Quantile Regressions via Method of Moments with multiple fixed effects

Fernando Rios-Avila[a,*], Leonardo Siles, Gustavo Canavire-Bacarreza[b]

[a]*Levy Economics Institute of Bard College,*

[b]*World Bank,*

Abstract

This paper proposes a new method to estimate quantile regressions with multiple fixed effects. The method expands on the strategy proposed by Machado and Santos Silva (2019), allowing for multiple fixed effects, and providing various alternatives for the estimation of Standard errors. We provide Monte Carlo simulations to show the finite sample properties of the proposed method in the presence of two sets of fixed effects. Finally, we apply the proposed method to estimate the determinants of the surplus of government as a share of GDP allowing for both time and country fixed effects.

Keywords: Fixed effects, Linear heteroskedasticity, Location-scale model, Quantile regression

1. Introduction

Quantile regression (QR), introduced by Koenker and Bassett (1978), is an estimation strategy used for modeling the relationships between explanatory variables X and the conditional quantiles of the dependent variable $q_\tau(y|x)$. Using QR one can obtain richer characterizations of the relationships between dependent and independent variables, by accounting for otherwise unobserved heterogeneity.

A relatively recent development in the literature has focused on extending quantile regressions analysis to include individual fixed effects in the framework of panel data. However, as described in Neyman and Scott (1948), and Lancaster (2000), when individual fixed effects are included in quantile regression analysis it generates an incident parameter problem. While many strategies have been proposed for estimating this type of model (see Galvao and Kengo (2017) for a brief review), neither has become standard because of their restrictive assumptions in regards to the individual effects, the computational complexity, and implementation.

More recently, Machado and Santos Silva (2019) (MSS hereafter) proposed a methodology based on a conditional location-scale model similar to the one described in He (1997) and Zhao (2000), for the estimation of quantile regressions models for panel data via a method of moments. This method allows individual fixed effects allowing to have heterogeneous effects on the entire conditional distribution of the

*Corresponding author

Email addresses: `friosavi@levy.org` (Fernando Rios-Avila), `leo@gmail.com` (Leonardo Siles), `gcanavire@worldbank.org` (Gustavo Canavire-Bacarreza)

outcome, rather constraining their effect to be a location shift only as in Canay (2011), Koenker (2004), and Lancaster (2000).

In principle, under the assumption that data generating process behind the data is based on a multiplicative heteroskedastic process that is linear in parameters (Cameron and Trivedi, 2005, Machado and Santos Silva (2019), He (1997), Zhao (2000)), the effect of a variable X on the q_{th} quantile can be derived as the combination of a location effect, and scale effect moderated by the quantile of an underlying i.i.d. error. For statistical inference, MSS derives the asymptotic distribution of the estimator, suggesting the use of bootstrap standard errors, as well.

While this methodology is not meant to substitute the use of standard quantile regression analysis, given the assumptions required for the identification of the model, it provides a simple and fast alternative for the estimation of quantile regression models with individual fixed effects.

In this framework, our paper expands on Machado and Santos Silva (2019), following some of the suggestions by the authors regarding further research. First, making use of the properties of GMM estimators, we derive various alternatives for the estimation of standard errors based on the empirical Influence functions of the estimators. Second, we reconsider the application of Frisch–Waugh–Lovell (FWL) theorem (Frisch and Waugh (1933) and Lovell (1963)) to extend the MSS estimator to allow for the inclusion of multiple fixed effects, for example, individual and year fixed effects.

The rest of the paper is restructured as follows. Section 2 presents the basic setup of the location-Scale model described in He (1997) and Zhao (2000), tying the relationship between the standard quantile regression model, and the location and scale model. It also revisits MSS methodology, proposing alternative estimators for the standard errors based on the properties of GMM estimators and the empirical influence functions. It also shows that FWL theorem can be used to control for multiple fixed effects. Section 3 presents the results of a small simulation study and Section 4 illustrates the application of the proposed methods with two empirical examples. Seccion 5 concludes.

2. Methodology

2.1. Quantile Regression: Location-Scale model

Quantile regressions are used to identify relationships between the explanatory variables x and the conditional quantiles of the dependent variable $Q(y|\tau, X)$. This relationship is commonly assumed to follow a linear functional form:

$$q(Y|X, \tau) = X\beta(\tau) \tag{1}$$

This allows for nonlinearities in the effect of X on Y across all values of τ. This formulation can also be related to a random coefficient model, where all coefficients are assumed to be some nonlinear function of τ, where τ follows a random uniform distribution.

An alternative formulation of quantile regressions is the location-scale model. This approach assumes that the conditional quantile of Y given X and τ can be expressed as a combination of two models: the location model, which describes the central tendency of the conditional distribution, and the scale model, which describes deviations from the central tendency:

$$q(Y|X,\tau) = X\beta + X\gamma(\tau) \tag{2}$$

Here, the location parameters β are typically identified using a linear regression model (as in Machado and Santos Silva (2019)), or a median regression (as in Melly (2005)), and the scale parameters $\gamma(\tau)$ can be estimated using standard approaches.

Both the standard quantile regression (Equation 1) and the location-scale specification (Equation 2) can be estimated as the solution to a weighted minimization problem:

$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \left(\sum_{i \in y_i \geq x_i'\beta} \tau(y_i - x_i'\beta) - \sum_{i \in y_i < x_i'\beta} (1-\tau)(y_i - x_i'\beta) \right) \tag{3}$$

One characteristic of this estimator is that the $\beta(\tau)$ coefficients are identified locally, and thus the estimated quantile coefficients will exhibit considerable variation when analyzed across τ. It is also implicit that if one requires an analysis of the entire distribution, it would be necessary to estimate the model for each quantile.[1]

One insightful extension to the location-scale parameterizations suggested by He (1997), Cameron and Trivedi (2005), and Machado and Santos Silva (2019) is to assume that the data-generating process (DGP) can be written as a linear model with a multiplicative heteroskedastic process that is linear in parameters.[2]

$$\begin{aligned} y_i &= x_i'\beta + \nu_i \\ \nu_i &= \varepsilon_i \times x_i'\gamma \end{aligned} \tag{4}$$

Under the assumption that ε is an independent and identically distributed (iid) unobserved random variable that is independent of X, the conditional quantile of Y given X and τ can be written as:

$$q(Y|X,\tau) = X\beta + q(\varepsilon|\tau) \times X\gamma \tag{5}$$

[1] There are other estimators that provide smoother estimates for the quantile regression coefficients using a kernel local weighted approach (Kaplan and Sun, 2017), as well as identifying the full set of quantile coefficients simultaneously assuming some parametric functional forms (Frumento and Bottai, 2016).

[2] Machado and Santos Silva (2019) also discuss a model where heteroskedasticity can be an arbitrary nonlinear function $\sigma(x_i'\gamma)$, but develop the estimator for the linear case, i.e., when $\sigma()$ is the identity function.

In this setup, the traditional quantile coefficients are identified as the location model coefficients, plus the scale model coefficients moderated by the τ_{th} unconditional quantile of the standardized error ε.

$$\beta(\tau) = \beta + q(\varepsilon|\tau) \times \gamma \tag{6}$$

While this specification imposes a strong assumption on the DGP, it has two advantages over the standard quantile regression model. First, because the location and scale model can be identified globally, with only a single paramater ($q(\varepsilon|\tau)$) requiring local estimation, this estimation approach would be more efficient than the standard quantile regression model (Zhao (2000)). Second, under the assumption that $X\gamma$ is strictly possitive, the model would produce quantile coefficients that do not cross.

Following MSS, the quantile regression model defined by Equation 5 can be estimated using a method of moments approach. And while its possible to identify all coefficients $(\beta, \gamma, q(\varepsilon|\tau))$ simultaneously, we describe and use the implementation approach advocated by MSS which identifies each set of coefficients separately.

1. The location model can be estimated using a standard linear regression model, where the dependent variable is the outcome Y, and the independent variables are the explanatory variables X (including a constant) with an error u, which is by definition heteroskedastic. In this case, the location model coefficients are identified under the following condition:

$$\begin{aligned} y_i &= x_i'\beta + \nu_i \\ E[x_i\nu_i] &= 0 \end{aligned} \tag{7}$$

2. After the location model is estimated, the scale coefficients can be identified by modeling heteroskedasticity as a linear function of characteristics X. For this we use the absolute value of the errors from the location model u as dependent variable, which would allow us to estimate the conditional standard deviation (rather than conditional variance) of the errors. In this case, the coefficients are identified under the following condition:

$$\begin{aligned} |\nu_i| &= x_i'\gamma + \omega_i \\ E[x_i\omega_i] &= 0 \\ E[x_i(|\nu_i| - x_i'\gamma)] &= 0 \end{aligned} \tag{8}$$

3. Finally, given the location and scale coefficients, the τ_{th} quantile of the error ε can be estimated using the following condition:

$$\begin{aligned} E\left[\mathbb{1}\left(x_i'(\beta + \gamma q(\varepsilon|\tau)) \geq y_i\right) - \tau\right] &= 0 \\ E\left[\mathbb{1}\left(q(\varepsilon|\tau) \geq \frac{y_i - x_i'\beta}{x_i'\gamma}\right) - \tau\right] &= 0 \end{aligned} \tag{9}$$

Where one identifies the quantile of the error ε using standardized errors $\frac{y_i - x_i'\beta}{x_i'\gamma}$, or by finding the values that identify the overall quantile coefficients $\beta(\tau) = \beta + \gamma q(\varepsilon|\tau)$. Afterwords, the conditional quantile coefficients is simply defined as the combination of the location and scale coefficients.

2.2. Standard Errors: GLS, Robust, Clustered

As discussed in the previous section, the estimation of quantile regression coefficients using the location-scale model with heteroskedstic linear errors can be estimated using a the following set of moments, which fits in the Generalized Method of Moments framework:

$$\begin{aligned} E[x_i \nu_i] &= E[h_{1,i}] = 0 \\ E[x_i(|\nu_i| - x_i\gamma)] &= E[h_{2,i}] = 0 \\ E\left[\mathbb{1}\left(q(\varepsilon|\tau) \geq \frac{y_i - x_i'\beta}{x_i'\gamma}\right) - \tau\right] &= E[h_{3,i}] = 0 \end{aligned} \tag{10}$$

Under the conditions described in Newey and McFadden (1994) (see section 7), Cameron and Trivedi (2005) (see chapter 6.3.9) or as shown in Machado and Santos Silva (2019), the location, scale and residual quantile coefficients are asymptotically normal.[3]

Call $\theta = [\beta' \quad \gamma' \quad q(\varepsilon|\tau)']'$ the set of coefficients that are identified by the modement conditions in Equation 10, a just identified model. And the function h_i is a vector function that stacks all the moments at the individual level described in Equation 10. Then $\hat{\theta}$ follows a normal distribution with mean θ and variance-covariance matrix $V(\theta)$ that is estimated as:

$$\hat{V}(\hat{\theta}) = \frac{1}{N}\bar{G}(\hat{\theta})^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} h_i h_i'\Big|_{\theta=\hat{\theta}}\right)\bar{G}(\hat{\theta})^{-1}$$

Which is equivalent to the Eicker-White Heteroskedastic-Consistent estimator for least-squares estimators.

Here, the inner product is the moment covariance matrix, and $\bar{G}(\theta)$ is the Jacobian matrix of the moment equations evaluated at $\hat{\theta}$.

$$\bar{G}(\theta) = -\frac{1}{N}\sum_{i=1} \frac{\partial h_i}{\partial \theta'}\Big|_{\theta=\hat{\theta}}$$

In this framework, the quantile regression coefficients, a combination of the location-scale-quantile estimates, will follow a normal distribution with mean $\beta(\tau) = \beta + q(\varepsilon|\tau)\gamma$ and variance-covariance matrix equal to:

[3] Zhao (2000) also shows that the quantile coefficients for the location-scale model also follows a normal distribution, but uses the assumption that the location model is derived using a least absolute deviation approach (median regression).

$$\hat{V}(\beta(\tau)) = \Xi \hat{V}(\hat{\theta})\Xi'$$

where Ξ is a $k \times (2k+1)$ matrix defined as:

$$\Xi = [I(k), \hat{q}(\varepsilon|\tau) \times I(k), \hat{\gamma}] \tag{11}$$

with $I(k)$ being an identity matrix of dimension k (number of explanatory variables in X including the constant).

While it is possible to estimate the variance-covariance matrix using simultaneous model estimation, for a just identified model, it is more efficient to estimate each set of coefficients separately. Afterwards, the variance-covariance matrix can be estimated using the empirical influence functions of the estimators (see Jann (2020) for an overview of the application, and Hampel et al. (2005) for an in-depth review).

Specifically, given an arbitrary vector of empirical influence functions $\lambda_i(\theta)$, the variance-covariance matrix can be estimated as:

$$\hat{V}(\theta) = \frac{1}{N^2}\sum_{i=1}^{N} \lambda_i(\theta)\lambda_i(\theta)' \tag{12}$$

where the influence functions are defined as:

$$\lambda_i(\theta) = \bar{G}(\theta)^{-1} h_i(\theta)$$

For the specific case of quantile regressions via momoments, the influence functions for the location, scale and quantile coefficients are:[4]

$$\begin{aligned}
\lambda_i(\theta) &= \begin{bmatrix} \lambda_i(\beta) \\ \lambda_i(\gamma) \\ \lambda_i(q(\varepsilon|\tau)) \end{bmatrix} \\
\lambda_i(\beta) &= N(X'X)^{-1} x_i (x_i'\gamma) \times \varepsilon_i \\
\lambda_i(\gamma) &= N(X'X)^{-1} x_i (x_i'\gamma) \times (\tilde{\varepsilon}_i - 1) \\
\lambda_i(q(\varepsilon|\tau)) &= \frac{\tau - \mathbb{1}(q(\varepsilon|\tau) \geq \varepsilon_i)}{f_\varepsilon(q(\varepsilon|\tau))} - \frac{x_i'\gamma \times \varepsilon_i}{\bar{x}_i'\gamma} - q(\varepsilon|\tau)\frac{x_i'\gamma \times (\tilde{\varepsilon}_i - 1)}{\bar{x}_i'\gamma}
\end{aligned} \tag{13}$$

The different types of Standard errors estimation, thus, depend on the assumptions imposed for the estimation of $V(\theta)$.

[4]The derivation of the influence functions can be found in Section 6.

2.2.1. *Robust Standard Errors*

The first, and most natural standard error estimator is given by equation Equation 12. This is equivalent to the Eicker-White Heteroskedastic-Consistent estimator for least-squares estimators. Considering the location-scale model, the variance-covariance matrix for the quantile coefficients can be estimated as:

$$\hat{V}_{robust}\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\\ \hat{q}(\varepsilon|\tau)\end{pmatrix} = \frac{1}{N^2}\begin{pmatrix}\sum\lambda_i(\beta)\lambda_i(\beta)' & \sum\lambda_i(\beta)\lambda_i(\gamma)' & \sum\lambda_i(\beta)\lambda_i(q(\varepsilon|\tau))'\\ \sum\lambda_i(\gamma)\lambda_i(\beta)' & \sum\lambda_i(\gamma)\lambda_i(\gamma)' & \sum\lambda_i(\gamma)\lambda_i(q(\varepsilon|\tau))'\\ \sum\lambda_i(q(\varepsilon|\tau)\lambda_i(\beta)' & \sum\lambda_i(q(\varepsilon|\tau))\lambda_i(\gamma)' & \sum\lambda_i(q(\varepsilon|\tau))\lambda_i(q(\varepsilon|\tau))'\end{pmatrix}$$

This estimator of Standard errors should be robust to arbitrary heteroskedasticity. However, because the location-scale specification relies on the correct specification of the model heteroskedasticity, large differences in Standard errors compared to GLS-standard errors may be an indication of misspecification of the model.

2.2.2. *Clustered Standard Errors*

Because one of the typical applications of quantile regressions is the analysis of panel data, allowing for clustered standard errors at the individual level is important. If the unobserved error ε is correlated within clusters, GLS-Standard errors could be severily biased. The standard recommendation has been to report block-bootstrap standard errors, clustering at the individual level.

Since we have access to the influence functions, it is straight forward to estimate one-way clustered standard errors.

Call N_G to be the total number of clusteres g, where $g = 1 \dots N_G$. The clustered variance covariance matrix is given by:[5]

$$\hat{V}_{clustered}\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\\ \hat{q}(\varepsilon|\tau)\end{pmatrix} = \frac{1}{N^2}\left(\sum_{g=1}^{N_G} S\lambda_i(\theta)S\lambda_i(\theta)'\right)$$

Where $S\lambda_i(\theta)$ is the sum of the influence functions over all observations within a given cluster g.

$$S\lambda_i(\theta) = \sum_{i\in g}\lambda_i(\theta)$$

2.2.3. *GLS Standard Errors*

The standard errors proposed by MSS can be understood as an application of generalized least squares (GLS), which will be valid as long as the model for heteroskedasticity is correctly specified.[6] To estimate

[5] It should be noted that one could just as well apply the insights of Cameron et al. (2011), allowing for multiway clustering.

[6] As discussed in most econometric textbooks, like Cameron and Trivedi (2005), one approach to correct for heteroskedasticity, when the heteroskedasticity functional form is known, or can be estimated, is to use weighted least squares. While

the GLS-Standard errors, we make use of the following property:

Consider the influence functions and robust variance-covariance matrix for the location coefficients:

$$\begin{aligned}\hat{V}(\hat{\beta}) &= \frac{1}{N}\sum_{i}^{N}\lambda_i(\beta)\lambda_i(\beta)' \\ &= \frac{1}{N}(X'X)^{-1}\sum_{i}^{N}x_ix_i'(x_i'\gamma\times\varepsilon_i)^2(X'X)^{-1}\end{aligned}$$

Under the assumption that the model for heteroskedasticity is correctly specified, we can apply the law of iterated expectations and rewrite the variance-covariance matrix as:

$$\begin{aligned}\hat{V}(\hat{\beta}) &= \frac{1}{N}\sum_{i}^{N}\lambda_i(\beta)\lambda_i(\beta)' \\ &= E(\varepsilon_i^2)\frac{1}{N}(X'X)^{-1}\sum_{i}^{N}x_ix_i'(x_i'\gamma)^2(X'X)^{-1} \\ &= \sigma_\varepsilon^2\frac{1}{N}(X'X)^{-1}\hat{\Omega}_{\beta\beta}(X'X)^{-1}\end{aligned}$$

This standard error estimator is an application of GLS that accounts for the heteroskedasticity the model uses to identify the quantile coefficients. We can apply the same principle to find the GLS-Standard errors for the system of location-scale and quantile coefficients. To do this, define the following modified influence functions:

$$\begin{aligned}\tilde{\lambda}_{1,i} &= \tilde{\lambda}_{2,i} = N(X'X)^{-1}x_i(x_i'\gamma) \\ \tilde{\lambda}_{3,i} &= x_i'\gamma \\ \tilde{\psi}_1 &= \varepsilon_i \\ \tilde{\psi}_2 &= \tilde{\varepsilon}_i - 1 \\ \tilde{\psi}_3 &= \frac{1}{x_i'\gamma}\frac{\tau - \mathbb{1}(q(\varepsilon|\tau)\geq\varepsilon_i)}{f_\varepsilon(q(\varepsilon|\tau))} - \frac{\varepsilon_i}{\bar{x}_i'\gamma} - q(\varepsilon|\tau)\frac{(\tilde{\varepsilon}_i - 1)}{\bar{x}_i'\gamma}\end{aligned}$$

Then, the GLS-Standard errors for the location-scale and quantile coefficients can be estimated as:

$$\hat{V}_{gls}\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\\ \hat{q}(\varepsilon|\tau)\end{pmatrix} = \frac{1}{N^2}\begin{pmatrix}\hat{\sigma}_{11}\hat{\Omega}_{11} & \hat{\sigma}_{12}\hat{\Omega}_{12} & \hat{\sigma}_{11}\hat{\Omega}_{13}\\ \hat{\sigma}_{21}\hat{\Omega}_{21} & \hat{\sigma}_{22}\hat{\Omega}_{22} & \hat{\sigma}_{21}\hat{\Omega}_{23}\\ \hat{\sigma}_{31}\hat{\Omega}_{31} & \hat{\sigma}_{32}\hat{\Omega}_{32} & \hat{\sigma}_{33}\hat{\Omega}_{33}\end{pmatrix}$$

where

feasible, however, this approach would defeat the purpose of identifying quantile effects exploiting the heteroskedasticity of the model.

$$\hat{\Omega}_{ij} = \frac{1}{N}\sum_{i}^{N}\tilde{\lambda}_i(\theta)\tilde{\lambda}_j(\theta)'$$
$$\hat{\sigma}_{ij} = \frac{1}{N}\sum_{i}^{N}\phi_i\phi_j$$

This estimator of Standard errors is equivalent to the one derived by MSS by Theorem 3.

2.3. Multiple Fixed Effects: Expanding on Machado and Santos Silva (2019)

Using the setup described in the previous section, Machado and Santos Silva (2019) proposes an extension to the model proposed by He (1997) that would allow for the estimation of quantile regression models with panel data, allowing for the inclusion of individual fixed effects. However, as the authors suggest, the methodology can be generalized to allow for the inclusion of multiple fixed effects. This type of analysis may be useful when considering data such as employer-employee linked data (Abowd et al., 2006), or teacher-student linked data (Harris and Sass, 2011). Or, in the most common case, allowing to control for both individual and time fixed effects.

Reconsider the original model, and assume there are sets of unobserved heterogeneity that are assumed to be constant across observations, if they belong to common groups. In panel data, the groups would be the individual fixed effects and the time fixed effects. Without loss of generality, we can assume that the data generating process is as follows:

$$y_i = x_i'\beta + \delta_{g1} + \delta_{g2} + \nu_i$$
$$\nu_i = \varepsilon_i \times (x_i'\gamma + \zeta_{g1} + \zeta_{g2})$$

where we assume x_i vary across groups g_1 and g_2, thus are not collinear, and that $\delta's$ and $\zeta's$ are the location and scale effects associated with groups fixed effects.[7]

If the dimension of groups g_k is low, this model could be estimated using a dummy inclussion approach following Section 2.1, and standard errors obtained as discussed in Section 2.2. However, if the dimensionality of g_k is high, the dummy inclusion approach may not be feasible. Instead, a more feasible approach is to apply the Frisch-Waugh-Lovell (FWL) theorem, and partial out the impact of the group fixed effects on the control variables x_i, the outcome of interest y_i, with a similar approach for the identification of $\sigma(x)$. In the case of unbalanced setups, with multiple groups, the estimation involves iterative processes for which various approaches have been suggested and implemented (see Correia (2016), Gaure (2013), Rios-Avila (2015), among others).

When applying the partialing out approach, some modifications to the approach described in Section 2.1 are needed.

[7]We could just as well consider multiple sets of fixed effects

1. For all dependent and independent variables in the model ($w = y, x$), we partial out the group fixed effects, and obtain the centered-residualized variables:

$$w_i = \delta_{g1}^w + \delta_{g2}^w + u_i^w$$
$$w_i^{rc} = E(w_i) + \hat{u}_i^w$$

2. We estimate the location model using the centered-residualized variables:[8]

$$y_i^{rc} = x_i^{rc'}\beta + \nu_i$$

3. Since $|\hat{\nu}_i|$ is the dependent variable for the scale model, we apply the partialling out and recentering to this expression ($|\hat{\nu}_i|^{rc}$), and use that to estimate the model:

$$|\hat{\nu}_i|^{rc} = x_i^{rc'}\gamma + \omega_i$$

4. Finally the standardized residuals ε_i can be obtained as follows

$$\hat{\varepsilon}_i = \frac{\nu_i}{|\hat{\nu}_i| - \hat{\omega}_i}$$

where $|\hat{\nu}_i| - \hat{\omega}_i$ is the prediction for the conditional standard deviation $\sigma(x_i) = x_i'\gamma + \zeta_{g1} + \zeta_{g2}$

The τ_{th} quantile of the error ε can be estimated as usual, and the variance-covariance matrices obtained in the same way as before (see Section 2.2), but using x_i^{rc} instead of x_i when estimating the influence functions for all estimated coefficients.

3. Simulation Evidence

To show the performance of the extended strategy, we implement a small simulation study. We consider a simple model with a single explanatory variable x. In contrast with MSS, we consider a two-way fixed effect structure that is more general than the panel structure used in by the authors. For this excercise, we consider the following data generating process:

$$y_i = \alpha_{1i} + \alpha_{2i} + x_i + (1 + x_i + \alpha_{1i} + \alpha_{2i})\varepsilon_i$$

where $\alpha_{1i} \sim \chi^2(1)$, $\alpha_{2i} \sim \chi^2(1)$ and $x_i = 0.5 * (\chi_i + 0.5(\alpha_{1i} + \alpha_{2i}))$ with $\chi_i \sim \chi^2(1)$. We only consider the case when the error term ε_i is assumed to be normally distributed with mean zero and standard

[8] Using centered-residualized variables allow us to include a constant in the model specification, which simplifies the derivation of the influence functions. However, as with other fixed effects models, the constant is not identified, and thus should not be interpreted.

Table 1: Bias, Simulated Standard error, and Mean Squared Error

N = 500	Bias	Sim SE	MSE
q25			
mmqreg	0.104	0.247	0.072
jkc	0.008	0.297	0.088
q75			
mmqreg	-0.091	0.251	0.071
jkc	0.002	0.303	0.092

N = 1000	Bias	Sim SE	MSE
q25			
mmqreg	0.050	0.173	0.032
jkc	-0.001	0.191	0.036
q75			
mmqreg	-0.044	0.176	0.033
jkc	0.006	0.195	0.038

N = 1000	Bias	Sim SE	MSE
q25			
mmqreg	0.025	0.122	0.016
jkc	0.003	0.129	0.017
q75			
mmqreg	-0.023	0.124	0.016
jkc	0.001	0.131	0.017

N = 2000	Bias	Sim SE	MSE
q25			
mmqreg	0.012	0.087	0.008
jkc	0.000	0.089	0.008
q75			
mmqreg	-0.011	0.087	0.008
jkc	0.000	0.089	0.008

deviation 0.5. We assume that there are 50 mutually exclusive fixed effect groups for each set of fixed effects. Observations are assigned to each subgroup randomly, using a uniform distribution between 1 and 50. We consider a sample sizes of 500, 1000, 2000 and 4000 observations, which implies an average of 10, 20, 40 and 80 observations per group. The model is estimated using the location-scale model with heteroskedastic linear errors, and we report the coefficients for the 25th and 75th quantiles. We run this excercise 5000 times. Table 1 reports the bias, simulated standard error and mean squared error. We also report the results obtained using an adaptation bias-corrected estimator based on the split-panel jackknife estimator proposed by Dhaene and Jochmans (2015).[9] These results are labeled JKC.

Similar to the findings in MSS, we find that the estimator presents a bias when the sample is small ($N = 500$), but this bias shrinks as the sample sizes increases. As MSS describes, the bias seems to be proportional to the sample size, or more precisely to N/N_{g1} and N/N_{g2}, where N_{g1} and N_{g2} are the average number of observations per group. Interestingly, the bias-corrected estimator presents an almost zero bias, even when the samples are is small (about 10 observations per sub-group). Also replicating the results of MSS, despite bias reduction obtained using the JKC estimator, the standard errors and mean squared error (MSE) are larger than without correction.

As described in section Section 2.2, GLS-Standard errors may not be appropriate if heteroskedasticity

[9] For the implementation, we estimate the model using the full sample, then, randomly assign each observation into one of two groups, and re-estimate the quantile coefficients for each group. The bias-corrected estimator is then obtained as $\hat{\beta}(\tau)_{jkc} = 2 * \hat{\beta}(\tau)_{full} - 0.5 * (\hat{\beta}(\tau)_{s1} + \hat{\beta}(\tau)_{s1})$.

in the model is misspecified. This could happen if the heteroskedasticity is not linear in parameters, or if there is correlations within groups. To show the performance of the GLS-Standard errors, we consider two additional simulations. To see the impact of heteroskedasticity misspecification, we assume the DGP is given by:

$$y_i = \alpha_{1i} + \alpha_{2i} + x_i + (1 + x_i + \sqrt{x_i} + \alpha_{1i} + \alpha_{2i})\varepsilon_i$$

To show the impact of ignoring correlation within groups, we consider the case where the error ε is correlated within groups defined by the first fixed effect. Specifically:

$$\varepsilon_i = 0.5 * (\sqrt{0.5}z_i + \sqrt{0.5}w_g)$$

where z_i $N(0,1)$ and w_g $N(0,1)$, with g being the g_{th} group defined by the first fixed effect.

For the simulation, we run the excercise 5000 times, using a sample of size $N = 2000$. The results from both simulations are reported in the Table 6. We report the average coefficient, simulation based Standard errors, and the standard errors using GLS, robust and clustered estimations.

Table 6: Heteroskedasticity misspecification and clustered errors

	Het:q25	Het:q75	Clust:q25	Clust:q75
Coefficient	0.550	1.442	0.775	1.229
SIM_SE	0.154	0.155	0.109	0.107
GLS_SE	0.476	0.451	0.288	0.221
Robust	0.140	0.139	0.087	0.088
Cluster			0.097	0.097

One of the conclusions from the results in Table 6 is that the results for the 25th and 75th quantiles are mostly identical, because of the symmetric error we used in the simulation. On the one hand, because the error structure is misspecified, we should expect to see a large bias in the estimated coefficients, especially in cases of misspecified heteroskedasticity. However, the purpose of this exercise is not to demonstrate the unbiasedness of the coefficients, but rather to assess the performance of the standard error estimation.

As expected, the GLS-standard errors are biased, and the bias is larger in the case of misspecified heteroskedasticity. In fact, with heteroskedastic misspecification, the average GLS-SE (.476) is almost three times as large as the simulation-based standard error (0.154). In this first case, robust standard errors perform better, but with a small difference (0.14 to 0.154). In the case of clustered correlation, GLS-SE also performs poorly. Robust standard errors perform better, but seem to underestimate the magnitude of the true standard errors. Finally, cluster standard errors perform the best, with a small difference between the simulation-based standard errors (0.097 to 0.109).

4. Illustrative application

In this section we replicate one of the excercises from MSS, allowing for time and individual fixed effects, as well as for different standard errors estimations. We use data from Persson and Tabellini (2005), to estimate the relationship between surplus of goverment as share of GDP, and a measure of quality of democracy (POLITY); log of real income per capita (LYP); trade volume as share of GDP (TRADE), Share of population between 15-65 uears of age (P1564), the share of the population 65 years or older (P65); one-year lag of the dependent variable (LSP); oil prices in US dollars differntiating importer and exported countries (OILIM and OILEX); and the output gap (YGAP). In addition to country fixed effects (as illustrated in MSS), we also show results allowing for time fixed effects. Table 7 and Table 8 provide the results for the model with and without time fixed effects, respectively. It show cases the location and scale coefficients, as well as the quantile coefficients for the 25th, 50th and 75th quantiles. We also report GLS-Standard errors, Robust standard errors (brackets) and clustered standard errors at the country level.

Table 7: The determinants of government surpluses: Individual Fixed effects

	polityt	lyp	trade	p1564	p65	lspl	oil_im	oil_ex	ygap
Location									
coeff	0.116	-0.715	0.030	0.121	0.028	0.691	-0.047	-0.006	0.010
se_gls	0.046	0.540	0.008	0.033	0.070	0.035	0.008	0.022	0.028
se_r	0.047	0.597	0.008	0.031	0.070	0.037	0.007	0.017	0.021
se_cl	0.046	0.465	0.007	0.032	0.071	0.035	0.010	0.020	0.023
Scale									
coeff	-0.097	-0.616	0.003	0.036	0.087	-0.085	0.013	0.016	-0.004
se_gls	0.032	0.371	0.005	0.023	0.048	0.024	0.006	0.015	0.019
se_r	0.031	0.398	0.005	0.020	0.049	0.025	0.005	0.010	0.015
se_cl	0.048	0.800	0.008	0.031	0.067	0.029	0.004	0.010	0.012
Q25									
coeff	0.191	-0.239	0.028	0.093	-0.039	0.756	-0.057	-0.018	0.013
se_gls	0.059	0.684	0.010	0.042	0.088	0.045	0.010	0.027	0.035
se_r	0.056	0.656	0.008	0.036	0.086	0.040	0.010	0.020	0.025
se_cl	0.073	0.687	0.006	0.041	0.098	0.023	0.010	0.021	0.029
Q50									
coeff	0.108	-0.765	0.030	0.124	0.035	0.684	-0.046	-0.005	0.009
se_gls	0.046	0.535	0.007	0.033	0.069	0.035	0.008	0.022	0.027
se_r	0.046	0.593	0.008	0.031	0.069	0.036	0.007	0.017	0.021

	polityt	lyp	trade	p1564	p65	lspl	oil_im	oil_ex	ygap
se_cl	0.043	0.484	0.008	0.032	0.070	0.036	0.010	0.020	0.023
Q75									
coeff	0.031	-1.258	0.033	0.153	0.104	0.616	-0.036	0.008	0.006
se_gls	0.048	0.551	0.008	0.034	0.071	0.036	0.008	0.022	0.028
se_r	0.049	0.696	0.009	0.034	0.075	0.043	0.007	0.018	0.023
se_cl	0.039	0.919	0.012	0.041	0.079	0.055	0.010	0.022	0.020

As expected, Table 7 shows that point estimates for the point estimates are identical to the ones reported in Machado and Santos Silva (2019) (Table 6), including analytical standard errors (GLS). With our estimator, however, we are able to also produce both robust and clustered standard errors for location and scale coefficients. Except for a few cases, robust and clustered standard errors are larger than GLS standard errors, which may be an indication of misspecification of the model. The GLS standard errors we report differ from the ones in MSS because they use panel standard errors, which are equivalent to our clustered standard errors, instead of the analytical standard errors we derive.

Considering the estimated effects across quantiles, we observe few differences in the reported GLS standard errors compared to the analytical standard errors reported MSS. Our clustered standard errors, however, are closer to the bootstrap based standard errors the authors report.[10]

In Table 8, we report the results including both individual and year fixed effects. Because oil prices only vary across years, the variable is excluded from the model specification. Accounting for time fixed effects does not change the general conclusions one could make based on the results from Table 7. The two largest differences are that the log of income per capita has a positive effect on Government Surpluses, but only for the 25th quantile, because of the largest impact on the Scale component. Similarly, we observe that the income gap now has an impact on Government surplus that is always negative, but increasing across quantiles. In both instances, the effects are not statistically significant.

Table 8: The determinants of government surpluses: Individual and Time Fixed effects

	polity	lyp	trade	prop1564	prop65	lspl	ygap
Location							
coeff	0.126	-0.418	0.028	0.108	0.042	0.693	-0.014
se_gls	0.087	1.157	0.015	0.072	0.136	0.066	0.053
se_r	0.047	0.703	0.008	0.038	0.068	0.038	0.022
se_cl	0.048	0.506	0.008	0.044	0.077	0.037	0.022

[10]There are two possible reasons that may explain the differences in the GLS standard errors. On the one hand, in our derivation, the influence function of the standardized τ_{th} quantile (see #eq-infs) does not have the same leading term as the one reported in MSS (see theorem 3, and the definition of W).

	polity	lyp	trade	prop1564	prop65	lspl	ygap
Scale							
coeff	-0.095	-1.255	0.005	0.033	0.040	-0.081	0.008
se_gls	0.081	1.073	0.014	0.067	0.126	0.061	0.049
se_r	0.031	0.452	0.005	0.025	0.045	0.025	0.017
se_cl	0.041	0.848	0.006	0.030	0.048	0.033	0.013
Q25							
coeff	0.201	0.576	0.025	0.082	0.010	0.757	-0.020
se_gls	0.154	2.070	0.024	0.118	0.219	0.121	0.085
se_r	0.058	0.751	0.008	0.049	0.080	0.040	0.026
se_cl	0.073	0.761	0.006	0.052	0.087	0.023	0.027
Q50							
coeff	0.119	-0.512	0.029	0.111	0.045	0.687	-0.013
se_gls	0.091	1.230	0.014	0.070	0.130	0.072	0.051
se_r	0.046	0.695	0.008	0.037	0.068	0.038	0.022
se_cl	0.045	0.529	0.008	0.044	0.077	0.039	0.021
Q75							
coeff	0.041	-1.555	0.033	0.138	0.078	0.619	-0.007
se_gls	0.067	0.898	0.011	0.053	0.098	0.052	0.038
se_r	0.048	0.827	0.009	0.037	0.075	0.046	0.026
se_cl	0.038	0.980	0.012	0.050	0.086	0.063	0.020

5. Conclusions

6. Appendix

6.1. Derivation of the influence functions

6.1.1. Model Identification

The estimation of quantile regression via moments assumes that the DGP is linear in parameters, with an heteroskedastic error term that is also linear function of parameters:

$$y_i = x_i'\beta + \nu_i$$
$$\nu_i = \varepsilon_i \times x_i'\gamma$$

where ε is an unobserved i.i.d. random variable that is independent of x, and such that $x\gamma$ is larger than zero for any x.

In this case, the τ_{th} conditional quantile model can be written as:

$$q(y|\tau,x) = x'(\beta + q(\varepsilon|\tau) \times \gamma)$$

This model is identified under the following conditions:

$$\begin{aligned}
E[(y_i - x_i'\beta)x_i] &= E[h_{1,i}] = 0 \\
E[(|y_i - x_i'\beta| - x_i'\gamma)x_i] &= E[h_{2,i}] = 0 \\
E\left[\mathbb{1}\left(q(\varepsilon|\tau)x_i'\gamma + x_i'\beta \geq y_i\right) - \tau\right] &= E[h_{3,i}] = 0
\end{aligned}$$

For simplicity, for the rest of the appendix, I will use q_τ^ε to represent $q(\varepsilon|\tau)$.

6.1.2. Estimation of the variance-covariance matrix

In this model, to estimate the variance-covariance matrix the set of coefficients $\theta' = [\beta' \ \gamma' \ q_\tau^\varepsilon]$, we need to obtain the influence functions of all coefficients, which are defined as:

$$\lambda_i = \bar{G}(\theta)^{-1} \begin{bmatrix} h_{1,i} \\ h_{2,i} \\ h_{3,i} \end{bmatrix}$$

where the Jacobian matrix $\bar{G}(\theta)$ is defined as:

$$\bar{G}(\theta) = \begin{bmatrix} \bar{G}_{11} & G_{12} & G_{13} \\ \bar{G}_{21} & \bar{G}_{22} & G_{23} \\ \bar{G}_{31} & \bar{G}_{32} & \bar{G}_{13} \end{bmatrix}$$

with

$$\bar{G}_{j,k} = -\frac{1}{N}\sum_{i=1}^{N} \frac{\partial h_{j,i}}{\partial \theta_k'} \ \ \forall j,k \in 1,2,3$$

6.1.2.1. First Moment Condition: Location Model.

$$h_{1,i} = x_i(y_i - x_i'\beta)$$

$$\begin{aligned}
\bar{G}_{1,1} &= -\frac{1}{N}\sum_{i=1}^{N} \frac{\partial h_{1,i}}{\partial \beta'} \\
&= -\frac{1}{N}\sum_{i=1}^{N}(-x_i x_i') \\
&= N^{-1}X'X
\end{aligned}$$

$$\bar{G}_{1,2} = \bar{G}_{1,3} = 0$$

6.1.2.2. Second Moment Condition: Scale model.

$$h_{2,i} = x_i(|y_i - x_i'\beta| - x_i'\gamma)$$

$$\begin{aligned}
\bar{G}_{2,1} &= -\frac{1}{N}\sum \frac{\partial h_{2,i}}{\partial \beta'} \\
&= \frac{1}{N}\sum x_i x_i' \frac{y_i - x_i'\beta}{|y_i - x_i'\beta|} \\
\frac{y_i - x_i'\beta}{|y_i - x_i'\beta|} &= sign(y_i - x_i'\beta)
\end{aligned}$$

Under the assumption $\varepsilon_i \times x\gamma$, or in this case $y_i - x_i'\beta$, is uncorrelated with x, we can simplify the expression as:

$$\begin{aligned}
\bar{G}_{2,1} &= N^{-1}\left(N^{-1}\sum sign(y_i - x_i'\beta)\right)\sum x_i x_i' \\
&= N^{-1}E[sign(y_i - x_i'\beta)]X'X
\end{aligned}$$

$$\begin{aligned}
\bar{G}_{2,2} &= -\frac{1}{N}\sum \frac{\partial h_{2,i}}{\partial \gamma'} \\
&= \frac{1}{N}\sum x_i x_i' \qquad = N^{-1}X'X
\end{aligned}$$

$$\bar{G}_{2,3} = 0$$

6.1.2.3. Third Moment Condition: Quantile of Standardized Residual.

$$\begin{aligned}
h_{3,i} &= \mathbb{1}\left(q_\tau^\varepsilon x_i'\gamma + x_i'\beta - y_i \geq 0\right) - \tau \text{ or} \\
h_{3,i} &= \mathbb{1}\left(q_\tau^\varepsilon \geq \frac{y_i - x_i'\beta}{x_i'\gamma}\right) - \tau = \mathbb{1}(q_\tau^\varepsilon \geq \varepsilon) - \tau
\end{aligned}$$

Because the indicator function $\mathbb{1}()$ is not differentiable, we borrow from the non-parametric literature, and approximate the indicator function with the integral of a kernel density function $I()$. This function $I()$, is monotonic and symetrical function around zero, with a domain over the real numbers, and a range between 0 and 1.

With an arbirarily small bandwidth h, this function will approximate the indicator function:

$$\lim_{h\to 0} I\left(\frac{z}{h}\right) \approx \mathbb{1}(z \geq 0)$$

Thus the function $h_{3,i}$ can be approximated as:

$$h_{3,i} \approx I\left(q_\tau^\varepsilon x_i'\gamma + x_i'\beta - y_i\right) - \tau$$

Now, we can obtain the Jacobian matrix $\bar{G}_{3,1}$ as:

$$\begin{aligned}\bar{G}_{3,1} &= -\frac{1}{N}\sum \frac{\partial h_{3,i}}{\partial \beta'} \\ &= -N^{-1}\sum K_h(q_\tau^\varepsilon x_i'\gamma + x_i'\beta - y_i)x_i'\end{aligned}$$

Under the assumption that we have enough observations within each combination of x, and of multiplicative heteroskedasticity, we have:

$$\begin{aligned}E(K_h(q_\tau^\varepsilon x_i'\gamma + x_i'\beta - y_i)|X) &= f_{y|X}(q_\tau^\varepsilon x_i'\gamma + x_i'\beta) \\ &= \frac{1}{x_i'\gamma} f_\varepsilon(q_\tau^\varepsilon)\end{aligned}$$

where $f_{y|X}$ is the conditional probability density function of y given X, and f_ε is the unconditional distribution of the standardized error. With this, we can rewrite the Jacobian matrix as:

$$\bar{G}_{3,1} = -N^{-1} f_\varepsilon(q_\tau^\varepsilon)\sum \frac{x_i'}{x_i'\gamma}$$

Asymptotically, however, the expression $\sum \frac{a_i}{b_i}$ can be approximated using taylor expansions by $N\frac{\bar{a}}{\bar{b}}$. [11] Thus, we can rewrite the last term as:

$$\bar{G}_{3,1} = -f_\varepsilon(q_\tau^\varepsilon)\frac{\bar{x}_i'}{\bar{x}_i'\gamma}$$

The Jacobian for the second matrix $\bar{G}_{3,2}$ can be derived similarly:

$$\begin{aligned}\bar{G}_{3,2} &= -\frac{1}{N}\sum \frac{\partial h_{3,i}}{\partial \gamma'} \\ &= -N^{-1}\sum K_h(q_\tau^\varepsilon x_i'\gamma + x_i'\beta - y_i)q_\tau^\varepsilon x_i' \\ &= -N^{-1}\sum f_{y|x}(q_\tau^\varepsilon x_i'\gamma + x_i'\beta)q_\tau^\varepsilon x_i' \\ &= -N^{-1} f_\varepsilon(q_\tau^\varepsilon)q_\tau^\varepsilon \sum \frac{x_i'}{x_i'\gamma} \\ &= -f_\varepsilon(q_\tau^\varepsilon)q_\tau^\varepsilon \frac{\bar{x}_i'}{\bar{x}_i'\gamma}\end{aligned}$$

and the Jacobian for the third matrix $\bar{G}_{3,3}$ is:

[11] This approximation will be useful when we consider the estimation of the influence functions.

$$
\begin{aligned}
\bar{G}_{3,3} &= -\frac{1}{N}\sum \frac{\partial h_{3,i}}{\partial q^{\varepsilon}_{\tau}} \\
&= -\frac{1}{N}\sum K_h(q^{\varepsilon}_{\tau}x_i'\gamma + x_i'\beta - y_i)x_i'\gamma \\
&= -\frac{1}{N}\sum f_{y|X}(q^{\varepsilon}_{\tau}x_i'\gamma + x_i'\beta)x_i'\gamma \\
&= -\frac{1}{N}\sum f_{\varepsilon}(q^{\varepsilon}_{\tau})\frac{x_i'\gamma}{x_i'\gamma} \\
&= -f_{\varepsilon}(q^{\varepsilon}_{\tau})
\end{aligned}
$$

6.1.3. Influence functions

6.1.3.1. Location coefficients.

$$\lambda_i(\beta) = \bar{G}^{-1}_{1,1}h_{1,i} = N(X'X)^{-1}(x_i(y_i - x_i'\beta)) = N(X'X)^{-1}(x_i\nu_i)$$

Which can also be written as a function of the standardized residuals:

$$\lambda_i(\beta) = \bar{G}^{-1}_{1,1}h_{1,i} = N(X'X)^{-1}(x_i(y_i - x_i'\beta)) = N(X'X)^{-1}(x_i(x_i'\gamma \times \varepsilon))$$

6.1.3.2. Scale Coefficients.

$$
\begin{aligned}
\lambda_i(\gamma) &= \bar{G}^{-1}_{2,2}\Big(h_{2,i} - \bar{G}_{2,1}\lambda_i(\beta)\Big) \\
&= N(X'X)^{-1}\Big(x_i(|\nu_i| - x_i'\gamma) - N^{-1}E[sign(\nu_i)]X'X[N(X'X)^{-1}(x_i\nu_i)]\Big) \\
&= N(X'X)^{-1}\Big(x_i(|\nu_i| - x_i'\gamma) - E[sign(\nu_i)](x_i\nu_i)\Big) \\
&= N(X'X)^{-1}x_i\Big(|\nu_i| - E[sign(\nu_i)]\nu_i - x_i'\gamma\Big)
\end{aligned}
$$

However,

$$
\begin{aligned}
|\nu_i| &= \nu_i \times \mathbb{1}(\nu_i \geq 0) - \nu_i \times \mathbb{1}(\nu_i < 0) \\
|\nu_i| &= \nu_i \times \mathbb{1}(\nu_i \geq 0) - \nu_i \times [1 - \mathbb{1}(\nu_i \geq 0)] \\
|\nu_i| &= 2\nu_i \times \mathbb{1}(\nu_i \geq 0) - \nu_i
\end{aligned}
$$

And

$$
\begin{aligned}
E[sign(\nu_i)] &= E[\mathbb{1}(\nu_i \geq 0)] - E[\mathbb{1}(\nu_i < 0)] \\
E[sign(\nu_i)] &= E[\mathbb{1}(\nu_i \geq 0)] - E[(1 - \mathbb{1}(\nu_i \geq 0))] \\
E[sign(\nu_i)] &= 2E[\mathbb{1}(\nu_i \geq 0)] - 1
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\lambda_i(\gamma) &= N(X'X)^{-1}x_i\Big(2\nu_i \times \mathbb{1}(\nu_i \geq 0) - \nu_i - (2E[\mathbb{1}(\nu_i \geq 0)] - 1)\nu_i - x_i'\gamma\Big) \\
&= N(X'X)^{-1}x_i\Big(2\nu_i \times \mathbb{1}(\nu_i \geq 0) - 2E[\mathbb{1}(\nu_i \geq 0)]\nu_i - x_i'\gamma\Big) \\
&= N(X'X)^{-1}x_i\Big(2\nu_i \times [\mathbb{1}(\nu_i \geq 0) - E[\mathbb{1}(\nu_i \geq 0)]] - x_i'\gamma\Big) \\
&= N(X'X)^{-1}x_i\Big(\tilde{\nu}_i - x_i'\gamma\Big)
\end{aligned}
$$

This last expression is the equivalent simplification used in Machado and Santos Silva (2019) and Im (2000). If the scale function is strictly possitive, it also follows that $\mathbb{1}(\nu_i \geq 0) = \mathbb{1}(\varepsilon_i \geq 0)$. Thus, it can be simplified as:

$$
\lambda_i(\gamma) = N(X'X)^{-1}x_i(x_i'\gamma) \times (\tilde{\varepsilon}_i - 1)
$$

6.1.3.3. Quantile of standardized residual.

$$
\begin{aligned}
\lambda_i(q_\tau^\varepsilon) &= \bar{G}_{3,3}^{-1}\Big(h_{3,i} - \bar{G}_{3,1}\lambda_i(\beta) - \bar{G}_{3,2}\lambda_i(\gamma)\Big) \\
&= -\frac{1}{f_\varepsilon(q_\tau^\varepsilon)} \times \Bigg(\Big(\mathbb{1}(q_\tau^\varepsilon \geq \varepsilon) - \tau\Big) \\
&\quad + f_\varepsilon(q_\tau^\varepsilon)\frac{\bar{x}_i'}{\bar{x}_i'\gamma}N(X'X)^{-1}x_i(x_i'\gamma \times \varepsilon) \\
&\quad + f_\varepsilon(q_\tau^\varepsilon)q_\tau^\varepsilon\frac{\bar{x}_i'}{\bar{x}_i'\gamma}N(X'X)^{-1}x_i(\tilde{\nu}_i - x_i'\gamma)\Bigg) \\
&= \frac{\tau - \mathbb{1}(q_\tau^\varepsilon \geq \varepsilon)}{f_\varepsilon(q_\tau^\varepsilon)} - \frac{x_i'\gamma \times \varepsilon_i}{\bar{x}_i'\gamma} - q_\tau^\varepsilon\frac{\tilde{\nu}_i - x_i'\gamma}{\bar{x}_i'\gamma}
\end{aligned}
$$

6.2. Implementation

The method described here can be implemented using any of the following packages

- `mmqreg` in Stata: Installed from Github
- `mmqreg` in R: Programs Available in Github
- `mmqreg` in Python:
- `mmqreg` in julia:

References

Abowd, J.M., Kramarz, F., Roux, S., 2006. Wages, mobility and firm performance: Advantages and insights from using matched worker–firm data*. The Economic Journal 116, F245–F285. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.2006.01099.x, doi:https://doi.org/10.1111/j.1468-0297.2006.01099.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0297.2006.01099.x.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust Inference With Multiway Clustering. Journal of Business & Economic Statistics 29, 238–249. URL: http://www.tandfonline.com/doi/abs/10.1198/jbes.2010.07136, doi:10.1198/jbes.2010.07136.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: methods and applications. Cambridge University Press, Cambridge ; New York.

Canay, I.A., 2011. A simple approach to quantile regression for panel data. The Econometrics Journal 14, 368–386. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1368-423X.2011.00349.x, doi:https://doi.org/10.1111/j.1368-423X.2011.00349.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1368-423X.2011.00349.x.

Correia, S., 2016. A Feasible Estimator for Linear Models with Multi-Way Fixed Effects. Unpublished Manuscript URL: http://scorreia.com/research/hdfe.pdf.

Dhaene, G., Jochmans, K., 2015. Split-panel Jackknife Estimation of Fixed-effect Models. The Review of Economic Studies 82, 991–1030. URL: https://doi.org/10.1093/restud/rdv007, doi:10.1093/restud/rdv007, arXiv:https://academic.oup.com/restud/article-pdf/82/3/991/5371725/rdv007.pdf.

Frisch, R., Waugh, F.V., 1933. Partial time regressions as compared with individual trends. Econometrica 1, 387–401. URL: http://www.jstor.org/stable/1907330.

Frumento, P., Bottai, M., 2016. Parametric modeling of quantile regression coefficient functions. Biometrics 72, 74–84. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12410, doi:https://doi.org/10.1111/biom.12410, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.12410.

Galvao, A.F., Kengo, K., 2017. Quantile regression methods for longitudinal data, in: Handbook of quantile regression. Chapman and Hall/CRC, pp. 363–380.

Gaure, S., 2013. OLS with multiple high dimensional category variables. Computational Statistics & Data Analysis 66, 8–18. URL: https://www.sciencedirect.com/science/article/pii/S0167947313001266, doi:https://doi.org/10.1016/j.csda.2013.03.024.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 2005. Robust Statistics: The Approach Based on Influence Functions. Wiley Series in Probability and Statistics. 1 ed., Wiley. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118186435, doi:10.1002/9781118186435.

Harris, D.N., Sass, T.R., 2011. Teacher training, teacher quality and student achievement. Journal of Public Economics 95, 798–812. URL: https://www.sciencedirect.com/science/article/pii/S0047272710001696, doi:https://doi.org/10.1016/j.jpubeco.2010.11.009.

He, X., 1997. Quantile curves without crossing. The American Statistician 51, 186–192. URL: https://www.tandfonline.com/doi/abs/10.1080/00031305.1997.10473959, doi:10.1080/00031305.1997.10473959, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/00031305.1997.10473959.

Im, K.S., 2000. Robustifying glejser test of heteroskedasticity. Journal of Econometrics 97, 179–188. URL: https://www.sciencedirect.com/science/article/pii/S0304407699000615, doi:https://doi.org/10.1016/S0304-4076(99)00061-5.

Jann, B., 2020. Influence functions continued. A framework for estimating standard errors in reweighting, matching, and regression adjustment. University of Bern Social Sciences Working Papers 35. University of Bern, Department of Social Sciences. URL: https://ideas.repec.org/p/bss/wpaper/35.html.

Kaplan, D.M., Sun, Y., 2017. Smoothed estimating equations for instrumental variables quantile regression. Econometric Theory 33, 105–157. URL: https://www.cambridge.org/core/journals/econometric-theory/article/abs/smoothed-estimating-equations-for-instrumental-variables-quantile-regression/59EE6581C077098DDBB87F3E98F9BF90#, doi:10.1017/S0266466615000407. publisher: Cambridge University Press.

Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89. URL: https://www.sciencedirect.com/science/article/pii/S0047259X04001113, doi:https://doi.org/10.1016/j.jmva.2004.05.006. special Issue on Semiparametric and Nonparametric Mixed Models.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50. URL: http://www.jstor.org/stable/1913643.

Lancaster, T., 2000. The incidental parameter problem since 1948. Journal of Econometrics 95, 391–413. URL: https://www.sciencedirect.com/science/article/pii/S0304407699000445, doi:https://doi.org/10.1016/S0304-4076(99)00044-5.

Lovell, M.C., 1963. Seasonal adjustment of economic time series and multiple regression analysis. Journal of the American Statistical Association 58, 993–1010. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10480682, doi:10.1080/01621459.1963.10480682, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1963.10480682.

Machado, J.A., Santos Silva, J., 2019. Quantiles via moments. Journal of Econometrics 213, 145–173. URL: https://www.sciencedirect.com/science/article/pii/S0304407619300648, doi:10.1016/j.jeconom.2019.04.009. annals: In Honor of Roger Koenker.

Melly, B., 2005. Decomposition of differences in distribution using quantile regression. Labour Economics 12, 577–590. URL: https://www.sciencedirect.com/science/article/pii/S0927537105000382, doi:https://doi.org/10.1016/j.labeco.2005.05.006.

Newey, W.K., McFadden, D., 1994. Chapter 36 Large sample estimation and hypothesis testing, in: Handbook of Econometrics. Elsevier. volume 4, pp. 2111–2245. URL: https://linkinghub.elsevier.com/retrieve/pii/S1573441205800054, doi:10.1016/S1573-4412(05)80005-4.

Neyman, J., Scott, E.L., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32. URL: http://www.jstor.org/stable/1914288.

Persson, T., Tabellini, G.E., 2005. The economic effects of constitutions. Munich lectures in economics. 1. mit press paperback ed ed., MIT Press, Cambridge, Mass.

Rios-Avila, F., 2015. Feasible Fitting of Linear Models with N Fixed Effects. The Stata Journal 15, 881–898. URL: https://doi.org/10.1177/1536867X1501500318, doi:10.1177/1536867X1501500318. publisher: SAGE Publications.

Zhao, Q., 2000. Restricted Regression Quantiles. Journal of Multivariate Analysis 72, 78–99. URL: https://www.sciencedirect.com/science/article/pii/S0047259X99918493, doi:https://doi.org/10.1006/jmva.1999.1849.