Counterfactual distributions: estimation and inference in Stata

Victor Chernozhukov — MIT
Iván Fernández-Val — Boston University
Blaise Melly — Bern University

November 17, 2016
Swiss Stata Users Group Meeting in Bern

Questions

- What would have been the wage distribution in 1979 if the workers had the same distribution of characteristics as in 1988?
- What would be the distribution of housing prices resulting from cleaning up a local hazardous-waste site?
- What would be the distribution of wages for female workers if female workers were paid as much as male workers with the same characteristics?

- In general, given an outcome Y and a covariate vector X. What is the effect on F_Y of a change in
  1. F_X (holding $F_{Y|X}$ fixed)?
  2. $F_{Y|X}$ (holding F_X fixed)?
- To answer these questions we need to estimate counterfactual distributions.

Counterfactual distributions

- Let 0 denote 1979 and 1 denote 1988.
- Y is wages and X is a vector of worker characteristics (education, experience, ...).
- $F_{X_k}(x)$ is worker composition in $k \in \{0,1\}$; $F_{Y_j}(y \mid x)$ is wage structure in $j \in \{0,1\}$.
- Define
$$F_{Y\langle j|k\rangle}(y) := \int F_{Y_j}(y \mid x) dF_{X_k}(x).$$
- $F_{Y\langle 0|0\rangle}$ is the observed distribution of wages in 1979; $F_{Y\langle 0|1\rangle}$ is the *counterfactual distribution* of wages in 1979 if workers have 1988 composition.
- Common support: $F_{Y\langle 0|1\rangle}$ is well defined if the support of X_1 is included in the support of X_0.

Effect of changing F_X

- We are interested in the effect of shifting the covariate distribution from 1979 to that of 1988.
- Distribution effects

$$\Delta^{DE}(y) = F_{Y\langle 0|1\rangle}(y) - F_{Y\langle 0|0\rangle}(y)$$

- The quantiles are often also of interest:

$$Q_{Y\langle j|k\rangle}(\tau) = \inf\{y : F_{Y\langle j|k\rangle}(y) \geq u\},\ 0 < \tau < 1.$$

  Quantile effects

$$\Delta^{QE}(\tau) = Q_{Y\langle 0|1\rangle}(\tau) - Q_{Y\langle 0|0\rangle}(\tau)$$

- In general, for a functional ϕ, the effects is

$$\Delta(w) := \phi(F_{Y\langle 0|1\rangle})(w) - \phi(F_{Y\langle 0|0\rangle})(w).$$

  Special cases: Lorenz curve, Gini coefficient, interquartile range, and more trivially the mean and the variance.

Types of counterfactual changes in F_X

1. Groups correspond to different subpopulations (different time periods, male vs. female, black vs. white).
2. Transformations of the population: $X_1 = g(X_0)$:
   - Unit change in location of one covariate: $X_1 = X_0 + 1$ where X is the number of cigarettes smoked by the mother and Y is the birthweight of the newborn.
   - Neutral redistribution of income: $X_1 = \mu_{X_0} + \alpha(X_0 - \mu_{X_0})$, where Y is the food expenditure (Engel curve).
   - Stock (1991): effect on housing prices of removing hazardous waste disposal site.

$$\text{In 1. and 2., } \widehat{F}_{X_1}(x) = n_1^{-1} \sum_{i=1}^{n_1} 1\{X_{1i} \leq x\}.$$

3. Change in some variable(s) but not in the other ones: unionization rate in 1988 and other characteristics from 1979.

$$\text{In 3., } d\widehat{F}_{X_1}(x) = d\hat{F}_{U_1|C_1}(u|c) d\hat{F}_{C_0}(c).$$

Effect of changing $F_{Y|X}$

- We are often interested in the effect of changing the conditional distribution of the outcome for a given population.
- Program evaluation: Group 1 is treated and group 0 is the control group. The quantile treatment effect on the treated is

$$QTET = Q_{Y\langle 1|1 \rangle}(\tau) - Q_{Y\langle 0|1 \rangle}(\tau).$$

- The counterfactual distributions are always statistically well-defined object. The effects are of interest even in 'non-causal' framework (e.g. gender wage gap).
- Causal interpretation under additional assumptions that give a structural interpretation to the conditional distribution. Selection on observables: the conditional distribution may be estimated using quantile or distribution regression. Endogenous groups: IV quantile regression (e.g. Chernozhukov and Hansen 2005).

Decompositions

- The counterfactual distributions that we analyze are the key ingredients of the decomposition methods often used in economics.
- Blinder/Oaxaca decomposition (parametric, linear decomposition of the mean difference):

$$\bar{Y}_0 - \bar{Y}_1 = (\bar{X}_0 \beta_0 - \bar{X}_1 \beta_0) + (\bar{X}_1 \beta_0 - \bar{X}_1 \beta_1).$$

  This fits in our framework (even if our machinery is not needed in this simple case) as

$$\overline{Y\langle 0|0\rangle} - \overline{Y\langle 1|1\rangle} = \left(\overline{Y\langle 0|0\rangle} - \overline{Y\langle 0|1\rangle}\right) + \left(\overline{Y\langle 0|1\rangle} - \overline{Y\langle 1|1\rangle}\right).$$

- Our results allow us to do similar decomposition of any functional of the distribution. E.g. a quantile decomposition

$$\left(Q_{Y\langle 0|0\rangle}(\tau) - Q_{Y\langle 0|1\rangle}(\tau)\right) + \left(Q_{Y\langle 0|1\rangle}(\tau) - Q_{Y\langle 1|1\rangle}(\tau)\right).$$

Estimation: plug-in principle

- We estimate the unknown elements in $\int F_{Y_0}(y \mid x) dF_{X_1}(x)$ by analog estimators.
- We estimate the distribution of X_1 by the empirical distribution for group 1.
- The conditional distribution can be estimated by:
    1. Location and location-scale shift models (e.g. OLS and independent errors),
    2. Quantile regression,
    3. Duration models (e.g. proportional hazard model),
    4. Distribution regression.
- Our results also cover other methods (e.g. IV quantile regression).

Outline of the algorithm for $F_{Y\langle 0|1\rangle}(y)$

1. Estimation
   1.1 Estimate $F_{X_1}(x)$ by $\widehat{F}_{X_1}(x)$.
   1.2 Estimate $F_{Y_0}(y \mid x)$ by $\widehat{F}_{Y_0|X_0}(y|x)$.
   1.3 $\widehat{F}_{Y\langle 0|1\rangle}(y) = \int \widehat{F}_{Y_0|X_0}(y|x) d\widehat{F}_{X_1}(x)$
   (in most cases: $n_1^{-1} \sum_{i=1}^{n_1} \widehat{F}_{Y_0|X_0}(y|X_{1i})$).
2. Pointwise inference
   2.1 Bootstrap $\widehat{F}_{Y\langle 0|1\rangle}(y)$ to obtain the pointwise s.e. $\hat{\Sigma}(y)$.
   2.2 Obtain a 95% CI as $\widehat{F}_{Y\langle 0|1\rangle}(y) \pm 1.96 \cdot \hat{\Sigma}(y)$.
3. Uniform inference
   Obtain the 95% confidence bands as $\widehat{F}_{Y\langle 0|1\rangle}(y) \pm \hat{t} \cdot \hat{\Sigma}(y)$, where $\hat{t}$ is the 95th percentile of the bootstrap draws of the maximal t statistic over y.

Conditional quantile models

- Location shift model (OLS with independent error term):

$$\begin{aligned} Y &= X'\beta + V, V \perp\!\!\!\perp X \\ Q_Y(u|x) &= x'\beta + Q_V(u). \end{aligned}$$

Parsimonious but restrictive, X only impact location of Y.

- Quantile regression (Koenker and Bassett 1978):

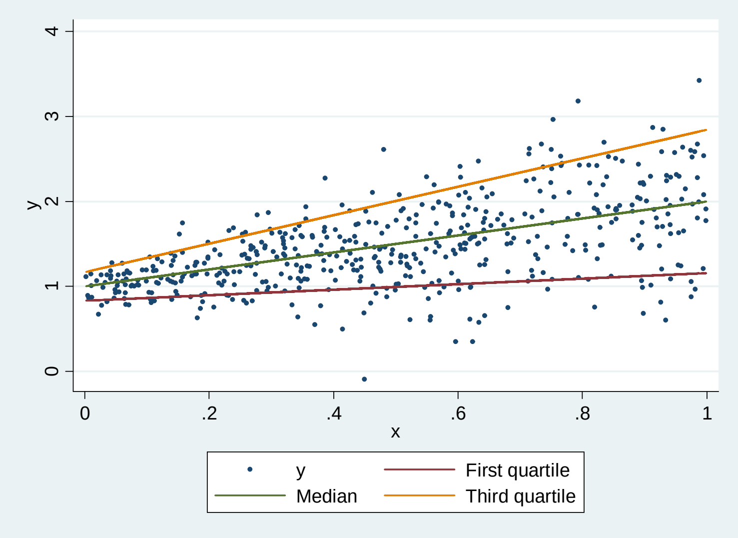
$$\begin{aligned} Y &= X'\beta(U), U \mid X \sim U(0,1) \\ Q_Y(u|x) &= x'\beta(u). \end{aligned}$$

X can change shape of entire conditional distribution.

- Connect the conditional distribution with the conditional quantile

$$F_{Y_0}(y|x) \equiv \int_0^1 1\{Q_{Y_0}(u|x) \leq y\} du.$$

Quantile regression



Conditional distribution models

- Distribution regression model (Foresi and Peracchi 1995):

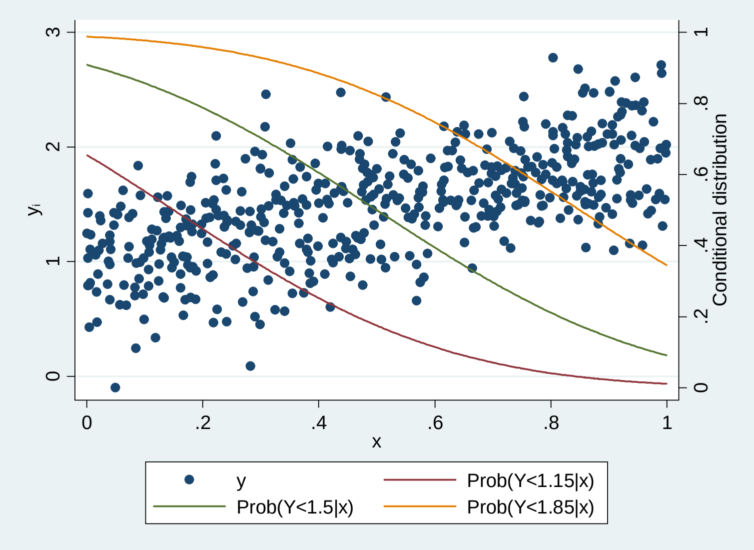
$$F_Y(y|x) = \Lambda(x'\beta(y)),$$

  where Λ is a link function (probit, logit, cauchit).
  X can have heterogeneous effects across the distribution.
- Cox (72) proportional hazard model is a special case with complementary log-log link and constant slope parameter

$$F_Y(y|x) = 1 - \exp(-\exp(\beta_0(y) - x'\beta_1))$$

  In other words: $\beta(y)$ is assumed to be constant.
- Estimate functional parameter vector $y \mapsto \beta(y)$ by MLE:
  1. Create indicators $1\{Y \leq y\}$,
  2. Probit/logit of $1\{Y \leq y\}$ on X.

Distribution regression



Comparison: QR vs DR

- QR and DR are flexible semiparametric models for the conditional distribution that generalize important classical models.
- Equivalent if X is saturated; but not nested otherwise. Choice cannot be made on the basis of generality.
- QR requires smooth conditional density of Y.
- QR usually overperforms DR under smoothness, but is less robust when Y has mass points.
- Different ability to deal with data limitations: censoring and rounding.

Pointwise and uniform inference

- The covariance function of $\widehat{F}_{Y\langle 0|1\rangle}(y)$ is cumbersome to estimate $\Longrightarrow$ *exchangeable bootstrap* (covers empirical bootstrap, weighted bootstrap and subsampling) provides the pointwise s.e. $\hat{\Sigma}(y)$.
- Many policy questions of interest involve functional hypotheses: no effect, constant effect, stochastic dominance. $\Longrightarrow$ uniform confidence bands:

$$\widehat{F}_{Y\langle 0|1\rangle}(y) \pm \widehat{t} \cdot \hat{\Sigma}(y).$$

The true t corresponds to the 95th percentile of the distribution of the maximum t-statistic

$$\sup_y \widehat{\Sigma}(y)^{-1/2} |\widehat{F}_{Y\langle 0|1\rangle}(y) - F_{Y\langle 0|1\rangle}(y)|,$$

which is unknown. We use the bootstrap to estimate it.

Theoretical results

- Under high level conditions we prove
  1. Functional central limit theorems

     $$\sqrt{n}\left(\widehat{F}_{\langle 0|1\rangle}(y) - F_{Y\langle 0|1\rangle}(y)\right) \rightsquigarrow Z_{\langle 0|1\rangle}(y)$$

     where $Z_{\langle 0|1\rangle}(y)$ is a tight zero-mean Gaussian process.
  2. The validity of the uniform confidence bands

     $$\lim_{n\to\infty} \Pr\left\{F_{Y\langle 0|1\rangle}(y) \in [\widehat{F}_{Y\langle 0|1\rangle}(y) \pm \widehat{t} \cdot \hat{\Sigma}(y)] \text{ for all } y\right\} = 0.95$$

- Under standard primitive conditions we show that QR and DR satisfy the high level conditions, i.e. functional central limit theorem and validity of the bootstrap for the coefficients processes.

The cdeco command

Quantile decomposition (go to definition):

`cdeco` *depvar indepvars* [*if*] [*in*] [*weight*], `group`(*varname*) [*options*]

- `group`(*varname*): binary variable defining the groups.
- `quantiles`(*numlist*): quantile(s) τ at which the decomposition will be estimated.
- `method`(*string*): estimator of the conditional distribution; available: qr (the default), loc, locsca, cqr, cox, logit, probit, and lpm.
- `nreg`(#): number of regressions estimated to approximate the conditional distribution; default is 100.
- `reps`(#): number of bootstrap replications.
- `noboot`: suppresses the bootstrap.

Application: private-public sector wage differences

- Data: Merged Outgoing Rotation Groups from the Current Population Survey in 2015.
- Sample: white males between 25 and 60 years old.
- Stata command and head of output:

```
. cdeco lwage education educ2 exp exp2 married widow divorced, group(public) metho
> d(probit)
(bootstrapping ...................................................................
> .................................)

Conditional model                          probit
Number of regressions estimated            90

The variance has been estimated by bootstraping the results 100 times.

No. of obs. in the reference group         25819
No. of obs. in the counterfactual group    3377
```

Total difference (private - public)

Differences between the observable distributions (based on the conditional model)

Quantile	Quantile effect	Pointwise Std. Err.	Pointwise [95% Conf.	Interval]	Functional [95% Conf.	Interval]
.1	-.248458	.022312	-.29219	-.204727	-.321072	-.175845
.2	-.22028	.013937	-.247597	-.192963	-.265638	-.174922
.3	-.183922	.013439	-.210262	-.157582	-.227658	-.140185
.4	-.146604	.011723	-.16958	-.123627	-.184754	-.108453
.5	-.117783	.012054	-.141409	-.094157	-.157013	-.078553
.6	-.053486	.015911	-.084671	-.022302	-.105266	-.001706
.7	0	.010962	-.021485	.021485	-.035674	.035674
.8	.064543	.011993	.041037	.088049	.025512	.103573
.9	.080044	.017911	.044939	.115148	.021755	.138333

Characteristics (private - public)

Effects of characteristics

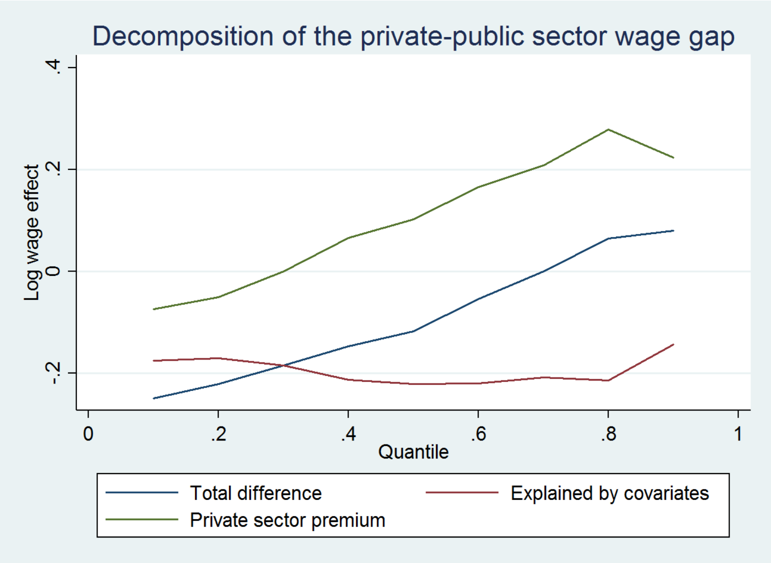
Quantile	Quantile effect	Pointwise Std. Err.	Pointwise [95% Conf.	Interval]	Functional [95% Conf.	Interval]
.1	-.174353	.011945	-.197764	-.150942	-.218815	-.129892
.2	-.169899	.00941	-.188342	-.151456	-.204926	-.134872
.3	-.183922	.014498	-.212337	-.155507	-.237887	-.129957
.4	-.212741	.011314	-.234917	-.190565	-.254857	-.170626
.5	-.220058	.012154	-.243879	-.196236	-.265299	-.174817
.6	-.219472	.013818	-.246555	-.192389	-.270907	-.168037
.7	-.207638	.004231	-.215931	-.199346	-.223387	-.191889
.8	-.213571	.009355	-.231906	-.195236	-.248393	-.178749
.9	-.143099	.007372	-.157549	-.128649	-.170542	-.115656

Private sector wage premium

Effects of coefficients

Quantile	Quantile effect	Pointwise Std. Err.	Pointwise [95% Conf.	Interval]	Functional [95% Conf.	Interval]
.1	-.074105	.023962	-.121069	-.027141	-.14338	-.00483
.2	-.050381	.012021	-.073943	-.02682	-.085136	-.015626
.3	0	.01703	-.033378	.033378	-.049235	.049235
.4	.066138	.012947	.040762	.091513	.028707	.103568
.5	.102275	.013464	.075887	.128663	.063351	.141199
.6	.165986	.017107	.132457	.199514	.116529	.215442
.7	.207638	.010464	.187128	.228148	.177385	.237892
.8	.278114	.014151	.250379	.305849	.237203	.319024
.9	.223143	.016952	.189918	.256367	.174134	.272151

Summary



Decomposition of the private-public sector wage gap
Log wage effect
-.2
0
.2
.4
0
.2
.4
.6
.8
1
Quantile
Total difference
Explained by covariates
Private sector premium

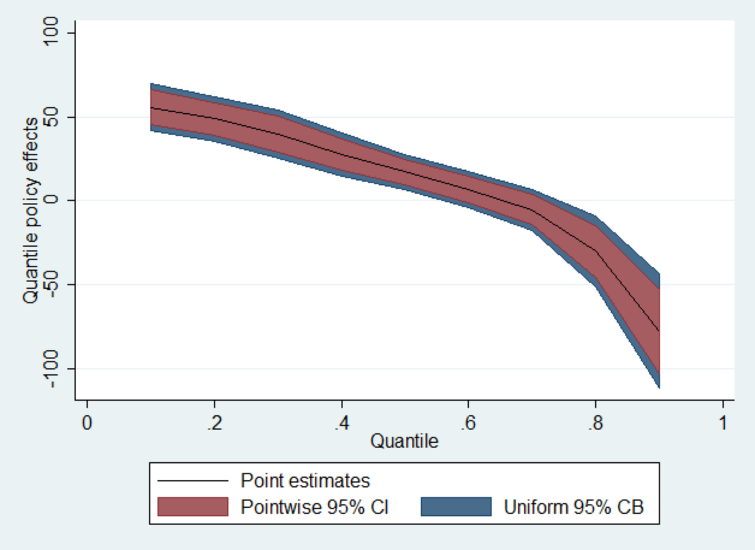
The counterfactual command

- The `counterfactual` command estimates the effect of changing the distribution of the covariates on the distribution of the outcome (link to definition).
- Syntax:
  `counterfactual` *depvar indepvars* [*if*] [*in*] [*weight*] [, `group`(*varname*) `counterfactual`(*varlist*) *other options*]
- Either the option `group` or `counterfactual` must be specified:
  - `group` if X_0 and X_1 correspond to different subpopulations,
  - `counterfactual` if X_1 is a transformations X_0. This option must provide a list of the counterfactual covariates that corresponds to the reference covariates given in indepvars. The order matters!
- The other options are the same as for `cdeco`.

Application: Engel curve

```
. summarize income

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      income |       235     982.473    519.2309   377.0584   4957.813

. generate counter_income = r(mean) + 0.75*(income-r(mean))

. counterfactual foodexp income, counterfactual(counter_income)
(bootstrapping ..................................................................
> ..................................)

Conditional model                          linear quantile regression
Number of regressions estimated            100

The variance has been estimated by bootstraping the results 100 times.

No. of obs. in the reference group         235
No. of obs. in the counterfactual group    235
```

Engel curve: effect of redistributing income



Summary

- Chernozhukov, Fernandez-Val and Melly (2013) suggest regression-based estimation and inference methods for counterfactual distributions.
- `cdeco` and `counterfactual` implement these methods in Stata.
- To do list:
  - Write an article to submit to the Stata Journal.
  - For non-continuous outcomes: implement the procedure in Chernozhukov, Fernandez-Val, Melly and Wüthrich (2016).
  - Detailed decomposition: work in progress with Philipp Ketz.
  - Faster algorithms for quantile and distribution regression.