Minimum Distance Estimation of Quantile Panel Data Models

Blaise Melly*and Martina Pons†

This version: February 11, 2022

Ling

Abstract

Abstract

To add somewhere:

Inference: clustered standard errors

1 Introduction

This paper suggests a minimum distance (MD) estimator for quantile panel data models. The estimator applies to quantile versions of many known panel data estimators such as the between, fixed effects (FE), random effects (RE), and Hausman Taylor. Depending on the underlying assumptions, the estimator can estimate coefficients of both time–varying and time–invariant regressors. Estimation is performed in two stages. The first stage consists of individual-level quantile regressions using time–varying covariates for each quantile. In the second stage, the first-stage fitted values are regressed on time–invariant and time–varying variables using a 2SLS regression or, more generally, GMM. Thus, including external or internal instrumental variables in the second stage is straightforward. In the special cases of a least-squares first stage, we show that our two-step estimators are algebraically identical to the one-step counterparts. The estimator is simple to implement, flexible, computationally fast, and can be used in various applied fields. For instance, while we present our estimator as a panel data estimator, it is of practical relevance in settings where a researcher has micro-data with some clustered structure, where some variables vary within clusters while others only between clusters. Further, simulations show that our estimator and inference procedure performs well in small samples with few time periods.

Like non-linear models, quantile regression with fixed effects is subject to the incidental parameter problem. A standard solution in the literature has been to allow the number of individuals and the number of time periods per individual to diverge to infinity jointly. Recently,

*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.

Galvao et al. (2020) weakened the requirements on N and T for asymptotic normality for FE quantile estimators. Using similar techniques, we show that our estimator is asymptotically normal under the condition that $N(\log T)^2/T \to 0$. To improve inference in finite samples, we keep higher-order terms coming from the first stage error in the asymptotic distribution. Ignoring these terms would yield the same asymptotic distribution as if we knew the first stage. Further, we suggest a consistent estimator of the covariance matrix.

This paper contributes to the literature on quantile panel data and IV models. A large share of the literature focused on fixed effects models (see, for example, Canay, 2011; Galvao and Kato, 2016; Gu and Volgushev, 2019; Zhang et al., 2019). Koenker (2004) introduced a Quantile FE estimator that treated the individual heterogeneity as a pure location shift. Kato et al. (2012) allow the individual effects to depend on the quantiles 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 a quantile regression RE estimator. Our RE estimator is different in two ways. First, we focus on the conditional quantile function that also conditions on the individual i, thus we estimate a different parameter. Further, we suggest how to implement an efficient estimator.

The estimator nests the minimum distance estimators of Chamberlain (1994) and Galvao and Wang (2015) as special cases. Our estimator differs from the estimator of Chamberlain (1994) for two main reasons. First, he considers a setting where there is a finite number of individuals (cells). Second, there are only time–invariant regressors (that is, regressors vary only across cells).[1] In the special case that we do not include time–varying regressors, our estimator reduces to the minimum distance estimator of Chamberlain (1994), with nonetheless crucial differences from a theoretical perspective. 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. Further, we allow for both internal and external instruments. A similar model and estimator that allows for instrumental variables are analyzed in Chetverikov et al. (2016). They focus on a special case where variables are divided between group-level and individual-level (instead of time–invariant and time–varying). 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. First, by keeping all variables in the second stage, we can impose equality of the coefficient on the time–varying regressors and, therefore, substantially increase precision at a minimal cost from a computational perspective. Second, since they use only the constant, their estimator is not invariant to reparametrization of the first stage regressors. Simulations using the data generating process in Chetverikov et al. (2016) show that our minimum distance estimator has substantially lower variance and MSE across all sample

[1] More precisely, Chamberlain (1994) studies the return on education by dividing observations into years of experience and years of education cells. Since only a finite combination of years of education and experience are feasible, there are only a finite number of cells.

sizes considered. Other instrumental variable quantile methods are suggested in Abadie et al. (2002) and Chernozhukov and Hansen (2005).

The remainder of the paper is structured as follows. Section 2 presents the model and the estimator and briefly discusses equivalent methods to estimate linear 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 the panel data model, and we present a 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 Estimator

2.1 Quantile Model

We want to learn the effects of the time–varying variables X_{it} and of the time–constant variables W_i on the distribution of an outcome Y_{it}. We observe these variables for the individuals $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$.[2] We use the traditional terminology of panel data, but the i index can define groups of any sort, and the time 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. Our model can be written using the Skorohod representation as

$$Y_{it} = \alpha_i(U_{it}) + X_{it}\beta(U_{it}) + W_i\gamma(U_{it})$$

where U_{it} is an individual rank variable responsible for the heterogeneity in responses given the observed variables X_{it} and W_i. For the within individual variation we assume

$$U_{it} \sim U(0,1)|Z_{it}, I,$$

$$U_{it}|\dot{X}_{it}, \alpha_i(\cdot), Z_{it} \sim U_{it}|\dot{X}_{it}, \alpha_i(\cdot)$$

and for the between group variation

$$E[\alpha_i(\tau)Z_{it}] = 0.$$

We assume that

$$Q_{it,\tau}(X_{it}, W_i) = X_{it}'\beta(\tau) + W_i'\gamma(\tau) + \alpha_i(\tau), \tag{1}$$

where $Q_{it,\tau}(X_{it}, W_i)$ is the τth conditional quantile function of the response variable Y_{it} for individual i in period t given a d_x-vector of time–varying regressors X_{it} and a d_z-vector of time invariant variables W_i. The parameters $\beta(\tau)$, $\gamma(\tau)$ and $\alpha_i(\tau)$ can depend on the quantile index τ. Depending on the setting, both the parameters $\beta(\tau)$ and $\gamma(\tau)$ might be of interest. We normalize $E[\alpha_i(\tau)] = 0$, which is not restrictive because we allow W_i to include a constant.

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

In this paper, we model the distribution conditional on the covariates and on the individuals i. This implies that the quantile function might take a different shape for different individuals depending on the unobserved $\alpha_i(\tau)$. An alternative notation used in the literature is to include α_i in the conditioning set. A one-step quantile regression of Y_{it} on X_{it} and z_i models the conditional quantile function without conditioning on the individual and therefore identifies different parameters. In the special case where $\alpha_i = 0$ for all i, both models would identify the same parameters. If the unconditional treatment effect is of interest, one can obtain the unconditional distribution function by integrating out the α_i (and possibly other variables) from the conditional distribution. Inverting the distribution function, then, yields the unconditional quantile function (see Chernozhukov et al. 2013).

In this paper, we consider cases where X_{it} and W_i are potentially correlated with $\alpha_i(\tau)$ but continue to be exogenous within individuals. To deal with the potential endogeneity caused by α_i, we assume that there is a vector of valid instruments Z_{it}. Alternatively, treating $\alpha_i(\tau)$ as a parameter to be estimated as in a fixed effects approach would allow to consistently estimate $\beta(\tau)$.

Throughout the paper, we will use the following notation. We denote the $T \times (d_X + 1)$ matrix of time–varying (or first-stage) regressors for individual i by $\tilde{X}_i = ((1, x'_{i1})', (1, x'_{i2})', \ldots, (1, x'_{in})')'$ and define the $NT \times (d_X + 1)$ matrix of time–varying regressors for all individuals by $\tilde{X} = (\tilde{X}'_1, \ldots, \tilde{X}'_N)'$. We also define the $T \times k$ matrix of regressors for individual i as $X_i = ((x'_{i1}, w'_i)', (x'_{i2}, w'_i)' \ldots, (x'_{iT},$ and the $NT \times k$ matrix of regressors for all individuals as $X = (X'_1, \ldots, X'_N)'$. Similarly, we define the $T \times l$ matrix of instruments for individual i as $Z_i = ((1, z'_{i1})', (1, z'_{i2})', \ldots, (1, z'_{iT})')'$ and the $NT \times l$ matrix of instruments for all individuals as $Z = (Z'_1, \ldots, Z'_N)'$. Let $\mathbf{l}_i$ be a $T \times 1$ vector of 1. Similarly, let Y be the $(NT \times 1)$ vector of response variables. 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 $P_i\tilde{X}_i$ gives the individual specific means, and $Q_i\tilde{X}_i$ yields a matrix of deviation for individual means. Finally, we define the $NT \times (d_X + 1)$ transformed matrix for all individuals as $P_D\tilde{X}$ and $Q_D\tilde{X}$ where $P_D = \text{diag}\{P_1, P_2, \ldots, P_N\}$ and $Q_D = \text{diag}\{Q_1, Q_2, \ldots, Q_N\}$.

Before introducing the estimator, we motivate our approach using linear models 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, as well as proofs and more formal results, are in the appendix.

2.2 Linear Estimators

Fixed effects models are routinely estimated by transforming dependent and independent variables in deviations from means. This within transformation is useful as it eliminates the potential endogeneity coming from α_i, and provides a consistent estimator of β without placing assumptions on the unobserved time–invariant heterogeneity. This approach is not feasible in quantile models, as there is no known transformation that eliminates the FE. A second possibility to estimate FE models is to include dummy variables for each individual in a least squares regression. This is algebraically identical to the within estimator. However, in quantile models, a dummy variables regression is computationally unattractive, as it requires estimating many parameters. A third possibility, which we use in this paper, is to divide 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 constant of each regression. The second stage aggregates the point estimates from the first step by regressing the fitted values from the first stage on x_{it} using demeaned x_{it} as an instrument. As with the within estimator, this instrument only exploits the variation within individuals. While this procedure does not provide any advantages in a linear model, it can be easily extended to quantile models, where it substantially reduces the computational burden of quantile FE estimation. This two-step procedure is not specific to the within case, but it applies to a wide range of estimators. More in general, define the two-step procedure as follows. The first stage consists of individual-level least squares regressions with 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 2 in the Appendix). Similarly, clustered standard errors of the one-step estimator are identical to clustered standard errors of the two-steps estimator as long as the clusters are at a level higher than the individuals (see Proposition 5 in the Appendix). Like with the fixed effects models, instrumental variables methods can be used to estimate a between regression, and random effect models. More precisely, regressing $P_D Y$ on $P_D X$ is identical to an IV regression of the first-stage fitted values on X using $P_D X$ as instrument. While FGLS is the most common estimator for the RE model, Im et al. (1999) show that the overidentified 3SLS estimator, with instruments $Z = (P_D X, Q_i \tilde{X})$, is identical to the RE 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 estimand. Alternatively, the RE estimator can be implemented using the theory on optimal instrument with a just identified 2SLS regression (see Im et al., 1999; Hansen, 2021). Finally, the Hausman-Taylor model (Hausman and Taylor, 1981) can be estimated by instrument variables regression with the first-stage fitted values or Y as dependent variable (see Hansen, 2021).

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 1. 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 and quantile, regress y_{it} on the time varying variables x_{it} and a constant using quantile regression. The intercept of the first stage regression, contains both the individual effect $\alpha_i(\tau)$ and the term $w_i'\gamma(\tau)$ as these vary only between individuals. Then, in a second step, regress the fitted values of the first stage on w_i and x_{it} using GMM with instruments z_{it}.

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

$$\hat{\beta}_i(\tau) \equiv \left(\hat{\beta}_{0,i}, \hat{\beta}_{1,i}'\right)' = \underset{(b_0,b_1)\in\mathbb{R}^{d_x+1}}{\arg\min} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_{it} - b_0 - x_{it}'b_1), \tag{2}$$

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 $\beta_i(\tau) = (\alpha_i(\tau) + z_i'\gamma(\tau), \beta(\tau)')'$. The fitted value for individual i in period t is $\hat{y}_{it}(\tau) = \hat{\beta}_{0,i}(\tau) + x_{it}'\hat{\beta}_{1,i}(\tau)$.

The second stage consists in a GMM regression using $\mathbb{E}[g_i(\delta)] = 0$ as a moment condition, with $g_i(\delta) = Z_i(\tilde{X}_i\beta - X_i\delta)$ is a $l \times 1$ matrix. 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 $\hat{y} = (\hat{y}_1', \ldots, \hat{y}_n')'$. The second stage estimator can be written as[3]

$$\hat{\delta}(\tau) = \left(X'Z\hat{W}Z'X\right)^{-1} X'Z\hat{W}Z'\hat{y}(\tau), \tag{3}$$

where $\hat{W}$ is a $l \times l$ symmetric positive definite weighting matrix and $\delta = (\beta', \gamma')'$ is the k-dimensional vector of coefficients. If $l = k$, the second step estimator in 3 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 $\hat{W} = \hat{S}^{-1}$, where $\hat{S}$ is a consistent estimator of $S = \mathbb{E}[g_i(\delta)g_i(\delta)']$.

In a variation of the second stage that we consider in this paper, we will use optimal instruments. Optimal instruments are relevant when a researcher has a conditional moment restriction of the form $\mathbb{E}[g_i(\delta)|Z_i] = 0$. If the moment conditions hold conditional on Z_i, an infinite set of valid moments exist, and additional moment conditions could be used to increase efficiency. The goal is to select the instrument that minimizes the asymptotic variance. The optimal instrument takes the form $Z_i^* = \mathbb{E}[g_i(\delta_0)g_i(\delta_0)'|Z_i]^{-1}R_i(\delta)$ where $R_i(\delta) = \mathbb{E}[\frac{\partial}{\partial\delta}g_i(\delta)|Z_i]$ (see for example Chamberlain, 1987 and Newey, 1993). Although not explored in this paper, it would be possible to consider a Chernozhukov and Hansen (2006) IV quantile regression first stage, followed by a 2SLS second stage.[4]

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

[3]Throughout the paper, we consider different second stage estimators. Here, we present the second stage as a GMM estimator, since most of the estimators will be special cases of GMM.

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

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 in subsection 2.2, if our two-step estimator is applied with OLS in the first stage, then our estimator is algebraically equivalent to a one-step estimator. Therefore, providing an intuitive justification for our approach.

The estimator can be written as a minimum distance 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 minimum distance estimator minimizes

$$\hat{\delta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{N} (\hat{\beta}_i(\tau) - R_i\delta)'\tilde{W}(\hat{\beta}_i(\tau) - R_i\delta) \tag{4}$$

where $\tilde{W}$ is a $k \times k$ weighting matrix, R_i is defined such that $\tilde{X}_i R_i = X_i$, that is

$$\underset{(d_x+1)\times K}{R_i} = \begin{pmatrix} 1 & w_i' & 0 \\ 0 & 0 & I_{d_x} \end{pmatrix}.$$

Similarly to Galvao and Wang (2015), the efficient minimum distance estimator can be implemented by substituting $\tilde{W}$ with the estimated first-stage covariance matrix $\hat{V}_i$. The solution to the minimization problem in equation 4 is

$$\hat{\beta}_{EMD} = \left(\sum_{i=1}^{N} R_i' \hat{V}_i^{-1} R_i \right)^{-1} \sum_{i=1}^{N} R_i' \hat{V}_i^{-1} \hat{\beta}_i. \tag{5}$$

If instead $\tilde{W} = \tilde{X}_i'\tilde{X}_i$ then the estimator in equation 5 is algebraically identical to using OLS in the second stage.

Two remarks about the efficient minimum distance estimator follow. First, except in the FE case, we are not treating α_i as a parameter to be estimated, but as part of the error term. For this reason, the efficient minimum distance estimator using the inverse of the first stage variance as a weighting matrix ignores α_i and is, thus, inefficient. This is not the case in Galvao and Wang (2015) as they estimate α_i. Second, efficient minimum distance can be implemented using optimal instruments when $\alpha_i(\tau) = 0$ for all i and τ (see Proposition 6 in the Appendix).[5] To improve efficiency of our estimator, we will exploit the covariance structure implied by α_i.

3 Asymptotic Theory

This section states the assumptions and presents the asymptotic results. Let $x_{it}' = (z_i', \tilde{x}_{it})'$

Assumption 1 (Assumption A0 in Galvao, Gu, and Volgushev (2020) (GGV)). *The processes $\{y_{it}, x_{it}\}$ are strictly stationary for each i and are independent across individuals.*

[5]The efficient minimum distance estimator of Galvao and Wang (2015) is numerically identical to our estimator using an IV second stage with instrument $Z_i^* = \tilde{X}_i(\tilde{X}_i'\tilde{X}_i V_i \tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i' Z_i = (\tilde{X}_i V_i \tilde{X}_i')^{+} Z_i$, where Z_i contains a constant, x_{it}, and individual dummies.

Assumption 2 (Assumption A1 in GGV)**.** *$||x_{it}|| \leq M \leq \infty$ a.s., and the eigenvalues of $\mathbb{E}[x_{it}x_{it}']$ are uniformly bounded away from zero and infinity.*

Assumption 3 (Assumption A2 in GGV)**.** *The conditional distribution $F_{y_{it}|x_{it}}(y|x)$ is twice differentiable w.r.t. y, with the corresponding derivatives $f_{y_{it}|x_{it}}(y|x)$ and $f'_{y_{it}|x_{it}}(y|x)$. Further, assume that*

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

and

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

Assumption 4 (Assumption A3 in GGV)**.** *Les $\mathcal{T}$ be an open neighborhood of τ. Assume that uniformly across groups, there exists a constant $f_{min} < f_{max}$ such that*

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

Assumption 5 (Assumption I in GGV)**.** *The observations (y_{it}, x_{it}) are i.i.d. within groups*

$$\frac{N(\log T)^2}{T} = o(1)$$

Assumption 6. ***Assumption MI in GGV***
Assume that $d_T = o(1)$, as $T \to \infty$ and

$$\frac{\log T}{T d_T^2} = o(1).$$

Assumption 7. ***Instruments***
(i) For all $i = 1, \ldots, N$ and all $t = 1, \ldots, T$, $E[z_{it}\alpha_i(\tau)] = 0$. (ii) For all $i = 1, \ldots, N$ and all $t = 1, \ldots, T$, y_{it} is independent of z_{it} conditional on $(x_{ig}, z_i, \alpha_i(\tau))$. (iii) $E[X_i'Z_i] \equiv \Sigma_{XZ}$ is of full column rank. (iv) $||z_{it}|| \leq M \leq \infty$ a.s. (v) $\mathbb{E}[z_{it}z_{it}'\alpha_i(\tau)^2]$ is nonsingular.

[Discuss the assumptions. They are quite standard in the literature] Assumption 1 excludes the possibility of temporal dependence within individuals.

[Compare to Chamberlain] Under correct specification (i.e. if $\alpha_i = 0$ for all i) Chamberlain (1994) finds that only the variance coming from the first stage error matter and $V_2 = 0$. He considers a setting with a finite N, so that the variance is of order $\sqrt{NT}$. In contrast, by considering a setting when both N and T go to infinity, the variance coming from the first stage will not show up in the asymptotic variance.

4 Quantile Panel Data Estimators

4.1 Fixed Effects, Random Effects and Between Estimators

The minimum distance 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 a time and is unaffected. Given that the second stage is linear, panel data estimators are straightforward to implement by regressing the fitted values on X_i using the desired estimator. For all the panel data models, we assume that

$$E\left[1\left(Y_{it} \leq X_{it}'\delta\left(\tau\right) + \alpha_i(\tau)\right) - \tau|X_{it}\right] = 0 \tag{6}$$

for $t = 1, \ldots, T$. Under this assumption, the FE estimator provides a consistent estimator of δ. Consistency of RE estimation requires additionally that

$$\mathbb{E}[\alpha_i(\tau)|X_i] = \mathbb{E}[\alpha_i(\tau)] = 0. \tag{7}$$

FE and Between models can be easily estimated by an instrumental variables regression of the first-stage fitted values $\hat{Y}(\tau)$ on $\tilde{X}$ or X with instrument $Q_D\tilde{X}$ or P_DX, respectively.[6]

Implementation of the random effect estimator is more involved. We present two different motivations for our two-step random effect estimator. The first one comes from analyzing the moment conditions. The second start from equivalence results of various estimators for the random effects estimator in linear model.

The moment restriction in 6 implies $Q_{it,\tau}(X_{it}) = X_{it}'\delta(\tau) + \alpha_i(\tau)$. It follows that the second moment (equation 7) can be written as

$$E[Q_{it,\tau}(X_{it}) - X_{it}'\delta(\tau)|X_{it}] = 0.$$

Thus we have two moment conditions, one corresponding to each of the steps of our estimator which is the sample analogue. Implementing efficient estimation is one of the main challenges with the quantile random effect estimator. Given the equality between optimal instruments and RE in linear models, we suggest using optimal instruments to attain an efficient RE estimator. Since applying optimal instruments on the moment $g(\delta) = Q_{it,\tau}(X_{it}) - X_{it}'\delta(\tau)$ is not feasible, in the second stage, we use the sample counterpart $\tilde{g}(\delta) = \hat{Q}_{it,\tau}(X_{it}) - X_{it}'\delta(\tau)$ as a moment. Note that under homoskedasticity with a least squares first stage, both moment conditions imply the same optimal instrument (see proposition 4). The resulting optimal instrument is $Z_i^*(\tau) = \mathbb{E}[\tilde{g}_i(\delta)\tilde{g}_i(\delta)|X_i]^{-1}X_i$. Under the additional assumption that $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$ the

[6] 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 for the time–invariant regressors W_i without changing the assumptions on the time varying regressors X_{it}. If W_i is uncorrelated with α_i, it is possible to consistently estimate γ by regressing the fitted values for each quantile τ on X using $(Q_D\tilde{X}_i, (W_1', W_2', \ldots, W_N')')$ as instrument. Therefore, our two-step approach allows to consistently estimate the effect of time–invariant regressors using the same approach as with linear regression.

optimal instrument simplifies to $Z_i^* = (\tilde{X}_i \frac{V_i(\tau)}{T} \tilde{X}_i' + \mathbf{l}_T' \mathbf{l}_T \sigma_\alpha^2(\tau))^+ X_i$ where $V_i(\tau)$ is the asymptotic variance from the first stage for an individual i and $^+$ is the Moore-Penrose inverse.[7] Under this assumption, our estimator is efficient. A few remarks about the optimal instrument follow. First, in a least squares first stage and under usual RE assumptions, the matrix $\tilde{X}_i \frac{V_i(\tau)}{T} \tilde{X}_i' + \mathbf{l}_T' \mathbf{l}_T \sigma_\alpha^2(\tau)$ simplifies to the usual RE structure. Second, the moment conditions $g(\delta)$ and $\tilde{g}(\delta)$ imply the same optimal instrument. These results are summarized in proposition 4 in the Appendix and its proof. Third, if $\sigma_\alpha = 0$, this estimator is identical to the efficient minimum distance estimator (see proposition 6 in the Appendix). Fourth, as T increases the first stage variance converges to zero, and the RE estimator convergences to the FE estimator (Baltagi, 2021; Ahn and Moon, 2014).

Compared to the classical random effects structure, we use the first stage variance.[8] This formula has two main advantages. First, it is straightforward to compute $\hat{V}_i$. Second, in computing $\hat{V}_i$ it is possible to allow for dependence in the errors in the first stage regressions.

In the following we assume that $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$. To make the RE estimator operational, we need estimators of $V_i(\tau)$ and $\sigma_\alpha^2(\tau)$. 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 $\hat{A}_i = \tau(1-\tau)\frac{1}{T}\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}'$ and $B_i(\tau)$ can be estimated by the Kernel Density estimator of Powell (1991):

$$\hat{B}_i(\tau) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{Y_{it} - \tilde{X}_{it}'\beta_i(\tau)}{h}\right) X_{it}X_{it}', \tag{8}$$

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 group separately. We estimate $\sigma_\alpha(\tau)$ by

$$\hat{\sigma}_\alpha^2(\tau) = \frac{1}{NT(T-1)/2 - K}\sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \nu_{it}(\tau)\nu_{is}(\tau), \tag{9}$$

where $\nu_{it}(\tau) = \hat{y}_{it}(\tau) - x_{it}'\hat{\delta}(\tau)$.

If the assumption that $\mathbb{E}[\alpha_i^2(\tau)|X_i] = \sigma_\alpha^2(\tau)$ were to be violated, we could improve efficiency by using efficient GMM with instruments $(P_T X_i, Q_T \tilde{X}_i)$.[9] Thus, it is possible to implement a RE estimator using weaker assumptions by efficient GMM in the second stage. Given the cluster dependence in the moment conditions, we use a clustered-robust weight matrix.

4.2 Hausman and Taylor Model

For the quantile extension of the Hausman-Taylor estimator, we consider the following model:

$$Q_{it,\tau}(\tilde{x}_{1,it}, \tilde{x}_{2,it}, z_{1,t}, z_{2,t}) = \tilde{x}_{1,it}\beta_1(\tau) + \tilde{x}_{2,it}\beta_2(\tau) + z_{1,t}\gamma_1(\tau) + z_{2,t}\gamma_2(\tau) + \alpha_i(\tau). \tag{10}$$

[7]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.

[8]Using the first stage variance will not impose equality on the estimated densities of the errors $\hat{f}_{Y_i - \tilde{X}_i\beta_i}(0)$ and in the second stage, observations will be weighted differently, depending on the first stage variance.

[9]With an OLS first stage, this GMM estimator is asymptotically as efficient as the RE estimator if the RE assumption are correct (Im et al., 1999). The GMM estimator will be inefficient relative to the GMM estimator using the full set of moment conditions. However, it may have better finite sample properties.

Like in the previous section, we assume that α_i is the only possible source of endogeneity, and that only $x_{1,it}$ and $z_{1,t}$ might be correlated with α_i. Define $Z = (Q_T X_1, Q_T X_2, P_T X_1, Z_1)$. Then, we have the following conditional moment restrictions:

$$E\left[1\left(y_{it} \le \tilde{x}_{it}'\beta(\tau) + \alpha_i(\tau)\right) - \tau | Z_{it}\right] = 0$$
$$E[Q_{it,\tau}(X_{it}) - x_{it}'\delta(\tau)|Z_{it}] = 0$$

where $x_{it} = (\tilde{x}_{1,it}, \tilde{x}_{2,it}, z_{1,t}, z_{2,t})$.

These moment conditions suggest estimating the Hausman-Taylor model using efficient instruments, as in the previous section. The optimal instrument is then $Z_i^* = \mathbb{E}[\tilde{x}_i(\hat{\beta}_i - \beta) + \alpha_i|Z_i]^{-1}\mathbb{E}[X_i|Z_i]$. Implementation of the optimal instrument is not straightforward as it requires the estimation of $\mathbb{E}[X_i|Z_i]$ usually estimated nonpametrically (see Newey (1993)). In this paper, we do not contribute in this direction. Instead, we note that it is possible to use efficient GMM with instruments $Z = (Q_T X_1, Q_T X_2, P_T X_1, Z_1)$. In the special case where there is no $x_{1,it}$, so that all time varying regressors are exogenous, then the optimal instrument approach can easily be implemented as the first stages include only exogenous variables.

4.3 Hausman Test

Consistency of the RE estimator requires stronger assumptions, compared to the FE estimator. Under these assumptions, both estimators are consistent, but the FE is inefficient. Hausman (1978) suggested a test for the null hypothesis of RE against the alternative of FE. This subsection present a variation of the Hausman Test for our two-step estimator (see Hansen, 1982). Some generalization of the Hausman test that have been proposed in the literature are not applicable in our context, as they would not converge when $T \to \infty$ (Ahn and Low, 1996; Arellano, 1993; Wooldridge, 2019). Ahn and Low (1996) show the equivalence between the Hausman Test and the Hansen GMM statistics in the 3SLS estimator. In section 4, we suggested Efficient GMM as a possibility to perform RE estimation. Here, we suggest using a Hansen overidentification test to check the validity of the moment conditions. Since we will be comparing the FE and the RE model, it makes sense to consider the smaller model only with time varying regressors. Consistency of the RE estimator requires that $\tilde{X}_i$ is uncorrelated with α_i and $\mathbb{E}[Q_i\tilde{X}_i'\alpha_i(\tau)] = 0$ and $\mathbb{E}[P_i\tilde{X}_i'\alpha_i(\tau)] = 0$ are a valid moment conditions. By contrast, the fixed effects rely only on the moment condition $\mathbb{E}[Q_i\tilde{X}_i'\alpha_i(\tau)] = 0$. We wish to test the null hypothesis $H_0 : \mathbb{E}[\tilde{Q}_i X_i'\alpha_i(\tau)] = 0$ and $\mathbb{E}[\tilde{P}_i X_i'\alpha_i(\tau)] = 0$ using a GMM overidentification test.

Define $Z_i = (P_T\tilde{X}_i, Q_T\tilde{X}_i)$ and let $\tilde{g}_i(\delta) = Z_i\left(\hat{y}_i(\tau) - \tilde{X}_i\delta(\tau)\right)$ and $\bar{g}_n(\delta) = \frac{1}{N}\sum_{i=1}^N g_i(\delta)$. Define

$$J\left(\hat{\delta}_{GMM}(\tau)\right) = n\bar{g}_n(\delta)'\hat{\Omega}^{-1}(\tau)\bar{g}_n(\delta) \tag{11}$$

Proposition 1. *Under Assumptions ??? as T and $N \to \infty$, $J(\hat{\delta}_{GMM}) \xrightarrow{d} \chi^2_{l-k}$.*

[Ahn and Low (1996) suggest a GMM test that can be regarded as a generalization of the Hausman Test. They include some additional (redundant) instruments and argued that these are relevant for the construction of an appropriate test statistic, since any evidence of correlation between these additional instruments and the error implies a violation of H0. Ahn and Low (1996) consider also another GMM test, which is not feasible in our case. This GMM statistics is asymptotically chi-squared with Tk degrees of freedom. Are our estimators efficient under heteroskedasticity? If heteroskedasticity is due to the idiosyncratic error yes, but if it is due to α_i we are not. Thus the variance of the difference is not the difference of the variances. This is not true. We are using efficient GMM for the test.]

4.4 Simulations

This section presents simulation results for the different panel data estimators and the Hausman Test discussed in the previous subsections. The estimators differ only in the second stage. The Pooled estimator regresses $\hat{Y}$ on X, the between and the FE use instruments $P_D X$ and $Q_D \tilde{X}$, respectively. The RE estimator is performed using optimal instruments and assuming $\mathbb{E}[\alpha_i(\tau)|X_i] = \sigma_\alpha^2$. Last, the GMM estimator uses the optimal weighing matrix and instruments $(P_D X, Q_D \tilde{X})$. The weighing matrix exploits the cluster-dependence. These simulations focus on the estimation of $\beta(\tau)$. The next section presents simulation for $\gamma(\tau)$. We consider the following data generating process where all variables are scalars:

$$y_{it} = \beta x_{it} + \alpha_i + (1 + 0.1 x_{it})\nu_{it}. \tag{12}$$

We let $\beta = 1$ and $\nu_{it} \sim \mathcal{N}(0, 1)$. The regressor is defined by $x_{it} = h_i + 0.5 u_{it}$ 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_i is correlated with α_i so that only the fixed effects estimator is consistent. For the simulation of the 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.

As shown in Table 1, the estimator performs relatively well also when both N and T are small. The GMM estimator performs similarly to the RE estimator, and in some cases even better. As expected, asymptotically, the GMM, the RE, and the FE estimators become indistinguishable as T increases. Whereas with small T there is a clear gain in using RE or GMM. The precision of the FE, RE, and GMM estimators increases in similar magnitude when N or T increases. In contrast, the standard deviation of the Pooled and Between estimator decreases only when N

Table 1: Bias and Standard Deviation

Quantile	Pooled	BE	FE	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:
Simulation performed using 10000 simulations. Standard errors in parenthesis.

Table 2: Standard Errors of $\hat{\beta}(\tau)$

Quantile	Pooled	BE	FE	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:
Simulation performed using 10000 simulations.

Table 3: Rejection Probabilities

	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:
Simulation performed using 10000 simulations.

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 RE and GMM 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 and only the FE estimator is consistent. 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. The power of the test is higher in large samples and clearly increases the larger the correlation between $P_T X_i$ and the unobserved heterogeneity α_i. An increase in N improves the power of the test substantially, 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 application.

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 would represent 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 individuals or firm in these counties. In these models, empirical researchers might include fixed effects be at the level of the city or the county. 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 1. The second model is as follows:

$$Q_{y_{it}}(\tau|x_{it}, z_i, \beta_i) = \beta_{0,i}(\tau) + x_{it}'\beta_i(\tau) \tag{13}$$

$$\beta_{0,i}(\tau) = z_i'\gamma(\tau) + \alpha_i(\tau). \tag{14}$$

Compared to the model 1, the coefficient on x_{it} is allowed to vary over i. Chetverikov et al. (2016) decide to study model 13-14 given its flexibility as it does not impose equality of the coefficients on $\tilde{x}_{it}$ and because it allows to study interaction effects (???). They suggest a two-step estimator. The first stage consists in regressing y_{it} on x_{it} 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 z_i. Their estimation focuses on estimating $\gamma(\tau)$ and does not directly provide an estimate of $\beta(\tau)$. The asymptotic distribution of Chetverikov et al. (2016) suggest that inference can be computed as if there were no first stage. Ignoring the first stage error, as we will show in the simulation later, provide a poor approximation in finite samples.
There are two main problems with model 13-14. First, if the true model follows equation 1, the estimator of Chetverikov et al. (2016) (henceforth CLP estimator) does not exploit the equality in β and the exogeneity of x_{it} between groups. Second, this estimator is consistent for the treatment effect at $x_{it} = 0$. If the true model corresponds to model 1, then their estimator consistently estimates the QTE for the whole population, but it is not invariant to reparametrization of $\tilde{x}_{it}$ and it may have poor finite sample properties. If the model is misspecified, their estimator will not converge in general to an interpretable parameter ($x_{it} = 0$ may be out of the support of x_{it}). If the treatment effect is heterogeneous in $\tilde{x}_{it}$, 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_{it}) QTE. Using our approach, we can also allow for heterogeneous effects by including interaction terms between x_{it} and z_i. 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_{it}, it requires different exogeneity assumption for consistency of $\hat{\gamma}(\tau)$. Their estimator remains consistent if x_{it} is endogenous between groups. To put it differently, the CLP estimator provides a consistent estimate of γ under the assumption that z_i *or* $z_i|x_{it}$ is exogenous. By contrast, the MD is consistent under the stronger assumption that $z_i|\tilde{x}_{it}$ is exogenous with respect to α_i. While this might look like a limitation, it is straightforward to recover consistency with our minimum distance estimator by using $(Q_T\tilde{X}_i, Z_i)$ as instruments in the second stage.

5.2 Simulations

This subsection presents Monte Carlo simulations comparing our estimator to the CLP estimator. The simulation as based on the same data generating process and sample sizes as in 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:

$$y_{it} = \beta_0(u_{it}) + x_{it}\beta(u_{it}) + w_i\gamma(u_{it}) + \alpha_i(u_{it}) \tag{15}$$

$$z_i = w_i + \eta_i + \nu_i \tag{16}$$

$$\alpha_i(u_{it}) = u_{it}\eta_i - \frac{u_{it}}{2} \tag{17}$$

where, x_{it}, w_i 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 $E[\alpha(u_{it})|w_i] = E[u_{it}\eta_i - \frac{u_{it}}{2}|w_i] = E[\frac{u_{it}}{2} - \frac{u_{it}}{2}|w_i] = 0$. 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 T is small.[10]. 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), $\alpha_i = 0$ for all i. In this case,

[10] With $T = 25$ 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 $T = 200$.

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	0.003	-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) = (25, 200)					
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) = (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:
Results based on 10000 simulations. Standard errors in parenthesis. The relative MSE reports the MSE of the MD estimator relative to the CLP estimator.

conditioning on the individual, does not affect the quantile function thus, quantile regression would be consistent. In the second case, there are individual specific effects ($\alpha_i(\tau) \neq 0$) and these are uncorrelated with the regressors. In the third case, z_i is endogenous as $\alpha_i(\tau)$ is correlated with the regressor of interest, and 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 $\gamma(\tau)$.

Table 4 shows the bias, standard deviation, and relative MSE of the CLP and MD estimators. The relative MSE reports the MSE of the MD estimator relative to that of the CLP estimator. Thus, a number smaller than 1 indicates that the MD estimator has a lower MSE. The CLP estimator seems to have a smaller bias than the MD estimator when $N = 25$. When N increases to 200, the difference disappears. There are 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

Table 5: Estimated Standard Errors and Empirical Sizes of $\hat{\gamma}(\tau)$

	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	0.073	0.247	0.068	0.116	0.078	0.249	0.067	0.116	11.327	26.243	0.039	0.070
0.5	0.063	0.212	0.074	0.116	0.141	0.238	0.072	0.111	19.497	21.399	0.058	0.074
0.9	0.031	0.103	0.065	0.119	0.229	0.217	0.076	0.108	21.143	16.644	0.068	0.083
					(N, T) = (25, 200)							
0.1	0.066	0.279	0.057	0.048	0.067	0.280	0.056	0.047	0.106	0.450	0.046	0.042
0.5	0.056	0.223	0.041	0.050	0.069	0.229	0.045	0.050	0.096	0.351	0.039	0.046
0.9	0.032	0.140	0.061	0.047	0.076	0.158	0.049	0.049	0.086	0.226	0.037	0.043
					(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:
Results based on 10000 simulations. Estimated standard errors of the MD estimator are clustered at the individual level.

difference is somewhat smaller. 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 $\alpha_i = 0$ and remains substantially larger in all scenarios.

Table 5 show the standard errors. 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 T. As a consequence, the empirical size for the 5% theoretical level of the CLP estimator can be as high as 12%. The standard errors of the MD distance estimator are remarkably close to the standard deviations of the simulation, even with small T. As T increase, the two become indistinguishable.

If $\alpha_i = 0$, 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 sample, the two estimator are indistinguishable in terms of bias and variance.

6 Conclusion

References

Abadie, A., J. D. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

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.*

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

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. (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.

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.

Hansen, B. E. (2021): *Econometrics.*

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

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.

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

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

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

Zhang, Y., H. J. Wang, and Z. Zhu (2019): "Quantile-Regression-Based Clustering for Panel Data," *Journal of Econometrics*, 213, 54–67.

A Linear Models

This section complements subsection 2.2. We consider a linear version of our estimator where OLS instead of quantile regression is used in the first stage. We consider the following standard panel data model:

$$Y_{it} = X_{it}\beta + W_i\gamma + \alpha_i + \varepsilon_{it},$$

where ε_{it} is the idiosyncratic error term. In this section, we show that linear models can be estimated using a two-step procedure. Notation is the same as in the paper, with the only difference is that the fitted values are computed using an OLS regression. More precisely, the vector of fitted values of individual i is

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

Let $Q_D = \text{diag}\{Q_1, Q_2, \ldots, Q_n\}$ and $P_D = \text{diag}\{P_1, P_2, \ldots, P_n\}$. Further, let $C(A)$ be the column space of the matrix A. The next Proposition states the equivalence of the two-step procedure using the fitted values and the conventional one-step estimator in linear models.

Proposition 2. *Denote $\hat{\delta}^{MD}_{GMM}$ the coefficient vector of a linear GMM regression of $\hat{Y}$ on X with instrument Z. Let $\hat{\delta}_{GMM}$ be the coefficient vector of the same GMM regression but with regressand Y. If $C(\tilde{X}_i) \subseteq C(Z_i)$, then $\hat{\delta}^{MD}_{GMM} = \hat{\delta}_{GMM}$.*

The proof of this Proposition and all subsequent proofs are in Appendix B.1. Proposition 2 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, $\tilde{X}_i$.[11] 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 $\hat{\delta}^{MD}_{FE}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on $\tilde{X}$ with instruments $Q_D\tilde{X}$. Let $\hat{\delta}_{FE}$ be the coefficient vector of the within estimator, i.e. a regression of Q_DY on $Q_D\tilde{X}$. Then $\hat{\delta}^{MD}_{FE} = \hat{\delta}_{FE}$.*

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

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

[11] 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. Further, it is possible to estimate between and fixed effects models using average or demeaned fitted values and regressors. In this paper, we concentrate on the IV approach as it applies to all the estimators we analyze here.

Pooled OLS and the between estimator can estimate both β and γ, but are not efficient. The RE estimator optimally combines between and the within variation to find a more efficient estimator. While FGLS is the most common estimator for the RE model, Im et al. (1999) show that the overidentified 3SLS estimator, with instruments $Z_i = (P_i X_i, Q_i \tilde{X}_i)$, is identical to the RE estimator. The 3SLS estimator is a special case of GMM with weighting matrix $W = \mathbb{E}[Z_i' \Omega Z_i]$ where Ω follows the usual RE covariance structure. Thus, by Proposition 2, the RE estimator can also be computed in two steps using the fitted values in the second stage.

Corollary 3. *Denote $\hat{\delta}_{RE}^{MD}$ the coefficient vector of a 3SLS regression of $\hat{Y}$ on X with instruments $(P_i X_i, Q_i \tilde{X}_i)$. Let $\hat{\delta}_{RE}$ be the coefficient vector of a RE regression Y on X. Then $\hat{\delta}_{RE}^{MD} = \hat{\delta}_{RE}$.*

Alternatively, the RE estimator can be implemented using the theory on 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 4 in Appendix **??**).

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 variables x_{it} and z_i are partitioned as $x_{it} = (x_{it,1}, x_{it,2})$ and $w_i = (w_{i,1}, w_{i,2})$ where $x_{it,1}$ and $z_{i,1}$ are assumed to be orthogonal to α_i. No assumption is placed on the relationship between α_i and $x_{it,2}$ and $w_{i,2}$. The model can be estimated by IV using instruments $Z_i = (Q_i \tilde{X}_i, P_i X_{i,1}, Z_{i,1})$ where $X_{i,1} = (x_{i1,1}', \ldots, x_{iT,1}')'$ and $Z_{i,1} = \mathbf{l}_i z_{i,1}$ (see, e.g. Hansen, 2021). Thus, it follows by Proposition 2 that the Hausman Taylor model can be estimated in two stages.

Proposition 3. *Denote $\hat{\delta}_{HT}^{MD}$ the coefficient vector of a 2SLS regression of $\hat{Y}$ on X with instruments $(Q_i \tilde{X}_i, P_i X_{i,1}, Z_{i,1})$. Let $\hat{\delta}_{HT}$ be the coefficient vector of the Hausman Taylor Estimator based on a regression Y on X. Then $\hat{\delta}_{HT}^{MD} = \hat{\delta}_{HT}$.*

B Proofs

B.1 Linear Models

Proof of Proposition 2. Define the projection matrix $P = \tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'$. Then since Z_i is in the column space of $\tilde{X}_i$,

$$PZ_i = Z_i \tag{18}$$

The MD estimator with a GMM second stage is:

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

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

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

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 18. Thus, it follows directly that $\hat{\delta}_{MD}$ is equal to $\hat{\delta}_{GMM}$. ■

Proof of Corollary 1. First, note that since $Q_i\tilde{X}_i = \dot{X}_i$, $\dot{X}_i$ lies in the column space of $\tilde{X}_i$. Then, we apply Proposition 2 and since $k = l$, the 2SLS estimator reduces to an IV estimator. It follows that a 2SLS (or IV) regression of $\hat{Y}$ on X_i with instrument Z_i is algebraically identical to a 2SLS regression with Y_i as dependent variable, i.e. $\hat{\delta}^{MD}_{IV} = \hat{\delta}_{IV}$. Denote $Q_i = I_i - \mathbf{l}_i(\mathbf{l}_i'\mathbf{l}_i)^{-1}\mathbf{l}_i'$ Then,

$$\begin{aligned}
\hat{\delta}_{IV} &= \left(\sum_{i=1}^{n} Z_i'\tilde{X}_i'\right)^{-1}\sum_{i=1}^{n} Z_i'Y_i \\
&= \left(\sum_{i=1}^{n} \dot{X}_i'\tilde{X}_i'\right)^{-1}\sum_{i=1}^{n} \dot{X}_i'Y_i \\
&= \left(\sum_{i=1}^{n} \tilde{X}_i'Q_i'\tilde{X}_i'\right)^{-1}\sum_{i=1}^{n} X_i'Q_iY_i \\
&= \left(\sum_{i=1}^{n} \dot{X}_i{}'\dot{X}_i'\right)^{-1}\sum_{i=1}^{n} \dot{X}_i'\dot{Y}_i = \hat{\delta}_{FE},
\end{aligned}$$

where the second line follows since $Z_i = \dot{X}_i$, the third and last line by $Q_iX_i = \dot{X}_i$, $Q_iY_i = \dot{Y}_i$ and since Q_i is idempotent. ∎

Proof of Corollary 2. First, note that since $P_i\tilde{X}_i = \dot{X}_i$, $\dot{X}_i$ lies in the column space of $\tilde{X}_i$. Then, we apply Proposition 2 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, i.e. $\hat{\delta}^{MD}_{IV} = \hat{\delta}_{IV}$. Denote $P_i = \mathbf{l}_i(\mathbf{l}_i'\mathbf{l}_i)^{-1}\mathbf{l}_i'$. Then,

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

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

Proposition 4. *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 $\mathbb{E}[y_i - X_i\delta|X_i] = 0$ and $\mathbb{E}[\hat{y}_i - X_i\delta|X_i] = 0$ imply the same optimal instrument.*

Proof. The optimal instrument takes the form $Z_i^* = \mathbb{E}[g_i(\delta_0)g_i(\delta_0)'|Z_i]^{-1}R_i(\delta)$, where $R_i(\delta) = \mathbb{E}[\frac{\partial}{\partial\delta}g_i(\delta)|Z_i]$. For both moment conditions, $R_i(\delta)$ is identical. Then,

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

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

$$\begin{aligned}\mathbb{E}[(y_i - X_i\delta)(y_i - X_i\delta)'|X_i]^{-1} &= \mathbb{E}[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})'|X_i]^{-1} \\ &= \left(\mathbf{I}_T\sigma_\varepsilon^2 + \mathbf{l}_T\mathbf{l}_T'\sigma_\alpha^2\right)^{-1}.\end{aligned}$$

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 $V_i = (\frac{1}{T}\tilde{X}_i'\tilde{X}_i)^{-1}\sigma_\varepsilon^2$ 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 5. *Denote $\hat{V}_\delta$ the clustered covariance matrix of $\hat{\delta}$ from a GMM regression of Y on X with instrument Z. Let $\hat{V}_{\hat{\delta}_{2SLS}^{MD}}$ be the clustered covariance matrix of $\hat{\delta}^{MD}$ from a GMM regression of $\hat{Y}$ on X with instrument Z. Let the clusters be at weakly higher level than i. Then, $\hat{V}_{\hat{\delta}_{2SLS}^{MD}} = \hat{V}_{\hat{\delta}_{2SLS}}$.*

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 higher than the individuals i. Let $g = 1, \dots, G$ index the clusters and assume that each of the clusters has n_g observations. There are G clusters. 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

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

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

Proof. Denote $Z_g = (Z_{1g}, \dots, Z_{n_g g})'$ and $Y_g = (y_{1g}, \dots y_{n_g g})'$. The first and third terms of the expression are identical for both estimators. Thus, we focus on the middle term. Let $\hat{u}_g = Y_g - \hat{Y}_g$ be the vector of residuals from the regression using Y as dependent variable, and let $\hat{u}_g^{MD} = \hat{Y}_g^{FS} - \hat{Y}_g$ 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 2, $\hat{\delta}^{MD} = \hat{\delta}$. Thus, the fitted values of both estimators are identical. Next, define $\tilde{X}_g = \text{diag}\{\tilde{X}_{1g}, \ldots, \tilde{X}_{n_g g}\}$ and recall that regressing y_g on $\tilde{X}_g$ is the same as performing n_g separate regressions. Note that Z_g is in the column space of $\tilde{X}_g$. Define the projection matrix $P_g = X_g(\tilde{X}_g'\tilde{X}_g)^{-1}\tilde{X}_g'$. Since Z_i is in the column space of $\tilde{X}_g$,

$$PZ_g = Z_g. \tag{19}$$

Then,

$$\begin{aligned}
Z_g'\hat{u}_g^{MD} &= Z_g'(\hat{Y}_g^{FS} - \hat{Y}_g) \\
&= Z_g'\tilde{X}_g\hat{\beta}_g - Z_g\hat{Y}_g \\
&= Z_g'\tilde{X}_g(\tilde{X}_g'\tilde{X}_g)^{-1}\tilde{Z}_g'Y_g - Z_g\hat{Y}_i \\
&= Z_g'(Y_g - \hat{Y}_g) = Z_g'\hat{u}_g,
\end{aligned}$$

where $\hat{\beta}_g = \left(\hat{\beta}_{1g}', \ldots, \hat{\beta}_{n_g g}'\right)'$ is the coefficient vector of a regression of y_g on $\tilde{X}_g$ and the fourth line follows by 19. Since this holds for all g, the result follows directly.

■

B.2 Optimal Instruments and Minimum Distance

Proposition 6. *The IV regression with instrument $Z_i^* = \tilde{X}_i(\tilde{X}_i'\tilde{X}_iV_i\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'X_i = (\tilde{X}_iV_i\tilde{X}_i')^{+}X_i$ equals the efficient minimum distance estimator.*

Proof.

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

where the last two lines are the expressions of the GIV estimator with the respective instruments. The third line follows by the relationship between $\tilde{X}_i$ and X_i: $\tilde{X}_iR_i = X_i$ and the fourth 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 Consistency of the estimated Covariance Matrix

Consider the following estimator of the clustered covariance matrix of $\hat{\delta}$

$$\hat{V}_\delta(\tau) = \left(\frac{1}{N}X'Z\hat{W}\frac{1}{N}Z'X\right)^{-1}\frac{1}{N}X'Z\hat{W}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i'\tilde{u}_i(\tau)\tilde{u}_i(\tau)'Z_i\right)\frac{1}{N}\hat{W}Z'X\left(\frac{1}{N}X'Z\hat{W}\frac{1}{N}Z'X\right)^{-1}$$
$$= \left(S_{XZ}\hat{W}S_{XZ}'\right)^{-1}S_{XZ}\hat{W}\left(\frac{1}{NT^2}\sum_{i=1}^{N}Z_i'\tilde{u}_i(\tau)\tilde{u}_i(\tau)'Z_i\right)\frac{1}{NT}\hat{W}S_{XZ}'\left(S_{XZ}\hat{W}S_{XZ}'\right)^{-1},$$

where $\tilde{u}_i$ is the vector of errors of individual i in the second stage.
First, by equation ????, $\left(S_{XZ}\hat{W}S_{XZ}'\right)^{-1}S_{XZ}\hat{W} \xrightarrow{p} (\Sigma_{XZ}W\Sigma_{XZ})^{-1}\Sigma_{XZ}W$. Next, we focus on the term in the middle. We can decompose the estimated error as follows:

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

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 [HERE THE T are missing!]

$$\begin{aligned}
\frac{1}{T^2}\frac{1}{N}\sum_{i=1}^{N}Z_i'\tilde{X}_i(\hat{\beta}_i-\beta_i)(\delta-\hat{\delta})'X_i'Z_i &= O_p\left(\frac{1}{NT^{1/2}}\right)\\
\frac{1}{T^2}\frac{1}{N}\sum_{i=1}^{N}Z_i'\tilde{X}_i(\hat{\beta}_i-\beta_i)\alpha_i'Z_i &= O_p\left(\frac{1}{NT^{1/2}}\right)\\
\frac{1}{T^2}\frac{1}{N}\sum_{i=1}^{N}Z_i'\tilde{X}_i(\delta-\hat{\delta})\alpha_i'Z_i &= O_p\left(\frac{1}{N^{3/2}}\right)\\
\frac{1}{T^2}\frac{1}{N}\sum_{i=1}^{N}X_i'X_i(\hat{\delta}-\delta)(\hat{\delta}-\delta)'X_i'X_i &= O_p\left(\frac{1}{N^{3/2}}\right).
\end{aligned}$$

This part if only to convince myself! First line:

$$
\begin{aligned}
&\frac{1}{T^2N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'(\hat{\beta}_i-\beta_i)\right)\left(\sum_{t=1}^{T}Z_{it}X_{it}'(\delta-\hat{\delta})\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'\right)(\hat{\beta}_i-\beta_i)(\delta-\hat{\delta})'\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}X_{it}'\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}O_p(1)O_p\left(\frac{1}{\sqrt{T}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)O_p\left(1\right)=O_p\left(\frac{1}{NT^{1/2}}\right)
\end{aligned}
$$

Second Line:

$$
\begin{aligned}
&\frac{1}{T^2N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'(\hat{\beta}_i-\beta_i)\right)\left(\sum_{t=1}^{T}Z_{it}\alpha_i\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'\right)(\hat{\beta}_i-\beta_i)\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}\alpha_i\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'\right)(\hat{\beta}_i-\beta_i)\left(\bar{Z}_i\alpha_i\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}O_P(1)O_P\left(\frac{1}{\sqrt{T}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)=O_P\left(\frac{1}{N\sqrt{T}}\right)
\end{aligned}
$$

There the last line follows, since $\mathbb{E}[Z_{it}\alpha_i]=0$. Line 3:

$$
\begin{aligned}
&\frac{1}{T^2N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'(\hat{\delta}-\delta)\right)\left(\sum_{t=1}^{T}Z_{it}\alpha_i\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'\right)(\hat{\delta}-\delta)\left(\bar{Z}_i\alpha_i\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}O_P(1)O_P\left(\frac{1}{\sqrt{T}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)=O_P\left(\frac{1}{N^{3/2}}\right)
\end{aligned}
$$

Fourth Line

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}\tilde{X}_{it}'\right)(\hat{\delta}-\delta)(\delta-\hat{\delta})'\left(\frac{1}{T}\sum_{t=1}^{T}Z_{it}X_{it}'\right)'\\
&\frac{1}{N}\sum_{i=1}^{N}O_p(1)O_p\left(\frac{1}{\sqrt{N}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)O_p\left(1\right)=O_p\left(\frac{1}{N^{3/2}}\right)
\end{aligned}
$$

For the first equation note that $(\delta - \hat{\delta}) = O_p(\frac{1}{N})$ and $(\hat{\beta}_i - \beta_i) = O_P\left(\frac{1}{T}\right)$ Thus, we can write the term in the middle as

$$\begin{aligned}
\frac{1}{NT^2}\sum_{i=1}^{N} Z_i'\tilde{u}_i\tilde{u}_i'Z_i =& \frac{1}{NT^2}\sum_{i=1}^{N}\left(Z_i'\tilde{X}_i(\hat{\beta}_i-\beta_i)(\hat{\beta}_i-\beta_i)'\tilde{X}_i'Z_i + Z_i'a_ia_i'Z_i\right) + O_p\left(\frac{1}{N^{3/2}}\right)\\
=& \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau))\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau))\right)'\\
&+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\alpha_i\right)' + O_p\left(\frac{1}{N^{3/2}}\right)\\
=& \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau))\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau))\right)'\\
&+\frac{1}{N}\sum_{i=1}^{N}\left(\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2\right) + O_p\left(\frac{1}{N}\right)\\
=& \,\mathbb{E}\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right] + \mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2] + O_P\left(\frac{1}{N^{3/2}}\right) + O_P\left(\frac{1}{N^{1/2}}\right),
\end{aligned}$$

where the last line follows since $\frac{1}{T}\sum_{t=1}^{T} z_{it}\tilde{x}_{it}'(\hat{\beta}_i(\tau)-\beta_i(\tau)) = \Sigma_{ZXi}\frac{1}{T}\sum_{t=1}^{T}\phi(\tilde{x}_{it},y_{it}) + o_p\left(\frac{1}{N^{3/2}T}\right)$ and from equation [in joint normality 3 ???] we know that $Var\left(\Sigma_{ZXi}\frac{1}{T}\sum_{i=1}^{T}\phi(\tilde{x}_{it},y_{it})\right) = E\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right]$. Thus, the first term converges to $E\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right]$ which is $O\left(\frac{1}{T}\right)$ and the second term converges to $\mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2]$. By the continuous mapping theorem, it follows that

$$\hat{V}_\delta(\tau) \xrightarrow{p} \left(\Sigma_{XZ}W\Sigma_{XZ}\right)^{-1}\Sigma_{XZ}W\left(\mathbb{E}\left[\Sigma_{ZXi}\frac{V_i}{T}\Sigma_{ZXi}'\right] + \mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2]\right)W\Sigma_{XZ}'\left(\Sigma_{XZ}W\Sigma_{XZ}'\right)^{-1}.$$

$$\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2\right) = \frac{1}{N}\mathrm{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2\right) = O_p\left(\frac{1}{N}\right)$$

THE TERM $\frac{1}{N}\sum_{i=1}^{N}\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2 - \mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2] = O_P(1/\sqrt{N})$:((By Chebyshev's inequality

$$\mathbb{P}\left(||\frac{1}{N}\sum_{i=1}^{N}\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2 - \mathbb{E}[\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2]|| \geq u\right) \leq \frac{\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\bar{z}_i\bar{z}_i'\alpha_i(\tau)^2\right)}{u^2}$$

∎