Estimating Quantile Regressions with Multiple Fixed Effects through Method of Moments

Fernando Rios-Avila

Leonardo Siles

Gustavo Canavire-Bacarreza

2024-09-10

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.

# 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 . 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 Cameron and Trivedi (2005), the effect of a variable on the 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 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.

# Methodology

## Quantile Regression: Location-Scale model

Quantile regressions are used to identify relationships between the explanatory variables and the conditional quantiles of the dependent variable . This relationship is commonly assumed to follow a linear functional form:

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

An alternative formulation of quantile regressions is the location-scale model. This approach assumes that the conditional quantile of given and 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:

Here, the location parameters 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 can be estimated using standard approaches.

Both the standard quantile regression ([Equation 1](#eq-eq1)) and the location-scale specification ([Equation 2](#eq-eq2)) can be estimated as the solution to a weighted minimization problem:

One characteristic of this estimator is that the coefficients are identified locally and thus the estimated quantile coefficients will exhibit considerable variation when analyzed across . 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]](#footnote-24)

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]](#footnote-25)

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

In this setup, the traditional quantile coefficients are identified as the location model coefficients plus the scale model coefficients moderated by the unconditional quantile of the standardized error . For simplicity we will use to denote in the rest of the paper.

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 () requiring local estimation, this estimation approach will be more efficient than the standard quantile regression model (Zhao 2000). Second, under the assumption that 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](#eq-eq5) can be estimated using a generalized method of moments approach. And while it is possible to identify all coefficients () 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 and the independent variables are the explanatory variables (including a constant) with an error , which is by definition heteroskedastic. In this case, the location-model coefficients are identified under the following condition:

Second, after the location model is estimated, the scale coefficients can be identified by modeling heteroskedasticity as a linear function of characteristics . For this we use the absolute value of the errors from the location model 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:

It should be noticed that the estimation of the standard errors (next section) requires that the Scale component prediction is strictly positive, because it represents the conditional standard deviation of the error . Because this component is identified using a linear model, some values for 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 quantile of the error can be estimated using the following condition:

where one identifies the quantile of the error using standardized errors or by finding the values that identify the overall quantile coefficients . Afterwards, the conditional quantile coefficients are simply defined as the combination of the location and scale coefficients.

## 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:

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]](#footnote-34)

Call the set of coefficients that are identified by the moment conditions in [Equation 10](#eq-eq10), a just identified model, and the function is a vector function that stacks all the moments described in [Equation 10](#eq-eq10) at the individual level. Then follows a normal distribution with mean and variance-covariance matrix that is estimated as

which is equivalent to the Eicker-White heteroskedasticity-consistent estimator for least-squares estimators.

Here, the inner product is the moment covariance matrix and is the Jacobian matrix of the moment equations evaluated at .

In this framework, the quantile regression coefficients, a combination of the location-scale-quantile estimates, will follow a normal distribution with mean and variance-covariance matrix equal to

where is a matrix defined as

with being an identity matrix of dimension (number of explanatory variables in 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 , the variance-covariance matrix can be estimated as

where the influence functions are defined as:

For the specific case of quantile regressions via moments, the influence functions for the location, scale, and quantile coefficients are[[4]](#footnote-37)

The different types of standard errors estimation thus depend on the assumptions imposed for the estimation of .

### Robust Standard Errors

The first and most natural standard error estimator is given by equation [Equation 12](#eq-eqvcv). 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

### 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 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 be the total number of clusters , where . The clustered variance covariance matrix is given by[[5]](#footnote-40)

where is the sum of the influence functions over all observations within a given cluster .

### 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]](#footnote-42) 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:

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

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:

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

where

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.

## 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, Kramarz, and Roux 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:

where we assume vary across groups and (and thus are not collinear) and that and are the location and scale effects associated with the groups fixed effects.[[7]](#footnote-45)

If the dimension of groups is low, this model could be estimated using a dummy inclusion approach following [Section 2.1](#sec-betas), and the standard errors obtained as discussed in [Section 2.2](#sec-se). However, if the dimensionality of 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 , and the outcome of interest , and using a similar approach for the identification of . 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](#sec-betas) are needed.

First, for all dependent and independent variables in the model (), we partial out the group fixed effects and obtain the centered-residualized variables:

Afterward, we estimate the location model using the centered-residualized variables:[[8]](#footnote-46)

Because is the dependent variable for the scale model, we apply the partialing out and recentering to this expression (), and use that to estimate the following model:

Finally, the standardized residuals can be obtained as follows:

where is the prediction for the conditional standard deviation

The quantile of the error can be estimated as usual and the variance-covariance matrices obtained in the same way as before (see [Section 2.2](#sec-se)) by using instead of when estimating the influence functions for all estimated coefficients.

# 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 . In contrast with MSS, we consider a two-way fixed effect structure. For this exercise, we consider the following data-generating process:

where , , and , with . We only consider the case when the error term is assumed to follow a centered distribution.[[9]](#footnote-49) 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:

where and , are errors that vary across individuals or across clusters . is the cumulative distribution function of a normal distribution, and 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]](#footnote-50)

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](#tbl-sim1) 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]](#footnote-51) 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 (), 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.

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

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](#tbl-sim2) considers the d.g.p. without intra-cluster correlation, while [Table 3](#tbl-sim3) considers the d.g.p. that induces intra-cluster correlations. All simulations consider 5000 repetitions.

The bias magnitude in [Table 2](#tbl-sim2) and [Table 3](#tbl-sim3) is comparable to those observed in [Table 1](#tbl-sim1), 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](#tbl-sim2)), 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](#tbl-sim2), 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.

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

When we consider the presence of Intra-Cluster correlation, [Table 3](#tbl-sim3), 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.

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

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.

# Illustrative applications

## Determinants of government surpluses

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](#tbl-tb1) and [Table 5](#tbl-tb2) 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 |

As expected, [Table 4](#tbl-tb1) 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]](#footnote-57)

In [Table 5](#tbl-tb2), 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](#tbl-tb1). 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 |

## The Effect of Criminal Activities on Labor Incomes

In this section, we investigate the effects of criminal activities on labor incomes using a quantile regression approach to discover whether crime has different impacts on the distribution of interest. We use data from the ENEMDU labor survey from Ecuador and combine it with administrative records of crime reports by Ecuadorian Cantons.

The combined data set covers the period 2019-2022. For the sake of brevity, we will only present here the results of quantile regressions of the log of total labor income on the log of annual reported murders (LMURDERS) and other covariates displayed in Table 1. However, our findings are robust to alternative specifications of the model where the dependent variable is either the log of per-capita household income or the log of total labor income from main occupation, both of which are displayed in [Section 7](#sec-app-crime).

|  |
| --- |
| Table 6: The Effect of Crime on Labor Income: Canton, Activity Sector and Time Fixed Effects |

The findings in [Table 6](#tbl-ecuador) support the idea that the effect of murders diminishes as we move forward in the distribution from low to high incomes: in the lowest decile, a 1% increase in reported murders is associated with a 3.08% reduction in total labor income; instead, for the top 10% labor income earners, murders are seen to have little effect on the dependent variable and, in addition, the coefficient is not significant at conventional levels. Quantile regression estimates are thus really useful to discover relationships between dependent and explanatory variables that are not seen in standard (mean) regression analysis. For this application, we have found a monotonic decreasing effect of the log of reported crimes on total labor income.

While the capability of estimating distinct coefficients of explanatory variables for each quantile is a feature present since Koenker and Bassett (1978), the method of moments approach allows us to include not one but multiple sets of fixed effects so that one can account for unobserved heterogeneity at different levels. Regarding this application, we have included in the estimation fixed effects for Canton, Activity Sector of the income earner and Year as is shown in [Table 6](#tbl-ecuador). Given that LMURDERS is measured at the Canton level, it may be important to control for unobserved factors affecting both crime and income earnings in the Cantons of Ecuador.

Finally, our approach also allows us to obtain clustered standard errors for estimated coefficients so that statistical inference can be performed properly. [Table 6](#tbl-ecuador) presents clustered standard errors at the Canton level, the rationale being the same as described above for the inclusion of Canton fixed effects in the specification. Had we not taken into account the correlated nature of our standard errors, we would have computed misleading (and smaller) estimates. For example, robust standard errors in this specification are, on average, 2.5 times smaller than clustered standard errors. It is important, however, to remark that the effects of murders on labor income at the lower half of the distribution are still statistically different from zero when clustering at the Canton level.

# 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.

Our example using data to analyze the impact of Crime on labor income in Ecuador also suggests the relevance of using clustered standard errors in the presence of unobserved heterogeneity at different levels. We find that the effect of murders on labor income diminishes as we move forward in the distribution from low to high incomes, which is a result that would not have been possible without the use of quantile regression.

# References

Abowd, John M., Francis Kramarz, and Sébastien Roux. 2006. “Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker–Firm Data\*.” *The Economic Journal* 116 (512): F245–85. https://doi.org/<https://doi.org/10.1111/j.1468-0297.2006.01099.x>.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. “Robust Inference With Multiway Clustering.” *Journal of Business & Economic Statistics* 29 (2): 238–49. <https://doi.org/10.1198/jbes.2010.07136>.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge ; New York: Cambridge University Press.

Canay, Ivan A. 2011. “A Simple Approach to Quantile Regression for Panel Data.” *The Econometrics Journal* 14 (3): 368–86. https://doi.org/<https://doi.org/10.1111/j.1368-423X.2011.00349.x>.

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

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

Frisch, Ragnar, and Frederick V. Waugh. 1933. “Partial Time Regressions as Compared with Individual Trends.” *Econometrica* 1 (4): 387–401. <http://www.jstor.org/stable/1907330>.

Frumento, Paolo, and Matteo Bottai. 2016. “Parametric Modeling of Quantile Regression Coefficient Functions.” *Biometrics* 72 (1): 74–84. https://doi.org/<https://doi.org/10.1111/biom.12410>.

Galvao, Antonio F., and Kato Kengo. 2017. “Quantile Regression Methods for Longitudinal Data.” In *Handbook of Quantile Regression*, 363–80. Chapman; Hall/CRC.

Gaure, Simen. 2013. “OLS with Multiple High Dimensional Category Variables.” *Computational Statistics & Data Analysis* 66: 8–18. https://doi.org/<https://doi.org/10.1016/j.csda.2013.03.024>.

Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. 2005. *Robust Statistics: The Approach Based on Influence Functions*. 1st ed. Wiley Series in Probability and Statistics. Wiley. <https://doi.org/10.1002/9781118186435>.

Harris, Douglas N., and Tim R. Sass. 2011. “Teacher Training, Teacher Quality and Student Achievement.” *Journal of Public Economics* 95 (7): 798–812. https://doi.org/<https://doi.org/10.1016/j.jpubeco.2010.11.009>.

He, Xuming. 1997. “Quantile Curves Without Crossing.” *The American Statistician* 51 (2): 186–92. <https://doi.org/10.1080/00031305.1997.10473959>.

Im, Kyung So. 2000. “Robustifying Glejser Test of Heteroskedasticity.” *Journal of Econometrics* 97 (1): 179–88. https://doi.org/<https://doi.org/10.1016/S0304-4076(99)00061-5>.

Jann, Ben. 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. <https://ideas.repec.org/p/bss/wpaper/35.html>.

Kaplan, David M., and Yixiao Sun. 2017. “Smoothed Estimating Equations for Instrumental Variables Quantile Regression.” *Econometric Theory* 33 (1): 105–57. <https://doi.org/10.1017/S0266466615000407>.

Koenker, Roger. 2004. “Quantile Regression for Longitudinal Data.” *Journal of Multivariate Analysis* 91 (1): 74–89. https://doi.org/<https://doi.org/10.1016/j.jmva.2004.05.006>.

Koenker, Roger, and Gilbert Bassett. 1978. “Regression Quantiles.” *Econometrica* 46 (1): 33–50. <http://www.jstor.org/stable/1913643>.

Lamarche, Carlos. 2010. “Robust Penalized Quantile Regression Estimation for Panel Data.” *Journal of Econometrics* 157 (2): 396–408. https://doi.org/<https://doi.org/10.1016/j.jeconom.2010.03.042>.

Lancaster, Tony. 2000. “The Incidental Parameter Problem Since 1948.” *Journal of Econometrics* 95 (2): 391–413. https://doi.org/<https://doi.org/10.1016/S0304-4076(99)00044-5>.

Lovell, Michael C. 1963. “Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis.” *Journal of the American Statistical Association* 58 (304): 993–1010. <https://doi.org/10.1080/01621459.1963.10480682>.

Machado, José A. F., and J. M. C. Santos Silva. 2019. “Quantiles via Moments.” *Journal of Econometrics* 213 (1): 145–73. <https://doi.org/10.1016/j.jeconom.2019.04.009>.

Newey, Whitney K., and Daniel McFadden. 1994. “Chapter 36 Large Sample Estimation and Hypothesis Testing.” In *Handbook of Econometrics*, 4:2111–2245. Elsevier. <https://doi.org/10.1016/S1573-4412(05)80005-4>.

Neyman, J., and Elizabeth L. Scott. 1948. “Consistent Estimates Based on Partially Consistent Observations.” *Econometrica* 16 (1): 1–32. <http://www.jstor.org/stable/1914288>.

Persson, Torsten, and Guido Enrico Tabellini. 2005. *The Economic Effects of Constitutions*. 1. MIT Press paperback ed. Munich Lectures in Economics. Cambridge, Mass.: MIT Press.

Rios-Avila, Fernando. 2015. “Feasible Fitting of Linear Models with N Fixed Effects.” *The Stata Journal* 15 (3): 881–98. <https://doi.org/10.1177/1536867X1501500318>.

Zhao, Quanshui. 2000. “Restricted Regression Quantiles.” *Journal of Multivariate Analysis* 72 (1): 78–99. https://doi.org/<https://doi.org/10.1006/jmva.1999.1849>.

# Derivation of the influence functions

## 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:

where is an unobserved i.i.d. random variable that is independent of and such that is larger than 0 for any .

In this case, the conditional quantile model can be written as

This model is identified under the following conditions:

For simplicity, the rest of the appendix uses to represent .

## Estimation of the Variance-Covariance Matrix

In this model, to estimate the variance-covariance matrix the set of coefficients , we need to obtain the influence functions of all coefficients, which are defined as

where the Jacobian matrix is defined as

with

First Moment Condition: Location Model

Second Moment Condition: Scale model

Under the assumption , or in this case , is uncorrelated with , we can simplify the expression as

Third Moment Condition: Quantile of Standardized Residual

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

Thus the function can be approximated as

Now, we can obtain the Jacobian matrix as:

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

Finally, this simplifies to:

where we use the fact that asymptotically, the expression can be approximated using Taylor expansions by . [[13]](#footnote-118) Thus, we can rewrite the last term as

The Jacobian for the second matrix can be derived similarly:

and the Jacobian for the third matrix is

## Influence Functions

Location Coefficients

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

Scale Coefficients

However,

and

Thus,

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 . Thus, it can be simplified as

Quantile of Standardized Residual

# Complementary analysis of [Section 4.2](#sec-crime)

In this appendix, we present the quantile regressions of both (log) per-capita household income and the (log) of labor income from main occupation on the (log) of annual reported murders. While in the first specification the coefficients of interest are not significant at conventional levels because of clustered standard errors (see [Table 7](#tbl-acrime1)), the same pattern of decreasing association between murders and measurements of well-being as we move to the right of the distribution is revealed using our proposed method.

However, in [Table 8](#tbl-acrime2) we get significant (at the 1% level) associations between reported murders and a slightly different measurement of labor income that only takes into account the main occupation of the individual. Compared to total labor income, the coefficients displayed in [Table 8](#tbl-acrime2) are smaller in magnitude: for the lowest decile of the distribution, we find a 1 percentage point difference in the coefficient of interest between the two alternative specifications, where the largest effect corresponds to total labor income. Nevertheless, and once again, the decreasing pattern of reported murders across quantiles is documented for this specification. What is even more interesting, compared to our estimations in [Table 6](#tbl-ecuador), is that crime is no longer statistically meaningful from the median of the distribution onwards.

|  |
| --- |
| Table 7: The Effect of Crime on Per Capita Household Income: Canton and Time Fixed Effects |

|  |
| --- |
| Table 8: The Effect of Crime on Labor Income in Main Occupation: Canton, Activity Sector and Time FE |

# 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)

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). [↑](#footnote-ref-24)
2. Machado and Santos Silva (2019) also discuss a model where heteroskedasticity can be an arbitrary nonlinear function , but develop the estimator for the linear case, i.e., when is the identity function. [↑](#footnote-ref-25)
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). [↑](#footnote-ref-34)
4. The derivation of the influence functions can be found in the appendix. [↑](#footnote-ref-37)
5. It should be noted that one could just as well apply the insights of Cameron, Gelbach, and Miller (2011), allowing for multiway clustering. [↑](#footnote-ref-40)
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. [↑](#footnote-ref-42)
7. We could just as well consider multiple sets of fixed effects. [↑](#footnote-ref-45)
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. [↑](#footnote-ref-46)
9. Specifically we assume , where follows a chi-squared distribution such that . Simulations under the assumption of normal errors are available upon request. [↑](#footnote-ref-49)
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 [↑](#footnote-ref-50)
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 . [↑](#footnote-ref-51)
12. The differences in the GLS standard errors may be due that in our derivation the influence function of the standardized quantile (see [Equation 13](#eq-ifs)) does not have the same leading term as the one reported in MSS (see Theorem 3, and the definition of ). [↑](#footnote-ref-57)
13. This approximation will be useful when we consider the estimation of the influence functions. [↑](#footnote-ref-118)