

Untitled

Abstract.

Keywords: st0001,

1 Introduction

2 The Basics

While standard regression techniques allow the researcher to determine the effect of covariates X on the expected value of a response variable Y , regression quantiles are capable of characterizing how changes in X affect the entire distribution of Y . In applied research, this may be useful when one seeks to determine how does a given percentile of the distribution of interest respond to changing values of explanatory variables. For example, consider the effects of smoking on birth weight, as [abrevaya2008] studied in the context of quantile regression. Instead of fixing an *unconditional threshold* so as to divide the sample in low birth weights and non-low birth weights, the authors followed a *conditional quantile* approach that enabled them to study the effects of birth inputs (e.g. smoking) in different parts of the birth weight distribution ([abrevaya2008]).

In its more common setting, quantile regressions are assumed to follow a linear form:

$$Q_{\tau}(Y|X) = X\beta(\tau)$$

Such that, while the parameters of interest change across quantiles, the effects of covariates X on the τ -th quantile of Y are assumed to be linear.

In addition to heterogeneous effects of explanatory variables on the outcome of interest, the researcher might be interested as well in properly addressing the problem of unobserved heterogeneity. Returning to the example of birth weights, it may be the case that unobservable characteristics of the mother are correlated with both the birth outcome and birth inputs, such as her smoking status during pregnancy or whether she attended prenatal care visits ([abrevaya2008]). In this case, a panel data approach may be useful for obtaining causal estimates of the effects of explanatory variables on quantiles of Y .

At the same time, panel data models with unobserved time-invariant effects depend upon demeaning techniques to eliminate such effects and yield consistent estimators of the parameters of interest. However, differentiating out individual effects is not a feasible approach within the context of regression quantiles, the reason being that quantiles are *not* linear operators. Many authors, for example [abrevaya2008], [canay2011] or

[@mss2019] have stated that this key limitation has prevented further progress in the estimation of regression quantiles with fixed effects.

In this paper, we review four estimators for regression quantiles in the presence of fixed effects. All of them are relatively simple to implement with our provided commands for Stata (see sections 3, 4 and 5 below). However, this easiness comes at the cost of making important assumptions regarding, among others, the data generating process (DGP) and the effect of unobserved heterogeneity α_i on the distribution of Y . The following sections describe each of the estimators, their key assumptions and provide examples for their implementation in Stata.

3 Correlated Random Effects

This estimator, described by [wooldridge2010], rests upon the Mundlak representation of a Correlated Random Effects (CRE) model for the estimation of regression quantiles with panel data. The underlying DGP for the outcome of interest Y_{it} is:

$$Y_{it} = X'_{it}\beta + \alpha_i + u_{it}$$

where α_i is the time-invariant individual effect and can be described in terms of $X_i \equiv (X_{i1}, \dots, X_{iT})$ as follows:

$$\alpha_i = \psi + \bar{X}_i' \lambda + v_i$$

where \bar{X}_i denotes the time average of characteristics across individuals. Then, the conditional τ -th quantile of Y_{it} is:

$$Q_\tau(Y_{it}|X_i) = \psi + X'_{it}\beta + \bar{X}_i' \lambda + Q_\tau(a_{it}|X_i)$$

where $a_{it} \equiv v_i + u_{it}$ is the composite error. The CRE estimator using the Mundlak representation imposes the following key assumption:

CRE1. a_{it} is independent of X_i .

Stringent as may be, this assumption allows to consistently estimate β and λ from a pooled quantile regression of Y_{it} on X_{it} and \bar{X}_i . As [wooldridge2010] remarks, CRE1 amounts to assuming that quantile functions are parallel or, in other words, that by imposing a relationship between X_i and α_i of the form displayed in equation [eq:mundlak_rep] we can only estimate the location shifts of quantile regressions. As it will be shown below, this is a feature the CRE estimator shares with Canay's method (see section 4) for estimating regression quantiles in a panel data context. Both estimators are constrained in the sense that only location effects on quantile functions are accounted

for, whereas the MMQREG method (see section 5) has the ability to estimate quantile coefficients that take into account both location and scale effects.

[@abrevaya2008] develop an alternative CRE estimator for regression quantiles with panel data drawing on the model of [@chamberlain1982]. However, its implementation is severely restricted because Abrevaya and Dahl’s estimator regresses Y_{it} on X_{it} and X_i for a given τ . Even if we have a balanced panel available for our application, we will have to deal with the large quantity of regressors included in the equation $[K(1 + T)]$ and its corresponding loss of degrees of freedom for statistical inference.

Introduction of multiple fixed effects in this setting is straightforward. First, modify equation [eq:dgp_cre] so as to include two or more sets of fixed effects $(\alpha^1, \alpha^2, \dots, \alpha^k)$ and proceed by computing the averages of explanatory variables X with respect to the variable identifying the fixed effect. Finally, run a regression of Y on the set of original control variables and the averages just calculated. In terms of the quantile function, equation [eq:qr_cre] requires to be modified accordingly:

$$Q_\tau(Y_{it}|X_i) = \Psi + X'_{it}\beta + \bar{X}_i^{1'}\lambda^1 + \bar{X}_i^{2'}\lambda^2 + Q_\tau(a_{it}|X_i)$$

For the case of twoway fixed effects and where $\Psi \equiv \psi^1 + \psi^2$.

Implementation of the CRE estimator for quantile regression in Stata can be done with the command `cre`, which is actually a prefix for the estimation of CRE models. Option `[abs(varlist)]` is used to provide the command with variables identifying groups of observations that will be subsequently used to estimate the fixed effects. By default, singleton groups are dropped from the estimation. Now, we present an application of the CRE estimator for regression quantiles with multiple fixed effects using the aforementioned command. Note that the same dataset will be used for all estimation methods surveyed in this paper, so it will be worth to outline the main features of the data before proceeding to display, comment and compare the estimation results.

We use the panel from [@persson2007] who study the effect of alternative types of government on its overall spending in the last year of the legislature. The authors consider both coalition and single-party as government types, leaving out from consideration minority governments. Here, we will only focus on the relationship a coalition government has with its overall spending and expand [@persson2007] results by estimating the .25, .5 and .75 quantiles of the government spending distribution.

Variables used for estimation include central government spending as percentage of GDP in the last year of legislature (*last_exp*), an indicator variable for coalition governments (*coalition*), population size in log scale (*lpop*), the percentage of population over 65 years of age (*prop65*), the log-deviation of output from the country trend (*ygap*) and the length (in years) of the legislature (*length*). Both country and period fixed effects are included in our estimations, in the same fashion as [@persson2007].

Table 1: Government spending and coalition government - CRE Estimator

VARIABLES	(1) 25th quantile	(2) 50th quantile	(3) 75th quantile
coalition	1.629 (1.370)	2.182 (1.567)	1.423 (1.662)
	0.227 (0.150)	0.110 (0.177)	0.167 (0.261)
	-14.01 (9.392)	-13.90 (11.52)	-18.10 (20.32)
	0.239 (0.603)	1.008 (0.773)	0.709 (0.912)
	-0.380* (0.211)	-0.538* (0.300)	-0.574 (0.407)
	-0.242 (0.533)	0.373 (0.423)	0.00656 (0.481)
	245.0 (152.7)	238.1 (189.4)	309.8 (331.5)
Observations Country FE Period FE Bootstrapped	133 Yes	133 Yes	133 Yes
Standard errors in parentheses *** p<0.01, **	Yes	Yes	Yes
p<0.05, * p<0.1 Mean (across FE) coefficients of independent variables are omitted.			

Estimation results from [persson2007] dataset using the CRE estimator for quantile regression are displayed in table 1 and show heterogeneous effects of a permanent switch from a single-party to a coalition government on government spending. The largest effect is obtained at the median of the distribution, and the estimated coefficient is close to the one [persson2007] computed using the within estimator (2.36). At the lower and upper quartile of the distribution we still find positive coefficients, which support the theoretical predictions of the model developed by the same authors. However, none of the coefficients for coalition across quantiles are statistically different from zero. In the context of CRE estimation, this fact may be partly due to the loss of degrees of freedom implied by the addition of 12 extra regressors to the equation. In general, CRE will add $W \times K$ explanatory variables, with W being the number of fixed effects we are accounting for.

4 Canay (2011) Estimator

Canay's estimator key assumption is that fixed effects are merely treated as location-shifters. That is, unobserved individual effects only change the distribution of the response variable in a parallel fashion. While this may be a rather strong assumption, it helps to consistently estimate the parameters of interest using a simple transformation of the data as detailed below [canay2011].

The underlying DGP of Y_{it} is:

$$Y_{it} = X'_{it}\beta(u_{it}) + \alpha_i$$

To consistently estimate the β_τ from the data, the following assumptions are in place:

CANAY1. Conditional on $X_{it} = x$ the random variables X_{it} and α are independent.

CANAY2. U_{it} is independent of X_i and α_i , with $U_{it} \sim U[0, 1]$

CANAY3. $\beta_\mu \equiv E[\beta(U_{it})]$. That is, β_μ exists.

Start defining the following equation, which is simply a restatement of equation [eq:dgp_canay], as follows:

$$Y_{it} = X'_{it}\beta(\tau) + \alpha_i + e_{it}(\tau)$$

Where $e_{it}(\tau) \equiv X'_{it}[\beta(U_{it}) - \beta(\tau)]$. Then, due to assumption CANAY2, in the above equation only $\beta(\tau)$ and $e_{it}(\tau)$ depend on τ . Next, write a conditional mean equation for Y_{it} as the following:

$$Y_{it} = X'_{it}\beta_\mu + \alpha_i + u_{it}$$

Where $u_{it} \equiv X'_{it}[\beta(U_{it}) - \beta_\mu]$. Assumption CANAY3 allows us to write this kind of mean equation. Then, the two-step estimator due to [canay2011] is defined next:

Step 1. Let $\hat{\beta}_\mu$ be a \sqrt{nT} -consistent estimator of β_μ . Define $\hat{\alpha}_i \equiv \mathbb{E}_T[Y_{it} - X'_{it}\hat{\beta}_\mu]$.

Step 2. Let $\hat{Y}_{it} = Y_{it} - \hat{\alpha}_i$ and define the two-step estimator as:

$$\hat{\beta}(\tau) \equiv \min_{\beta \in \beta} \mathbb{E}_{nT}[\rho(\hat{Y}_{it} - X'_{it}\beta)]$$

Where $\rho(\cdot)$ is the check function [koenker1978], $\mathbb{E}_T(\cdot) \equiv T^{-1} \sum_{t=1}^T (\cdot)$ and $\mathbb{E}_{nT}(\cdot) \equiv (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n (\cdot)$. [canay2011] defines this estimator as “consistent and asymptotically normal”. This stated, Canay’s estimator stands out for its simplicity: Step 1 may be conveniently restated as obtaining the known within estimator of β and then subtracting the time-average residuals from the dependent variable. Next, using this simple transformation of the data, Step 2 tells us to obtain regression quantiles of \hat{Y}_{it} that are guaranteed to be consistent, as long as key assumptions hold.

It is important to remark that consistency of Canay’s estimator is obtained when $T \rightarrow \infty$ [canay2011]. Thus, the estimator at hand may perform poorly in contexts of panels with few time periods. In addition to this, the key assumption of the approach relies on the fact that the α_i ’s are pure location shifts: in other words, that

the unobservable time-invariant characteristics grouped in α_i have coefficients that are constant across τ [Canay2011]. As [MSS2019] point out using simulated data, the estimator's performance deteriorates as the DGP features fixed effects that alter the entire distribution of interest, not only its location.

Regarding statistical inference, [Canay2011] provides a simple algorithm to obtain bootstrapped standard errors. The key in this procedure is to compute the two step estimator for each bootstrap sample, avoiding the mistake of just sampling from the transformed data \hat{Y}_{it} . As [Canay2011] remarks, the algorithm should be such that it computes the first step estimators $\hat{\beta}_\mu$ and $\hat{\alpha}_i$ for each repetition.

Finally, we propose a Modified Canay estimator in which, instead of subtracting out the estimated fixed effects $\hat{\alpha}_i$ from the dependent variable Y_{it} , they are included as regressors in equation [eq:greg_canay]. In stark contrast to the CRE estimator (see section 3), the Modified Canay only adds W variables to the estimation, which makes it preferable in contexts of a small sample size when our goal is to estimate only location shift effects. However, we must be aware of the fact that the requirement of $T \rightarrow \infty$ for consistency may not be met, especially for short panels.

Describe the implementation in Stata here.

Table 2: Government spending and coalition government - Canay Estimator

VARIABLES	(1) 25th quantile	(2) 50th quantile	(3) 75th quantile
coalition	2.142 (1.478)	2.319* (1.339)	1.779 (1.391)
	0.235 (0.174)	0.258 (0.176)	0.294* (0.178)
	-10.75 (12.95)	-10.75 (12.96)	-10.76 (12.97)
	0.902 (0.599)	0.910 (0.602)	0.797 (0.588)
	-0.494* (0.290)	-0.430 (0.290)	-0.664** (0.333)
	-0.0973 (0.414)	0.103 (0.406)	0.169 (0.418)
	184.9 (212.3)	184.8 (212.3)	186.8 (212.6)
Observations Country FE Period FE Bootstrapped	133 Yes	133 Yes	133 Yes
Standard errors in parentheses *** p<0.01, **	Yes	Yes	Yes
p<0.05, * p<0.1			

Tables 2 and 3 display the results of Canay and Modified Canay estimations using the [Persson2007] dataset of government expenditures. Once again, the median coefficients are close to the mean coefficient from the original paper, with the Canay estimator

delivering the closest coefficient to 2.36. Even further, statistical significance is achieved for the median coefficient in table 2, strengthening our reasoning on degrees of freedom above. Otherwise, we still obtain positive coefficients of the coalition variable at the 25th and 75th percentile of government spending.

It is also worth stating the increasing monotonic pattern of quantile coefficients of coalition in table 3, conveying us that the effect of a permanent switch from single-party to coalition governments increases as we move toward greater government spending. Finally, note that even after explicitly accounting for country and period of legislature fixed effects in the Modified Canay estimator, the median coefficient remains stable, although it is no longer significant.

Table 3: Government spending and coalition government - Modified Canay Estimator

VARIABLES	(1) 25th quantile	(2) 50th quantile	(3) 75th quantile
coalition	1.806 (1.422) 0.294* (0.170) -9.272 (12.01) 0.801 (0.553) -0.446 (0.292) -0.218 (0.412) 0.855*** (0.0956) 1.014*** (0.258) 161.0 (196.8)	2.180 (1.367) 0.255 (0.190) -10.95 (13.39) 0.912 (0.618) -0.468 (0.299) -0.0964 (0.417) 1.020*** (0.0794) 1.319*** (0.238) 188.9 (219.4)	2.281 (1.460) 0.251 (0.211) -11.36 (14.51) 0.887 (0.655) -0.741** (0.350) 0.194 (0.441) 1.055*** (0.103) 1.313*** (0.256) 196.4 (237.8)
Observations Bootstrapped Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1	133	133	133

5 Method of Moments Quantile Regression

This approach of estimating regression quantiles with multiple fixed effects distinguishes itself from the other estimators reviewed above in the sense that [mss2019] introduce location-scale effects of fixed effects upon the distribution of interest. Compared to Canay's estimator, the Method of Moments Quantile Regression (MMQREG) not only

allows the α_i 's to affect Y_{it} through location shifts, but rather MMQREG is able to identify the scale shifts that alter different points of the distribution belonging to Y .

We begin by defining the DGP of the location scale model:

$$Y_{it} = \alpha_i + X'_{it}\beta + (\delta_i + X'_{it}\gamma)u_{it}$$

Where parameters α_i and δ_i capture the individual fixed effects. Note that, compared to equation [eq:dgp_canay], the fixed effects not only enter the model in an additive fashion, instead they also have a multiplicative effect upon the error term. In addition, the U_{it} are i.i.d. across i and t , statistically independent of X_{it} and satisfy $E(U) = 0$ and $E(|U|) = 1$ both of which normalize the random variable.

Our location scale model in equation [eq:dgp_mmqreg] implies that:

$$Q_\tau(Y_{it}|X_i) = [\alpha_i + \delta_i q(\tau)] + X'_{it}\beta + X'_{it}\gamma q(\tau)$$

Where the scalar coefficient $\alpha_i(\tau) \equiv \alpha_i + \delta_i q(\tau)$ is the quantile- τ fixed effect for individual i , which represent how time-invariant variables have *different impacts on different regions* of the conditional distribution of Y_{it} . However, our real interest is in the regression quantile coefficients:

$$\beta_\tau = \beta + q(\tau)\gamma$$

Which are simply a linear combination of the location coefficients (β) and the scale coefficients (γ), where the second vector of coefficients is weighted by the value of the τ -th quantile of the variable of interest Y_{it} . [mss2019] develop the following algorithm for implementing the MMQREG estimator:

1. Obtain $\hat{\beta}_k$ by regressing time-demeaned Y_{it} on time-demeaned controls X_{it} , i.e. obtain $\hat{\beta}$ by the within estimator.
2. Estimate the $\hat{\alpha}_i$'s and calculate the residuals $\hat{R}_{it} = Y_{it} - \hat{\alpha}_i - X'_{it}\hat{\beta}$.
3. Obtain $\hat{\gamma}_k$ by the within estimator using $|\hat{R}_{it}|$ as the dependent variable.
4. Estimate $\hat{\delta}_i$ by taking the time average of $|\hat{R}_{it}| - X'_{it}\hat{\gamma}$.
5. Estimate $q(\tau)$ by \hat{q} , which corresponds to the regression quantile of standardized residuals $[\hat{R}_{it}/(\hat{\gamma}_i + X'_{it}\hat{\gamma})]$ upon an intercept term.

Note that Steps 1 and 2 from the MMQREG algorithm are the same as those performed in Canay's estimator. However, the location-scale model used in [mss2019] adds three additional steps that are required to estimate $\hat{\gamma}$ and \hat{q} , so that regression quantile coefficients β_τ are allowed to affect not only the location of the distribution, but also its shape.

Add notes on the assumptions of MMQREG here *if* necessary.

Statistical inference can be performed using the asymptotic distribution of the estimator derived in [mss2019]. Expanding on this literature, [riosavila2024] propose methods for computing alternative standard errors using the empirical influence functions of the estimators. Robust and clustered standard errors can be estimated following this approach, and are readily available as options in the `mmqreg` command in Stata. Another extension of the MMQREG estimator due to [riosavila2024] is to allow for the inclusion of multiple fixed effects using an application of the Frisch-Waugh-Lovell (FWL) theorem to partial out the effect of variables capturing unobserved heterogeneity from both dependent and explanatory variables. Likewise, the command `mmqreg` allows for multiple fixed effects, as will be shown next.

Table 4: Government spending and coalition government - MMQREG Estimator

VARIABLES	(1) q25	(2) q50	(3) q75
mcoalition	2.690***	2.313**	1.703
	(1.043)	(0.950)	(1.207)
	0.259***	0.258**	0.255
	(0.0975)	(0.114)	(0.173)
	-9.875	-10.58	-11.73
	(6.454)	(8.887)	(14.78)
	0.934**	0.896**	0.835
	(0.372)	(0.386)	(0.539)
	-	0.518***	0.875***
	0.299**	(0.175)	(0.285)
	(0.131)	0.139	0.266
	0.0615	(0.243)	(0.338)
	(0.239)	181.9	203.3
	168.8	(145.9)	(241.8)
	(106.6)		
Observations	133	133	133
Country FE	Yes	Yes	Yes
Period FE	Yes	Yes	Yes
Robust	Yes	Yes	Yes
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

As we allow for both location and scale shifts in our estimation method, the monotonic pattern observed in the first row of table 3 is reversed. Now, the farther we move from the median of the distribution in the direction of greater government spending, the lower is the effect of a permanent change from single-party to coalition upon spending. The median coefficient, once again, remains close to the mean coefficient (2.36) under MMQREG estimation. Results from table 4 support the conclusions from [persson2007] regarding types of government and its spending during the last year of legislature, expanding the results to the response of the distribution of spending upon changes in explanatory variables.

We must also note that statistical inference —with robust standard errors displayed

in table 4— for the [persson2007] dataset show that many coefficients that with the previous estimators were not statistically different from zero, are now significant even at the 99% confidence level. This is due to the ability to compute robust standard errors for the MMQREG estimator, contrasting with the CRE and Canay estimators, for which we estimated the covariance matrix using the bootstrap. Although not displayed here, our findings are robust to clustering standard errors for country so that the coefficients of 25th and 50th quantiles corresponding to the indicator variable for coalition government remain statistically significant.

6 Conclusions

About the authors