Comments on "IV quantile regression for group-level treatments, with an application to the distributional effects of trade"
by Denis Chetverikov, Brad Larsen, and Christopher Palmer

Manuel Arellano
CEMFI

Interactions: Bringing Together Econometrics and Applied Microeconomics
University of Chicago

September 25, 2015
Introduction

- This paper discusses econometric methods that relate to a substantial section of applied work, so it is a good choice for an Interactions workshop.

- It provides a practical framework for estimating distributional effects, asymptotic properties for their method, and evidence of computational and empirical applicability.

- A tour de force around a simple theme, all accomplished with great skill.

- I will provide a summary of the paper and some general remarks.
Summary

- Concerned with applications where the interest is in the effect of a group-level policy on the group-level distribution of some individual characteristic.

- An example of their approach is a linear regression of a quantile $q^\tau_g$ on a variable $x_g$:
  \[ q^\tau_g = \gamma(\tau) + \beta(\tau) x_g + \varepsilon_g(\tau) \]
  where $\varepsilon_g(\tau)$ is an error term orthogonal to $x_g$.

- In their setting $x_g$ is observed but $q^\tau_g$ is not. So it is replaced by a sample quantile $\hat{q}^\tau_g$ obtained from individual-level data (of size $N_g$).

- Since $\hat{q}^\tau_g$ is not an unbiased estimate of $q^\tau_g$, consistency for $\beta(\tau)$ of a regression of $\hat{q}^\tau_g$ on $x_g$ requires large $G$ and large $N_1, \ldots, N_G$.

- In an IV version of the problem $\varepsilon_g(\tau)$ is correlated with $x_g$ but not with an instrument $w_g$ that satisfies the rank and exclusion conditions.

- In a more general version $\hat{q}^\tau_g$, instead of a sample quantile, is a sample QR coefficient (intercept or slope) involving individual-level covariates.

- The paper provides an asymptotic normality result for $\sqrt{G} \left( \hat{\beta}(.) - \beta(.) \right)$ as long as $N_g$ grows sufficiently fast as $G \to \infty$ (a mild requirement).

- It also provides an estimator of the asymptotic covariance function.

- These are all useful results.
1. Group-level causality and individual-level causality

- A natural context for the current framework is group-level causality.

- A formulation for a population of groups is as follows. A potential outcome is a random function (a cdf) rather than a r.v. and treatment takes place at group-level.

- Larsen (2014) and Autor, Dorn, and Hanson (2013)’s extension are nice examples.

- In Larsen a group is a state-year cell, treatment is a teacher certification law, and a potential outcome is the cdf of teacher quality in a state-year under some licensing law.

- Group-level issues are eg the exogeneity of licensing laws across states & years, or possible spillover effects. But the group-level perspective is silent about what goes on at the individual level.

- One could think of each group as a separate market for teacher quality, the equilibrium outcome of which is a cdf of teacher quality that is affected by licensing laws through supply and demand channels.
Individual-level causality

- The same model can also be regarded as a model for an individual potential outcome:

\[ y_{ig} = \gamma(u_{ig}) + \beta(u_{ig})x_g + \varepsilon(u_{ig}, \eta_g) \]

where \( u_{ig} \sim U(0,1) \) indep. of \((x_g, \eta_g)\) and \( \varepsilon_g(\tau) = \varepsilon(\tau, \eta_g) \) for arbitrary dim(\( \eta_g \)).

- \( y_{ig} \) is the individual outcome whose group quantile function is \( q^\tau_g \) (teacher’s i quality).

- There is a single individual unobservable \( u_{ig} \) and many group unobservables \( \eta_g \).

- \( x_g \) is always exogenous w.r.t. to \( u_{ig} \) but may or may not be exogenous w.r.t. \( \eta_g \).

- As a model of individual potential outcomes, \( \beta(u_{ig}) \Delta x \) would measure the causal effect of a change \( \Delta x \) on individual’s i outcome.

- The model assumes comonotonicity (rank invariance) at individual and group levels.

- This is an unappealing model of individual response in applications such as those in Larsen or ADH where occupational entry/exit is a relevant aspect of the response.
Point-wise comparison of quantiles

- A point-wise comparison of quantiles gives the shape of individual treatment gains under comonotonicity, but it is not an obvious metric in a group-level comparison of distributions from different populations.

- The function $\widehat{\beta}(\tau)$ is not necessarily an interpretable distributional treatment effect at group level, so is not to be taken for granted as the focus of empirical reporting.

- It may be more natural to look at changes in distributional measures motivated in substantive considerations, such as inequality indices, polarization, or probabilities of exceeding a preestablished threshold.

- A situation where the interest is in comparing distributions of outcomes of different populations, rather than in the distribution of individual changes in a fixed population.
2. Reducing the dimensionality of group effects

- There is a trade-off between allowing for large or low dimensional unobservable group-effects and the scope of nonparametric identification.

- Application of the present method is straightforward as it proceeds in a quantile-by-quantile fashion allowing for a different error at each quantile.

- However, if the number of individual observations per group is small the incidental parameter problem is a challenge.

- Moreover, while being agnostic about the group-factor dimension is attractive, often substantive knowledge suggests that only a small no. of underlying factors play a role.

- Whether one uses a model with a different group effect at each quantile or one with a small number of group effects may have implications for fixed-$N_g$ identification.

- For example, Rosen (2010) shows that a fixed-effects panel model for a single quantile is not point identified.
Group-level analysis when the number of observations per group is small

- Arellano & Bonhomme (2013) show that a QR model with a scalar group effect is nonparametrically identified in panel data with $N_g = 3$ under completeness conditions.

- They consider the fixed-$N_g$ identification and estimation of functions $Q_{yi}$ and $Q_\alpha$:
  
  $$y_{ig} = Q_{yi}(z_{ig}, \alpha_g, u_{ig})$$
  
  $$\alpha_g = Q_\alpha(z_g, v_g)$$

  where $\alpha_g$ is a group-effect, $z_g = \{z_{ig}\}_{i=1}^{N_g}$, and $u_{ig} \mid (z_g, \alpha_g)$ and $v_g \mid z_g$ are $\mathcal{U}(0, 1)$.

- A centered measure of location on the pdf of $y_{ig} \mid z_{ig}, \alpha_g$ for some $i$ is imposed.

- If $z_{ig}$ includes an $i$-invariant $x_g$, the result may hold for a reparameterization that subsumes $x_g$.

- Exploring conditions (such as within variability in $Q_{yi}$) under which derivative effects of $Q_{yi}$ w.r.t. $x_g$ can be disentangled is an interesting question.
Group-level analysis when $N_g$ is small (continued)

- If the interest is in individual-level effects, $Q_{yi}$ is the main response function and $Q_{\alpha}$ is a nuisance function.

- If the interest is in group-level effects, $Q_{yi}$ is an aggregator that produces the factors $\alpha_g$ and $Q_{\alpha}$ is a group-level response function.

- In the IV situation, a similar reinterpretation of the micro setup conditionally on $(x_g, w_g)$ leads to identification of the joint density of $(\alpha_g, x_g, w_g)$.

- The question here is how to control the small-$N_g$ noise in a nonparametric way so that one can still say something about the effect of $x_g$ on the group-level cdf of $y_{ig}$.
3. Standard errors

- Chamberlain (1994) considered a version of the estimator in this paper motivated in the distributional analysis of censored earnings data.
- He analyzed the properties of $\hat{\beta}(\tau)$ when $G$ is fixed, $N_g$ is large, and there are no group-level unobservables (except for model misspecification).
- Chamberlain’s standard errors are the mirror image of those considered here.
- In his case all sampling error comes from the discrepancy between $\hat{q}_g^\tau$ and $q_g^\tau$ whereas this is ignored in the asymptotics here leading to standard errors driven by $\epsilon_g(\tau)$.
- Chamberlain’s situation is similar to DiD case-studies where $G$ and/or the number of treated groups are small.
- A reinterpretation of $\hat{\beta}(\tau)$ in those cases is as an estimate of the infeasible fixed-$G$ sample statistic that uses group-level population quantiles.
- Desirable to report standard errors that are robust to alternative asymptotic plans.
- For example, standard errors for fixed-$N_g$ pseudo-true values that retain double asymptotic validity.
- Ignoring the first stage may come at a cost in finite samples.
4. Hausman-Taylor internal instruments

- The setting for the Hausman-Taylor estimator is a model of the form
  \[ y_{ig} = \gamma z_{ig} + \beta x_g + \eta_g + v_{ig} \]
  where \( z_{ig} \) is uncorrelated with \( \eta_g \) but \( x_g \) is not, and both are \( v_{ig} \)-exogenous.

- The idea is to use \( z_{ig} \) as an instrument for itself and \( \bar{z}_g \) as an instrument for \( x_g \).

- The method works as long as \( \bar{z}_g \) is correlated with the observed component of the group-effect (\( x_g \)) but not with the unobserved one (\( \eta_g \)).

- In panel data there are not many applications of this idea due to difficulty in finding time-varying covariates that can be thought a priori as being fixed-effect exogenous.

- A related common practice is to look at predictive effects on group-level unobservables, be intercepts or slopes.

- This paper emphasizes the benefits from being able to use internal instruments, but I was under the impression that they did not feature prominently in their examples.

- Perhaps they have in mind exploiting the micro-data to construct aggregate instruments more generally, but it would be nice to see more examples where internal instruments determine the empirical force of the results.
Concluding remarks

- In group-level comparisons of potential cdf-outcomes is not obvious that we want to focus on point-wise comparisons of quantiles, which are reminiscent of comonotonic individual-level effects.

- There is a trade-off between double asymptotic approaches with high dimensional group-level unobservables and approaches with fewer group-level unobservables that deliver nonparametric identification when the number of units per group is small.

- Standard errors that exhibit robustness to alternative asymptotic plans are attractive in applications that combine group-level data and micro data.