

# **Quantile Response and Panel Data**

**Manuel Arellano**  
CEMFI

Africa Region Training Workshop  
Econometric Society

Lusaka, July 22, 2015

## Introduction

- In this lecture I provide an introduction to quantile regression and discuss three empirical applications of quantile techniques to panel data.
- Quantile regression is a useful tool for studying conditional distributions.
- The application of quantile techniques to panel data is interesting because it offers opportunities for identifying nonlinear models with unobserved heterogeneity and relaxing exogeneity assumptions.
- Importantly it also offers the opportunity to consider conceptual experiments richer than a static cross-sectional treatment, such as dynamic responses.

## Introduction (continued)

- The first application looks at the effect of child maturity on academic achievement using group data on students and their schools.
- The second application examines the effect of smoking during pregnancy on the birthweight of children.
- The third application examines the persistence of permanent income shocks in a nonlinear model of household income dynamics.
- The applications are based on the results of joint research:
  - Arellano and Weidner (2015)
  - Arellano and Bonhomme (2015)
  - Arellano, Blundell, and Bonhomme (2015).

## **Part 1**

### **Quantile regression**

## Conditional quantile function

- Econometrics deals with relationships between variables involving unobservables.
- Consider an empirical relationship between two variables  $Y$  and  $X$ .
- Suppose that  $X$  takes on  $K$  different values  $x_1, x_2, \dots, x_K$  and that for each of those values we have  $M_k$  observations of  $Y$ :  $y_{k1}, \dots, y_{kM_k}$ .
- If the relationship between  $Y$  and  $X$  is exact, the values of  $Y$  for a given value of  $X$  will all coincide, so that we could write

$$Y = q(X).$$

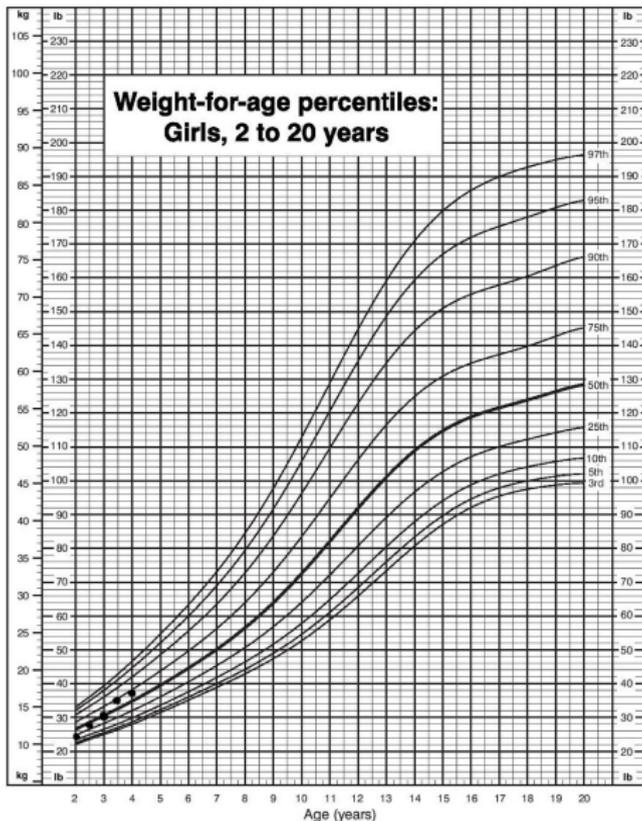
- However, in general units having the same value of  $X$  will have different values of  $Y$ .
- Suppose that  $y_{k1} \leq y_{k2} \leq \dots \leq y_{kM_k}$ , so the fraction of observations that are less than or equal to  $y_{km}$  is  $u_{km} = m/M_k$ .
- It can then be said that a value of  $Y$  does not only depend on the value of  $X$  but also on the rank  $u_{km}$  of the observation in the distribution of  $Y$  given  $X = x_k$ .
- Generalizing the argument:

$$Y = q(X, U)$$

## Conditional quantile function (continued)

- The distribution of the ranks  $U$  is always the same regardless of the value of  $X$ , so that  $X$  and  $U$  are statistically independent.
- Also note that  $q(x, u)$  is an increasing function in  $u$  for every value of  $x$ .
- An example is a growth chart where  $Y$  is body weight and  $X$  is age (Figure 1).
- In this example  $U$  is a normalized unobservable scalar variable that captures the determinants of body weight other than age, such as diet or genes.
- The function  $q(x, u)$  is called a conditional quantile function.
- It contains the same information as the conditional cdf (it is its inverse), but is in the form of a statistical equation for outcomes that may be related to economic models.
- $Y = q(X, U)$  is just a statistical statement: e.g. for  $X = 15$  and  $U = 0.5$ ,  $Y$  is the weight of the median girl aged 15, but one that can be given substantive content.

## CDC Growth Charts: United States



Published May 30, 2000.

SOURCE: Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).



SAFER • HEALTHIER • PEOPLE™

### *Quantile function of normal linear regression*

- If the distribution of  $Y$  conditioned on  $X$  is the normal linear regression model of elementary econometrics:

$$Y = \alpha + \beta X + V \text{ with } V | X \sim \mathcal{N}(0, \sigma^2),$$

the variable  $U$  is the rank of  $V$  and it is easily seen that

$$q(x, u) = \alpha + \beta x + \sigma \Phi^{-1}(u)$$

where  $\Phi(\cdot)$  is the standard normal cdf.

- In this case all quantiles are linear and parallel, a situation that is at odds with the growth chart example.

## Linear quantile regression (QR)

- The linear QR model postulates linear dependence on  $X$  but allows for a different slope and intercept at each quantile  $u \in (0, 1)$

$$q(x, u) = \alpha(u) + \beta(u)x \quad (1)$$

- In the normal linear regression  $\beta(u) = \beta$  and  $\alpha(u) = \alpha + \sigma\Phi^{-1}(u)$ .
- In linear regression one estimates  $\alpha$  and  $\beta$  by minimizing the sum of squares of the residuals  $Y_i - a - bX_i$  ( $i = 1, \dots, n$ ).
- In QR one estimates  $\alpha(u)$  and  $\beta(u)$  for fixed  $u$  by minimizing a sum of absolute residuals where (+) residuals are weighted by  $u$  and (-) residuals by  $1 - u$ .
- Its rationale is that a quantile minimizes expected asymmetric absolute value loss.
- For the median  $u = 0.5$ , so estimates of  $\alpha(0.5)$ ,  $\beta(0.5)$  are least absolute deviations.
- All observations are involved in determining the estimates of  $\alpha(u)$ ,  $\beta(u)$  for each  $u$ .
- Under random sampling and standard regularity conditions, sample QR coefficients are  $\sqrt{n}$ -consistent and asymptotically normal.
- Standard errors can be easily obtained via analytic or bootstrap calculations.
- The popularity of linear QR is due to its computational simplicity: computing a QR is a linear programming problem (Koenker 2005).

## Linear quantile regression (QR) (continued)

- One use of QR is as a technique for describing a conditional distribution. For example, QR is a popular tool in wage decomposition studies.
- However, a linear QR can also be seen as a semiparametric random coefficient model with a single unobserved factor:

$$Y_i = \alpha(U_i) + \beta(U_i) X_i$$

where  $U_i \sim \mathcal{U}(0, 1)$  independent of  $X_i$ .

- For example, this model determines log earnings  $Y_i$  as a function of years of schooling  $X_i$  and ability  $U_i$ , where  $\beta(U_i)$  represents an ability-specific return to schooling.
- This is a model that can capture interactions between observables and unobservables.
- A special case of model with an interaction between  $X_i$  and  $U_i$  is the heteroskedastic regression  $Y | X \sim \mathcal{N}[\alpha + \beta X, (\sigma + \gamma X)^2]$ .
  - In this case  $\alpha(u) = \alpha + \sigma\Phi^{-1}(u)$  and  $\beta(u) = \beta + \gamma\Phi^{-1}(u)$ .
- As a model for causal analysis, linear QR faces similar challenges as ordinary linear regression. Namely, linearity, exogeneity and rank invariance.
- Let us discuss each of these aspects in turn.

## Flexible QR

- Linearity is restrictive. It may also be at odds with the monotonicity requirement of  $q(x, u)$  in  $u$  for every value of  $x$ .
- Linear QR may be interpreted as an approximation to the true quantile function (Angrist, Chernozhukov, and Fernández-Val 2006).

- An approach to nonparametric QR is to use series methods:

$$q(x, u) = \theta_0(u) + \theta_1(u)g_1(x) + \dots + \theta_P(u)g_P(x).$$

- The  $g$ 's are anonymous functions without an economic interpretation. Objects of interest are derivative effects and summary measures of them.
- In practice one may use orthogonal polynomials, wavelets or splines (Chen 2007).
- This type of specification may be seen as an approximating model that becomes more accurate as  $P$  increases, or simply as a parametric flexible model of the quantile function.
- From the point of view of computation the model is still a linear QR, but the regressors are now functions of  $X$  instead of the  $X$ s themselves.

## Exogeneity and rank invariance

- To discuss causality it is convenient to use a single 0 – 1 binary treatment  $X_i$  and a potential outcome notation  $Y_{0i}$  and  $Y_{1i}$ .
- Let  $U_{0i}, U_{1i}$  be ranks of potential outcomes and  $q_0(u), q_1(u)$  the quantile functions.
- Note that unit  $i$  may be ranked differently in the distributions of the two potential outcomes, so that  $U_{0i} \neq U_{1i}$ . The causal effect for unit  $i$  is given by

$$Y_{1i} - Y_{0i} = q_1(U_{1i}) - q_0(U_{0i}).$$

- Under exogeneity  $X_i$  is independent of  $(Y_{0i}, Y_{1i})$ .
- The implication is that the quantile function of  $Y_i | X_i = 0$  coincides with  $q_0(u)$  and the quantile function of  $Y_i | X_i = 1$  coincides with  $q_1(u)$ , so that

$$\beta(u) = q_1(u) - q_0(u).$$

- This quantity is often called a quantile treatment effect (QTE). In general it is just the difference between the quantiles of two different distributions.
- It will only represent the gain or loss from treatment of a particular unit under a rank invariance condition. i.e. that the ranks of potential outcomes are equal to each other.
- Under rank invariance treatment gains may still be heterogeneous but a single unobservable variable determines the variation in the two potential outcomes.
- Next we introduce IV endogeneity in a quantile model with rank invariance.

## Instrumental variable QR

- The linear instrumental variable (IV) model of elementary econometrics assumes

$$Y_i = \alpha + \beta X_i + V_i$$

where  $X_i$  and  $V_i$  are correlated, but there is an instrumental variable  $Z_i$  that is independent of  $V_i$  and a predictor of  $X_i$ .

- Potential outcomes are of the form  $Y_{x,i} = \alpha + \beta x + V_i$  so that rank invariance holds.
- If  $x$  is a 0 – 1 binary variable,  $Y_{0,i} = \alpha + V_i$  and  $Y_{1,i} = \alpha + \beta + V_i$ .
- A QR generalization subject to rank invariance is to consider

$$Y_{x,i} = q(x, U_i).$$

- A linear version of which is

$$Y_{x,i} = \alpha(U_i) + \beta(U_i) x.$$

## Instrumental variable QR (continued)

- Chernozhukov and Hansen (2006) propose to estimate  $\alpha(u)$  and  $\beta(u)$  for given  $u$  by directly exploiting the IV exclusion restriction.
- Specifically, if we write the model as

$$Y_i = \alpha(U_i) + \beta(U_i) X_i + \gamma(U_i) Z_i,$$

the IV assumption asserts that  $Z_i$  only affects  $Y_i$  via  $X_i$  so that  $\gamma(u) = 0$  for each  $u$ .

- Now let  $\hat{\gamma}_u(b)$  be the estimated slope coefficient in a  $u$ -quantile regression of  $(Y_i - bX_i)$  on  $Z_i$  and a constant term.
- The idea, which mimics the operation of 2SLS, is to choose as estimate of  $\beta(u)$  the value of  $b$  that minimizes  $|\hat{\gamma}_u(b)|$ , hence enforcing the exclusion restriction.
- In the absence of rank invariance the treatment effects literature (e.g. Abadie 2003) has focused on QTEs for compliers in the context of a binary treatment that satisfies a monotonicity assumption.

## **Part 2**

### **QR with fixed effects in large panels**

## Basics

- The most popular tool in panel data analysis is a linear regression model with common slope parameters and individual specific intercepts:

$$Y_{it} = \beta X_{it} + \alpha_i + V_{it} \quad (i = 1, \dots, N; t = 1, \dots, T),$$

in which  $X_i = (X_{i1}, \dots, X_{iT})$  is independent of  $V_{it}$  but possibly correlated with  $\alpha_i$ .

- This is seen as a way of allowing for a special form of non-exogeneity (fixed-effect endogeneity) but also a way of introducing heterogeneity and persistence.
- The estimator of  $\beta$  is OLS including individual dummies, or equivalently OLS of  $Y$  on  $X$  in deviations from individual-specific means (within-group estimation).
- Observations may be from actual panel data, in which units are followed over time, or from data with a group structure, in which case  $i$  denotes groups and  $T$  is group size.
- In practice group size will be group specific ( $T_i$ ) and techniques will be adapted accordingly.

## QR with fixed effects

- A QR version of the within-group model specifies

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

where  $U_{it} \sim \mathcal{U}(0, 1)$  independent of  $X_i$  and  $\alpha_i(\cdot)$ .

- The term  $\alpha_i(U_{it})$  can be regarded as a function of  $U_{it}$  and a vector  $W_i$  of unobserved individual effects of unspecified dimension:  $\alpha_i(U_{it}) = r(W_i, U_{it})$ .
- Thus, the model allows for multiple individual characteristics that affect differently individuals with different error rank  $U_{it}$ .
- For example, there may be a multiplicity of school characteristics, some of which are only relevant determinants of academic achievement for high ability students while others are only relevant for low ability students.
- In QR one estimates  $\beta(u)$  and  $\alpha_1(u), \dots, \alpha_N(u)$  for fixed  $u$ .
- The large sample properties of these estimates are those of standard QR if  $T$  is large in absolute terms and relative to  $N$ .
- However, if  $T$  is small relative to  $N$  or if  $T$  and  $N$  are of similar size, estimates of the common parameter  $\beta(u)$  may be biased or even underidentified.
- The reason is too much sample noise due to estimating too many parameters relative to sample size. This situation is known as the incidental parameter problem.

## Dealing with incidental parameters: fixed $T$ and large $T$ approaches

- In the static linear model, within-group estimates of the slope parameter are free from incidental parameter biases, but in nonlinear models the opposite is true in general.
- In situations where  $T$  is very small relative to  $N$  one reaction is to consider models and estimators of those models that are fixed- $T$  consistent for large  $N$ .
- An example is the second application on the effect of smoking on birthweight, which uses a sample of  $N = 12360$  women with  $T = 3$  children each.
- There are also panels in which  $T$  is not negligible and not negligible relative to  $N$ , even if  $N$  still is much larger than  $T$ .
- An example, is the dataset in our first application that contains  $N = 389$  schools with an average of  $\bar{T} = 40$  students per school.
- An alternative approach in those situations has been to approximate the sampling distribution of the fixed effects estimator as  $T/N$  tends to a constant.
- For smooth objective functions this approach leads to a bias correction that can be easily implemented by analytical or numerical methods.
- A simple implementation is Jackknife bias correction (delete-one Jackknife in Hahn and Newey 2004; split-panel Jackknife in Dhaene and Jochmans 2015).

## Bias reduction in QR

- The existing techniques are not applicable to QR due to the non-smoothness of the sample moment conditions of quantile models.
- Arellano and Weidner (2015) characterize the incidental parameter bias of QR and instrumental-variable QR estimators.
- They also find bias correcting moment functions that are first-order unbiased, that is, whose expected value is of order  $1/T^2$ .
- Moment functions within their class depend on the choice of a weight sequence. Some weight sequences are bias reducing while others are not.
- They uncover a bias-variance trade-off when attempting to correct bias, and provide bias corrected estimators that balance this trade-off.
- Interestingly their discussion of bias correction around choice of weight sequence is similar to bias reduction in nonparametric Kernel regression.
- Arellano and Weidner show that delete-one Jackknife is not first-order bias correcting in QR due to the fact that the second-order bias has a non-standard structure.
- They find that a permutation-invariant version of split-panel Jackknife is bias-correcting and exhibits good variance properties.

## Interpreting the incidental parameter bias

- Arellano and Weidner (2015) find that the leading-order bias term vanishes in the special case where  $\beta(u) = \beta$  is constant over  $u$ .
- This result is of limited interest if the goal is to estimate nonlinear models, although it may be useful in testing for linearity.
- They also provide an approximation to the leading order bias in the case where  $\beta(u)$  is almost constant, so that  $\beta(u) - \bar{\beta}$  is small.
- Under this approximation the leading order bias can be interpreted as resulting from measuring  $\beta(u)$  at the wrong quantile  $u + \Delta u$  and from smoothing out  $\beta(u)$  around this wrong quantile with a density whose standard deviation shrinks at the rate  $T^{-1/2}$ .
- The implication is that the incidental parameter bias would tend to average effects across quantiles.

## The effect of child maturity on academic achievement

- Arellano and Weidner study the effect of age on academic achievement of school children following Bedard and Dhuey (2006).
- Bedard and Dhuey consider multiple countries and students of different age groups. Their question is whether initial maturity differences in kindergarten and primary school have long-lasting effects.
- Here we only consider data from Canada for third and fourth graders (9 year old in 1995) from the Trends in International Mathematics and Science Study (TIMSS).
- There are 389 schools with an average of 40 students per school. Therefore, it is an unbalanced pseudo-panel or dataset with a group structure.
- The outcome variable is the math test score of student  $t$  in school  $i$  normalized to have mean 50 and standard deviation 10 over the whole sample.
- The main regressor is observed age measured in months.
- Age is potentially endogenous because of grade retention and early or late school enrolment (which are not observed).

## The effect of child maturity on academic achievement (continued)

- Following Bedard and Dhuey we use age relative to the school cutoff date to instrument for age.
- The school cutoff date in Canada is January 1. So we define relative or assigned age as  $z = 0$  for children born in December and  $z = 11$  for children born in January.
- Relative age is a strong instrument.
- We only require exogeneity of relative age conditional on school effects, which for example will capture the age distribution at school level.
- Quantile analysis is interesting, because age effects might be different for low- and high-performing students.
- Whether maturity and academic ability are substitutes or complements is an empirical question that may have implications for school policy.
- Controlling for school fixed effects turns out to be important for the results. Age composition may vary across schools, so age is likely fixed-effect endogenous.

Table 1  
 Effect of Age on Math Test Scores at 3rd & 4th Grade  
 Canadian TIMSS 15549 students  $N = 394$  schools

| OLS     | IV      | OLS+FE  | IV+FE    |
|---------|---------|---------|----------|
| 0.017   | 0.184   | -0.0332 | 0.178    |
| (0.010) | (0.026) | (0.009) | (0.0241) |

Number in brackets are standard errors

IV uses assigned age to instrument for observed age

Controls: sex, grade, rural, mother native, father-native both parents, calculator, computer, +100books, hh size

std(Y)=10, i.e. age effect of 0.18 is a 1.8% st dev per month effect or 22% st deviations per year

- Table 1 reproduces results in Bedard and Dhuey (2006).
- IV estimates with and without school fixed effects are very similar, i.e. the instrument appears to be uncorrelated with school effects.

Table 2  
 Effect of Age on Math Test Scores at 3rd & 4th Grade  
 Quantile IV, no fixed effects

| $u = 0.1$ | $u = 0.3$ | $u = 0.5$ | $u = 0.7$ | $u = 0.9$ |
|-----------|-----------|-----------|-----------|-----------|
| 0.14      | 0.16      | 0.18      | 0.24      | 0.19      |
| (0.01)    | (0.01)    | (0.01)    | (0.07)    | (0.03)    |

IV uses assigned age to instrument for observed age  
 Controls: sex, grade, rural, mother native, father-native  
 both parents, calculator, computer, +100books, hh size

- Without controlling for school fixed effects, one finds a significant difference in age effects across quantiles.
- Age effects are increasing.
- The results in Table 2 would point to maturity and ability as complements in the production of test scores.

Table 3  
Effect of Age on Math Test Scores at 3rd & 4th Grade  
Quantile IV with fixed effects, no bias correction

| $u = 0.1$  | $u = 0.3$ | $u = 0.5$ | $u = 0.7$ | $u = 0.9$ |
|--|-----------|-----------|-----------|-----------|
| 0.18   | 0.15      | 0.18      | 0.19      | 0.16      |
| (0.05)   | (0.03)    | (0.03)    | (0.04)    | (0.04)    |
| IV uses assigned age to instrument for observed age<br>Controls: sex, grade, rural, mother native, father-native<br>both parents, calculator, computer, +100books, hh size |           |           |           |           |

- Table 3: Once we control for school fixed effects, we do not find a significant difference in age effects across quantiles.
- Age effects are relatively constant in  $u$ . But is this because there is really no effect, or because the incidental parameter bias tends to average effects across quantiles?

Table 4  
 Effect of Age on Math Test Scores at 3rd & 4th Grade  
 Quantile IV with fixed effects, bias correction

| $u = 0.1$  | $u = 0.3$ | $u = 0.5$ | $u = 0.7$ | $u = 0.9$ |
|--|-----------|-----------|-----------|-----------|
| 0.21   | 0.15      | 0.18      | 0.18      | 0.09      |
| (0.05)   | (0.03)    | (0.04)    | (0.04)    | (0.05)    |
| IV uses assigned age to instrument for observed age<br>Controls: sex, grade, rural, mother native, father-native<br>both parents, calculator, computer, +100books, hh size |           |           |           |           |

- Table 4: After bias correction age effects are decreasing in  $u$ .
- There seems to be evidence that maturity and ability are substitutes in academic achievement.

## **Part 3**

### **QR with random effects in short panels**

## *Dimensionality reduction of fixed effects*

- Application of QR with fixed effects is straightforward as it proceeds in a quantile-by-quantile fashion allowing for a different fixed effect at each quantile.
- However, in short panels the incidental parameter problem is a challenge.
- Moreover, while being agnostic about the number of the unobserved group factors affecting outcomes is attractive, sometimes substantive reasons suggest that only a small number of underlying factors play a role.
- Whether one uses a quantile model with a different individual effect at each quantile or a model with a small number of unobserved effects also has implications for identification.
- Rosen (2010) shows that a fixed-effects model for a single quantile may not be point identified.
- Arellano and Bonhomme (2015) show that a QR model with a scalar fixed effect is nonparametrically identified in panel data with  $T = 3$  subject to completeness assumptions (Newey and Powell 2003; Hu and Schennach 2008).

## Flexible quantile modelling with random effects

- Arellano and Bonhomme aim to estimate models of the form:

$$Y_{it} = \beta(U_{it}) X_{it} + \gamma(U_{it}) \eta_i + \alpha(U_{it}) \quad (2)$$

where  $U_{it} \sim \mathcal{U}(0, 1)$  independent of  $X_i$  and  $\eta_i$ , but  $X_i$  and  $\eta_i$  may be correlated.

- Model (2) is a special case of a series-based specification that allows for nonlinearities and interactions between  $X_{it}$  and  $\eta_i$ :

$$Y_{it} = \sum_{k=1}^{K_1} \theta_k(U_{it}) g_k(X_{it}, \eta_i) \quad (3)$$

- The dependence of  $\eta_i$  on  $X_i$  is also specified as a flexible quantile model:

$$\eta_i = \sum_{k=1}^{K_2} \delta_k(V_i) h_k(X_i) \quad (4)$$

where  $V_i$  is a uniform random variable independent of  $U_{it}$  and  $X_{it}$  for all  $t$ .

- This is a correlated random effects approach in the sense that a model for the dependence between  $\eta_i$  and  $X_i$  is specified.
- However, it is more flexible than alternative specifications in the literature and can be seen as an approximation to the conditional quantile function as  $K_2$  increases.
- If  $\eta_i$  is a vector of individual effects a triangular structure is assumed in place of (4).

## Simulation-based estimation

### *Basic intuition behind the Arellano and Bonhomme method*

- If  $\eta_i$  were observed, one would simply run an ordinary QR of  $Y_{it}$  on  $X_{it}$  and  $\eta_i$ .
- But since  $\eta_i$  is not observed they construct some imputations, say  $M$  imputed values  $\eta_i^{(m)}$ ,  $m = 1, \dots, M$  for each individual in the panel. Having got those, one can get estimates by computing a QR averaged over imputed values.
- For the imputed values to be valid they have to be draws from the distribution of  $\eta_i$  conditioned on the data, which depends on the parameters to be estimated ( $\theta$ 's and  $\delta$ 's in the flexible model).
- This is therefore an iterative approach.
- They start by selecting initial values for a grid of conditional quantiles of  $Y_{it}$  and  $\eta_i$ , which then allows them to generate imputes of  $\eta_i$ , which can be used to update the quantile parameter estimates and so on.
- To deal with the complication that  $\theta_k(u)$  and  $\delta_k(v)$  are functions, they use a finite-dimensional approximation to those functions based on interpolating splines with  $L$  knots (similar to Wei and Carroll 2009).
- The resulting method is a stochastic EM algorithm.

## Simulation-based estimation (continued)

### *Stochastic EM algorithm*

- A difference with most applications of EM algorithms is that parameters are not updated in each iteration using maximum likelihood but QR.
- This is important because once imputes for  $\eta_i$  are available, QR estimates can be calculated in a quantile-by-quantile fashion, which together with the convexity of QR minimization make each parameter update fast and reliable.
- Arellano and Bonhomme obtain the asymptotic properties of the estimator based on the stochastic EM algorithm for a fixed number of draws  $M$  in the case where the parametric model is assumed correctly specified (extending results in Nielsen 2000).
- That, is  $K_1$ ,  $K_2$  and  $L$  are held fixed as  $N$  tends to infinity for fixed  $T$ .
- They also establish consistency as  $K_1$ ,  $K_2$  and  $L$  tend to infinity with  $N$  in the large- $M$  limit.

### *Other approaches*

- Other recent approaches to quantile panel data models include Chernozhukov, Fernández-Val, Hahn & Newey (2013), and Graham, Hahn, Poirier & Powell (2015).
- These approaches are non-nested with the previous model and will recover different quantile effects.

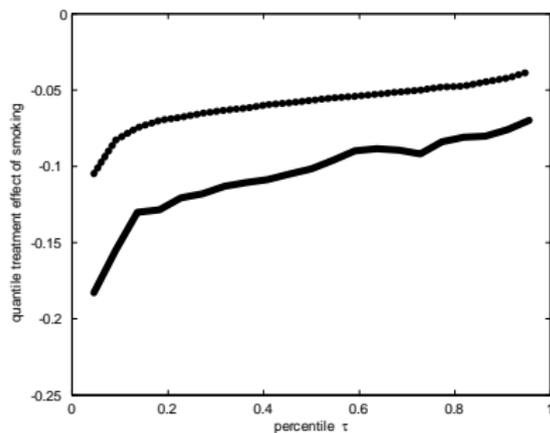
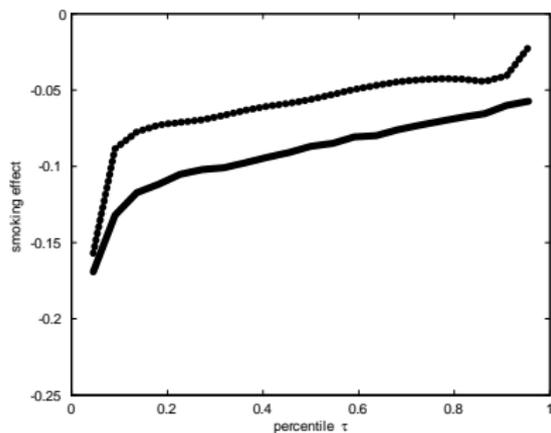
## The effect of smoking on birth weight

- We revisit the effect of maternal inputs on children's birth outcomes. Specifically, we study the effect of smoking during pregnancy on children's birthweights.
- Abrevaya (2006) uses a mother-FE approach to address endogeneity of smoking.
- We use QR with mother-specific effects to allow for both unobserved heterogeneity and nonlinearities in the relationship between smoking and birthweight.
- We use a balanced subsample from the US natality data used in Abrevaya (2006), which comprises 12360 women with 3 children each. Our outcome is log-birthweight.
- The main covariate is a binary smoking indicator. Age of the mother and gender of the child are used as additional controls.
- An OLS regression yields a negative point estimate of the smoking coefficient:  $-.095$ . The fixed-effects estimate is also negative, but it is twice as small:  $-.050$  (significant).
- Moreover, running a standard (pooled) QR suggests that the effect of smoking is more negative at lower quantiles of birthweights.
- However, these results might be subject to an endogeneity bias, which may not be constant along the distribution.

## The effect of smoking on birth weight (continued)

- The left graph of Figure 2 shows the smoking coefficient in a pooled QR (solid line), and the REQR estimate of the smoking effect (dashed line).
- REQR estimates use  $L = 21$  knots. The stochastic EM algorithm is run for 100 iterations, with 100 random walk Metropolis-Hastings draws within each iteration.
- Parameter estimates are averages of the 50 last iterations of the algorithm.
- The smoking effect becomes less negative when correcting for time-invariant endogeneity through the introduction of mother-specific fixed-effects.
- At the same time, the effect remains sizable, and is increasing along the distribution.
- The right graph shows the QTE of smoking as the difference in log-birthweight between a sample of smoking women, and a sample of non-smoking women, keeping all other characteristics (observed,  $X_i$ , and unobserved,  $\eta_i$ ) constant.
- This calculation illustrates the usefulness of estimating a complete model of the joint distribution of outcomes and unobservables, to compute counterfactual distributions that take unobserved heterogeneity into account.
- The solid line shows the empirical difference between unconditional quantiles, while the dashed line shows the QTE that accounts for both observables and unobservables.
- The results are broadly in line with those reported on the left graph of Figure 2.

**Figure 2: QR coefficient of smoking and QTE (difference in potential outcomes)**

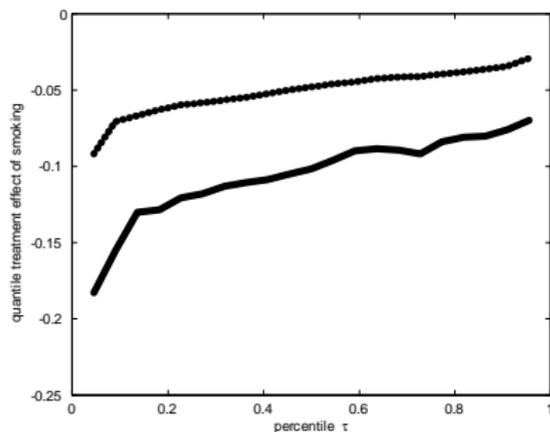
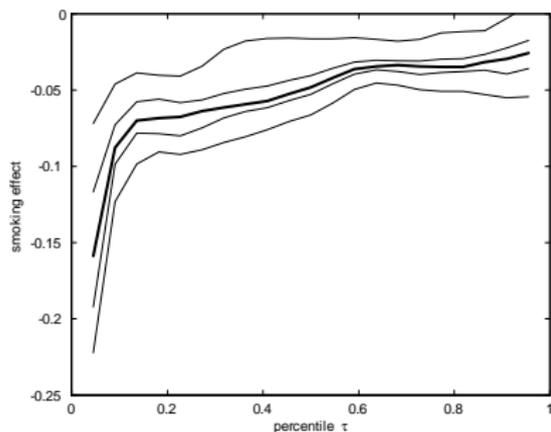


- Data from Abrevaya (2006).
- Left: Solid line is the pooled QR smoking coefficient; dashed line is the panel QR smoking coefficient.
- Right: Solid line is the raw QTE of smoking; dashed line is the QTE estimate based on panel QR.

## QR with smoking interacted with mother heterogeneity and baby heterogeneity

- Lastly, we report the results of an interacted quantile model, where the specification allows for all first-order interactions between covariates and the unobserved mother-specific effect.
- In this model the quantile effect of smoking is mother-specific.
- The results on the right graph in Figure 3 show the unconditional QTE of smoking. Results are similar to the ones obtained for the linear specification.
- However, on the left graph we see substantial mother-specific heterogeneity in the conditional quantile treatment effect of smoking.
- For some mothers smoking appears particularly detrimental to children's birthweight, whereas for other mothers the smoking effect, while consistently negative, is much smaller.
- This evidence is in line with the results of a linear random coefficients model reported in Arellano and Bonhomme (2012).

**Figure 3: Quantile effects of smoking and QTE (interacted specification)**



- Data from Abrevaya (2006).
- Left: lines represent the percentiles .05, .25, .50, .75, and .95 of the heterogeneous smoking effect across mothers, at various percentiles  $u$ .
- Right: Solid line is the raw QTE of smoking; dashed line is the QTE estimate based on panel QR with interactions.

## **Part 4**

### **Dynamic quantile models**

## Autoregressive models and predetermined variables

- The Arellano-Bonhomme approach covers dynamic autoregressive models and models with general predetermined variables of the form:

$$Y_{it} = Q_Y (Y_{i,t-1}, X_{it}, \eta_i, U_{it})$$

- If the  $X$ s are strictly exogenous variables, the quantile model for the individual effect is as before except for the inclusion of the initial outcome variable:

$$\eta_i = Q_\eta (Y_{i1}, X_i, V_i)$$

- In the case of general predetermined variables the model is incomplete.
- To complete the specification a Markov feedback process is assumed:

$$X_{it} = Q_X (Y_{i,t-1}, X_{i,t-1}, \eta_i, A_{it})$$

and the quantile model of the effects is conditioned only on initial values:

$$\eta_i = Q_\eta (Y_{i1}, X_{i1}, V_i)$$

## Models with time-varying unobservables

- The framework also extends to models with time-varying unobservables, such as the following nonlinear permanent-transitory model:

$$Y_{it} = \eta_{it} + V_{it} \quad (5)$$

$$\eta_{it} = Q_Y(\eta_{i,t-1}, U_{it}) \quad (6)$$

where  $V_{it}$  and  $U_{it}$  are i.i.d. distributed.

- Arellano, Blundell and Bonhomme (2014) use a quantile-based approach to document nonlinear relationships between earnings shocks to households and their lifetime profiles of earnings and consumption.
- They estimate model (5)-(6) using PSID household labor income data for the years 1998–2008.

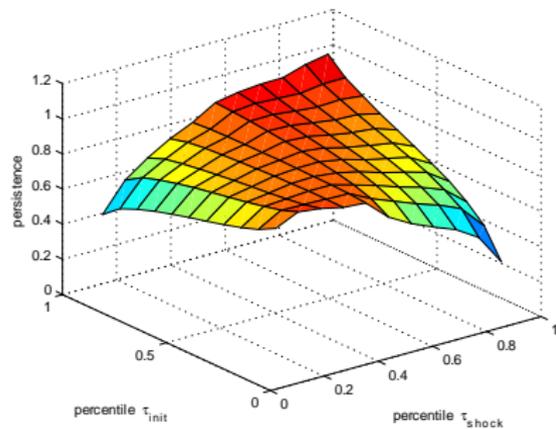
### *Persistence of permanent income shocks*

- Evidence of nonlinearity in the persistence of earnings can be seen from Figure 4.
- This figure plots estimates of the average derivative of the conditional quantile function of current income with respect to lagged income.
- The graphs show strong similarity in the patterns of the nonlinearity of household earnings in the PSID survey data and in the population register data from Norway.
- They also show a clear difference in the impact of past shocks according to the percentile of the shock and the percentile of the past level of income.
- A large positive shock for a low income family or a large negative shock for a high income family appears to reduce the persistence of past shocks.

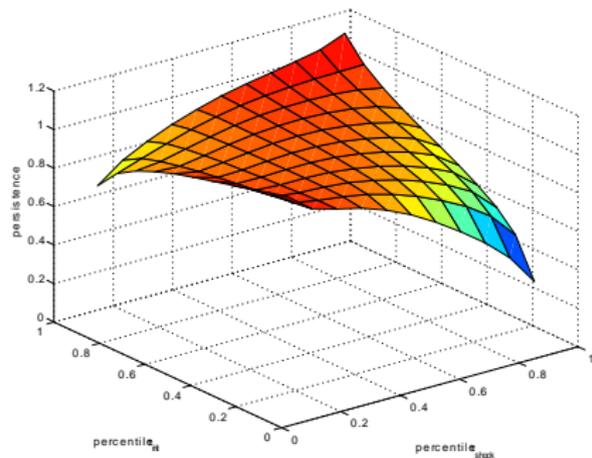
Figure 4: Quantile autoregressions of log-earnings

$$\frac{\partial Q_{y_t|y_{t-1}}(y_{i,t-1}, \tau)}{\partial y}$$

PSID data



Norwegian administrative data



Note: Residuals of log pre-tax household labor earnings, Age 35-65 1999-2009 (US), Age 25-60 2005-2006 (Norway). Estimates of the average derivative of the conditional quantile function of  $y_{it}$  given  $y_{i,t-1}$  with respect to  $y_{i,t-1}$ .

*Persistence of permanent income shocks (continued)*

- Arellano, Blundell, and Bonhomme find that in the central range of the distribution, measured persistence of  $\eta_{i,t-1}$  is of similar magnitude and close to unity, so that the unit root model would be an acceptable description for this part of the distribution.
- However, a very negative shock reduces the persistence of a “positive history” (a positive lagged level of  $\eta$ ) but preserves the persistence of a negative history.
- At the other end, a very positive shock reduces the persistence of a negative history but preserves the persistence of a good history.
- These results suggest a richer view of persistence, away from the conventional unit root versus mean reversion dichotomy, and help explain household consumption behavior.