Estimating Quantile Regressions with Multiple Fixed Effects through Method of Moments*

Fernando Rios-Avila Leonardo Siles Gustavo Canavire-Bacarreza

2024-02-08

Abstract

This paper proposes a new method to estimate quantile regressions with multiple fixed effects. The method, which expands on the strategy proposed by Machado and Santos Silva [2019], allows for the inclusion of multiple fixed effects and provides 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.

*The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the World Bank, its Board of Directors, or the countries it represents.

The authors would like to thank Enrique Pinzon an Joao Santos-Silva for their helpful comments and suggestions.

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 $Q_y(\tau|x)$. Using QR one can obtain richer characterizations of the relationships between dependent and independent variables, by exploring how the variables relate along the entire conditional distribution.

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 an incidental parameter problem is generated. 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 regard to the inclusion of individual and multiple fixed 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 to have heterogeneous effects on the entire conditional distribution of the outcome, rather than constraining their effect to be a location shift only, as in Canay [2011], Koenker [2004], and Lamarche [2010].

In principle, under the assumption that data-generating process behind the data is based on a multiplicative heteroskedastic process that is linear in parameters [Machado and Santos Silva, 2019, He [1997], Zhao [2000], Cameron and Trivedi [2005]], 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.

This methodology is not meant to substitute for the use of standard quantile regression analysis. That said, 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] in two ways. First, making use of the properties of generalized method of moments (GMM) estimators, we derive various alternatives for the estimation of standard errors based on the empirical influence functions of the estimators. Even if the model is correctly specified, robust standard errors perform better than GLS standard errors due small violations of the model assumptions due to sampling variability. Furthermore, clustered standard errors may help to further account for typically unobserved correlations across observations. Second, we reconsider the application of Frisch-Waugh-Lovell (FWL) theorem [Frisch and Waugh, 1933, and Lovell [1963]] to extend the MSS estimator and allow for the inclusion of multiple fixed effects. This extension may be useful for empirical analysis, as it is common to control for multiple fixed effects such as individual and time fixed effects.

The rest of the paper is structured 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-scale model. It also revisits the methodology of MSS, proposing alternative estimators for the standard errors based on the properties of GMM estimators and the empirical influence functions. It also shows that the 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 one empirical example. Section 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:

$$Q_y(\tau|X) = X\beta(\tau) \tag{1}$$

This allows for a linear effect of X on Y, but that could vary across values of τ.

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:

$$Q_y(\tau|X) = X\beta + X\gamma(\tau) \tag{2}$$

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 He [1997] and Zhao [2000]) 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], Zhao [2000], 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]

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

[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 while 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.

Under the assumption that ε is an i.i.d. unobserved random variable that is independent of X, the conditional quantile of Y given X and τ can be written as

$$Q_y(\tau|X) = X\beta + Q_\varepsilon(\tau) \times X\gamma \tag{5}$$

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. For simplicity we will use q_τ to denote $Q_\varepsilon(\tau)$ in the rest of the paper.

$$\beta(\tau) = \beta + q_\tau \times \gamma \tag{6}$$

While this specification imposes a strong assumption on the DGP, it has two advantages over the standard quantile regression model. First, because the location-scale model can be identified globally, with only a single parameter (q_τ) requiring local estimation, this estimation approach will be more efficient than the standard quantile regression model [Zhao, 2000]. Second, under the assumption that $X\gamma$ is strictly positive, the model will produce quantile coefficients that do not cross [He, 1997].

Following MSS, the quantile regression model defined by Equation 5 can be estimated using a generalized method of moments approach. And while it is possible to identify all coefficients (β, γ, q_τ) simultaneously, we describe and use the implementation approach advocated by MSS, which identifies each set of coefficients separately.

First, 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:

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

Second, 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 allows 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:

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

It should be noticed that the estimation of the standard errors (next section) requires that the Scale component prediction $x_i'\gamma$ is strictly positive, because it represents the conditional standard deviation of the error ν_i. Because this component is identified using a linear model, some values for $x_i'\gamma$ may be negative, which will affect the estimation of the standard errors, as shown in the simulation study.

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

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

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 $\beta(\tau) = \beta + \gamma q_\tau$. Afterwards, the conditional quantile coefficients are 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 heteroskedastic linear errors can be estimated using a the following set of moments, which fits within the GMM framework:

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

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 $\theta = [\beta' \quad \gamma' \quad q_\tau]'$ the set of coefficients that are identified by the moment conditions in Equation 10, a just identified model, and the function h_i is a vector function that stacks all the moments described in Equation 10 at the individual level. 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 heteroskedasticity-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 $\beta(\tau) = \beta + \gamma q_\tau$ and variance-covariance matrix equal to

[3] Zhao [2000] also shows that the quantile coefficients for the location-scale model follow 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

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

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. Afterward, 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 moments, the influence functions for the location, scale, and quantile coefficients are[4]

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

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 the appendix.

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 heteroskedasticity-consistent estimator for least-squares estimators. Considering the location-scale model, the variance-covariance matrix for the quantile coefficients can be estimated as

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

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, generalized least squares (GLS) standard errors could be severely biased. The standard recommendation has been to report block-bootstrap standard errors, clustered at the individual level.

Because we have access to the influence functions, it is straightforward to estimate one-way clustered standard errors.

Let N_G be the total number of clusters g, where $g = 1 \ldots N_G$. The clustered variance covariance matrix is given by[5]

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

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 GLS, which will be valid as long as the model for heteroskedasticity is correctly specified.[6] To estimate the GLS standard errors, we make use of the following property:

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

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

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

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

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

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, we define the following modified influence functions:

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

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

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

where

$$
\begin{aligned}
\hat{\Omega}_{jk} &= \frac{1}{N}\sum_{i=1}^{N}\tilde{\lambda}_{i,j}(\theta)\tilde{\lambda}_{i,k}(\theta)' \\
\hat{\sigma}_{jk} &= \frac{1}{N}\sum_{i=1}^{N}\psi_{i,j}\psi_{i,k}
\end{aligned}
$$

This estimator of standard errors is equivalent to the one derived by MSS using Theorem 3. Empirically, the simulation study shows that this estimator may be very sensitive to predictions of the scale model. In the simulation study, we show that when the sample small, and the likelihood of predicting zero or negative values for the scale model is high, the GLS standard errors may be unreasonably large.

2.3 Multiple Fixed Effects: Expanding on Machado and Santos Silva [2019]

Using the setup described in the previous section, MSS proposes an extension to the model proposed by He [1997] that enables the estimation of quantile regression models with panel data, allowing for the inclusion of individual fixed effects. However, the methodology can also be generalized to allow for the inclusion of multiple fixed effects. This type of analysis can 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.

We return to the original model and now assume there are sets of unobserved heterogeneity that are constant across observations, if they belong to common groups. Without loss of generality, we can assume that the data-generating process is as follows:

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

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

If the dimension of groups g_k is low, this model could be estimated using a dummy inclusion approach following Section 2.1, and the 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 computationally feasible. A more feasible approach is to apply the FWL theorem and partial out the impact of the group fixed effects on the control variables x_i, and the outcome of interest y_i, and using 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 for example, 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.

First, 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:

$$
\begin{aligned}
w_i &= \delta^w_{g1} + \delta^w_{g2} + u^w_i \\
w^{rc}_i &= E(w_i) + \hat{u}^w_i
\end{aligned}
$$

Afterward, we estimate the location model using the centered-residualized variables:[8]

$$
y^{rc}_i = x^{rc'}_i \beta + \nu_i
$$

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

[8] Using centered-residualized variables allows 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.

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

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

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) by 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. For this exercise, we consider the following data-generating process:

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

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 follow a centered χ^2 distribution.[9] We assume that there are 50 mutually exclusive groups for each set of fixed effects. Observations are assigned to each subgroup randomly using a uniform distribution between 1 and 50.

To assess the alternative Standard error estimators, we consider a second data-generating process where the error term is correlated within clusters. For this, we assume observations are assigned ramdomly to 100 mutually exclusive groups, which are independent from the fixed effect groups. In this setup the data-generating process is:

$$\begin{aligned}
y_i &= \alpha_{1i} + \alpha_{2i} + x_i + (2 + x_i + \alpha_{1i} + \alpha_{2i}) * \kappa_i \\
\kappa_i &= inv - \chi_5^2(r_i)/5 - 1 \\
r_i &= \Phi((\sqrt{.25} * s_i + \sqrt{.75} * s_g))
\end{aligned}$$

where $s_i \sim N(0,1)$ and $s_g \sim N(0,1)$, are errors that vary across individuals i or across clusters g. $\Phi()$ is the cumulative distribution function of a normal distribution, and $inv - \chi_5^2()$ is the inverse function for a Chi-2 distribution with 5 degrees of freedom. With this setup, we generate an error structure with a strong intra-cluster correlation, but without affecting the model specification assumption.[10]

[9] Specifically we assume $\varepsilon = \frac{r_i}{5} - 1$, where r_i follows a chi-squared distribution such that $r_i \sim \chi^2(5)$. Simulations under the assumption of normal errors are available upon request.

[10] We do not consider the case of misspecification of the scale model because that change would not only have an impact on the standard errors, but also affect the bias of the estimated coefficients

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

	N = 500			N = 1000		
	Mean Bias	SE	MSE	Mean Bias	SE	MSE
q25						
mmqreg	0.169	0.267	0.099	0.092	0.172	0.038
jkc	0.048	0.318	0.104	0.014	0.189	0.036
q75						
mmqreg	-0.050	0.446	0.202	-0.010	0.310	0.096
jkc	0.048	0.546	0.301	0.018	0.339	0.115
	N = 2000			N = 4000		
	Mean Bias	SE	MSE	Mean Bias	SE	MSE
q25						
mmqreg	0.050	0.119	0.017	0.026	0.084	0.008
jkc	0.006	0.126	0.016	0.003	0.087	0.008
q75						
mmqreg	0.001	0.215	0.046	0.003	0.151	0.023
jkc	0.006	0.222	0.049	0.002	0.154	0.024

Note: mmqreg - The proposed estimator. JKC-Jacknife Bias Corrected Estimator. SE- Simulated Standard Error. MSE - Mean Squared Error. Mean bias is the difference between the estimated coefficient and the analytical true value.

We consider 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, using the first data generation structure only. 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].[11] These results are labeled JKC.

Similar to the findings in MSS, we find that while the estimator presents a substantial bias when the sample is small ($N = 500$), this bias shrinks as the sample size increases. As MSS describes, the bias seems to be proportional to the sample size, or more precisely to the average number of observations per sub-group. Interestingly, the bias-corrected estimator presents an almost 0 bias for the 25th percentile, even when the samples are small. In contrast, when considering the 75th percentile, the simple estimator shows smaller bias than the Jacknife estimator.[^88: This is also consistent with simulations with a single fixed effect] In either case, despite the bias reduction obtained using the JKC estimator, the standard errors are larger than without correction. For the 25th quantile, the reduction in bias is large enough to produce a smaller Mean Squared Error (MSE) than the simple estimator. This is similar to the results of MSS.

To evaluate the performance of the different standard errors, we present bias, 95% coverage of the biased corrected estimates, as well as the simulated standard errors, average and median of the standard errors obtained using the GLS, robust SE, and clustered standard SE. Table 2 considers the d.g.p. without intra-cluster correlation, while Table 3 considers the d.g.p. that induces intra-

[11] For the implementation, we first estimate the model using the full sample, then randomly assign each observation into one of two groups, and finally reestimate the quantile coefficients for each group. The bias-corrected estimator is then obtained as $\hat{\beta}(\tau)_{jkc} = 2 * \hat{\beta}(\tau)_{full} - 0.5 * (\hat{\beta}(\tau)_{s1} + \hat{\beta}(\tau)_{s2})$.

Table 2: 95% Coverage Ratio, and Standard Error Estimation: No Intra-cluster Correlation

	Q25				Q75			
N	=500	=1000	=2000	=4000	=500	=1000	=2000	=4000
Bias	0.173	0.096	0.050	0.026	-0.040	-0.008	0.003	0.003
Sim SE	0.268	0.171	0.118	0.084	0.443	0.314	0.219	0.153
Mean GLS SE	1.9e7	0.830	0.401	0.084	7.3e7	1.876	4.259	0.156
Median GLS SE	0.495	0.215	0.123	0.083	1.037	0.429	0.225	0.152
CR95%	0.988	0.980	0.958	0.948	0.991	0.977	0.967	0.952
Mean Rbst SE	0.224	0.159	0.112	0.080	0.353	0.269	0.199	0.144
CR95%	0.892	0.928	0.939	0.932	0.875	0.904	0.927	0.936

Table 3: 95% Coverage Ratio, and Standard Error Estimation: With Intra-cluster Correlation

	Q25				Q75			
N	=500	=1000	=2000	=4000	=500	=1000	=2000	=4000
Bias	0.180	0.091	0.053	0.027	-0.022	-0.008	0.003	0.003
Sim SE	0.299	0.200	0.141	0.100	0.503	0.352	0.253	0.182
Mean GLS SE	1.5e7	1.053	0.190	0.093	6.0e7	1.332	0.309	0.167
Median GLS SE	0.476	0.215	0.133	0.092	0.913	0.384	0.236	0.165
CR95%	0.984	0.963	0.939	0.928	0.986	0.963	0.940	0.925
Mean Rbst SE	0.253	0.179	0.126	0.089	0.402	0.303	0.222	0.160
CR95%	0.896	0.916	0.917	0.915	0.880	0.905	0.917	0.906
Mean CLS SE	0.252	0.180	0.130	0.096	0.397	0.304	0.228	0.172
CR95	0.892	0.915	0.923	0.935	0.875	0.902	0.919	0.928

cluster correlations. All simulations consider 5000 repetitions.

The bias magnitude in Table 2 and Table 3 is comparable to those observed in Table 1, with simulated standard errors that show to be slighly larger when we allow for intra-cluster correlations, just as expected. When assuming there is no need to clustered standard errors (Table 2), the coverage associated to GLS standard errors is above 95% when the samples are small, but it approximates to 95% as the sample size increases. On the other and, the coverage rates associated to Robust Standard errors are closer to 90%, albeit increasing slighly for larger samples.

One of the main reason that the GLS-Standard errors achieve higher than expected rates of coverage may be related to the fact that the Standard error estimator is very sensitive to near zero predictions from the scale model. As shown in Table 2, average and median GLS-SE are considerably larger than the simulated standard errors when the sample are small. While robust standard errors are less sensitive to this problem, producing more stable results, they tend to underestimate the magnitude of the true standard errors in small samples. This translates into the lower coverage rates.

When we consider the presence of Intra-Cluster correlation, Table 3, we still observe similar problems with the GLS-SE, albeit with high coverage rates. While Robust and Clustered Standard errors are in average smaller than the Simulated Standard errors, coverage rates are above 90%. Clustered standard errors perform the best only when the sample size is large, with Robust standard error producing low Standard errors estimates.

Overall, the simulation study shows that using GLS-SE estimator may be appropriate when the sample is relatively large. When there is suspected presence of intra-cluster correlation, the use of clustered standard errors is recommended in larger samples. However, when the sample is small, the use of robust standard errors may be appropriate if GLS standard errors when there is the risk of near-zero predictions from the scale model.

4 Illustrative application

In this section we replicate one of the exercises 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 government 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 and 65 years of age (P1564); the share of the population 65 years and older (P65); one-year lag of the dependent variable (LSP); oil prices in US dollars, differntiating between importer and exporter countries (OILIM and OILEX); and the output gap (YGAP). In addition to country fixed effects (as illustrated in MSS), we show results allowing for time fixed effects. Table 4 and Table 5 provide the results for the model with and without time fixed effects, respectively. The tables showcase 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 4: 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
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 4 shows that the point estimates are identical to those reported in Machado and Santos Silva [2019] (Table 6), including analytical standard errors (GLS). With our estimator, however, we are also able to produce both robust and clustered standard errors for location and scale coefficients. Except for a few cases, the robust and clustered standard errors are larger than the GLS standard errors, which may be an indication of misspecification of the model. The GLS standard errors we report differ from those 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 in MSS. Our clustered standard errors, however, are closer to the bootstrap-based standard errors the authors report.[12]

In Table 5, 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 that can be drawn, based on the results from Table 4. 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 at this point the largest impact on the Scale component is felt. Similarly, we observe that the income gap has an impact on government surplus that is always negative and increasing across quantiles. In both instances, the effects are not statistically significant.

May be worth noting that the GLS SE are almost twice as large as the robust and clustered standard errors. This is consistent with the results from the simulation study. Examining the predictions of the scale model, we find that there are 9 observations with a negative predicted scale, and 9 observations that could be considered outliers. This may be the reason for the large GLS standard errors.

Table 5: 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
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							

[12]The differences in the GLS standard errors may be due that in our derivation the influence function of the standardized τ_{th} quantile (see #eq-ifs) 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
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

In this paper, we have extended the methodology proposed by Machado and Santos Silva [2019] in order to estimate quantile regression models with multiple sets of fixed effects as well as with alternative standard errrors. This methodology will allow researchers to implement more-comprehensive analyses of data sets characterized by complex hierarchies and unobserved heterogeneity. This extension is particularly valuable in contexts where group specific effects vary across the conditional distribution of the outcome of interest.

Using a small simulation study, we show that our extended approach is as effective in identifying the parameters of interest as that of Machado and Santos Silva [2019], even in contexts with two sets of fixed effects. Notably, the bias-corrected estimator based on the split-panel jackknife estimator exhibits promising results, mitigating biases when samples are small but increasing standard errors.

Furthermore, we have assessed the impact of intracluster correlation on the performance of standard error estimations. Our findings emphasize the importance of using appropriate standard error estimators to ensure accurate inference. In particular, we find that GLS standard errors are biased when when the scale model predictions are close to zero or negative, and that Robust and clustered standard errors are more stable in those scenarios. Clustered Standard errors perform the best in the presence of intra-cluster correlation, but the advantage is only evident when the sample size is large.

Finally, we have illustrated the application of our extended methodology using data from Persson and Tabellini [2005]. Our results are consistent with those reported in Machado and Santos Silva [2019], and we are able to provide robust and clustered standard errors for the location and scale coefficients. We find that the GLS standard errors are almost twice as large as the robust and clustered standard errors, which is consistent with the results from our simulation study. Nevertheless, there is no drastic change in the original conclusions.

References

John M. Abowd, Francis Kramarz, and Sébastien Roux. Wages, mobility and firm performance: Advantages and insights from using matched worker–firm data*. *The Economic Journal*, 116 (512):F245–F285, 2006. doi: https://doi.org/10.1111/j.1468-0297.2006.01099.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.2006.01099.x.

A. Colin Cameron and Pravin K. Trivedi. *Microeconometrics: methods and applications*. Cambridge University Press, Cambridge ; New York, 2005. ISBN 978-0-521-84805-3.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, April 2011. ISSN 0735-0015, 1537-2707. doi: 10.1198/jbes.2010.07136. URL http://www.tandfonline.com/doi/abs/10.1198/jbes.2010.07136.

Ivan A. Canay. A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386, 2011. doi: https://doi.org/10.1111/j.1368-423X.2011.00349.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1368-423X.2011.00349.x.

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

Geert Dhaene and Koen Jochmans. Split-panel Jackknife Estimation of Fixed-effect Models. *The Review of Economic Studies*, 82(3):991–1030, 02 2015. ISSN 0034-6527. doi: 10.1093/restud/rdv007. URL https://doi.org/10.1093/restud/rdv007.

Ragnar Frisch and Frederick V. Waugh. Partial time regressions as compared with individual trends. *Econometrica*, 1(4):387–401, 1933. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907330.

Paolo Frumento and Matteo Bottai. Parametric modeling of quantile regression coefficient functions. *Biometrics*, 72(1):74–84, 2016. doi: https://doi.org/10.1111/biom.12410. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12410.

Antonio F. Galvao and Kato Kengo. Quantile regression methods for longitudinal data. In *Handbook of quantile regression*, pages 363–380. Chapman and Hall/CRC, 2017.

Simen Gaure. OLS with multiple high dimensional category variables. *Computational Statistics & Data Analysis*, 66:8–18, 2013. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2013.03.024. URL https://www.sciencedirect.com/science/article/pii/S0167947313001266.

Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Statistics. Wiley, 1 edition, March 2005. ISBN 978-0-471-73577-9 978-1-118-18643-5. doi: 10.1002/9781118186435. URL https://onlinelibrary.wiley.com/doi/book/10.1002/9781118186435.

Douglas N. Harris and Tim R. Sass. Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7):798–812, 2011. ISSN 0047-2727. doi: https://doi.org/10.1016/j.jpubeco.2010.11.009. URL https://www.sciencedirect.com/science/article/pii/S0047272710001696.

Xuming He. Quantile curves without crossing. *The American Statistician*, 51(2):186–192, 1997. doi: 10.1080/00031305.1997.10473959. URL https://www.tandfonline.com/doi/abs/10.1080/00031305.1997.10473959.

Kyung So Im. Robustifying glejser test of heteroskedasticity. *Journal of Econometrics*, 97(1):179–188, 2000. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(99)00061-5. URL https://www.sciencedirect.com/science/article/pii/S0304407699000615.

Ben Jann. 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, March 2020. URL https://ideas.repec.org/p/bss/wpaper/35.html.

David M. Kaplan and Yixiao Sun. Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory*, 33(1):105–157, February 2017. ISSN 0266-4666, 1469-4360. doi: 10.1017/S0266466615000407. URL https://www.cambridge.org/core/journals/econometric-theory/article/abs/smoothed-estimating-equations-for-instrumental-variables-quantile-regression/59EE6581C077098DDBB87F3E98F9BF90#. Publisher: Cambridge University Press.

Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2004.05.006. URL https://www.sciencedirect.com/science/article/pii/S0047259X04001113. Special Issue on Semiparametric and Nonparametric Mixed Models.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913643.

Carlos Lamarche. Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, 157(2):396–408, 2010. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2010.03.042. URL https://www.sciencedirect.com/science/article/pii/S0304407610000850.

Tony Lancaster. The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413, 2000. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(99)00044-5. URL https://www.sciencedirect.com/science/article/pii/S0304407699000445.

Michael C. Lovell. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010, 1963. doi: 10.1080/01621459.1963.10480682. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10480682.

José A.F. Machado and J.M.C. Santos Silva. Quantiles via moments. *Journal of Econometrics*, 213(1):145–173, 2019. ISSN 0304-4076. doi: 10.1016/j.jeconom.2019.04.009. URL https://www.sciencedirect.com/science/article/pii/S0304407619300648. Annals: In Honor of Roger Koenker.

Whitney K. Newey and Daniel McFadden. Chapter 36 Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, 1994. ISBN 978-0-444-88766-5. doi: 10.1016/S1573-4412(05)80005-4. URL https://linkinghub.elsevier.com/retrieve/pii/S1573441205800054.

J. Neyman and Elizabeth L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32, 1948. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1914288.

Torsten Persson and Guido Enrico Tabellini. *The economic effects of constitutions*. Munich lectures in economics. MIT Press, Cambridge, Mass., 1. mit press paperback ed edition, 2005. ISBN 978-0-262-66192-8 978-0-262-16219-7.

Fernando Rios-Avila. Feasible Fitting of Linear Models with N Fixed Effects. *The Stata Journal*, 15(3):881–898, October 2015. ISSN 1536-867X. doi: 10.1177/1536867X1501500318. URL https://doi.org/10.1177/1536867X1501500318. Publisher: SAGE Publications.

Quanshui Zhao. Restricted Regression Quantiles. *Journal of Multivariate Analysis*, 72(1):78–99, 2000. ISSN 0047-259X. doi: https://doi.org/10.1006/jmva.1999.1849. URL https://www.sciencedirect.com/science/article/pii/S0047259X99918493.

A Derivation of the influence functions

A.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 a 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 0 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, the rest of the appendix uses q_τ to represent $Q_\varepsilon(\tau)$.

A.2 Estimation of the Variance-Covariance Matrix

In this model, to estimate the variance-covariance matrix the set of coefficients $\theta' = [\beta' \; \gamma' \; q_\tau]$, 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

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

with

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

First Moment Condition: Location Model

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

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

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

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

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

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

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

Third Moment Condition: Quantile of Standardized Residual

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

Because the indicator function $\mathbb{1}()$ is not differentiable, we borrow from the nonparametric literature to approximate this function with a kernel function. Call $k()$ a well behaved kernel function that is symetric around 0, and $K()$ its integral, with range between 0 and 1. With an arbirarily small bandwidth h, we can use the function $K()$ to approximate the indicator function:

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

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

$$h_{3,i} \approx K\left(\frac{1}{h}\Big(q_\tau x_i'\gamma + x_i'\beta - y_i\Big)\right) - \tau$$

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

$$\begin{aligned}
\bar{G}_{3,1} &= -\frac{1}{N}\sum \frac{\partial h_{3,i}}{\partial \beta'} \\
&= -N^{-1}\sum \frac{1}{h} k\left(\frac{1}{h}\Big(q_\tau x_i'\gamma + x_i'\beta - y_i\Big)\right) x_i' \\
&= -N^{-1}\sum \frac{1}{h} k\left(\frac{1}{h}(q_\tau x_i'\gamma - \nu_i)\right) x_i' \\
&= -N^{-1}\sum \frac{1}{h} k\left(\frac{1}{h}(q_\tau - \varepsilon_i) x_i'\gamma\right) x_i'
\end{aligned}$$

If we condition on X, use the law of iterated expectations, and assume that ε_i is homoskedastic, we can obtain:

$$\begin{aligned}
\bar{G}_{3,1} &= -N^{-1}\sum \frac{1}{h} k\left(\frac{1}{h}(q_\tau - \varepsilon_i) x_i'\gamma\right) x_i' \\
&= -N^{-1}\sum \frac{1}{h} k\left(\frac{q_\tau - \varepsilon_i}{h}\right) \frac{x_i'}{x_i'\gamma} \\
&= -N^{-1} f_\varepsilon(q_\tau) \sum \frac{x_i'}{x_i'\gamma}
\end{aligned}$$

Finally, this simplifies to:

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

where we use the fact that asymptotically, the expression $\frac{1}{N}\sum \frac{a_i}{b_i}$ can be approximated using Taylor expansions by $\frac{\bar{a}}{\bar{b}}$. [13] Thus, we can rewrite the last term as

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

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

$$
\begin{aligned}
\bar{G}_{3,2} &= -\frac{1}{N}\sum \frac{\partial h_{3,i}}{\partial \gamma'} \\
&= -N^{-1}\sum \frac{1}{h}k\left(\frac{1}{h}(q_\tau - \varepsilon_i)x_i'\gamma\right) q_\tau x_i' \\
&= -N^{-1}\sum \frac{1}{h}k\left(\frac{q_\tau - \varepsilon_i}{h}\right) q_\tau \frac{x_i'}{x_i'\gamma} \\
&\simeq -f(q_\tau)q_\tau \frac{\bar{x}'}{\bar{x}'\gamma}
\end{aligned}
$$

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

$$
\begin{aligned}
\bar{G}_{3,3} &= -\frac{1}{N}\sum \frac{\partial h_{3,i}}{\partial q_\tau} \\
&= -N^{-1}\sum \frac{1}{h}k\left(\frac{1}{h}(q_\tau - \varepsilon_i)x_i'\gamma\right) x_i'\gamma \\
&= -N^{-1}\sum \frac{1}{h}k\left(\frac{q_\tau - \varepsilon_i}{h}\right) \frac{x_i'\gamma}{x_i'\gamma} \\
&\simeq -f(q_\tau)
\end{aligned}
$$

A.3 Influence Functions

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:

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

Scale Coefficients

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

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 positive, it also follows that $\mathbb{1}(\nu_i \geq 0) = \mathbb{1}(\varepsilon_i \geq 0)$. Thus, it can be simplified as

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

Quantile of Standardized Residual

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

B Implementation

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

- `mmqreg` in Stata: `net install mmqreg, from(https://friosavila.github.io/stpackages)`
- `mmqreg` in R: Available on GitHub (https://github.com/friosavila/mmqreg)