Distribution regression methods

Philippe Van Kerm

CEPS/INSTEAD, Luxembourg
philippe.vankerm@ceps.lu

Ninth Winter School on Inequality and Social Welfare Theory
“Public policy and inter/intra-generational distribution”

January 13–16 2014, Alba di Canazei, Italy

Outline

- Non-technical toolbox review: quantile regression, distribution regression, parametric income distribution models, RIF regression, Machado-Mata simulations
- Relate some distributional statistics $v(F)$ to some 'explanatory' variableS X
  - F is a (univariate) income distribution function
  - $v(F)$ generic 'welfare functional': some quantile, Gini coefficient, poverty index, welfare index, polarization measure,...

Outline

- Non-technical toolbox review: quantile regression, distribution regression, parametric income distribution models, RIF regression, Machado-Mata simulations
- Relate some distributional statistics $\upsilon(F)$ to some 'explanatory' variableS X
  - F is a (univariate) income distribution function
  - $\upsilon(F)$ generic 'welfare functional': some quantile, Gini coefficient, poverty index, welfare index, polarization measure,...

Objectives

Methods address two related but distinct questions:

1. How does $v(F)$ vary with X?
   That is, calculate and summarize $v(F_x)$ (remember $dim(X) > 1$), 'partial effects')
   - EOp, Intergenerational mob, Educ choices, Income risk and vulnerability, wage gap and glass ceilings. etc.
2. How much does X contributes to $v(F)$?
   - How much can a change in some element in X affect $v(F)$? ('policy effects')
   - How much do differences in X account for differences in $v(F)$ between A and B (across time, countries, gender, race, ...)?

Objectives

Methods address two related but distinct questions:

1. How does $\upsilon(F)$ vary with X?
   That is, calculate and summarize $\upsilon(F_x)$ (remember $dim(X) > 1$), 'partial effects')
   - EOp, Intergenerational mob, Educ choices, Income risk and vulnerability, wage gap and glass ceilings. etc.
2. How much does X contributes to $\upsilon(F)$?
   - How much can a change in some element in X affect $\upsilon(F)$? ('policy effects')
   - How much do differences in X account for differences in $\upsilon(F)$ between A and B (across time, countries, gender, race, ...)?

Objectives

Methods address two related but distinct questions:

1. How does $\upsilon(F)$ vary with X?
   That is, calculate and summarize $\upsilon(F_x)$ (remember $dim(X) > 1$), 'partial effects')
   - EOp, Intergenerational mob, Educ choices, Income risk and vulnerability, wage gap and glass ceilings. etc.

2. How much does X contributes to $\upsilon(F)$?
   - How much can a change in some element in X affect $\upsilon(F)$? ('policy effects')
   - How much do differences in X account for differences in $\upsilon(F)$ between A and B (across time, countries, gender, race, ...)?

Objectives

NB:

1. If $v(F)$ is the mean: easy!
   - (1.) is just the regression and
   - (2.) is Oaxaca-Blinder type of analysis.

   Complication is of course non-linearities and non-additivity in $v(F)$ that makes exercise more challenging

2. 'Descriptive' rather than 'causal' inference:
   sidestep in this lecture issues of 'exogeneity' of X (and feedback/general equilibrium issues).
   Is X determined by F? Is a difference in X independent of outcome distribution? etc.

Objectives

NB:

1. If $v(F)$ is the mean: easy!
   - (1.) is just the regression and
   - (2.) is Oaxaca-Blinder type of analysis.

   Complication is of course non-linearities and non-additivity in $v(F)$ that makes exercise more challenging

2. 'Descriptive' rather than 'causal' inference:
   sidestep in this lecture issues of 'exogeneity' of X (and feedback/general equilibrium issues).
   Is X determined by F? Is a difference in X independent of outcome distribution? etc.

Two main approaches

Two main approaches in recent literature

1. Recentered influence function regression (Firpo et al., 2009): Linearization approach: $\upsilon(F) = \mathrm{E}_X \, \mathrm{E}[\mathrm{RIF}(y; \upsilon, F)|X]$
   - Back to simple regression framework!
   - very easy to implement
   - specific to particular $\upsilon(F)$
   - interpretation and local approximation issues?

2. Indirect distribution function modelling (e.g., Machado and Mata, 2005, Chernozhukov et al., 2013):
   - plug a model for F in $\upsilon(F)$ (one model for any set of distributional statistics)
   - model $F(y) = \int F_x(y)h(x)dx$: essentially involves modelling the conditional distribution $F_x(y)$ (or the quantile function)
   - counterfactual effects
   - straightforward Monte Carlo integration principles

Two main approaches

Two main approaches in recent literature

1. Recentered influence function regression (Firpo et al., 2009): Linearization approach: $\upsilon(F) = \mathrm{E}_x \mathrm{E}[\mathrm{RIF}(y;\upsilon,F)|X]$
   - Back to simple regression framework!
   - very easy to implement
   - specific to particular $\upsilon(F)$
   - interpretation and local approximation issues?
2. Indirect distribution function modelling (e.g., Machado and Mata, 2005, Chernozhukov et al., 2013):
   - plug a model for F in $\upsilon(F)$ (one model for any set of distributional statistics)
   - model $F(y) = \int F_x(y)h(x)dx$: essentially involves modelling the conditional distribution $F_x(y)$ (or the quantile function)
   - counterfactual effects
   - straightforward Monte Carlo integration principles

CEPS INSTEAD

Two main approaches

Two main approaches in recent literature

1. Recentered influence function regression (Firpo et al., 2009): Linearization approach: $\upsilon(F) = \mathrm{E}_x \, \mathrm{E}[\mathrm{RIF}(y; \upsilon, F)|X]$
   - Back to simple regression framework!
   - very easy to implement
   - specific to particular $\upsilon(F)$
   - interpretation and local approximation issues?

2. Indirect distribution function modelling (e.g., Machado and Mata, 2005, Chernozhukov et al., 2013):
   - plug a model for F in $\upsilon(F)$ (one model for any set of distributional statistics)
   - model $F(y) = \int F_x(y)h(x)dx$: essentially involves modelling the conditional distribution $F_x(y)$ (or the quantile function)
   - counterfactual effects
   - straightforward Monte Carlo integration principles

Outline

Part I:
Recentered influence function regression

Part II:
Conditional distribution models

Part III:
Simulating counterfactual distributions

Part IV:
Envoi

Definition of (recentered) influence function

The influence function of an index $\upsilon(F)$ on distribution F at income y is

$$IF(y;\upsilon,F) = \lim_{\epsilon \downarrow 0} \frac{\upsilon((1-\epsilon)F + \epsilon\Delta_y) - \upsilon(F))}{\epsilon}$$

(e.g., for the variance, $IF(y;\sigma^2,F) = (y-\mu(F))^2 - \sigma^2$)
In a sample, one can evaluate the empirical IF: $if_i = IF(y_i;\upsilon,\hat{F})$ (cf. jackknife equivalence!)

$\mu(if_i) = 0$ by construction, so adding back $\upsilon(F)$ (recentering) gives

$$rif_i = if_i + \upsilon(\hat{F})$$

so $\mu(rif_i) = \upsilon(\hat{F})$

RIF regression

Now, by the law of iterated expectations:

$$E(rif_i) = \upsilon(F) = E_X(E(rif_i|X_i))$$

Problem boils down to estimation a conditional expectation. Easy!

RIF-OLS specifies:

$$E(rif_i|X_i) = X_i\beta$$

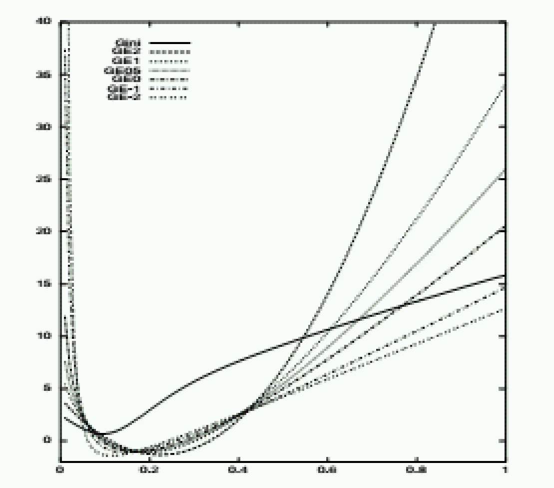
so RIF regression is OLS of sample of empirical rif_i on the covariates X

Interpretation of RIF regression coefficients

- Remember the RIF at y gives the influence on $\upsilon(F)$ of an infinitesimal increase in the density of the data at y:
- Regression coefficients reveal how much the average influence of observations vary with X (holding other covariates constant)
- It also reveals how much $\upsilon(F)$ would respond to a change in the distribution of X in the population holding distribution of other covariates constant
  - linear approximation valid only for *marginal* changes in X!

Example: Gini coefficient

Influence functions for Gini and GE indices (Cowell and Flachaire, 2007)



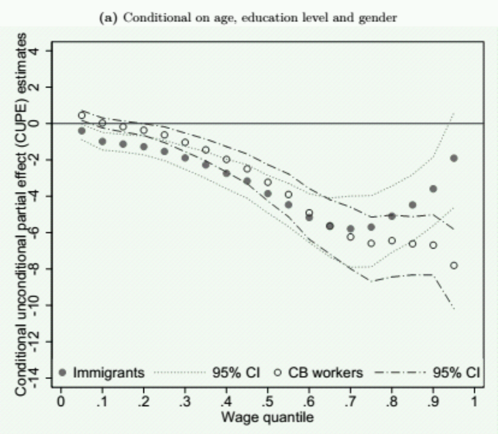
Example: Gini coefficient

RIF regression coefficients on Gini coefficient of private sector wages in Luxembourg (Choe and Van Kerm, 2013)

Immigrant	-0.003	-0.004	0.002
Non-resident	-0.054*	-0.052*	-0.030*
Female		-0.005	-0.023†
Age		-0.018*	-0.021*
Age squared/100		0.028*	0.030*
Secondary education		0.033*	0.012‡
Tertiary education		0.255*	0.126*
Years at current employer		-0.001	-0.000
Manager			0.028*
10-49 employees in firm			-0.003
50-249 employees in firm			0.022
500-999 employees in firm			0.017
1000+ employees in firm			-0.009
Part time contract			0.067*

Example: Quantiles

RIF regression coefficients on foreign workers on 19 quantiles of (unconditional) distibution of private sector wages in Luxembourg (Choe and Van Kerm, 2013)



Outline

Part I:
Recentered influence function regression

Part II:
Conditional distribution models

Part III:
Simulating counterfactual distributions

Part IV:
Envoi

Inference via counterfactual distributions

Analysis via counterfactual distributions is typically three-stage:

- model and estimate conditional distribution functions $F_x(y)$ (or conditional quantile functions)
- recover prediction for F by averaging over covariate distribution: $F(y) = \int F_x(y)h(x)dx$:
- simulate counterfactual distributions $\tilde{F}$ by manipulating
  - the conditional distribution functions: $\tilde{F}(y) = \int G_x(y)h(x)dx$:
  - the covariate distributions: $\tilde{F}(y) = \int F_x(y)g(x)dx$:
  - (typical analysis swaps either component across, say countries, gender, etc.)

Estimating conditional distribution functions

Many estimators available:

- quantile regression (Koenker and Bassett, 1978)
- distribution regression (Foresi and Peracchi, 1995)
- parametric income distribution models (Biewen and Jenkins, 2005)
- (duration models (Donald et al., 2000, **?**), ordered probit model (Fortin and Lemieux, 1998))

Quantile regression

Distribution regression

Parametric models

Quantiles as check function minimizer

Quantile regression (QR) is fully analogous to mean regression

The trick is to express Q_τ as similar minimization problem:

$$\hat{Q}_\tau = \arg\min_\xi \sum_i \rho_\tau(y_i - \xi)$$

where

$$\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0))$$

Quantiles as check function minimizer

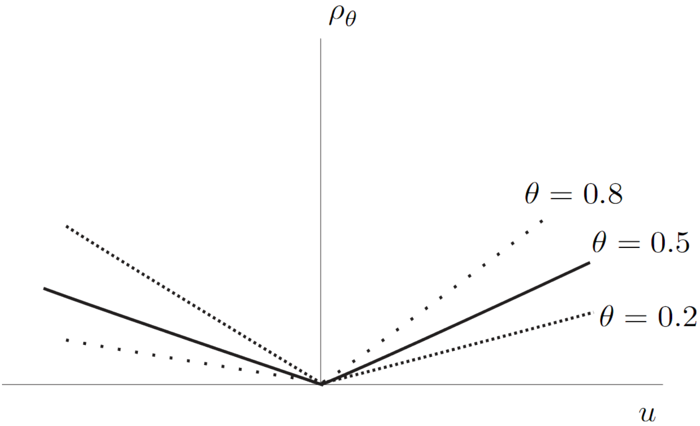


Figure: Check function $\rho_\theta(u) = (\theta - 1_{\{u<0\}})u$.

Linear quantile regression model

Focus on conditional quantile now and assume a particular relationship (linear) between conditional quantile and x:

$$Q_\tau(y|x) = x\beta_\tau$$

(Or equivalently $y_i = x_i\beta_\tau + u_i$ where $F^{-1}_{u_i|x_i}(\tau) = 0$)

$$\hat{\beta}_\tau = \arg\min_\beta \sum_i \rho_\tau(y_i - x_i\beta)$$

Estimate of the conditional quantile (given linear model):

$$\hat{Q}_\tau(y|x) = x\hat{\beta}_\tau$$

Interpretation of linear QR

- Estimation of $\hat{Q}_\tau(y|x)$ for a continuum of τ in $(0, 1)$ provides a model for the entire conditional quantile function of Y given X
- $\hat{\beta}_\tau$ can be interpreted as the marginal change in the τ conditional quantile for a marginal change in x ...
- (... However, this does not imply that it is really the effect of a change in x for a person at the τ quantile of the distribution! This would require a “rank-preservation” assumption.)
- Beware of quantile crossing within the support of X! (Simple solution is re-arrangement of quantile predictions (Chernozhukov et al., 2007))

Recovering $\upsilon(F_x)$

After estimation of the quantile process $(0, 1)$, estimation of the distributional statistic conditional on X is straightfoward:

- The set of predicted conditional quantile values $\{x_i \hat{\beta}_\theta\}_{\theta \in (0,1)}$ is a pseudo-random draw from F_x (if grid for θ is equally-spaced) (Autor et al., 2005)
- So a simple estimator for υ from unit-record data can be used to estimate $\upsilon(F_{X_i})$
- (no need to use $\upsilon(\hat{F}_x)$ which would often require numerical integration)

Flexible estimation

Linearity of the model $Q_\tau(y|x) = x\beta_\tau$ may possibly be problematic in some situations

- Highly non-linear relationship between x and y
- Discovering non-linearities is the objective (e.g., growth curves)

Options:

- Add covariates in non-linear way (but still linearly in parameters).
  - series estimator (complexity with order of polynomial in x)
  - spline functions (complexity with order of spline and number of knots)
  - interactions

Smoothing techniques attempt to balance flexibility and sampling variance (variability)

Kernel smoothing and locally weighted regression

Idea: To estimate $q_\tau(x)$, first define a neighborhood around x $(x+/-h)$ and compute quantile using all x_i that fall in this neighborhood.
Refinement: run a quantile regression $Q_\tau(y|x) = x\beta_\tau$ using data *within neighborhood* and with kernel weight

$$
\begin{aligned}
(\hat{\alpha}, \hat{\beta}) &= \arg\min_{\alpha,\beta} \sum_i \rho_\tau(y_i - \alpha - (x_i - x)\beta)\ K_h(x_i - x) \\
\hat{q}_\tau(x) &= \hat{\alpha}
\end{aligned}
$$

Issue 1: set the size of bandwidth h (too small –too variable; too large then too linear)
Issue 2: not tractable when $\dim(X)$ is large (curse of dimensionality)

Quantile regression

Distribution regression

Parametric models

‘Distribution regression’

$F_x(y) = \Pr\{y_i \leq y|x\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$)

Idea is to estimate $F_x(y)$ on a grid of values for y spanning the domain of definition of Y by running repeated standard binary choice models, e.g. a logit:

$$F_x(y) = \Pr\{y_i \leq y|x\} = \Lambda(x\beta_y) = \frac{\exp(x\beta_y)}{1+\exp(x\beta_y)}$$

or a probit $F_x(y) = \Phi(x\beta_y)$ or else ...

‘Distribution regression’

$F_x(y) = \Pr\{y_i \leq y|x\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$)
Idea is to estimate $F_x(y)$ on a grid of values for y spanning the domain of definition of Y by running repeated standard binary choice models, e.g. a logit:

$$\begin{aligned} F_x(y) &= \Pr\{y_i \leq y|x\} \\ &= \Lambda(x\beta_y) \\ &= \frac{\exp(x\beta_y)}{1+\exp(x\beta_y)} \end{aligned}$$

or a probit $F_x(y) = \Phi(x\beta_y)$ or else ...

‘Distribution regression’

- Estimation of these models is well-known and straightforward!
- Repeating estimations at different values of y makes little assumptions about the overall shape of distribution
- Beware of crossing of predictions here too!
- Non-parametric, kernel-based approaches feasible too
- Conditional statistic $\upsilon(F_x)$ potentially less easy to recover from the predictions than with quantile regression

Quantile regression

Distribution regression

Parametric models

Parametric distribution fitting

Assume that the conditional distribution has a particular parametric form: e.g., (log-)normal (2 parameters – quite restictive), Gamma (2 params), Singh-Maddala (3 param.), Dagum (3 param.), GB2 (4 param.), ... or any other distribution that is likely to fit the data at hand (think domain of definition, fatness of tails, modality)

Let parameters (say vector θ) depend on x in a particular fashion, typically linearly (up to some transformation), e.g., $\theta_1 = \exp(x\beta_1)$, $\theta_2 = \exp(x\beta_2)$ and $\theta_3 = x\beta_3$

This gives a fully specified parametric model which can be estimated using maximum likelihood.

Parametric distribution fitting

- With parameters estimates, you can recover conditional quantiles, CDF, PDF, means, ... often with closed-from expression
- Typically much less computationally expensive than estimating full processes
- Price to pay is stronger parametric assumptions! (Look at goodness-of-fit statistics (KS, KL, of predicted dist – contrast with non-parametric fit also useful)
  - compare fit with quantile regression with ...
- log-normal distribution often much worse fit than 3-parameters alternative (while 4-parameter GB2 difficult to estimate reliably)

Endogenous sample selection in QR?

- “Buchinsky” two-step approach (Buchinsky, 2002):
  - first step: estimate probability of selection (using non-parametric model) given x and some exogenous z, $p(x,z)$
  - second step: add a flexible function $g(p(x,z))$ (polynomial?) in quantile regression of interest
- ... but the constant in the second step is difficult to identify separately from the constant in the polynomial $g(p(x,z))$.
- Structural assumptions about this model are questioned. (Severe identification problems being discovered in recent research (Huber and Melly, 2011)

Parametric distribution fitting wth endogenous selection

Parametric model can be extended to model endogenous participation explicitly: this is a big advantage!
Let s denote binary participation (outcome y only observed if $s = 1$).
Assume $s = 1$ if $s^* > 0$ and $s = 0$ otherwise. s^* is latent propensity to be observed.
Assume pair (y, s^*) is jointly distributed H and express H using its copula formulation

$$H(y, s^*) = \Psi(F(w), G(s^*))$$

where F is outcome distribution, G is latent participation distribution (typically Gaussian), and Ψ is a parametric copula function.
Everything is parametric and can be estimated using maximum likelihood (Van Kerm, 2013)

Outline

Part I:
Recentered influence function regression

Part II:
Conditional distribution models

Part III:
Simulating counterfactual distributions

Part IV:
Envoi

Unconditional distributions from conditional distributions

Law of iterated expectations:

$$F(y) = E_x(F_x(y)) = \int_{\Omega_x} F_x(y)\ h(x)\ dx$$

$\Longrightarrow$ Decompose the unconditional CDF ($F(y)$) as weighted sum of conditional CDFs ($F_x(y)$) with weights being population share with x ($h(x)$)

In a sample, take the F_x predictions averaged over observations (Monte Carlo integration over covariates):

$$\hat{F}(y) = \frac{1}{N}\sum_{i=1}^{N} \hat{F}_{x_i}(y) = \frac{1}{N}\sum_{i=1}^{N} \Lambda(x_i \hat{\beta}_y)$$

Counterfactual distributions

Decompose differences in two CDFs as differences attributed to F_x ('structural' part) and $h(x)$ ('compositional' part).

$$\tilde{F}(y) = \int_{\Omega_x} F^m_x(y)\, h^f(x)\, dx$$

and

$$(F^f(y) - F^m(y)) = (F^f(y) - \tilde{F}(y)) + (\tilde{F}(y) - F^m(y))$$

Sample analog:

$$\tilde{F}^f(y) = \frac{1}{N^f} \sum_{i=1}^{N^f} \hat{F}^m_{x_i} = \frac{1}{N^f} \sum_{i=1}^{N^f} \Lambda(x_i \hat{\beta}^m_y)$$

(Alternative counterfactual constructions conceivable. See Chernozhukov et al. (2013) for inferential theory.)

Counterfactual distributional statistics?

Counterfactual estimates of $\upsilon(F)$ obtained by plugging-in $\hat{F}$ and $\tilde{F}$ in υ.

However, many functionals are not necessarily easily derived from F (require some form of numerical integration)
$\Longrightarrow$ Derivation from (countefactual) conditional quantile function!

- but there is no law of 'iterated quantiles'

$$Q_\tau(y) \neq \int_{\Omega_x} Q_\tau(y|x)\ h(x)\ dx$$

- simple simulations and Monte Carlo integration

Unconditional and counterfactual quantile functions

Simulation consists in generating a simulated sample from F on the basis of conditional quantile estimates.

Machado and Mata (2005) algorithm:

- pick a random value $\theta \in (0, 1)$ and calculate conditional quantile regression for the θ-th quantile
- select a random observation x_j from the sample and calculate predicted value $Q_{x_j} = x_j \hat{\beta}_\theta$
- repeat steps above B times to generate a simulated sample from F based on the conditional quantile model
- $\upsilon(F)$ calculated with standard formulae from the simulated sample

Decomposition of quantile differences

Machado-Mata very computationally intensive, especially since large B required for accurate estimation of $\upsilon(F)$.

Simplified version (Autor et al., 2005, Melly, 2005):

- estimate uniform (equally-spaced) sequence of conditional quantile predictions for each observations—pseudo-random sample from the conditional distribution F_x
- stack vectors of predictions for all observations into one long vector V—pseudo-random sample from the unconditional distribution!

Decomposition of quantile differences

Simplified version of Machado-Mata leads to very large simulated sample with smaller number of quantile regression estimation... but still computationally intensive (at least 99 quantile regressions recommended)!

$\Longrightarrow$ Replace estimation of 99 conditional quantile regressions by estimation of 1 parametric model (Van Kerm et al., 2013)

- predict conditional quantiles from parameter estimates—closed-form expressions often available

(NB: random subsampling from V useful for speeding-up calculations)

Decomposition of differences in $\upsilon(F)$

Model-based predictions allow generation of various counterfactual constructs:

- calculate conditional quantile predictions in group *m*
- aggregate over observations from group *f*

Decomposition of differences in $\upsilon(F)$

Let $\{x_i\hat{\beta}^f_\theta\}^f_{\theta\in(0,1)}$ be the predictions from women model

Let $\{x_i\hat{\beta}^m_\theta\}^f_{\theta\in(0,1)}$ be the $K \times N^f$ predictions from men model

And similarly $\{x_i\hat{\beta}^f_\theta\}^m_{\theta\in(0,1)}$ and $\{x_i\hat{\beta}^m_\theta\}^m_{\theta\in(0,1)}$ be the $K \times N^m$ predictions on the men sample

$$
\begin{aligned}
\hat{\upsilon}^f - \hat{\upsilon}^m &= \upsilon\left(\{x_i\hat{\beta}^f_\theta\}^f_{\theta\in(0,1)}\right) - \upsilon\left(\{x_i\hat{\beta}^m_\theta\}^m_{\theta\in(0,1)}\right) \\
&= (\hat{\upsilon}^f - \hat{\upsilon}^*) + (\hat{\upsilon}^* - \hat{\upsilon}^m) \\
&= \left(\upsilon\left(\{x_i\hat{\beta}^f_\theta\}^f_{\theta\in(0,1)}\right) - \upsilon\left(\{x_i\hat{\beta}^m_\theta\}^f_{\theta\in(0,1)}\right)\right) \\
&- \left(\upsilon\left(\{x_i\hat{\beta}^m_\theta\}^f_{\theta\in(0,1)}\right) - \upsilon\left(\{x_i\hat{\beta}^m_\theta\}^m_{\theta\in(0,1)}\right)\right)
\end{aligned}
$$

(Again, see Chernozhukov et al. (2013) for inferential theory and bootstrap confidence intervals.)

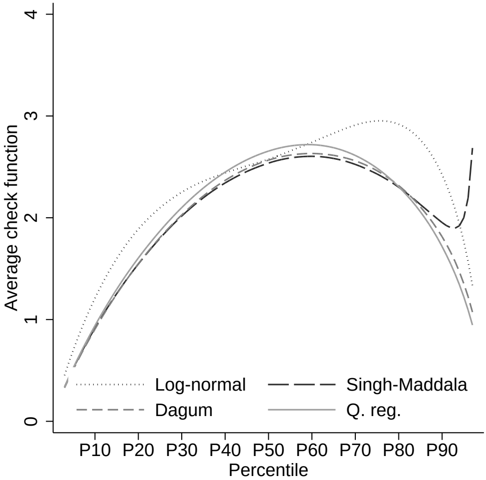
CEPS INSTEAD

Comparing fit of alternative conditional quantile estimators

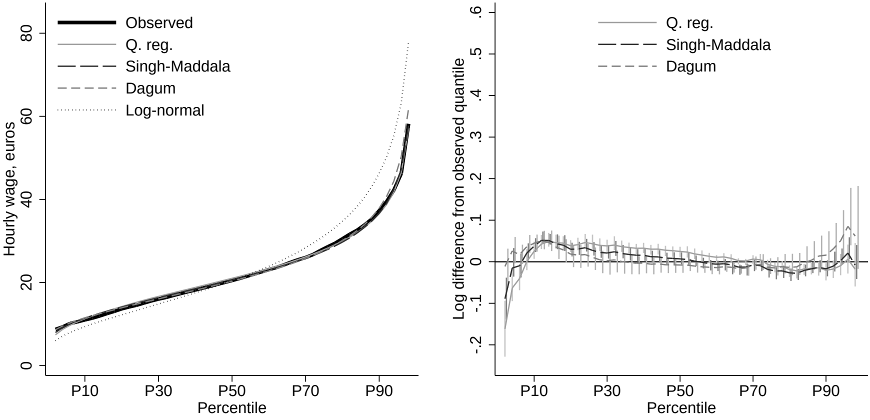
Look at average check function residual by conditional quantile:

$$\sum_i (y_i - \hat{Q}_{x_i}(\theta))(\theta - \mathbf{1}(y_i - \hat{Q}_{x_i}(\theta)) < 0))$$

Comparing fit of alternative conditional quantile estimators



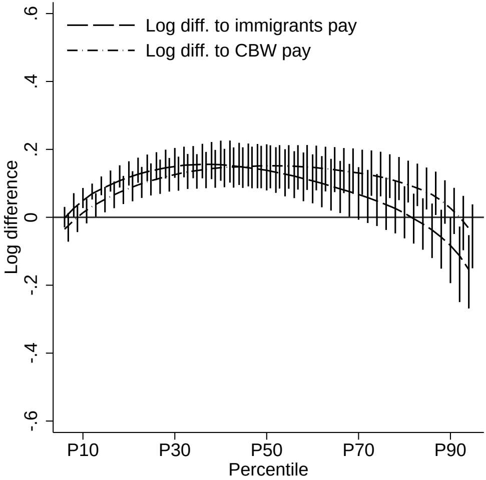
Comparing fit of alternative conditional quantile estimators (ctd.)



(Van Kerm et al., 2013)

Native-foreign workers quantile wage gap

Wage gap at unconditional quantiles



Outline

Part I:
Recentered influence function regression

Part II:
Conditional distribution models

Part III:
Simulating counterfactual distributions

Part IV:
Envoi

Two main approaches: Pros and Cons

1. Recentered influence function regression:
   - easy
   - regression coefficient interpretation direct but not necessarily intuitive
   - just as easy with F multivariate
2. Distribution function modelling:
   - accurate and valid for many potential counterfactual policy simulations/comparisons
   - allow for endogenous sample selection correction
   - no unique easy-to-interpret coefficient on each variable
   - computationally heavy (in computing time—think parametric distribution models!)
   - multivariate F brings in substantial complication

Support from the Luxembourg ‘Fonds National de la Recherche’ (contract C10/LM/785657) is gratefully acknowledged.

Autor, D. H., Katz, L. F. and Kearney, M. S. (2005), Rising wage inequality: The role of composition and prices, NBER Working Paper 11628, National Bureau of Economic Research, Cambridge MA, USA.

Biewen, M. and Jenkins, S. P. (2005), 'Accounting for differences in poverty between the USA, Britain and Germany', *Empirical Economics* **30**(2), 331–358.

Buchinsky, M. (2002), Quantile regression with sample selection: Estimating women's return to education in the U.S., *in* B. Fitzenberger, R. Koenker and J. A. F. Machado, eds, 'Economic applications of quantile regression', Physica-Verlag, Heidelberg, Germany, pp. 87–113.

Chernozhukov, V., Fernàndes-Val, I. and Galichon, A. (2007), Improving estimates of monotone functions by rearrangement, CEMMAP working paper CWP09/07, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, University College of London.

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2013), 'Inference on counterfactual distributions', *Econometrica* **81**(6), 2205–2268.

Choe, C. and Van Kerm, P. (2013), Foreign workers and the wage distribution: Where do they fit in? Unpublished manuscript, CEPS/INSTEAD, Esch/Alzette, Luxembourg.

Cowell, F. A. and Flachaire, E. (2007), 'Income distribution and inequality measurement: The problem of extreme values', *Journal of Econometrics* **141**(2), 1044–1072.

Donald, S. G., Green, D. A. and Paarsch, H. J. (2000), 'Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates', *Review of Economic Studies* **67**(4), 609–633.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009), 'Unconditional quantile regressions', *Econometrica* **77**(3), 953–973.

Foresi, S. and Peracchi, F. (1995), 'The conditional distribution of excess returns: An empirical analysis', *Journal of the American Statistical Association* **90**(430), 451–466.

Fortin, N. M. and Lemieux, T. (1998), 'Rank regressions, wage distributions, and the gender gap', *Journal of Human Resources* **33**(3), 610–643.

Huber, M. and Melly, B. (2011), Quantile regression in the presence of sample selection, Economics Working Paper 1109, School of Economics and Political Science, Department of Economics, University of St. Gallen.

Koenker, R. and Bassett, G. (1978), ‘Regression quantiles’, *Econometrica* **46**(1), 33–50.

Machado, J. A. F. and Mata, J. (2005), ‘Counterfactual decomposition of changes in wage distributions using quantile regression’, *Journal of Applied Econometrics* **20**, 445–465.

Melly, B. (2005), ‘Decomposition of differences in distribution using quantile regression’, *Labour Economics* **12**, 577–90.

Van Kerm, P. (2013), ‘Generalized measures of wage differentials’, *Empirical Economics* **45**(1), 465–482.

Van Kerm, P., Yu, S. and Choe, C. (2013), Wage differentials between natives, immigrants and cross-border workers: Evidence and model comparisons. Unpublished manuscript, CEPS/INSTEAD, Esch/Alzette, Luxembourg.