# Lecture 1: Introduction and General Theory of Decompositions

Nicole Fortin and Thomas Lemieux, UBC

CLSRN Summer School

HÉC Montréal, June 27 2011

# Two lectures on decomposition methods

- Heavily relies on our chapter "*Decomposition Methods in Economics*" (joint with Nicole Fortin and Sergio Firpo) in the recent *Handbook of Labor Economics* (Volume 4A, 2011)
- Provides some examples to motivate the use of decomposition methods
- Uses the classic work of Oaxaca (1973) and Blinder (1973) for the **mean** as its point of departure
- Focuses on recent developments (last 15 years) on how to go beyond the mean
  - Connection with the treatment effect litterature
  - Procedures specific to the case of decompositions
- Provides empirical illustrations and discusses applications
- Suggests a "user guide" of best practices

# Plan for this first lecture

- Introduction
  - A few motivating examples
  - Oaxaca decompositions: a refresher
  - Issues involved when going “beyond the mean”
- A formal theory of decompositions
  - Clarify the question being asked
  - Decompositions as counterfactual exercises
  - Formal connection with the treatment effect literature:
    - Identification: the central role of the ignorability (or unconfoundedness) assumption
    - Estimators for the “aggregate” decomposition: propensity score matching, reweighting, etc.

# Introduction: motivating examples

1. Growth accounting (Denison, Solow, etc.)
2. Gender wage gap (Oaxaca, Blinder)
3. Union wage gap and the connection with treatment effects
4. Understanding the causes of changes in wage inequality

# Growth accounting

- One of the earliest example of a decomposition
- Question: what fraction of per capita growth (in the United States) is due to changes in production factors (capital and labour)?
- The “residual” or “unexplained” growth that cannot be accounted for by capital and labour is thought to reflect technological progress (growth in total factor productivity, TFP)
- With estimates of the parameters of the production function one can compute this decomposition
- While the basic motivation for the decomposition is similar to what we will see below, the estimation approach is a bit different. So let’s cover these other cases in more detail.

# Oaxaca (1973): Gender wage gap

- First study (along with Blinder, 1973) aimed at understanding the sources of the large gender gap in wage.
- Gender gap (difference in mean wages) of 43 percent in the 1967 Survey of Economics Opportunity (precursor to the PSID)
- Question: how much of the gender gap can be “explained” by male-female differences in human capital (education and labour market experience), occupational choices, etc?
  - The “unexplained” part of the gender gap is often interpreted as representing labour market discrimination, though other interpretations (unobserved skills) are possible too
- Estimate OLS regressions of (log) wages on covariates/charactistics
- Use the estimates to construct a **counterfactual** wage such as “what would be the average wage of women if they had the same characteristics as men?”
- This forms the basis of the decomposition.

# Union wage gap

- Workers covered by collective bargaining agreement (minority of works in Canada, the U.S. or U.K.), or “union workers”, earn more on average than other workers.
- One can do a Oaxaca-type decomposition to separate the “true” union effect from differences in characteristics between union and nonunion workers. The “unexplained” part of the decomposition is now the union effect.
- Unlike the case of the gender wage gap, we can think of the union effect as a **treatment effect** since union status is **manipulable** (can in principle be switched on and off for a given worker)
- Useful since tools and results from the treatment effect literature can be used in the context of this decomposition.
- Results from the treatment effect can also be used in cases where the group indicator is not manipulable (e.g. men vs. women)

# Changes in wage inequality

- Wage/earnings inequality has increased in many countries over the last 30 years
- Decomposition methods have been used to look for explanations for these changes, such as:
  - Decline in unions and in the minimum wage
  - Increase in the rate of return to education
  - Technological change, international competition, etc.
- We need to go “beyond the mean” which is more difficult than performing a standard Oaxaca decomposition for the mean.
- Active area of research over the last 20 years. A number of procedures are now available, though they tend to have their strengths and weaknesses. A few examples:
  - Juhn, Murphy, and Pierce (1993): Imputation method
  - DiNardo, Fortin, and Lemieux (1996): Reweighting method
  - Machado and Mata (2005): Quantile regression-based method

# Refresher on Oaxaca decomposition

- We want to decompose the difference in the mean of an outcome variable *Y* between two groups *A* and *B*
- Groups could also be periods, regions, etc.
- Postulate linear model for *Y*, with conditionally independent errors ($E(\upsilon|X)=0$):

$$Y_{gi} = \beta_{g0} + \sum_{k=1}^{K} X_{ik}\beta_{gk} + \upsilon_{gi}, \quad g = A, B,$$

- The estimated gap

$$\widehat{\Delta}_O^{\mu} = \overline{Y}_B - \overline{Y}_A,$$

- Can be decomposed as

$$\widehat{\Delta}_O^{\mu} = \underbrace{(\widehat{\beta}_{B0} - \widehat{\beta}_{A0}) + \sum_{k=1}^{K} \overline{X}_{Bk}\left(\widehat{\beta}_{Bk} - \widehat{\beta}_{Ak}\right)}_{\widehat{\Delta}_S^{\mu} \text{ (Unexplained)}} + \underbrace{\sum_{k=1}^{K} \left(\overline{X}_{Bk} - \overline{X}_{Ak}\right)\widehat{\beta}_{Ak}}_{\widehat{\Delta}_X^{\mu} \text{ (Explained)}}$$

# Technical note (algebra)

To get the Oaxaca decomposition start with (in simplified notation):

$$\Delta_O = Y_B - Y_A$$

Now use the regression models to get:

$$\Delta_O = X_B\beta_B - X_A\beta_A$$

and then add and subtract $X_B\beta_A$ :

$$\Delta_O = (X_B\beta_B - X_B\beta_A) + (X_B\beta_A - X_A\beta_A)$$

This yields

$$\Delta_O = X_B(\beta_B - \beta_A) + (X_B - X_A)\,\beta_A$$

where

$$\Delta_S = X_B(\beta_B - \beta_A)$$

and

$$\Delta_X = (X_B - X_A)\,\beta_A$$

# Some terminology

- The lectures will focus on wage decompositions
- The “explained” part of the decomposition will be called the **composition effect** (X) since it reflects differences in the distribution of the X’s between the two groups
- The “unexplained” part of the decomposition will be called the **wage structure** (S) effect as it reflects differences in the β’s, i.e. in the way the X’s are “priced” (or valued) in the labour market.
- Studies sometimes refer to “price” (β’s) and “quantity” (X’s) effects instead.
- In the “aggregate” decomposition, we only divide Δ into its two components $\Delta_S$ (wage structure effect) and $\Delta_X$ (composition effect).
- In the “detailed” decomposition we also look at the contribution of each individual covariate (or corresponding β)

# A few remarks

- We focus on this particular decomposition, but we could also change the order

$$\widehat{\Delta}_O^{\mu} = \left\{ (\widehat{\beta}_{B0} - \widehat{\beta}_{A0}) + \sum_{k=1}^{K} \overline{X}_{Ak} \left( \widehat{\beta}_{Bk} - \widehat{\beta}_{Ak} \right) \right\} + \left\{ \sum_{k=1}^{K} \left( \overline{X}_{Bk} - \overline{X}_{Ak} \right) \widehat{\beta}_{Bk} \right\}$$

- Or add an interaction term:

$$\begin{aligned} \widehat{\Delta}_O^{\mu} &= \left\{ (\widehat{\beta}_{B0} - \widehat{\beta}_{A0}) + \sum_{k=1}^{K} \overline{X}_{Ak} \left( \widehat{\beta}_{Bk} - \widehat{\beta}_{Ak} \right) \right\} + \\ & \qquad \left\{ \sum_{k=1}^{K} \left( \overline{X}_{Bk} - \overline{X}_{Ak} \right) \widehat{\beta}_{Ak} \right\} + \left\{ \sum_{k=1}^{K} \left( \overline{X}_{Bk} - \overline{X}_{Ak} \right) \left( \widehat{\beta}_{Bk} - \widehat{\beta}_{Ak} \right) \right\} \end{aligned}$$

- Does not affect the substance of the argument in most cases.
- The "intercept" component of $\Delta_S$, $\beta_{B0}$- $\beta_{A0}$, is the wage structure effect for the base group. Unless the other β's are the same in group *A* and *B*, $\beta_{B0}$- $\beta_{A0}$ will arbitrarily depend on the base group chosen.
- Well known problem, e.g. Oaxaca and Ransom (1999).

# Problems when going beyond the mean

- Things get more complicated in the case of the variance

$$\begin{aligned} Var(Y) &= \mathbb{E}[Var(Y|X)] + \mathbb{E}\left\{[\mathbb{E}(Y|X) - \mathbb{E}(Y)]^2\right\} \\ &= \mathbb{E}[Var(Y|X)] + \mathbb{E}\left\{[X\beta - \mathbb{E}(X)\beta]^2\right\} \\ &= \mathbb{E}[Var(Y|X)] + \beta' Var(X)\beta, \end{aligned}$$

- Where we now need to also model the conditional variance (Var(Y|X))
- Becomes even more complicated for more general distributional statistics such as quantiles, the Gini coefficient, etc.

# Identification: what can we estimate using decomposition?

- Computing a Oaxaca decomposition is easy enough
  - Run OLS and compute mean values of X for each of the two groups
  - In Stata, “oaxaca.ado” does that automatically and also yields standard errors that reflect the fact both the β’s and the mean values of the X’s are being estimated.
- But practitioners often wonder what we are really estimating for the following reasons (to list a few):
  - What if some the X variables (e.g. years of education) are endogenous?
  - What is there is some selection between (e.g. choice of being unionized or not) or within (participation rates for men and women) the two groups under consideration
  - What about general equilibrium effects? For example if we increase the level of human capital (say experience) of women to the level of men, this may depress the return to experience and invalidate the decomposition
- Key question: Can we construct a valid counterfactual?

# Estimating counterfactuals

Simplify the notation as follows:

$\Delta_O = Y_B - Y_A$

Now write:

$\Delta_O = Y_B - Y_A = (Y_B - Y^C) + (Y^C - Y_A)$,

where $Y^C$ is the counterfactual wage members of group B (say women) would earn if they were paid like men (group A). Under the linearity assumption (of the regression model), we get:

$Y_A = X_A\beta_A$, $Y_B = X_B\beta_B$, and $Y^C = X_B\beta_A$.

We then obtain the Oaxaca decomposition as follows:

$$\begin{aligned}\Delta_O &= (X_B\beta_B - X_B\beta_A) + (X_B\beta_A - X_A\beta_A) \\ &= X_B(\beta_B - \beta_A) + \beta_A(X_B - X_A)\end{aligned}$$

So with an estimate of the counterfactual $Y^C$ one can compute the decomposition. This is a general point that holds for all decompositions, and not only for the mean

# Counterfactual and treatment effects

- Generally speaking, one can think of treatment effects as the difference between how people are actually paid, and how they would be paid under a different "treatment".
- Take the case of unions.
  - $Y_A$ : Average wage of non-union workers
  - $Y_B$ : Average wage of union workers
  - $Y^C$ : Counterfactual wage union members would earn if they were not unionized
  - $\Delta_S = Y_B - Y^C$ : "union effect" on union workers is an average treatment effect on the treated (ATET)
- This means that the identification/estimation issue involved in the case of decomposition/counterfactual are the same as in the case of treatment effects.
- This is very useful since we can then use results from the treatment effects literature.

# The wage structure effect ($\Delta_S$) can be interpreted as a treatment effect

- The conditional independence assumption ($E(\varepsilon|X)=0$) usually invoked in Oaxaca decompositions can be replaced by the weaker ignorability assumption to compute the aggregate decomposition
- For example, ability ($\varepsilon$) can be correlated with education (X) as long as the correlation is the same in groups A and B.
- This is the standard assumption used in “selection on observables” models where matching methods are typically used to estimate the treatment effect.
- Main result: If we have $Y_G=m_G(X, \varepsilon)$ and ignorability, then:
  - $\Delta_S$ solely reflects changes in the m(.) functions (ATET)
  - $\Delta_X$ solely reflects changes in the distribution of X and $\varepsilon$ (ignorability key for this last result).

# The wage structure effect ($\Delta_S$) can be interpreted as a treatment effect

- A number of estimators for ATET= $\Delta_S$ have been proposed in the treatment effect literature
  - Inverse probability weighting (IPW), matching, etc.
- Formal results exist, e.g. IPW is efficient for
  - ATET (Hirano, Imbens, and Ridder, 2003)
  - Quantile treatment effects (Firpo, 2007)
- This has been widely used in the decomposition literature since DiNardo, Fortin, and Lemieux (1996).

# Formal derivation of the identification result (Handbook chapter)

ASSUMPTION 1 [***Mutually Exclusive Groups***] *The population of agents can be divided into two mutually exclusive groups, denoted $A$ and $B$. Thus, for an agent $i$, $D_{Ai}+D_{Bi}=1$, where $D_{gi}=\mathbb{I}\{i \text{ is in } g\}$, $g=A,B$, and $\mathbb{I}\{\cdot\}$ is the indicator function.*

ASSUMPTION 2 [***Structural Form***] *A worker $i$ belonging to either group $A$ or $B$ is paid according to the wage structure, $m_A$ and $m_B$, which are functions of the worker's observable $(X)$ and unobservable $(\varepsilon)$ characteristics:*

$$Y_{Ai}=m_A\left(X_i,\varepsilon_i\right) \quad \text{and} \quad Y_{Bi}=m_B\left(X_i,\varepsilon_i\right), \tag{3}$$

*where $\varepsilon_i$ has a conditional distribution $F_{\varepsilon|X}$ given $X$, and $g=A,B$.*

# Formal derivation (2)

ASSUMPTION 3 [***Simple Counterfactual Treatment***] *A counterfactual wage structure, $m^C$, is said to correspond to a simple counterfactual treatment when it can be assumed that $m^C(\cdot,\cdot) \equiv m_A(\cdot,\cdot)$ for workers in group $B$, or $m^C(\cdot,\cdot) \equiv m_B(\cdot,\cdot)$ for workers in group $A$.*

ASSUMPTION 4 [***Overlapping Support***]: *Let the support of all wage setting factors $[X', \varepsilon']'$ be $\mathcal{X} \times \mathcal{E}$. For all $[x', e']'$ in $\mathcal{X} \times \mathcal{E}$, $0 < \Pr[D_B = 1 | X = x, \varepsilon = e] < 1$.*

ASSUMPTION 5 [***Conditional Independence/Ignorability***]: *For $g = A, B$, let $(D_g, X, \varepsilon)$ have a joint distribution. For all $x$ in $\mathcal{X}$: $\varepsilon$ is independent of $D_g$ given $X = x$ or, equivalently, $D_g \perp\!\!\!\perp \varepsilon | X$.*

# Formal derivation (3)

Proposition 1 [***Identification of the Aggregate Decomposition***]**:**
*Under assumptions 3 (simple counterfactual), 4 (overlapping support), and 5 (ignorability), the overall* $\nu - gap$*,* $\Delta_O^\nu$*, can be written as*

$$\Delta_O^\nu = \Delta_S^\nu + \Delta_X^\nu,$$

*where*
*(i) the wage structure term* $\Delta_S^\nu = \nu(F_{Y_B|D_B}) - \nu(F_{Y_A^C:X=X|D_B})$ *solely reflects difference between the structural functions* $m_B(\cdot,\cdot)$ *and* $m_A(\cdot,\cdot)$
*(ii) the composition effect term* $\Delta_X^\nu = \nu(F_{Y_A^C:X=X|D_B}) - \nu(F_{Y_A|D_A})$ *solely reflects the effect of differences in the distribution of characteristics (*$X$ *and* $\varepsilon$*) between the two groups*

# Some intuition about Proposition 1

- The wage setting model we use, $y_g = m_g(x, \varepsilon)$ is very general
  - Includes the linear model $y_g = x\beta_g + \varepsilon$ as a special case
- There are three reasons why wages can be different between groups g=a and g=b:
  - Differences in the wage setting equations $m_a(.)$ and $m_b(.)$
  - Differences in the distribution of X for the two groups
  - Differences in the distribution of $\varepsilon$ for the two groups
- The ignorability assumption states that the distribution of $\varepsilon$ given X is the same for the two groups, though this does not mean that $E(\varepsilon|X)=0$.
- So once we control for differences between the X's in the two groups, we also implicitly control for differences in the $\varepsilon$'s.
- Only source of difference left is, thus, differences in the wage structures $m_a(.)$ and $m_b(.)$.

# A few caveats discussed in the chapter

- This general result (Proposition 1) only works for the aggregate decomposition. More assumptions have to be imposed to get at the detailed decomposition.
- We are implicitly ruling out general equilibrium effects under assumption (simple counterfactual treatment). For instance, in the absence of unions, wages in the nonunion sector may change as firms are no longer confronted with the threat of unionization
- The wage structure observed among nonunion workers, $m_a$ (.), is no longer a valid counterfactual for union workers.
- This is closely related to the difference between union wage gap (no general equilibrium effects) and union wage gain (possible general equilibrium effects) discussed by H. Gregg Lewis decades ago

# Lecture 2: Decomposing Changes (or Differences) in Distributions

Nicole Fortin and Thomas Lemieux, UBC

CLSRN Summer School

HÉC Montréal, June 27 2011

# Plan of the lecture

- Examples of distributional questions
- Quantile regressions: analogy with standard regressions doesn’t work
- Constructing counterfactual distributions
- Aggregate decomposition: reweighting procedure
- Detailed decomposition:
  - Conditional reweighting
  - Quantile-regression based decomposition (Machado-Mata)
  - Distributional regressions (Chernozhukov et al.)
  - RIF-regression (Firpo, Fortin, and Lemieux)
- Empirical application and implementation in Stata

# Example of distributional questions

- Glass ceiling: Is the gender gap higher in the upper part of the distribution (e.g. Albrecht and Vroman)?
- What explains the growth in earnings inequality in the United States (large literature) and other countries?
- Changes over time in the world distribution of income (twin peaks, etc.)
- Changes or differences in the distribution of firm size.

# The analogy with quantile regressions does not work

- Tempting to run quantile regressions (say for the median) and perform a decomposition as in the case of the mean (Oaxaca)
- Does not work because there are two interpretations to β for the mean
  - Conditional mean: $E(Y|X) = X\beta$
  - Uncond. mean (LIE): $E(Y) = E_X(E(Y|X)) = E_X(X)\beta$
- But the LIE does not work for quantiles
  - Conditional quantile: $Q_\tau(Y|X) = X\beta_\tau$
  - Uncond. quantile: $Q_\tau \neq E_X(Q_\tau(Y|X)) = E_X(X)\beta_\tau$
- Only the first interpretation works for $\beta_\tau$, which is not useful for decomposing unconditional quantiles (unless one estimates quantiles regressions for all quantiles as in Machado-Mata)

# Constructing counterfactual distributions

The distribution of $Y_g|D_g$ is defined using the law of iterated probabilities, that is, after we integrate over the observed characteristics we obtain

$$F_{Y_g|D_g}(y) = \int F_{Y_g|X,D_g}(y|X=x) \cdot \mathrm{d}F_{X|D_g}(x), \quad g = A, B. \tag{4}$$

We can construct the counterfactual distribution as follows:

$$F_{Y_A^C:X=X|D_B} = \int F_{Y_A|X,D_A}(y|X=x) \cdot \mathrm{d}F_{X|D_B}(x). \tag{5}$$

Or in simpler notation:

$$F_{Y_A^C}(y) = \int F_{Y_A|X_A}(y|X) dF_{X_B}(X). \tag{27}$$

The idea is to integrate group A's conditional distribution of Y given X over group B's distribution of X

# Constructing counterfactual distributions

- There are two general ways of estimating the counterfactual distribution
- First approach is to start with group B and replace the conditional distribution $F_{Y_B|X_B}(y|X)$ with $F_{Y_A|X_A}(y|X)$
  - This requires estimating the whole conditional distribution.
  - Machado and Mata (2005) do so by estimating quantile regressions for all quantiles, and inverting back
- Second approach is the opposite. Start with group A but replace the distribution of X for group A ($F_{X_A}(X)$) by the distribution of X for group B ($F_{X_B}(X)$).
- This is simpler since this only depends on X, while the conditional distribution depends on both X and Y.
- Even better, all we need to do is to compute a reweighting factor

# Constructing counterfactual distributions

$$F_{Y_A^C}(y) = \int F_{Y_A|X_A}(y|X)\Psi(X)dF_{X_A}(X), \tag{28}$$

where $\Psi(X) = dF_{X_B}(X)/dF_{X_A}(X)$ is a reweighting factor

- This is the approach suggested by DiNardo, Fortin and Lemieux (1996)
- The reweighting factor can be estimated using a simple logit (or probit) model since after some manipulations we get

$$\Psi(X) = \frac{\Pr(X|D_B = 1)}{\Pr(X|D_B = 0)} = \frac{\Pr(D_B = 1|X)/\Pr(D_B = 1)}{\Pr(D_B = 0|X)/\Pr(D_B = 0)}.$$

- To estimate Prob($D_B$=1|X), pool the two groups and estimate a logit for the probability of belonging to group B as a function of X

# Reweighting procedure

The reweighting decomposition procedure can be implemented in practice as follows:

1. Pool the data for group $A$ and $B$ and run a logit or probit model for the probability of belonging to group $B$:

$$\Pr(D_B = 1|X) = 1 - \Pr(D_B = 0|X) = 1 - \Pr(\varepsilon > -h(X)\beta) = \Lambda(-h(X)\alpha) \qquad (30)$$

where $\Lambda()$ is either a normal or logit link function, and $h(X)$ is a polynomial in $X$.

2. Estimate the reweighting factor $\widehat{\Psi}(X)$ for observations in group $A$ using the predicted probability of belonging to group $B$ $(\widehat{\Pr}(D_B = 1|X))$ and $A$ $(\widehat{\Pr}(D_B = 0|X) = 1 - \widehat{\Pr}(D_B = 1|X))$, and the sample proportions in group $B$ $(\widehat{\Pr}(D_B = 1))$ and $A$ $(\widehat{\Pr}(D_B = 0))$:

$$\widehat{\Psi}(X) = \frac{\widehat{\Pr}(D_B = 1|X)/\widehat{\Pr}(D_B = 1)}{\widehat{\Pr}(D_B = 0|X)/\widehat{\Pr}(D_B = 0)}.$$

3. Compute the counterfactual statistic of interest using observations from the group A sample reweighted using $\widehat{\Psi}(X)$.

# Reweighting procedure

- Very easy to use regardless of the distribution statistics being considered (variance, Gini coefficient, inter-quartile range, etc.)
- Just compute the statistic for group A, group B, and group A reweighted to have the distribution of X of group B.
- Example: in the case of the Gini coefficient, this yields estimates of $Gini_A$, $Gini_B$, and $Gini_A^C$
- We can then estimate the composition effect as:

$$\Delta_X = Gini_A^C - Gini_A$$

- And the wage structure effect as

$$\Delta_S = Gini_B – Gini_A^C$$

- Standard errors can be computed using a bootstrap procedure (bootstrap the whole procedure starting with the logit)

# Going beyond the mean is a "solved" problem for the aggregate decomposition

- Can directly apply non-parametric methods (reweighting, matching, etc.) from the treatment effect literature.
- Ignorability is crucial, but $m_G(X, \varepsilon)$ does not need to be linear
- Inference by bootstrap or analytical standard errors in the case of reweighting ("generated regressor" correction required)
- Reweighting very easy to use with large and well behaved (no support problem) data sets.

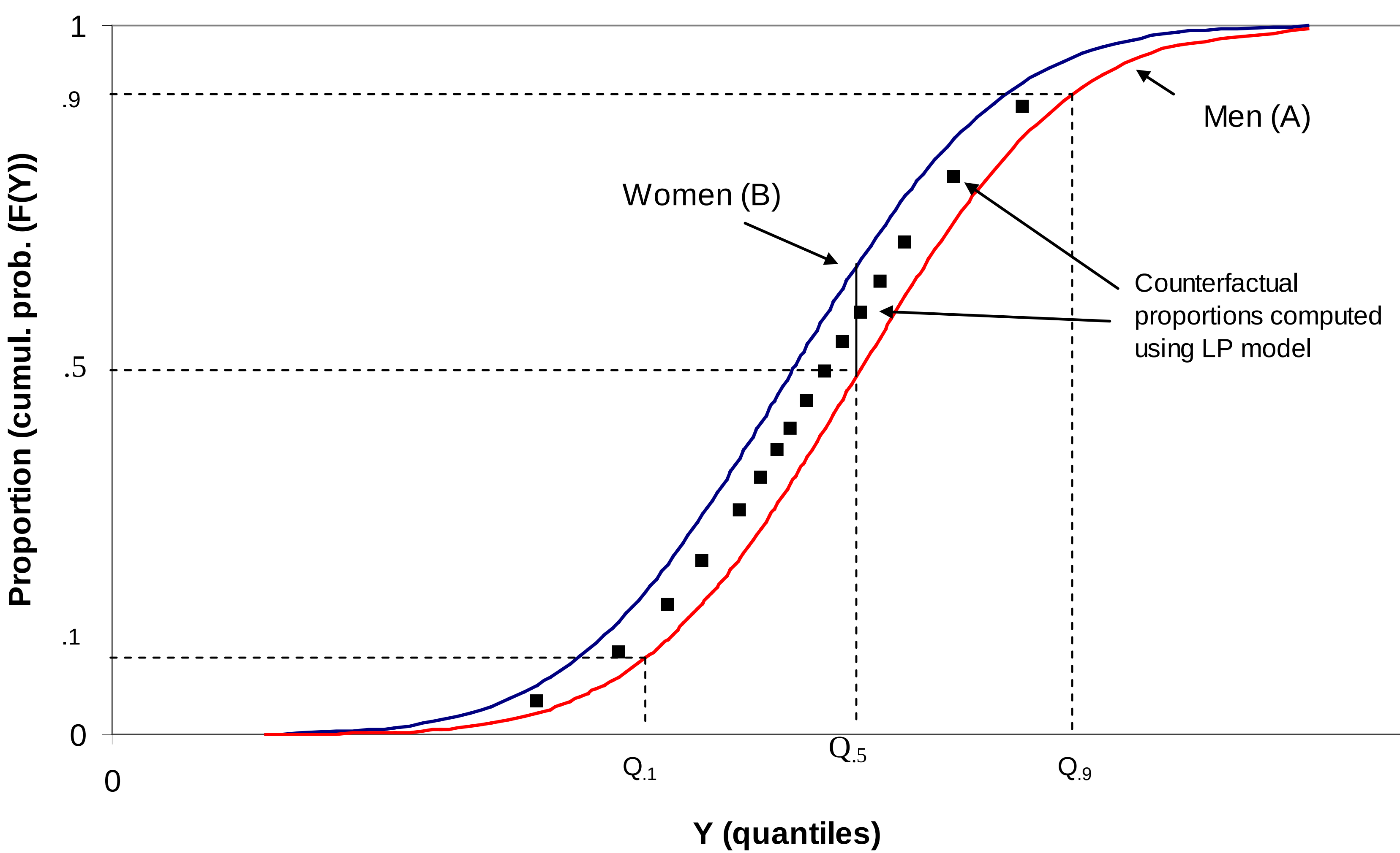
# Going beyond the mean is more difficult for the detailed decomposition

- Until recently, there were only a few partial (and not always satisfactory) ways of performing a detailed decomposition for general distributional measures (quantiles in particular):
  - DFL conditional reweighting for the components of $\Delta_X$ linked to dummy covariates (e.g. unions)
  - Machado-Mata quantile regressions for components of $\Delta_S$.
  - Sequential DFL-type reweighting, adding one covariate at a time. Sensitive to order used as in a simple regression.
- A more promising approach is to estimate for proportions, and invert back to quantiles. RIF regression of Firpo, Fortin and Lemieux (2009) is one possible way of doing so

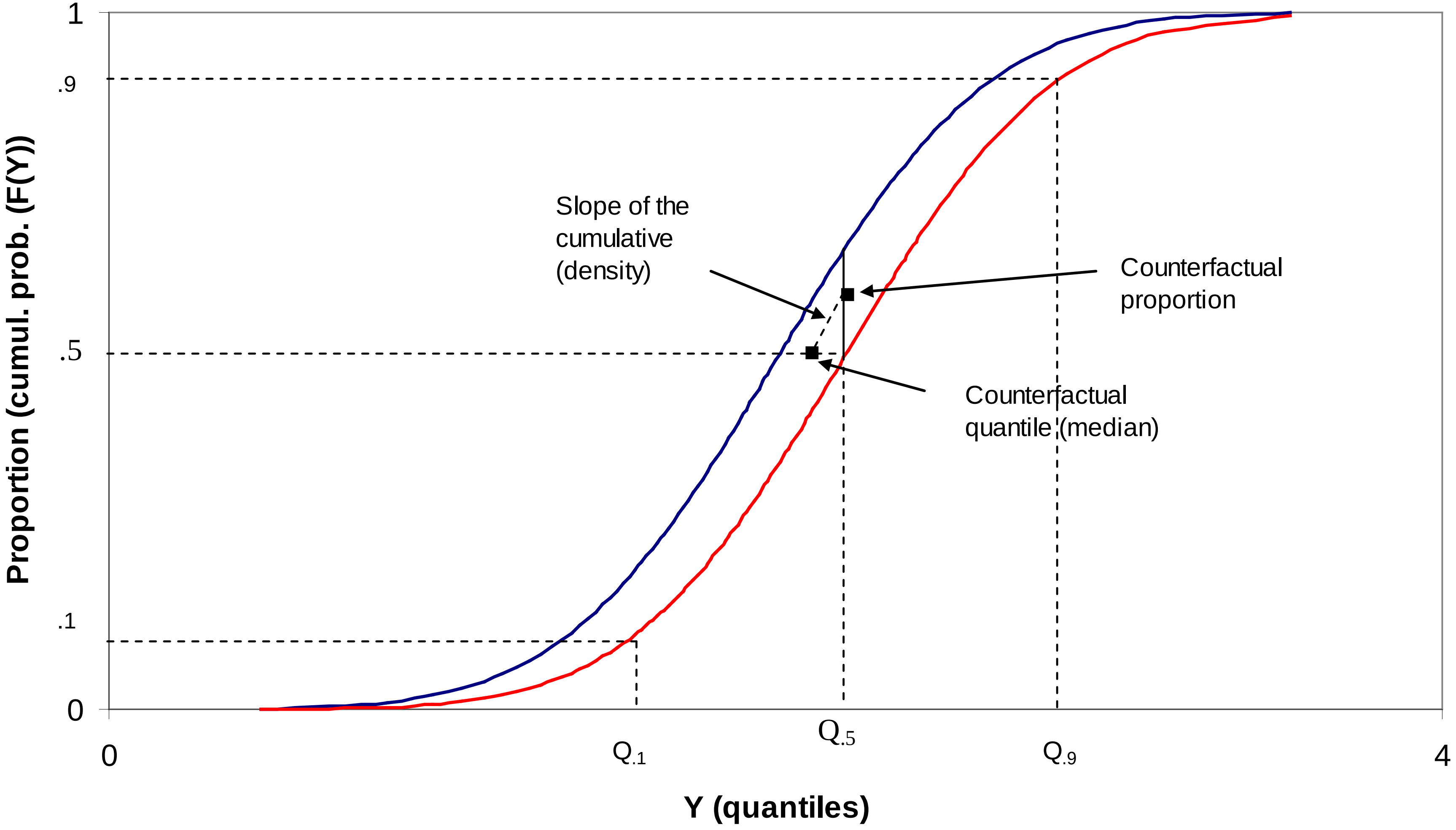
# Decomposing proportions is easier than decomposing quantiles

- Example: 10 percent of men earn more than 80K a year, but only 5 percent of women do so.
- Easy to do a decomposition by running LP models for the probability of earning less (or more) than 80K, and perform a Oaxaca decomposition on the proportions.
- By contrast, much less obvious how to decompose the difference between the 90$^{th}$ quantile for men (80K) and women (say 65K)
- But function linking proportions and quantiles is the cumulative distribution.
- Counterfactual proportions → Counterfactual cumulative → Counterfactual quantiles
- Can be illustrated graphically

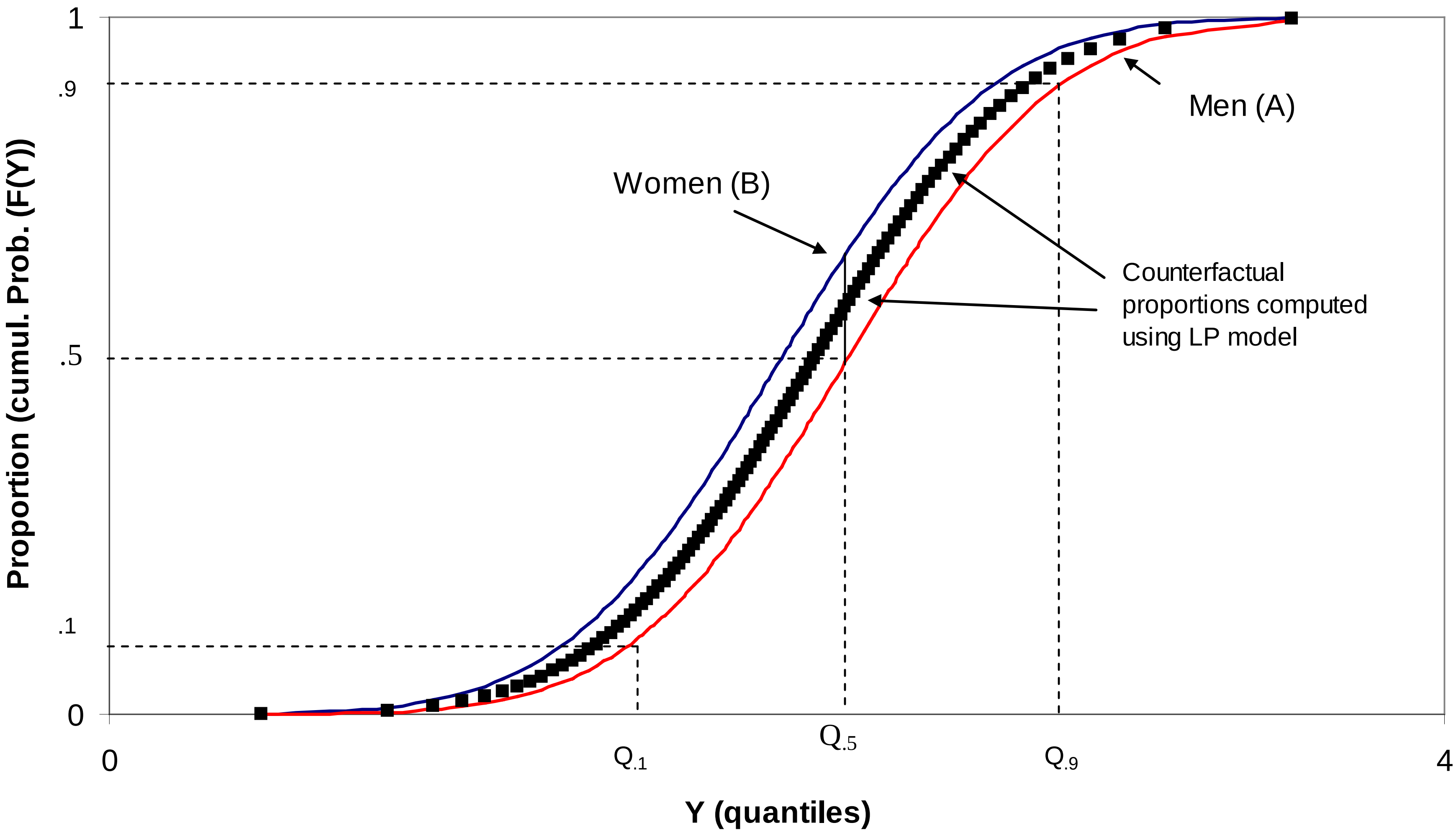
# Figure 1: Relationship Between Proportions and Quantiles



## Figure 2: RIF Regressions: Inverting Locally



## Figure 3: Inverting Globally



# Decomposing proportions is easier than decomposing quantiles

- FFL recentered influence function (RIF) regressions
  - Run LP models (or logit/probit) for being below a given quantiles, and divide by density (slope of cumulative) to locally invert.
  - Dependent variable is dummy $1(Y<Q_\tau)$ divided by density → influence function for the quantile.
  - RIF approach works for other distributional measures (Gini, variance, etc.)
- Chernozhukov et al. (2009)
  - Estimate “distributional regressions” (LP, logit or probit) for each value of Y (say at each percentile)
  - Invert back globally to recover counterfactual quantiles

# More on RIF regressions

- In the case of quantiles, the RIF is:

$$\mathrm{RIF}(y; Q_\tau) = Q_\tau + \frac{\tau - \mathbb{1}\{y \leq Q_\tau\}}{f_Y(Q_\tau)}$$

- Similar RIF can be obtained for other distributional statistics such as a Gini coefficient
- Unlike quantile regressions, an important property is that:

$$E(\mathrm{RIF}(y,Q_\tau)) = Q_\tau$$

- So if we have a regression model like

$$E[\mathrm{RIF}(y,Q_\tau)|X] = X\gamma$$

- We can do a standard Oaxaca decomposition using the fact that

$$Q_\tau = E(\mathrm{RIF}(y,Q_\tau)) = E_X[E[\mathrm{RIF}(y,Q_\tau)|X]] = E[X]\gamma$$