is the same as in Chetverikov et al. (2016) but we were able to weaken the growth rate condition from $\frac{N^{2/3}\log T}{T} \to 0$ to $\frac{N^{1/2}\log T}{T} \to 0$ by exploiting new results in Galvao et al. (2020). In the appendix we actually show the more general result that the whole estimated quantile regression process converges to a mean zero Gaussian process.

The parameters $\beta(\tau)$ are identified by the time-varying instruments z_{1it} and by the time-constant instruments z_{2i}. Under the stronger growth rate condition 7(c), all sample moment conditions are asymptotically normally distributed and centered at zero. Define the sample moment condition $\bar{g}_{NT}(\delta, \tau) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}(\delta, \tau)$.

[16] In the presence of a constant, the estimator is numerically identical if we recenter the instruments that do not vary between groups.

Lemma 2 (**Asymptotic distribution of the sample moments**). *Under Assumptions 1-6 and 7(c),*

$$\sqrt{N}\bar{g}_{NT}(\delta,\tau) \underset{d}{\to} N\left(0, \frac{\Omega_1}{T} + \Omega_2\right) \tag{9}$$

where $\Omega_1 = \mathbb{E}\left[\Sigma_{ZX_i} V_i \Sigma'_{ZX_i}\right]$, $\Sigma_{ZX_i} = \mathbb{E}[z_{it}x'_{it}]$, *and* $\Omega_2 = \mathbb{E}[\bar{z}_i \bar{z}'_i \alpha_i(\tau)^2]$. *Only the* $L_2 \times L_2$ *bottom-right submatrix of* Ω_2 *contains elements that are different* 0. *It follows that the first-order asymptotic distribution of* $\bar{g}_{NT}$ *is*

$$\Lambda_{NT} \cdot \bar{g}_{NT}(\delta,\tau) \underset{d}{\to} N(0,\Omega) \tag{10}$$

where

$$\Lambda_{NT} = \begin{pmatrix} \sqrt{NT} I_{L_1} & 0 \\ 0 & \sqrt{N} I_{L_2} \end{pmatrix}$$

and

$$\Omega = \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{pmatrix}$$

Ω_{11} *is the* $L_1 \times L_1$ *upper-left submatrix of* Ω_1 *and* Ω_{22} *is the* $L_2 \times L_2$ *bottom-right submatrix of* Ω_2.

Thus, there are L_1 sample moments that converge at the $\sqrt{NT}$ rate (the fast moments) and L_2 sample moments that converge at the $\sqrt{N}$ rate (the slow moments). The rate of convergence of the estimator $\hat{\delta}$ is determined by the rate of convergence of the moments that are used asymptotically to identify this parameter. If a parameter is asymptotically estimated using only the fast moments, then this estimator will converge at the $\sqrt{NT}$ rate, otherwise it will converge at the $\sqrt{N}$ rate. Remember that the first L_1 moments converge at a fast rate because they vary over time. Naturally, only the coefficients of time-varying regressors can be identified by these moments. To simplify notation, we assume in the following that the coefficients $\beta(\tau)$ on all time-varying regressors x_{1it} are identified using only instruments varying only within individuals z_{1it}.

In general, if we do not restrict the weighting matrix $\hat{W}$, then all elements of $\hat{\delta}$ will converge at the slow rate because the slow moments will contaminate even the estimators of parameters that would be identified using only the fast moments. To avoid this, we impose the following restrictions on the weighting matrix:

Assumption 8 (**Weighting matrix**). $\hat{W} \underset{p}{\to} W$ *such that*

$$W = \begin{pmatrix} W_1 T & 0 \\ 0 & 0 \end{pmatrix} + W_2 = \begin{pmatrix} W_1 T & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$$

where the $L_1 \times L_1$ *matrix* W_1 *and the* $L_2 \times L_2$ *matrix* W_{22} *are symmetric and strictly positive-definite.*

Asymptotically, the fast moments get weighted infinitely more than the slow moments such that the parameters identified by the fast moments will converge at the $\sqrt{NT}$ rate. The parameters that are identified only by the slow moments will not be affected by W_1 such that there asymptotic distribution will depend only on W_2.[17]

Theorem 3 (First-order asymptotic distribution). *Let Assumptions 1-6, 7(c) and 8 hold. Then,*

$$\sqrt{NT}(\hat{\beta}(\tau) - \beta(\tau)) \underset{d}{\rightarrow} N(0, \left(\Sigma_{11}' W_1 \Sigma_{11}\right)^{-1} \Sigma_{11}' W_1 \Omega_{11} W_1 \Sigma_{11} \left(\Sigma_{11}' W_1 \Sigma_{11}\right)^{-1})$$
$$\sqrt{N}(\hat{\gamma}(\tau) - \gamma(\tau)) \underset{d}{\rightarrow} N(0, \left(\Sigma_{22}' W_{22} \Sigma_{22}\right)^{-1} \Sigma_{22}' W_{22} \Omega_{22} W_{22} \Sigma_{22} \left(\Sigma_{22}' W_{22} \Sigma_{22}\right)^{-1})$$

$\hat{\beta}(\tau)$ and $\hat{\gamma}(\tau)$ are first-order independent.

This first-order asymptotic distribution is not completely satisfactory. Consider for example the random effects case where all the regressors are time-varying. The vector of instruments consists of $x_{it} - \bar{x}_i$ and $\bar{x}_i$ (This corresponds to the random effects estimator). If we consider only the first-order asymptotic distribution, then we can obtain a first-order efficient by giving zero weights to the slow moments. In other words, the instruments $\bar{x}_i$ are simply not used because their contribution is asymptotically negligible. Note that this is not specific to quantile models and also affect least squares models with large T (see Ahn and Moon, 2014). However, the slow moments can improve finite sample performance since these moments are still informative. For this reason, we do not implement the first-order efficient weighting matrix but the second order optimal weighting matrix, which is

$$W^* = \left(\Omega_1/T + \Omega_2\right)^{-1}.$$

Define $\Lambda_T = \Lambda_{NT}/\sqrt{N}$. The optimal weight matrix can be estimated by $\hat{W}(\tau) = \Lambda_T \hat{\Omega}(\tau)^{-1} \Lambda_T$ where

$$\hat{\Omega}(\tau) = \Lambda_T \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \hat{u}_i(\tau) \hat{u}_i(\tau)' Z_i \right) \Lambda_T.$$

The $T \times 1$ vector $\hat{u}_i(\tau)$ contains the residuals from a preliminary second stage using some inefficient weight matrix, that is $\hat{u}_i(\tau) = \tilde{X}_i \hat{\beta}_i(\tau) - X_i \tilde{\delta}(\tau)$ where $\tilde{\delta}$ is a preliminary estimate of δ. The proof that $\hat{\Omega}(\tau) \xrightarrow{p} \Omega(\tau)$ follows directly by the proof of proposition 1 below, and is therefore omitted.

The covariance matrix can be easily estimated in the second stage with a cluster robust covariance matrix estimator. To estimate the asymptotic covariance matrix, define the $T \times 1$ vector of residuals $\tilde{u}_i(\tau) = \tilde{X}_i \hat{\beta}(\tau) - X_i \hat{\delta}(\tau)$. Then the covariance matrix of $\hat{\delta}(\tau)$ denoted $\hat{V}_\delta(\tau) = \widehat{\mathrm{Var}}(\hat{\delta}(\tau))$ is estimated by

$$\hat{V}_\delta(\tau) = \left(X' Z \hat{W} Z' X\right)^{-1} X' Z \hat{W} \left(\sum_{i=1}^{N} Z_i' \tilde{u}_i(\tau) \tilde{u}_i(\tau)' Z_i \right) \hat{W} Z' X \left(X' Z \hat{W} Z' X\right)^{-1}.$$

[17] As we can see in the proof of proposition 3, the fact that W_1 is multiplied with T does not cause a problem because the weighting matrix matters only up to scale.

Proposition 1 (Consistency of the covariance matrix). *Let Assumptions 1-6, 7(c) and 8 hold. Further, assume that* $\frac{T}{N^{3/2}} \to 0$. *Then* $\Lambda_{NT}\hat{V}_{\delta}(\tau)\Lambda_{NT} \xrightarrow{p} G\Omega G'$,
where $G = \left(\Lambda_T^{-1}S_{ZX}'\hat{W}S_{ZX}\Lambda_T^{-1}\right)^{-1}\Lambda_T^{-1}S_{ZX}'\hat{W}\Lambda_T^{-1}$.

The covariance matrix estimator, does not require estimation of the density of the first stage, and it is computationally easy to compute. Further, the clustering will implicitly take the first stage error into account.
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 can test the hypothesis $\mathbb{H}_0 : \mathbb{E}[Z_i'\alpha_i(\tau)] = 0$. Compared to a traditional GMM, our overidentification test has to deal with the potential different convergences rate of the elements of $\hat{\delta}$. We solve this issue by rescaling the estimated variance by Λ_T. Let $g_i(\delta,\tau) = Z_i'\left(\hat{Y}_i(\tau) - X_i\delta(\tau)\right)$ and $\bar{g}_N(\delta,\tau) = \frac{1}{N}\sum_{i=1}^{N} g_i(\delta,\tau)$. Define the GMM criterion function

$$J\left(\hat{\delta}(\tau)\right) = N\bar{g}_N(\hat{\delta},\tau)'\hat{S}^{-1}(\tau)\bar{g}_N(\hat{\delta},\tau) \tag{11}$$

where $\hat{S} = \Lambda_T^{-1}\hat{\Omega}\Lambda_T^{-1}$ the inverse of the second order optimal weighting matrix.

Proposition 2. *Under the* $\mathbb{H}_0$ *and Assumptions 1-6, 7(c) and 8 as* T *and* $N \to \infty$, $J(\hat{\delta}(\tau)) \xrightarrow{d} \chi^2_{L-K}$.

This result shows that 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 to assess as a generalization of the Hausman Test for random effects.

4 Traditional Quantile Panel Data Estimators

4.1 Fixed Effects, Random Effects and Between Estimators

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 estimation uses only data for one individual at the time and is unaffected. In the second stage, as for least squares estimation (see section 2.1), we compute panel data estimators by selecting different instruments. Depending on the model, the instrument z_{it} will be defined so that the orthogonality condition holds. More precisely, for fixed effects estimation, the instrument z_{it} will contain the demeaned regressor $\dot{x}_{1it}$ and varies only within i. For between estimator z_{it} equals the individual mean of the regressors $\bar{x}_{it}$. Finally, for the pooled estimator, $z_{it} = x_{it}$.[18]

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

Implementing efficient estimation is one of the main challenges of the quantile random effects estimator as the model is overidentified. We suggest two different random effects estimators. The first is an efficient GMM estimator, while the second uses optimal instruments. Given the first stage, we have the following moment restriction:

$$\mathbb{E}[Z_i'(\tilde{X}_i\hat{\beta}_i(\tau) - X_i\delta(\tau)] = 0. \quad (12)$$

If the instrument Z_i contains both the mean and the demeaned regressors, the efficient GMM will optimally weight the within and between variation to obtain a random effect estimator. The moment condition in equation 12 contains both fast and slow moments, but the fast moments are sufficient to identify the coefficients on the time-varying regressors. Consequently, 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 the second-order 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 $T \to \infty$, 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 12 hold conditional on Z_i, we can use the theory of optimal instrument to derive a random effects estimator. Optimal instruments are relevant when a researcher has a conditional moment restriction of the form $\mathbb{E}[g_i(\delta, \tau)|Z_i] = 0$. When a moment condition holds conditional on Z_i, 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 $Z_i^* = \mathbb{E}[g_i(\delta,\tau)g_i(\delta,\tau)'|Z_i]^{-1}R_i(\delta,\tau)$ where $R_i(\delta,\tau) = \mathbb{E}[\frac{\partial}{\partial\delta}g_i(\delta,\tau)|Z_i]$ (see, e.g., Chamberlain, 1987 and Newey, 1993).

To implement the random effect estimator with optimal instrument we set $Z_i = X_i$. Under the additional assumption that $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$, the optimal instrument simplifies to $Z_i^*(\tau) = \left(\tilde{X}_i\frac{V_i(\tau)}{T}\tilde{X}_i' + \mathbf{l}_T'\mathbf{l}_T\sigma_\alpha^2(\tau)\right)^+ X_i$, where $V_i(\tau)$ is the asymptotic variance from the first stage for an individual i and $^+$ denotes the Moore-Penrose inverse.[19] If $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$, 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 with the true outcome or the first stage fitted values imply the same optimal instrument. To put it differently, with a least squares first stage and under usual random effects assumptions, the matrix $\tilde{X}_i\frac{V_i}{T}\tilde{X}_i' + \mathbf{l}_T'\mathbf{l}_T\sigma_\alpha^2$ simplifies to the usual random effects structure. These results are summarized in proposition 5 in Appendix B.1. Third, if $\sigma_\alpha = 0$, this estimator is identical to

[19]Since the matrix $(\tilde{X}_i\frac{V_i(\tau)}{T}\tilde{X}_i' + \mathbf{l}_T'\mathbf{l}_T\sigma_\alpha^2(\tau))$ is singular, we use the Moore-Penrose inverse.

the efficient MD estimator (see proposition 7 in Appendix B.2). Fourth, the optimal instrument depends on T analogously to the second-order optimal weighting matrix of the GMM estimator. As T increases, the first stage variance converges to zero, and the generalized inverse will give infinitely more weights to the within variation, and asymptotically also this estimator converges to the fixed effects estimator.

To make the optimal instrument approach operational, we need a consistent estimator of Z_i^*. In the following, we assume that $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$ and we suggest estimators for $V_i(\tau)$ and $\sigma_\alpha^2(\tau)$. Compared to the classical random effects structure, we use the first stage variance.[20] This formula has two main advantages. First, it is straightforward to compute $\hat{V}_i$. Second, it is possible to allow for dependence in the errors in the first stage regressions.
The first stage variance can be estimated by $\hat{V}_i(\tau) = \hat{A}_i^{-1}(\tau)\hat{B}_i(\tau)\hat{A}_i^{-1}(\tau)$ where $\hat{A}_i(\tau) = \tau(1-\tau)\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'$ and $B_i(\tau)$ can be computed using the Kernel Density estimator of Powell (1991):

$$\hat{B}_i(\tau) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{y_{it} - \tilde{x}_{it}'\beta_i(\tau)}{h}\right)\tilde{x}_{it}\tilde{x}_{it}', \tag{13}$$

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

$$\hat{\sigma}_\alpha^2(\tau) = \frac{N}{N-1}\sum_{i=1}^{N}(\hat{\alpha}_i - \bar{\hat{\alpha}}_i)^2, \tag{14}$$

where $\bar{\hat{\alpha}}_i = \frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i$ and the α_i are estimated by a preliminary least squares dummy variable regression of $\hat{y}_{it}(\tau)$ on x_{it}.[21]

If the assumption that $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$ is violated, we could improve efficiency by using efficient GMM with instruments $(\dot{X}_{1i}, \bar{X}_i)$ as described above. With an OLS first stage, this GMM estimator is asymptotically as efficient as the random effects estimator if the random effects assumptions are correct (Im et al., 1999). When the assumptions are violated, this GMM estimator will be inefficient relative to the GMM estimator using the complete set of moment restrictions. However, it may have better finite sample properties. Conditional on the first stage, the same conclusion holds for our quantile model. Thus, the random effects estimator implemented by efficient GMM, despite not relying on the stronger conditional moment condition, is asymptotically as efficient as the optimal instrument approach. A second advantage of the GMM approach over the optimal instrument is that it does not require a direct estimation of $V_i(\tau)$. Instead, it requires only the consistent estimation of the efficient weighting matrix, which is simpler to estimate.

[20]Using the first stage variance will not impose equality on the estimated densities of the errors $\hat{f}_{Y_i - \tilde{X}_i\beta_i}(0)$ in the second stage. Thus, observations will be weighted differently, depending on the first stage variance.
[21]The estimator can be modified in the case of unbalanced panels.

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 the fixed effects and the random effects. On the one hand, the random effects estimator relies on the orthogonality between $\alpha_i(\tau)$ and x_{it}. On the other hand, the fixed effects estimator only identifies the effect of time-varying variables. To this end, Hausman and Taylor (1981) assume that some elements of X might be correlated with $\alpha_i(\tau)$. We consider model 3 but we partition X into four types of variables, $X = [X_1^x \; X_1^n \; X_2^x \; X_2^n]$, where the superscript x indicates that the variable is exogenous, and the superscript n indicates that it might be endogenous. Thus,

$$\mathbb{E}[X_1^x \alpha_i(\tau)] = 0$$
$$\mathbb{E}[X_2^x \alpha_i(\tau)] = 0.$$

This assumption implies that we can estimate $\delta(\tau)$ using the instrument $Z = (\dot{X}_1^x, \dot{X}_1^n, \bar{X}_1^x, X_2^x)$. While X_2^n is potentially endogenous, the within variation is uncorrelated with $\alpha_i(\tau)$ as it varies only between i. Identification requires that there are at least as many instruments as parameters to estimate. Thus, we need $dim(x_{1it}^x) \geq dim(x_{2i}^n)$. If the model is overidentified, it is possible to implement efficient GMM, and if the conditional moment restrictions are available, optimal instruments can be implemented to obtain a more precise estimator. The optimal instrument is then $Z_i^*(\tau) = \mathbb{E}\left[\left(\tilde{X}_i(\hat{\beta}_i(\tau) - \beta(\tau)) + \alpha_i(\tau)\right)\left(\tilde{X}_i(\hat{\beta}_i(\tau) - \beta(\tau)) + \alpha_i(\tau)\right)' | Z_i\right]^{-1} \mathbb{E}[X_i|Z_i]$. Implementation of the optimal instrument is not straightforward as it requires the estimation of $\mathbb{E}[X_i|Z_i]$ and 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_{1it}^n, so that all time varying 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 than 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 overidentication test presented in section 3 as a quantile version of the Hausman test for our two-step model. 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. Further, he shows that Chamberlain-type tests can also be computed as Wald test, testing KT moment restrictions.[22] Ahn and Low (1996) propose a GMM test based on a 3SLS regression

[22] These type of test are not applicable in our context, as $T \to \infty$.

as an equivalent method for the Hausman test. In section 4, we suggest second-order efficient GMM as a possibility to perform random effects estimation. One worry in implementing a Hausman test for our estimator is that the test would converge to a degenerate distribution as the random effect estimator converges asymptotically to the fixed effects estimator. However, results in Ahn and Moon (2014) show that this is not the case for mean models. Thus, the overidentification test suggested in section 3 directly extends to the random effect estimator. 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_{it} is uncorrelated with $\alpha_i(\tau)$ so that $\mathbb{E}[\dot{X}_{1i}'\alpha_i(\tau)] = 0$ and $\mathbb{E}[\bar{X}_i'\alpha_i(\tau)] = 0$ are a valid moment conditions. By contrast, the fixed effects rely only on the moment condition $\mathbb{E}[\dot{X}_{1i}'\alpha_i(\tau)] = 0$. The overidentification test suggested in section 3 can be used as a test of the $H_0 : \mathbb{E}[\dot{X}_{1i}'\alpha_i(\tau)] = 0$ and $\mathbb{E}[\bar{X}_i'\alpha_i(\tau)] = 0$ where $Z_i = (\dot{X}_{1i}, \bar{X}_i)$.

Thus, the random effects assumption can be tested using an extension of the Hausman Test based on an overidentification GMM test. 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. We present simulation results for both random effects estimator we suggest. These simulations focus on the estimation of $\beta(\tau)$, while the next section includes simulation for $\gamma(\tau)$. We consider the following data generating process where all variables are scalars:

$$y_{it} = \beta x_{1it} + \alpha_i + (1 + 0.1 x_{1it})\nu_{it}. \tag{15}$$

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

$$\begin{pmatrix} h_i \\ \alpha_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \Lambda \\ \Lambda & 1 \end{pmatrix} \right).$$

If $\Lambda \neq 0$, x_{1it} is correlated with α_i. 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 let $\Lambda = \{0, 0.1, 0.2, 0.3, 0.4\}$. The true coefficient takes the values $\beta(\tau) = \beta + 0.1 F^{-1}(\tau)$ where F is the standard normal CDF.

We consider the samples with $T = \{10, 25, 200\}$ and $N = \{25, 200\}$ 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 standard errors. Simulation results of the rejection probabilities of the Hausman test are in Table 3.

As shown in Table 1, the estimator performs well also when both N and T are small. The RE-GMM (random effects implemented by GMM) estimator performs similarly to the RE-OI

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

Quantile	Pooled	BE	FE	RE–OI	RE–GMM
			(N, T) = (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)
			(N, T) = (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)
			(N, T) = (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)
			(N, T) = (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)
			(N, T) = (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)
			(N, T) = (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 2: Standard Errors of $\hat{\beta}(\tau)$

Quantile	Pooled	BE	FE	RE–OI	RE–GMM
		(N, T) = (25, 10)			
0.1	0.201	0.215	0.254	0.158	0.159
0.5	0.188	0.204	0.166	0.147	0.125
0.9	0.201	0.215	0.254	0.158	0.159
		(N, T) = (200, 10)			
0.1	0.067	0.079	0.091	0.064	0.059
0.5	0.063	0.075	0.060	0.059	0.046
0.9	0.067	0.079	0.091	0.064	0.059
		(N, T) = (25, 25)			
0.1	0.183	0.203	0.138	0.112	0.111
0.5	0.177	0.198	0.100	0.099	0.088
0.9	0.183	0.203	0.138	0.113	0.111
		(N, T) = (200, 25)			
0.1	0.061	0.074	0.049	0.042	0.041
0.5	0.060	0.072	0.036	0.036	0.032
0.9	0.061	0.074	0.049	0.042	0.041
		(N, T) = (25, 200)			
0.1	0.171	0.194	0.048	0.046	0.047
0.5	0.170	0.194	0.035	0.034	0.036
0.9	0.171	0.194	0.048	0.046	0.047
		(N, T) = (200, 200)			
0.1	0.058	0.071	0.017	0.016	0.017
0.5	0.057	0.071	0.013	0.012	0.012
0.9	0.058	0.071	0.017	0.016	0.017

Note:
The table reports standard errors of the simulations for $\beta(\tau)$ from 10,000 Monte Carlo simulations. The standard errors are clustered at the individual level.

(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 T increases. Whereas with small T there is a clear gain in using a random effects estimator. The standard deviations show the different rate of convergence of the estimators. The precision of the fixed effects, random effects estimators increases in similar magnitude when N or T increases. In contrast, the standard deviation of the Pooled and between (BE) estimator decreases only when N increases. The pooled and the between estimators have the smallest bias and in most cases also the largest variance.

The standard errors in Table 2 are close to the standard deviations of the simulations, suggesting that our inference procedure performs well also in finite samples. With $T = 10$ the standard errors tent to be slightly undersized in the random effects estimators. The difference is small and decreases quickly as the sample size increase.

Table 3 shows the rejection probabilities of the overidentification test for different values of λ. When $\lambda = 0$, the H_0 is satisfied so that we should be rejecting the null at a rate close to the theoretical size of 5%. If $\lambda \neq 0$, X_i is correlated with α_i some moment conditions used

Table 3: Hausman Test

	λ				
Quantile	0.0	0.1	0.2	0.3	0.4
		(N, T) = (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
		(N, T) = (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
		(N, T) = (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
		(N, T) = (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
		(N, T) = (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
		(N, T) = (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.

by the RE-GMM estimator are not valid. 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 with most sample sizes. The power of the test is higher in large samples and increases the larger the correlation between $P_i X_i$ and the unobserved heterogeneity α_i. An increase in N substantially improves the power of the test, while a larger number of time periods T improves the results to a lesser extent. In general, the test performs better both in terms of size and power when N is large, which is most often the case in empirical applications. While as T increases, the RE-GMM estimator converges to the fixed effects estimator so that RE-GMM coefficients on time-varying regression will be consistent even if $\lambda \neq 0$, the size and power of the test do not deteriorate as T increases. This result is consistent with the findings in Ahn and Moon (2014).

5 Grouped (IV) Quantile Regression Model

In this section, we discuss a special case of our model in which i indexes groups and t indexes any ordering between groups. The model is of practical relevance when a researcher has micro-data on a sample that can be divided into groups. To give an illustration, groups could be schools and students in these schools are the individuals. Variables are divided between group-level and individual-level, instead of time-varying and time-constant. Individual-level variables include students' characteristics, while school facilities is a group-level variable. Similarly, we might define group as county-year combination and individuals observations could be economic agents in these counties. In these models, empirical researchers might include fixed effects at the level of the city or the county in the second stage. An estimator for these models was suggested by Chetverikov et al. (2016).

5.1 Chetverikov et al. (2016)

Chetverikov et al. (2016) considers two different models. The first one is identical to model 3. The second model is as follows:

$$Q_{it,\tau}(x_{1it}, x_{2i}) = \beta_{0,i}(\tau) + x'_{1it}\beta_i(\tau) \tag{16}$$

$$\beta_{0,i}(\tau) = x'_{2i}\gamma(\tau) + \alpha_i(\tau). \tag{17}$$

Compared to the model 3, the coefficient on x_{1it} is allowed to vary over i. Chetverikov et al. (2016) decide to study model 16-17 given its flexibility as it does not impose equality of the coefficients on x_{1it} and because it allows to study interaction effects. They suggest a two-step estimator. The first stage consists in regressing y_{it} on x_{1it} and a constant using quantile regression separately for each group i and quantile τ. In the second stage, they regress the *intercept* from the first stage on x_{2i}. Their estimator focuses on estimating $\gamma(\tau)$ and does not directly provide an estimate of $\beta(\tau)$. Further, the inference procedure of Chetverikov et al. (2016), as we will show in the simulation later, provides a poor approximation in finite samples.

There are two main problems with model 16-17. First, if the true model follows equation 3, the estimator of Chetverikov et al. (2016) (henceforth CLP estimator) does not exploit the equality in β and the exogeneity of x_{1it} between groups. Second, this estimator is consistent for the treatment effect at $x_{1it} = 0$. If the true model corresponds to model 3, then their estimator consistently estimates the QTE for the whole population, but it is not invariant to reparametrization of x_{1it} and it may have poor finite sample properties. If the model is misspecified, their estimator will not converge in general to an interpretable parameter ($x_{1it} = 0$ may be out of the support of x_{1it}). If the treatment effect is heterogeneous in x_{1it}, then the QTE at 0 may not be of particular interest. In such a case, one could parametrize the treatment effect on the random slope and estimate separately the effect of the intercept and the effect on the slope. In a second step, both estimates could be combined to get, for instance, an average (in x_{1it}) QTE.

Using our approach, we can also allow for heterogeneous effects by including interaction terms between x_{1it} and x_{2i}. By estimating simultaneously all the parameters and imposing all the assumptions, we obtain a more efficient estimator.
As the CLP estimator does not exploit the between variation in x_{1it}, it requires different exogeneity assumption for consistency of $\hat{\gamma}(\tau)$. Their estimator remains consistent if x_{1it} is endogenous between groups. To put it differently, the CLP estimator provides a consistent estimate of $\gamma(\tau)$ under the assumption that x_{2i} *or* $x_{2i}|x_{1it}$ is exogenous. By contrast, the MD is consistent under the stronger assumption that $x_{2i}|x_{1it}$ is exogenous with respect to α_i. While this might seem a limitation, it is straightforward to recover consistency with our MD estimator by using $(Q_i X_{1i}, \mathbf{1}_i x_{2i})$ as instruments in the second stage.

5.2 Simulations

This subsection presents Monte Carlo simulations comparing the MD estimator with the CLP estimator. The simulation as based on the same data generating process and sample sizes of Chetverikov et al. (2016). That is, $(T, N) = \{(25, 25), (200, 25), (25, 200), (200, 200)\}$. The data generating processes includes one time-invariant regressor, one time-varying regressors and one instrument. Heterogeneity is introduced via a rank variable u_{it}. Since the effect of the individual-level covariates is constant across groups, $\beta(u) = (\beta_{i,0}(u), \beta_i(u)')' = (\beta_0(u), \beta(u)')'$, where $\beta_{i,0}(u) = \beta_0(u)$ is the constant of the first stage. The data is generated as follows:

$$y_{it} = \beta_0(u_{it}) + x_{1it}\beta(u_{it}) + x_{2i}\gamma(u_{it}) + \alpha_i(u_{it}) \tag{18}$$

$$z_i = x_{2i} + \eta_i + \nu_i \tag{19}$$

$$\alpha_i(u_{it}) = u_{it}\eta_i - \frac{u_{it}}{2} \tag{20}$$

where, x_{1it}, x_{2i} and ν_i are distributed $\exp(0.25 \cdot N[0, 1])$ and η_i as well as the rank variable u_{it} are $U[0, 1]$ distributed. The data generating process implies that $\mathbb{E}[\alpha(u_{it})|x_{2i}] = \mathbb{E}[u_{it}\eta_i - \frac{u_{it}}{2}|x_{2i}] = \mathbb{E}[\frac{u_{it}}{2} - \frac{u_{it}}{2}|x_{2i}] = 0$. At quantiles $\tau \in (0, 1)$, the true parameters $\gamma(\tau)$ and $\beta(\tau)$ equal $\sqrt{\tau}$ and, $\alpha_1(\tau)$ equals $\frac{\tau}{2}$. Consequently, $\gamma(u_{it}) = \beta(u_{it}) = \sqrt{u_{it}}$ and $\beta_0(u_{it}) = \frac{u_{it}}{2}$. It is worth mentioning that the data generating process of Chetverikov et al. (2016) has a weak instrument when N is small.[23] For this reason, one should keep this in mind when looking at the simulation results. In empirical research, it is straightforward to construct Anderson-Rubin confidence intervals.

The simulations consider three cases. In the first one (baseline), $\alpha_i(\tau) = 0$ for all i and all τ. In this case, conditioning on the individual, does not affect the quantile function thus, quantile regression would be consistent for the same parameter. Further, as $\alpha_i(\tau) = 0$ our estimator is $\sqrt{NT}$-consistent. In the second case, there are individual specific effects ($\alpha_i(\tau) \neq 0$) and these are uncorrelated with the regressors. The DGP multiplies the individual heterogeneity with the

[23] With $N = 25$ in over 40% of the draws, the F-statistics of the first stage of the 2SLS estimation of both estimators is below 10. The issue disappears when $N = 200$.

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

	Baseline			Exogenous			Endogenous		
Quantile	MD	CLP	Rel. MSE	MD	CLP	Rel. MSE	MD	CLP	Rel. MSE
				(N, T) = (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)	
				(N, T) = (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)	
				(N, T) = (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)	
				(N, T) = (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 bias and standard deviation (in parentheses) of the simulations for $\gamma(\tau)$ from 10,000 Monte Carlo simulations using the minimum distance (MD) estimator and the grouped IV quantile regression (CLP) estimator.

rank variable u_{it}. Thus, in the lower tail of the distribution we see a faster convergence rate of the MD estimator in the simulations. In the third case, x_{2i} is endogenous as $\alpha_i(\tau)$ is correlated with the regressor of interest. Thus, we use 2SLS in the second stage. 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 $\gamma(\tau)$. 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 $T = 25$. When T increases to 200, the difference disappears. There are remarkable differences in the standard deviation 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. In the exogenous and endogenous cases, the difference is somewhat

Table 5: Standard Errors and Emprical Sizes

	Baseline				Exogenous				Endogenous			
	s.e.		Empirical Size		s.e.		Empirical Size		s.e.		Empirical Size	
Quantile	MD	CLP	MD	CLP	MD	CLP	MD	CLP	MD	CLP	MD	CLP
					(N, T) = (25, 25)							
0.1	0.201	0.715	0.059	0.106	0.204	0.717	0.061	0.104	10.498	8.897	0.034	0.060
0.5	0.172	0.578	0.060	0.100	0.215	0.592	0.058	0.098	10.102	9.881	0.036	0.060
0.9	0.097	0.362	0.059	0.094	0.243	0.415	0.060	0.096	10.550	8.956	0.044	0.060
					(N, T) = (200, 25)							
0.1	0.066	0.279	0.068	0.061	0.067	0.280	0.068	0.058	0.106	0.450	0.058	0.053
0.5	0.056	0.223	0.054	0.064	0.069	0.229	0.055	0.061	0.096	0.351	0.048	0.057
0.9	0.032	0.140	0.075	0.059	0.076	0.158	0.059	0.062	0.086	0.226	0.047	0.053
					(N, T) = (25, 200)							
0.1	0.073	0.247	0.057	0.103	0.078	0.249	0.057	0.101	11.327	26.243	0.032	0.057
0.5	0.063	0.212	0.063	0.100	0.141	0.238	0.062	0.098	19.497	21.399	0.051	0.062
0.9	0.031	0.103	0.052	0.101	0.229	0.217	0.066	0.094	21.143	16.644	0.062	0.072
					(N, T) = (200, 200)							
0.1	0.024	0.096	0.056	0.063	0.025	0.096	0.053	0.062	0.039	0.155	0.049	0.057
0.5	0.020	0.081	0.054	0.062	0.045	0.091	0.048	0.058	0.052	0.132	0.043	0.054
0.9	0.010	0.040	0.058	0.053	0.072	0.081	0.050	0.056	0.075	0.094	0.046	0.051

Note:
The table reports standard errors and empirical sized of the simulations for $\gamma(\tau)$ from 10000 Monte Carlo simulations using the minimum distance (MD) estimator and the grouped IV quantile regression (CLP) estimator. Standard errors of the MD estimator are clustered at the individual level. Robust standard errors are used for the CLP estimator.

smaller but remains substantial.[24] This difference in precision explains the large discrepancies in MSE between the two estimators. The MSE of the CLP estimator is over 10 times larger than that of the MD estimator when $\alpha_i(\tau) = 0$ and remains substantially larger in all scenarios. If $\alpha_i(\tau) = 0$, quantile regression is a consistent estimator for $\beta(\tau)$. Although not shown here, simulation results comparing our estimator with traditional quantile regression show that in large samples, the two estimators are indistinguishable in terms of bias and variance.

Table 5 shows the standard errors and empirical sizes. Comparing the standard errors with the standard deviations of table 4 it is visible that the inference procedure suggested in Chetverikov et al. (2016) underestimates the true variance mostly with small N. Consequently, the empirical size for the 5% theoretical level of the CLP estimator can be as high as 10%. The standard errors of the MD distance estimator are remarkably close to the standard deviations of the simulation, even with small T. As T increases, the two become indistinguishable.

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

6 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 time-varying and time-invariant variables can be estimated. The estimator is computationally fast and straightforward to compute. We show that our two-step procedure applied to linear estimators is algebraically identical to traditional one-step estimators. We derive the asymptotic distribution of the estimator. Since T diverges to infinity, the first stage variance would not appear in the first-order asymptotic variance. To improve the approximation in finite samples, we keep higher-order terms in the asymptotic variance. Monte Carlo simulations show that our estimator and the suggested standard errors perform well in finite samples and that compared to the estimator of Chetverikov et al. (2016), the MD estimator has a much smaller MSE. Further, we suggest an overidentification test to test for the random effects assumption.

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.

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

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.

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

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.

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.

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.

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.

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.

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

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.

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

A Traditional Panel Data Models

This section complements subsection ??. We consider a linear version of our estimator, where OLS instead of quantile regression is used in the first stage. We consider the following traditional panel data model for average effects:

$$y_{it} = x_{1it}'\beta + x_{2i}'\gamma + \alpha_i + \varepsilon_{it},$$

where ε_{it} is the idiosyncratic error term. 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 individual i is

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

Let $C(A)$ be the column space of a matrix A. The next 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 $\hat{\delta}_{GMM}^{MD}$ 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 $C(\tilde{X}_i) \subseteq C(Z_i)$, $\hat{\delta}_{GMM}^{MD} = \hat{\delta}_{GMM}$.*

The proof of this Proposition and all subsequent proofs are in Appendix B.1. Proposition 3 implies that any linear model can be computed by a two-step estimator, as long as the matrix of instruments of individual i, Z_i lies in the column space of the matrix of first-stage regressors of individual i, $\tilde{X}_i$.[25] 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 results for the within estimator is summarized in the following Corollary.

Corollary 1. *Denote $\hat{\delta}_{FE}^{MD}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on $\tilde{X}$ with instruments QX_1. Let $\hat{\delta}_{FE}$ be the coefficient vector of the within estimator, that is, of a regression of QY on QX_1. Then $\hat{\delta}_{FE}^{MD} = \hat{\delta}_{FE}$.*

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

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

[25] Since $\tilde{X}_i$ includes a constant, the presence of time-invariant variables in Z_i will not affect its column space.

It is worth noting that the IV approach to these panel data estimators also work in one stage with Y as dependent variable. Further, it is possible to estimate between (within) models using average (demeaned) fitted values and regressors.
Pooled OLS and the between estimator 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 $Z_i = (Q_i X_{1i}, P_i X_i)$, is identical to the random effects estimator. 3SLS estimator is a special case of GMM with weighting matrix $W = \mathbb{E}[Z_i' \tilde{\Omega} Z_i]$ 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 $\hat{\delta}^{MD}_{RE}$ the coefficient vector of a 3SLS regression of $\hat{Y}$ on X with instruments $(Q_i X_{1i}, P_i X_i)$. Let $\hat{\delta}_{RE}$ be the coefficient vector of a random effects regression of Y on X. Then $\hat{\delta}^{MD}_{RE} = \hat{\delta}_{RE}$.*

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 $\mathbb{E}[Y_i - X_i\delta | X_i] = 0$ and $\mathbb{E}[\hat{Y}_i - X_i\delta | X_i] = 0$ imply the same optimal instrument (see Proposition 5 in Appendix B.1).

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_i, whereas 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 assumed to be orthogonal to α_i. No assumption is placed on the relationship between α_i and X_1^n and X_2^n. The model can be estimated by IV using instruments $Z = (QX_1^x, QX_2^n, \; PX_1^x, X_2^x)$ (see, for example, Hansen, 2022). Thus, it follows by Proposition 3 that the Hausman Taylor model can be estimated in two stages.

Proposition 4. *Denote $\hat{\delta}^{MD}_{HT}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on X with instruments $(QX_1^x, QX_2^n, \; PX_1^x, X_2^x)$. Let $\hat{\delta}_{HT}$ be the coefficient vector of the Hausman Taylor Estimator based on a regression Y on X. Then $\hat{\delta}^{MD}_{HT} = \hat{\delta}_{HT}$.*

B Proofs

B.1 Linear Models

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

$$\tilde{P} Z_i = Z_i \tag{21}$$

The MD estimator with a GMM second stage is:

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

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

$$\begin{aligned}
Z'\hat{Y} &= \sum_{i=1}^{N} Z_i \hat{Y}_i \\
&= \sum_{i=1}^{N} Z_i \tilde{X}_i \hat{\beta}_i \\
&= \sum_{i=1}^{N} Z_i \tilde{X}_i (\tilde{X}_i' \tilde{X}_i)^{-1} \tilde{X}_i' Y_i \\
&= \sum_{i=1}^{N} (\tilde{P} Z_i)' Y_i \\
&= \sum_{i=1}^{N} Z_i' Y_i = Z'Y,
\end{aligned}$$

where the third line uses $\hat{Y}_i = \tilde{X}_i \hat{\beta}_i$, the fourth line uses the definition of the OLS estimator in the first stage and the last line uses equation 21. Thus, it follows directly that $\hat{\delta}_{MD}$ equals $\hat{\delta}_{GMM}$. ∎

Proof of Corollary 1. First, note that since $Q_i X_{1i} = \dot{X}_{1i}$, $\dot{X}_{1i}$ lies in the column space of X_{1i}. 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_{1i} with instrument Z_i is algebraically identical to a 2SLS regression with Y_i as dependent variable. Then,

$$\begin{aligned}
\hat{\delta}_{FE}^{MD} &= \left(\sum_{i=1}^{N} Z_i' X_{1i}'\right)^{-1} \sum_{i=1}^{N} Z_i' Y_i \\
&= \left(\sum_{i=1}^{N} \dot{X}_i' X_{1i}'\right)^{-1} \sum_{i=1}^{N} \dot{X}_{1i}' Y_i \\
&= \left(\sum_{i=1}^{N} X_{1i}' Q_i' X_{1i}'\right)^{-1} \sum_{i=1}^{N} X_{1i}' Q_i Y_i \\
&= \left(\sum_{i=1}^{N} \dot{X}_{1i}'{}' \dot{X}_{1i}'\right)^{-1} \sum_{i=1}^{N} \dot{X}_{1i}' \dot{Y}_i = \hat{\delta}_{FE},
\end{aligned}$$

where the second line follows since $Z_i = \dot{X}_{1i}$, the third and last line by $Q_i X_{1i} = \dot{X}_{1i}$, $Q_i Y_i = \dot{Y}_i$ and since Q_i is idempotent. ∎

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

$$\begin{aligned}
\hat{\delta}^{MD}_{BE} &= \left(\sum_{i=1}^{N} Z_i' X_i\right)^{-1} \sum_{i=1}^{N} Z_i' Y_i \\
&= \left(\sum_{i=1}^{N} \bar{X}_i' X_i\right)^{-1} \sum_{i=1}^{N} \bar{X}_i' Y_i \\
&= \left(\sum_{i=1}^{N} X_i P_i' X_i\right)^{-1} \sum_{i=1}^{N} X_i' P_i Y_i \\
&= \left(\sum_{i=1}^{N} \bar{X}_i' \bar{X}_i\right)^{-1} \sum_{i=1}^{N} \bar{X}_i' \bar{Y}_i = \hat{\delta}_{BE}
\end{aligned}$$

where the second line follows since $Z_i = \bar{X}_i$, the third and last line by $P_i X_i = \bar{X}_i$, $P_i Y_i = \bar{Y}_i$ and, since P_i is idempotent. ∎

Proposition 5. *Assume $\mathbb{E}[\varepsilon_{it}^2|X_i] = \sigma_\varepsilon^2$ and $\mathbb{E}[\alpha_i^2|X_i] = \sigma_\alpha^2$. The conditional moments $\mathbb{E}[\hat{Y}_i - X_i\delta|X_i] = 0$ and $\mathbb{E}[Y_i - X_i\delta|X_i] = 0$ imply the same optimal instrument.*

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

$$\begin{aligned}
\mathbb{E}[(\hat{Y}_i - X_i\delta)(\hat{Y}_i - X_i\delta)'|X_i] &= \mathbb{E}[(\tilde{X}_i(\hat{\beta}_i - \beta) + \tilde{X}_i\beta - X_i\delta)(\tilde{X}_i(\hat{\beta}_i - \beta) + \tilde{X}_i\beta - X_i\delta)'|X_i] \quad (22)\\
&= \mathbb{E}[(\tilde{X}_i(\hat{\beta}_i - \beta) + \alpha_i)(\tilde{X}_i(\hat{\beta}_i - \beta) + \alpha_i)'|X_i] \\
&= \tilde{X}_i \frac{V_i}{T} \tilde{X}_i' + \mathbf{l}_T\mathbf{l}_T'\sigma_\alpha^2.
\end{aligned}$$

The matrix $\tilde{X}_i\frac{V_i}{T}\tilde{X}_i' + \mathbf{l}_T\mathbf{l}_T'\sigma_\alpha^2$ is singular, so that we suggest using the Moore-Penrose inverse to construct the optimal instrument.

For the second moment restriction, we have:

$$\begin{aligned}
\mathbb{E}[(Y_i - X_i\delta)(Y_i - X_i\delta)'|X_i] &= \mathbb{E}[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})'|X_i] \\
&= \left(\mathbf{I}_T\sigma_\varepsilon^2 + \mathbf{l}_T\mathbf{l}_T'\sigma_\alpha^2\right).
\end{aligned}$$

Then note that $\left(\mathbf{I}_T\sigma_\varepsilon^2 + \mathbf{l}_T\mathbf{l}_T'\sigma_\alpha^2\right)^{-1} = (\tilde{X}_i\tilde{X}_i^+\sigma_\varepsilon^2 + \mathbf{l}_T'\mathbf{l}_T\sigma_\alpha^2)^+ = (\tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'\sigma_\varepsilon^2 + \mathbf{l}_T'\mathbf{l}_T\sigma_\alpha^2)^+ = (\tilde{X}_i\frac{V_i}{T}\tilde{X}_i' + \mathbf{l}_T'\mathbf{l}_T\sigma_\alpha^2)^+X_i$, where $V_i = (\frac{1}{T}\tilde{X}_i'\tilde{X}_i)^{-1}\sigma_\varepsilon^2$ and since for a full column rank matrix $\tilde{X}_i$, $\tilde{X}_i\tilde{X}_i^+ = I_T$ and $\tilde{X}_i^+ = (\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'$. ∎

Proposition 6. *Denote $\hat{V}_{\hat{\delta}}$ the clustered covariance matrix of $\hat{\delta}$ estimated by a GMM regression of Y on X with instrument Z. Let $\hat{V}_{\hat{\delta}^{MD}}$ 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, $\hat{V}_{\hat{\delta}^{MD}} = \hat{V}_{\hat{\delta}}$.*

We show that correct standard errors can be obtained using a two-stage approach by clustering the standard errors in the second stage at a level weakly higher than the individuals 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

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

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

Proof. Define $Z_g = (z_{1g}, \ldots, z_{n_g g})'$, $X_g = (x_{1g}, \ldots, x_{n_g g})'$, $Y_g = (y_{1g}, \ldots, y_{n_g g})'$ and $\hat{Y}_g = (\hat{y}_{1g}, \ldots, \hat{y}_{n_g g})'$. 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 $\breve{X}_g = \operatorname{diag}\{\tilde{x}_{1g}, \ldots, \tilde{x}_{n_g g}\}$ 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_i is in the column space of $\breve{X}_g$,

$$\breve{P}Z_g = Z_g. \tag{23}$$

Then,

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

where the fourth line follows by 23. Since this holds for all g, the desired result follows directly. ∎

B.2 Optimal Instruments and Minimum Distance

In this subsection, we show that if $\alpha_i(\tau) = 0$ for all i and τ, efficient minimum distance can be implemented by optimal instruments. From equation 22 we have that if $\alpha_i(\tau) = 0$ for all i and all τ, $\mathbb{E}[(\tilde{X}_i\hat{\beta}_i(\tau) - X_i\delta(\tau))(\tilde{X}_i\hat{\beta}_i(\tau) - X_i\delta(\tau))'|X_i] = \tilde{X}_i\frac{V_i(\tau)}{T}\tilde{X}_i'$. This implies the optimal instrument $Z_i^* = (\tilde{X}_i\frac{V_i(\tau)}{T}\tilde{X}_i')^+X_i$. Since T is a scalar, using $Z_i^*(\tau) = (\tilde{X}_iV_i(\tau)\tilde{X}_i')^+X_i$ leads to the same results.

Proposition 7. *The IV regression with instrument $Z_i^*(\tau) = (\tilde{X}_iV_i(\tau)\tilde{X}_i')^+X_i$ equals the efficient MD estimator.*

Proof.

$$\begin{aligned}
\hat{\delta}_{EMD}(\tau) &= \left(\sum_{i=1}^N R_i'\hat{V}_i^{-1}(\tau)R_i\right)^{-1}\left(\sum_{i=1}^N R_i'\hat{V}_i^{-1}(\tau)\hat{\beta}_i(\tau)\right)\\
&= \left(\sum_{i=1}^N X_i'\tilde{X}_i\left(\tilde{X}_i'\tilde{X}_i\hat{V}_i(\tau)\tilde{X}_i'\tilde{X}_i\right)^{-1}\tilde{X}_i'X_i\right)^{-1}\left(X_i'\tilde{X}_i\left(\tilde{X}_i'\tilde{X}_i\hat{V}_i(\tau)\tilde{X}_i'\tilde{X}_i\right)^{-1}\tilde{X}_i'\hat{Y}_i(\tau)\right)\\
&= \left(\sum_{i=1}^N X_i'\left(\tilde{X}_i\hat{V}_i(\tau)\tilde{X}_i'\right)^{+}X_i\right)^{-1}\left(X_i'\left(\tilde{X}_i\hat{V}_i(\tau)\tilde{X}_i'\right)^{+}\hat{Y}_i(\tau)\right) = \hat{\delta}_{OI}(\tau),
\end{aligned}$$

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

B.3 Asymptotic Results

B.3.1 Proof of Theorems 1-3 and Lemmas 1-2

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

$$\begin{aligned}
\hat{\delta}(\tau) &= \left(X'Z\hat{W}Z'X\right)^{-1}X'Z\hat{W}Z'\hat{y}(\tau)\\
&= \left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{it}\tilde{x}_{it}'\hat{\beta}_i(\tau)\\
&= \left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{it}\left(\tilde{x}_{it}'\left(\hat{\beta}_i(\tau) - \beta_i(\tau)\right) + \tilde{x}_{it}'\beta_i(\tau)\right)\\
&= \left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{it}\left(\tilde{x}_{it}'\left(\hat{\beta}_i(\tau) - \beta_i(\tau)\right) + x_{it}'\delta(\tau) + \alpha_i(\tau)\right)\\
&= \delta(\tau) + \left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{it}\left(\tilde{x}_{it}'\left(\hat{\beta}_i(\tau) - \beta_i(\tau)\right) + \alpha_i(\tau)\right)
\end{aligned}$$

∎

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 2(i) and 5(i) and Chebyshev's inequality that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}x_{it}z_{it}' - \mathbb{E}[x_{it}z_{it}']\right) \underset{p}{\rightarrow} 0$$

In addition, by Assumption 5(iii), $N^{-1}\sum_{i=1}^{N}\mathbb{E}[x_{it}'] \rightarrow \Sigma_{ZX}'$. It follows that

$$S_{ZX} \underset{p}{\rightarrow} \Sigma_{ZX} \tag{24}$$

Together with $\hat{W} \underset{p}{\rightarrow} W$ and the invertibility of $\Sigma_{ZX}'W\Sigma_{ZX}$, it follows by the continuous mapping theorem that

$$\left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W} \underset{p}{\rightarrow} \left(\Sigma_{ZX}'W\Sigma_{ZX}'\right)^{-1}\Sigma_{ZX}'W \tag{25}$$

By lemma 1 in Galvao and Wang (2015), $\hat{\beta}_i(\tau)$ is uniformly consistent for $\beta_i(\tau)$ across i. Together with the boundedness of x_{it} in Assumption 2(i) and of z_{it} in Assumption 5(i), it follows that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau)) \underset{p}{\rightarrow} 0 \tag{26}$$

By Assumption 5(ii) $\mathbb{E}[z_{it}\alpha_i(\tau)] = 0$, by Assumption 5(i) z_{it} is bounded, and by Assumption 6)i) $\mathrm{Var}(\alpha_i)$ is finite for all i. It follows that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}\alpha_i(\tau) \underset{p}{\rightarrow} 0 \tag{27}$$

The result of the proposition follows from equations 25, 26, and 27. ∎

Proof of lemma 2. From the proof of Lemma 1 we know that the sample moments are the sum of a term coming from the first stage estimation and a term from the second stage estimation

$$\sqrt{N}\bar{g}_{NT}(\delta,\tau) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}z_{it}\left(\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau)) + \alpha_i(\tau)\right) \tag{28}$$

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

$$\hat{\beta}_i(\tau) - \beta_i(\tau) = \frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it},y_{it}) + R_{it}^{(1)}(\tau) + R_{it}^{(2)}(\tau)$$

where

$$\phi(\tilde{x}_{it},y_{it}) = -B_i^{-1}\tilde{x}_{it}(1(y_{it} \le \tilde{x}_{it}\beta_i(\tau)) - \tau)$$

is i.i.d. with $\mathbb{E}[\phi(\tilde{x}_{it},y_{it})|\tilde{x}_{it}] = 0$, $B_i = \mathbb{E}[f_{y|x}(Q_{y|x}(\tau|\tilde{x}_{it})\tilde{x}_{it}\tilde{x}_{it}']$, $R_{it}^{(2)}(\tau)$ and $\mathbb{E}[R_{it}^{(1)}(\tau)]$ are both uniformly $O_p\left(\frac{\log T}{T}\right)$. It follows that the individual level quantile regression coefficient are normally distributed when $T \rightarrow \infty$:

$$\sqrt{T}\left(\hat{\beta}_i(\tau) - \beta(\tau)\right) \underset{d}{\rightarrow} N\left(0, B_i^{-1}A_iB_i^{-1}\right) \overset{def}{=} N(0, V_i)$$

where $A_i = \tau(1-\tau)\mathbb{E}[\tilde{x}_{it}\tilde{x}_{it}']$.

Plugging-in this representation and defining $\overline{zx}_i = \frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'$ and $\Sigma_{ZXi} \equiv \mathbb{E}[z_{it}\tilde{x}_{it}']$we obtain

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau)) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{ZXi}\frac{1}{T}\sum_{t=1}^{T}\phi(\tilde{x}_{it}, y_{it}) + o_p\left(\frac{1}{\sqrt{T}}\right)$$

where the rate of convergence follows from equation (49) in Galvao et al. (2020) by replacing their weights with $\overline{zx'}_i$, which does not change the rate of convergence since both z_{it} and $\tilde{x}_{it}$ are bounded by our assumptions. Note also that $\mathbb{E}[\phi(\tilde{x}_{it}, y_{it})|Z_i, X_i] = 0$ because $\tilde{x}_{it} \in X_i$, the quantile regression is correctly specified and Z_i is independent of y_{it} conditionally on x_{it} and $\alpha_i(\tau)$. It follows that

$$E\left[\Sigma_{ZXi}\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it}, y_{it})\right] = 0$$

and

$$\begin{aligned}
Var\left(\Sigma_{ZXi}\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it}, y_{it})\right) &= E\left[\Sigma_{ZXi}\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it}, y_{it})\left(\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it}, y_{it})\right)'\Sigma_{XXi}'\right] \\
&= E\left[\Sigma_{ZXi}E\left[\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it}, y_{it})\left(\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it}, y_{it})\right)'|Z_i, \tilde{X}_i\right]\Sigma_{ZXi}'\right] \\
&= E\left[\Sigma_{ZXi}\frac{B_i^{-1}A_iB_i^{-1}}{T}\Sigma_{ZXi}'\right] \\
&= E\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right] \\
&= \frac{\Omega_1}{T}
\end{aligned}$$

We consider now the second term in equation (28).

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i(\tau) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\alpha_i(\tau)\frac{1}{T}\sum_{t=1}^{T} z_{it} \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\alpha_i(\tau)\bar{z}_i.
\end{aligned}$$

By Assumption 5,

$$\mathbb{E}\left[\bar{z}_i\alpha_i(\tau)\right] = 0 \text{ and } Var\left(\bar{z}_i\alpha_i(\tau)\right) = \mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2].$$

We can now derive the asymptotic distribution of the sample moments by combining both terms, noting that they are uncorrelated, and applying the Lindeberg central limit theorem:

$$\begin{aligned}
\sqrt{N}\bar{g}_{NT}(\delta, \tau) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}(\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau))+\alpha_i(\tau)) \\
&\underset{d}{\to} N\left(0, E\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right] + E\left[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2\right]\right) \\
&= N\left(0, \frac{\Omega_1}{T} + \Omega_2\right).
\end{aligned}$$

The first L_1 elements of z_{1it} are such that $\bar{z}_i = 0$ for all i. It follows that only the bottom-right $L_2 \times L_2$ submatrix of $\Omega_2 \neq 0$:

$$\Omega_2 = \mathbb{E}[\bar{z}_i \bar{z}_i' \alpha_i^2]) = \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{E}[\bar{z}_{2i} \bar{z}_{2i}' \alpha_i^2] \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \Omega_{22} \end{pmatrix}.$$

Thus,

$$\Lambda_{NT} \bar{g}_{NT}(\delta, \tau) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{1[12]}/\sqrt{T} \\ \Omega_{1[21]}/\sqrt{T} & \Omega_{1[22]}/T + \Omega_{22} \end{pmatrix} \right).$$

The first-order asymptotic distribution is, thus,

$$\Lambda_{NT} \bar{g}_{NT}(\delta, \tau) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{pmatrix} \right). \tag{29}$$

■

Proof of Theorem 2. The result follows by Theorem 4. The pointwise assumptions are enough for the pointwise result stated in the paper. ■

Proof of Theorem 3. Define the matrix

$$\Lambda_T = \frac{\Lambda_{NT}}{\sqrt{N}} = \begin{pmatrix} \sqrt{T} I_{L_1} & 0 \\ 0 & I_{l_2} \end{pmatrix}$$

From the definition of the estimator,

$$\begin{aligned} \Lambda_{NT}(\hat{\delta}(\tau) - \delta(\tau)) &= \sqrt{N} \Lambda_T \left(S_{ZX}' \hat{W} S_{ZX} \right)^{-1} S_{ZX}' \hat{W} \bar{g}_{NT}(\delta, \tau) \\ &= \left(\Lambda_T^{-1} S_{ZX}' \hat{W} S_{ZX} \Lambda_T^{-1} \right)^{-1} \Lambda_T^{-1} S_{ZX}' \hat{W} \Lambda_T^{-1} \Lambda_T \sqrt{N} \bar{g}_{NT}(\delta, \tau) \\ &= \left(\Lambda_T^{-1} S_{ZX}' \hat{W} S_{ZX} \Lambda_T^{-1} \right)^{-1} \Lambda_T^{-1} S_{ZX}' \hat{W} \Lambda_T^{-1} \Lambda_{NT} \bar{g}_{NT}(\delta, \tau) \end{aligned}$$

Consider the matrix S_{ZX} and separate the z_{1it} from the z_{2it} components as well as the x_{1it} from the x_{2it} 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 $S_{12} = 0$:

$$S_{12} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{2it} z_{1it} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{2i} z_{1it} = \frac{1}{NT} \sum_{i=1}^{N} x_{2i} \sum_{t=1}^{T} z_{1it} = \frac{1}{N} \sum_{i=1}^{N} x_{2i} \bar{z}_{1i} = 0$$

This means that the fast moments (time-varying-instruments) cannot identify the coefficients on the time-constant covariates.

It follows that

$$\begin{aligned} \Lambda_T^{-1} S_{ZX}' \hat{W} S_{ZX} \Lambda_T^{-1} &= \begin{pmatrix} S_{11}'/\sqrt{T} & S_{21}'/\sqrt{T} \\ 0 & S_{22}' \end{pmatrix} \left(\begin{pmatrix} W_1 T & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \right) \begin{pmatrix} S_{11}/\sqrt{T} & 0 \\ S_{21}/\sqrt{T} & S_{22} \end{pmatrix} \\ &= \begin{pmatrix} S_{11}' \hat{W}_1 S_{11} & 0 \\ 0 & S_{22}' \hat{W}_{22} S_{22} \end{pmatrix} + o_p(1) \end{aligned}$$

and

$$\Lambda_T^{-1} S'_{ZX} \hat{W} \Lambda_T^{-1} = \begin{pmatrix} S'_{11}/\sqrt{T} & S'_{21}/\sqrt{T} \\ 0 & S'_{22} \end{pmatrix} \left(\begin{pmatrix} W_1 T & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \right) \begin{pmatrix} I_{L_1}/\sqrt{T} & 0 \\ 0 & I_{L_2} \end{pmatrix}$$
$$= \begin{pmatrix} S'_{11}\hat{W}_1 & 0 \\ 0 & S'_{22}\hat{W}_{22} \end{pmatrix} + o_p(1)$$

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

$$\left(\Lambda_T^{-1} S'_{ZX} \hat{W} S_{ZX} \Lambda_T^{-1} \right)^{-1} \Lambda_T^{-1} S'_{ZX} \hat{W} \Lambda_T^{-1}$$
$$\underset{p}{\rightarrow} \begin{pmatrix} \left(\Sigma'_{11} W_1 \Sigma_{11}\right)^{-1} \Sigma'_{11} W_1 & 0 \\ 0 & \left(\Sigma'_{22} W_{22} \Sigma_{22}\right)^{-1} \Sigma'_{22} W_{22} \end{pmatrix} = G \qquad (30)$$

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

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

∎

B.3.2 Weak convergence of the estimated quantile process

In this appendix, we prove a stronger result (weak convergence of the estimated quantile regression process) that implies Theorem 2. $\mathcal{T} \in (0,1)$ is a compact set of quantile indices of interest, e.g. $\mathcal{T} = [0.1, 0.9]$. The symbol $\ell^\infty(\mathcal{T})$ denotes the set of component-wise bounded vector values function of $\mathcal{T}$ and $\Rightarrow$ denotes weak convergence. To derive this result, we assume that assumptions 4 and 5 hold uniformly over $\tau \in \mathcal{T}$, we add below two mild continuity conditions for the individual effects and coefficients.[26] Let d_M, D_L and D_M be strictly positive constants.

Assumption 9 (**Individuals effects uniform in** τ)**.**
(i) For all $i = 1, \ldots, N$, $\mathbb{E}\left[\sup_{\tau \in \mathcal{T}} |\alpha_i(\tau)|^{4+d_M}\right] \leq D_M$. (ii) For some (matrix-valued) function $J : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}^{L \times L}$, $N^{-1}\sum_{i=1}^N \mathbb{E}[\alpha_i(\tau_1)\alpha_i(\tau_2) z_{it} z'_{it}] \rightarrow J(\tau_1, \tau_2)$ uniformly over $\tau_1, \tau_2 \in \mathcal{T}$. (iii) For all $\tau_1, \tau_2 \in \mathcal{T}$, $|\alpha_i(\tau_2) - \alpha_i(\tau_1)| \leq D_L |\tau_2 - \tau_1|$.

Assumption 10 (**Coefficients**)**.** *For all $\tau_1, \tau_2 \in \mathcal{T}$ and $i = 1, \ldots, N$, $||\beta_i(\tau_2) - \beta_i(\tau_1)|| \leq D_L |\tau_2 - \tau_1|$.*

Theorem 4 (**Asymptotic distribution of** $\hat{\gamma}$)**.** *Assume that conditions 1-6, 7(b), 9, and 10 hold. Then*

$$\sqrt{N}\left(\hat{\gamma}(\cdot) - \gamma(\cdot)\right) \Rightarrow \mathbb{G}(\cdot) \quad \text{in} \quad \ell^\infty(\mathcal{T})$$

where $\mathbb{G}$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $QJ(\tau_1, \tau_2)Q'$, where $Q = \left(\Sigma'_{22} W_{22} \Sigma_{22}\right)^{-1} \Sigma'_{22} W_{22}$.

[26] The pointwise assumptions are enough for the pointwise result stated in the paper.

Proof of Theorem 4. From equation 28 we have

$$
\begin{aligned}
\sqrt{N}\bar{g}_{NT}(\delta,\tau) =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\left(\tilde{x}_{it}'\left(\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it},y_{it}) + R_{it}^{(1)}(\tau) + R_{it}^{(2)}(\tau)\right) + \alpha_i(\tau)\right)\\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it},y_{it}) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'R_{it}^{(1)}(\tau)\\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'R_{it}^{(2)}(\tau)) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i(\tau)
\end{aligned}
$$

For the last term, by Lemma 3 in Chetverikov et al. (2016) follows that

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i(\cdot) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{z}_i\alpha_i(\cdot) \to \mathbb{G}^0(\cdot), \quad \text{in} \quad l^\infty(\mathcal{T})
$$

where $\mathbb{G}^0$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $J(\tau_1,\tau_2)$.

For the third term, the Lemma 3 in Galvao et al. (2020) implies $\sup_i \sup_{\tau\in\mathcal{T}} ||R_{iT}^{(2)}|| = O_p\left(\frac{\log T}{T}\right)$, and since $\tilde{x}_{it}$ and z_i are bounded,

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'R_{it}^{(2)}(\tau)) = O_p\left(\frac{\log T\sqrt{N}}{T}\right).
$$

uniformly over τ.

For the second term, note that by Lemma 3 in Galvao et al. (2020), $\text{Var}(R_{it}^{(1)}(\tau)) = o\left(\frac{1}{T}\right)$ and $\mathbb{E}[\sqrt{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'R_{it}^{(1)}(\tau)] = O\left(\frac{\log T\sqrt{N}}{T}\right)$ uniformly over τ. Thus,

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'R_{it}^{(1)}(\tau) = O\left(\frac{\log T\sqrt{N}}{T}\right) + o_p\left(\frac{1}{\sqrt{T}}\right)
$$

uniformly over τ. Then,

$$
\sqrt{N}\bar{g}_{NT}(\delta,\tau) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it},y_{it}) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i(\tau) + o_p(1)
$$

uniformly over τ. Using the results of the proof of Lemma 2 it follows that

$$
\sqrt{N}\left(\hat{\gamma}(\cdot) - \gamma(\cdot)\right) \Rightarrow \mathbb{G}(\cdot) \quad \text{in} \quad \ell^\infty(\mathcal{T})
$$

where $\mathbb{G}(\cdot)$ is a zero-mean Gaussian process with uniformly continuous sample path and covariance function $QJ(\tau_1,\tau_2)Q'$ where $Q = \left(\Sigma_{22}'W_{22}\Sigma_{22}\right)^{-1}\Sigma_{22}'W_{22}$. Whereas, the bias will dominate the asymptotic behavior of $\hat{\beta}$.

∎

B.3.3 Covariance Matrix

In the following proofs, we drop the dependence on τ for notational simplicity.

Proof of proposition 1. First, we rewrite the covariance matrix estimator as follows:

$$
\begin{aligned}
\Lambda_{NT}\hat{V}_\delta\Lambda_{NT} =& \Lambda_{NT}\left(X'Z\hat{W}Z'X\right)^{-1}X'Z\hat{W}\left(\sum_{i=1}^{N} Z_i'\tilde{u}_i\tilde{u}_i'Z_i\right)\hat{W}Z'X\left(X'Z\hat{W}Z'X\right)^{-1}\Lambda_{NT}\\
=&\Lambda_T\left(\frac{1}{N}X'Z\frac{1}{N}\hat{W}Z'X\right)^{-1}\frac{1}{N}X'Z\hat{W}\left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\tilde{u}_i\tilde{u}_i'Z_i\right)\\
&\frac{1}{N}\hat{W}Z'X\left(\frac{1}{N}X'Z\frac{1}{N}\hat{W}Z'X\right)^{-1}\Lambda_T\\
=&\Lambda_T\left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W}\left(\frac{1}{NT^2}\sum_{i=1}^{N} Z_i'\tilde{u}_i\tilde{u}_i'Z_i\right)\hat{W}S_{ZX}\left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}\Lambda_T\\
=&\left(\Lambda_T^{-1}S_{ZX}'\hat{W}S_{ZX}\Lambda_T^{-1}\right)^{-1}\Lambda_T^{-1}S_{ZX}'\hat{W}\Lambda_T^{-1}\Lambda_T\left(\frac{1}{NT^2}\sum_{i=1}^{N} Z_i'\tilde{u}_i\tilde{u}_i'Z_i\right)\\
&\Lambda_T\Lambda_T^{-1}\hat{W}S_{ZX}\Lambda_T^{-1}\left(\Lambda_T^{-1}S_{ZX}'\hat{W}S_{ZX}\Lambda_T^{-1}\right)^{-1},
\end{aligned}
$$

where for the second equality we use $\sqrt{N}\Lambda_T = \Lambda_{NT}$ and the fourth equality follows since $\left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W} = \Lambda_T^{-1}\left(\Lambda_T^{-1}S_{ZX}'\hat{W}S_{ZX}\Lambda_T^{-1}\right)^{-1}\Lambda_T^{-1}S_{ZX}'\hat{W}\Lambda_T^{-1}\Lambda_T$.
Then, by equation 30,

$$
\left(\Lambda_T^{-1}S_{ZX}'\hat{W}S_{ZX}\Lambda_T^{-1}\right)^{-1}\Lambda_T^{-1}S_{ZX}'\hat{W}\Lambda_T^{-1} \xrightarrow{p} \begin{pmatrix} \left(\Sigma_{11}'W_1\Sigma_{11}\right)^{-1}\Sigma_{11}'W_1 & 0\\ 0 & \left(\Sigma_{22}'W_{22}\Sigma_{22}\right)^{-1}\Sigma_{22}'W_{22}\end{pmatrix} = G
$$

Next, we can decompose the estimated error as follows:

$$
\begin{aligned}
\tilde{u}_i &= \tilde{X}_i\hat{\beta}_i - X_i\hat{\delta}\\
&= \tilde{X}_i(\hat{\beta}_i-\beta_i) + X_i\delta - X_i\hat{\delta} + \alpha_i.
\end{aligned}
$$

Inserting this expression in $Z_i'\tilde{u}_i\tilde{u}_i'Z_i$ yields:

$$
\begin{aligned}
Z_i'\tilde{u}_i\tilde{u}_i'Z_i =& Z_i'\left(\tilde{X}_i(\hat{\beta}_i-\beta_i)+X_i(\delta-\hat{\delta})+\alpha_i\right)\cdot\left(\tilde{X}_i(\hat{\beta}_i-\beta_i)+X_i(\delta-\hat{\delta})+\alpha_i\right)'Z_i\\
=& Z_i'\Big(\tilde{X}_i(\hat{\beta}_i-\beta_i)(\hat{\beta}_i-\beta_i)'\tilde{X}_i' + \tilde{X}_i(\hat{\beta}_i-\beta_i)(\delta-\hat{\delta})'X_i' + \tilde{X}_i(\hat{\beta}_i-\beta_i)\alpha_i'\\
&+ X_i(\delta-\hat{\delta})(\hat{\beta}_i-\beta_i)'\tilde{X}_i' + X_i(\hat{\delta}-\delta)(\hat{\delta}-\delta)'X_i' + X_i(\delta-\hat{\delta})\alpha_i'\\
&+ \alpha_i(\hat{\beta}_i-\beta_i)'\tilde{X}_i' + \alpha_i(\delta-\hat{\delta})'X_i' + \alpha_i\alpha_i'\Big)Z_i.
\end{aligned}
$$

Then, we want to show that the average of all but two terms converge fast to zero. Note that

$(\delta - \hat{\delta}) = O_p\left(\frac{1}{N^{1/2}}\right)$ and $(\hat{\beta}_i - \beta_i) = O_P\left(\frac{1}{T^{1/2}}\right)$. Thus, we can write the term in the middle as:

$$
\begin{aligned}
\Lambda_T \left(\frac{1}{NT^2}\sum_{i=1}^{N} Z_i'\tilde{u}_i\tilde{u}_i'Z_i\right)\Lambda_T =& \Lambda_T \frac{1}{NT^2}\sum_{i=1}^{N}\left(Z_i'\tilde{X}_i(\hat{\beta}_i-\beta_i)(\hat{\beta}_i-\beta_i)'\tilde{X}_i'Z_i + Z_i'\alpha_i\alpha_i'Z_i\right)\Lambda_T + o_p(1)\\
=& \Lambda_T \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i-\beta_i)\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i-\beta_i)\right)'\Lambda_T\\
&+ \Lambda_T \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i\right)'\Lambda_T + o_p(1)\\
=& \Lambda_T \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i-\beta_i)\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i-\beta_i)\right)'\Lambda_T\\
&+ \frac{1}{N}\sum_{i=1}^{N}\left(\bar{z}_i\bar{z}_i'\alpha_i^2\right) + o_p(1)\\
&= \Lambda_T\mathbb{E}\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right]\Lambda_T + \mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i^2] + o_p(1)\\
&= \begin{pmatrix}\Omega_{11} & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & \Omega_{22}\end{pmatrix} + o_p(1)\\
&\xrightarrow{p} \Omega,
\end{aligned}
$$

where the third equality follows, since

$$
\frac{1}{N}\sum_{i=1}^{N}\left(\bar{z}_i\bar{z}_i'\alpha_i^2\right) = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}0 & 0\\ 0 & \bar{z}_{2i}\bar{z}_{2i}'\alpha_i^2\end{pmatrix},
$$

which implies that

$$
\Lambda_T\frac{1}{N}\sum_{i=1}^{N}\left(\bar{z}_i\bar{z}_i'\alpha_i^2\right)\Lambda_T = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{z}_i\bar{z}_i'\alpha_i^2\right).
$$

Combining the results yields

$$
\Lambda_T\hat{V}_\delta\Lambda_T \xrightarrow{p} G\Omega G'.
$$

■

B.3.4 Hausman 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.
Consider the objective function:

$$
\begin{aligned}
J(\hat{\delta}) &= N\bar{g}_{NT}(\hat{\delta})'\hat{S}^{-1}\bar{g}_{NT}(\hat{\delta})\\
&= N\left(\Lambda_{NT}\bar{g}_{NT}(\hat{\delta})\right)'\Lambda_{NT}^{-1}\hat{S}^{-1}\Lambda_{NT}^{-1}\Lambda_{NT}\bar{g}_{N}(\hat{\delta})\\
&= \left(\Lambda_{NT}\bar{g}_{NT}(\hat{\delta})\right)'\Lambda_{T}^{-1}\hat{S}^{-1}\Lambda_{T}^{-1}\Lambda_{NT}\bar{g}_{NT}(\hat{\delta}).\\
&= \left(\Lambda_{NT}\bar{g}_{NT}(\hat{\delta})\right)'\hat{\Omega}^{-1}\Lambda_{NT}\bar{g}_{NT}(\hat{\delta}).
\end{aligned}
$$

where $\hat{\Omega}^{-1} = \left(\Lambda_T \hat{S} \Lambda_T\right)^{-1}$.

Second, we want to show that $\bar{g}_{NT}(\hat{\delta}) = \hat{B}\left(\bar{g}_{NT}(\delta) + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i - \beta_i)\right)$. Recall that $\hat{Y}_i = X_i\delta + \alpha_i + \tilde{X}_i(\hat{\beta}_i - \beta_i)$. Hence, we can write

$$
\begin{aligned}
Z_i'\hat{Y}_i &= Z_i'X_i\delta + Z_i'\alpha_i + Z_i'\tilde{X}_i(\hat{\beta}_i - \beta) \\
S_{Z\hat{Y}} &= S_{ZX}\delta + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i - \beta_i) \\
S_{Z\hat{Y}} &= S_{ZX}\delta + \bar{g}_{NT}(\delta).
\end{aligned}
$$

Then, note that

$$
\begin{aligned}
\bar{g}_{NT}(\hat{\delta}) &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} z_{it}(\hat{y}_{it} - x_{it}'\hat{\delta}) \\
&= S_{Z\hat{Y}} - S_{ZX}\hat{\delta} \\
&= S_{Z\hat{Y}} - S_{ZX}\left(S_{ZX}\hat{S}^{-1}S_ZX\right)^{1} S_{ZX}\hat{S}^{-1}S_{Z\hat{Y}} = \hat{B}S_{Z\hat{Y}},
\end{aligned}
$$

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

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

Since ΩD is positive definite there exist a matrix C such that $\hat{\Omega}^{-1} = C'C$.

We define $A \equiv C\Lambda_T S_{ZX}'$ and $M \equiv I_L - A(A'A)^{-1}A'$.

In this third part, we will show that

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

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

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

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

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

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

and the probability limit of C is Q. We define $v \equiv C\Lambda_{NT}\bar{g}_{NT}(\delta)$.
By equation 29 in the proof of Lemmma 2,

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

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:

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

Since M is idempotent with rank L, it follows that

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

■