

Sources of Inequality: Additive Decomposition of the Gini Coefficient.

Carlos Hurtado

Econometrics Seminar

Department of Economics University of Illinois at Urbana-Champaign hrtdmrt2@illinois.edu

Feb 24th, 2016

C. Hurtado (UIUC - Economics)

Sources of Inequality

Motivation

- Many different factors are behind the distribution of economic welfare:
 - The distribution of initial endowments and sociodemographic characteristics.
 - The returns (prices or weights) of those initial endowments and individual characteristics.
 - The interaction between the previous two factors.
- Those factors are not independent of each other, and it is difficult to measure their impact on the inequality of the distribution. Example: Wages.

Motivation

- Using statistical decomposition techniques to identify the main causes of distributional differences in wages started with the methods proposed by [Oaxaca, 1973] and [Blinder, 1973].
- Considerable research has been devoted to go beyond the analysis of the mean differences, e.g. [Juhn et al., 1993], [DiNardo et al., 1996] or [Machado and Mata, 2005].
- The use of the entire distribution is important to understand the differences on the bottom or top part of the distribution of study.
- One step further is to understand the impact of individual characteristics on measurements of inequality of the distribution of study.

Key Question and Methodology

Key Question:

The aim of this paper is to propose a method to measure the contribution of various factors to the disparity of the distribution of wages and propose a decomposition of the changes on the distribution of wages using counterfactual scenarios.

Methodology:

- Base on the relation between the Lorentz curve and the conditional quantile function, it is possible to relate the Gini index with the quantile regression.
- For a given year, I propose a methodology to estimate the impact of each covariate on the Gini index by a polynomial approximation of the estimates of quantile regression coefficients.
- Using the estimates of the Gini index for different years, I propose a decomposition of the changes on the distribution of wages using counterfactual scenarios.
- An empirical application is developed using the Ongoing Rotation Group of the Current Population Survey (CPS) for 1986 and 1995.

Main Results

- Using the proposed method it is possible identify by quartiles the individual characteristics that contribute most in order to increase or decrease the inequality of the distribution (log) wages.
- The method complements other approaches that have been developed to answer the same question.
- Returns to schooling change in the US between 1986 and 1995, specially on the top of the distribution.
- ► The wage gap between man and woman has been reduced in the US between 1986 and 1995, except on the top of the distribution
- The change in the proportion of High School graduates between 1986 and 1995 has increased the inequality in the distribution of wages whereas the change in the proportion of people with Thechnical Careers has reduced the inequality of wages.

On the Agenda

- Ourly Wage Series From the CPS
- 3 Additive Decomposition of the Gini Coefficient
- 4 Estimation Procedure
- 5 Empirical Application



On the Agenda

- 2 Hourly Wage Series From the CPS
- 3 Additive Decomposition of the Gini Coefficient
- 4 Estimation Procedure
- 5 Empirical Application
- 6 Conclusions

- It is well documented that during the 1970's the differences in wage by education and occupation narrowed in the US, followed by a constant increase in wage inequality starting in the early 1980's up to the present e.g. [Katz, 1999], [Levy and Murnane, 1992].
- There have been an increasing number of approaches trying to disentangle the factors that contribute to the differences in wages. For a review see [Fortin et al., 2011].
- Given the classical model, part of the literature has focus its attention on the average wage, after controlling for *individual and institutional characteristics*, e.g. [Bound and Johnson, 1992], [Card and Lemieux, 2001],[Blau and Kahn, 1996].

Going Beyond the Analysis of the Mean

- Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach [DiNardo et al., 1996]
 - Semiparametric procedure to estimate the effect of institutional and labor market factors on the distribution of wages.
 - Applying kernel density methods to appropriately weighted samples, the authors estimate the effects of these factors.
 - The authors find *visual and quantitative* evidence that the decline in the real value of the minimum wage explains the increase in wage inequality in the US, particularly for women, of the wages from 1979 to 1988.

Going Beyond the Analysis of the Mean

- Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression [Machado and Mata, 2005]
 - Method to decompose the changes in the wage distribution over a period of time in several factors contributing to those changes.
 - Based on the estimation of marginal wage distributions consistent with a conditional distribution estimated by quantile regression as well as with the distribution of the covariates.
 - The authors apply this method to Portuguese data for the period 1986-1995, and find that the increase in educational levels contributed towards greater wage inequality.

Going Beyond the Analysis of the Mean

- The visual evidence presented by kernel estimates, or the analysis of some quantiles of the distribution, may be hard to interpret or may miss information that a measurement of inequality summarizes.
- It is possible to estimate the impact of various factors on the conditional quantile of (log) wages using quantile regression and link those estimates with the Gini index.
- The previous relationship makes possible to measure the effect of the factors that impact the distribution of wages <u>without modeling the density</u>, but rather directly measuring its inequality through the Gini index.

On the Agenda

Introduction

O Hourly Wage Series From the CPS

3 Additive Decomposition of the Gini Coefficient

4 Estimation Procedure

5 Empirical Application

6 Conclusions

- Starting in 1979, workers in the Ongoing Rotation Group (ORG) of the CPS are asked detailed questions related to earnings from work.
- Using the answers to hourly earnings, or weekly earnings divided by usual hours work per week, it is possible to compute hourly wages as a good measure of the price of labor.
- ► A difficulty of using the ORG is that the CPS classifies and processes differently the earnings of *hourly paid* and *non-hourly paid* workers throughout the time span.
- In an effort to construct consistent wages using the ORG of the CPS, the Center for Economic and Policy Research (CEPR) has developed publicly available code that I updated to create a consistent hourly wage series form 1980 to 2015.

- Wages updated to constant dollars of 2015 using the CPI reported by the Bureau of Labor Statistics
- Keep workers with hourly wage between \$1 and \$100 (in 1979 dollars) and with ages between 16 and 65 years.
- Potential experience: computed using the individuals age and subtracting years of education and five years before elementary school.
- Consistent classification for twenty industries for all the years of analysis.
- Classification of years of education:
 - non-school or dropouts: between zero and eleven years of education
 - high school: twelve years of education
 - some college: between thirteen and fifteen years of education
 - college: sixteen or more years of education

Hourly Wage Series From the CPS



Kernel density estimates for women's real log wages (\$2015)

Sources of Inequality



C. Hurtado (UIUC - Economics)

Sources of Inequality

| | | Men | | | | | | | | Women | | | | |
|------|---------------------|------------------|-----------|----------|-----------|------------------|----------|------------------|-----------|----------|-----------|------------------|----------|--|
| Year | Log Real Minimum | Log Real Wage | $Union^B$ | Nonwhite | Education | $Experience^{C}$ | No. Obs. | Log Real Wage | $Union^B$ | Nonwhite | Education | $Experience^{C}$ | No. Obs. | |
| | $Wage^{A}$ | | | | | | | | | | | | | |
| 1980 | 2.19 | 3.11 | | 0.17 | 12.75 | 18.26 | 106,936 | 2.70 | | 0.18 | 12.76 | 17.64 | 87,887 | |
| 1981 | 2.17 | 3.09 | | 0.18 | 12.82 | 18.19 | 99,530 | 2.69 | | 0.18 | 12.80 | 17.68 | 83,220 | |
| 1982 | 2.11 | 3.09 | | 0.18 | 12.92 | 18.23 | 92,249 | 2.71 | | 0.19 | 12.91 | 17.68 | 79,586 | |
| 1983 | 2.08 | 3.09 | 0.28 | 0.18 | 12.99 | 18.11 | 91,049 | 2.73 | 0.18 | 0.19 | 12.99 | 17.59 | 79,013 | |
| 1984 | 2.03 | 3.08 | 0.26 | 0.18 | 13.02 | 17.91 | 92,729 | 2.73 | 0.17 | 0.19 | 13.03 | 17.53 | 80,787 | |
| 1985 | 2.00 | 3.09 | 0.25 | 0.20 | 13.03 | 18.02 | 93,763 | 2.75 | 0.16 | 0.20 | 13.07 | 17.58 | 82,817 | |
| 1986 | 1.98 | 3.10 | 0.24 | 0.20 | 13.07 | 17.95 | 92,081 | 2.77 | 0.16 | 0.20 | 13.12 | 17.64 | 83,452 | |
| 1987 | 1.94 | 3.09 | 0.23 | 0.21 | 13.08 | 18.00 | 92,005 | 2.78 | 0.15 | 0.21 | 13.15 | 17.69 | 84,676 | |
| 1988 | 1.90 | 3.08 | 0.23 | 0.22 | 13.11 | 17.99 | 88,084 | 2.78 | 0.15 | 0.21 | 13.19 | 17.78 | 81,193 | |
| 1989 | 1.86 | 3.10 | 0.22 | 0.22 | 13.14 | 18.09 | 89,459 | 2.79 | 0.15 | 0.22 | 13.23 | 18.01 | 82,979 | |
| 1990 | 1.93 | 3.08 | 0.21 | 0.23 | 13.12 | 18.05 | 93,500 | 2.79 | 0.15 | 0.23 | 13.27 | 18.01 | 87,322 | |
| 1991 | 2.00 | 3.07 | 0.21 | 0.24 | 13.18 | 18.25 | 90,127 | 2.80 | 0.15 | 0.23 | 13.33 | 18.25 | 85,313 | |
| 1992 | 1.97 | 3.06 | 0.21 | 0.24 | 13.00 | 18.55 | 88,358 | 2.81 | 0.15 | 0.23 | 13.14 | 18.64 | 84,513 | |
| 1993 | 1.94 | 3.05 | 0.20 | 0.24 | 13.06 | 18.59 | 86,804 | 2.82 | 0.15 | 0.23 | 13.20 | 18.78 | 83,902 | |
| 1994 | 1.92 | 3.05 | 0.20 | 0.24 | 13.09 | 18.62 | 82,354 | 2.83 | 0.15 | 0.24 | 13.24 | 18.81 | 80,342 | |
| 1995 | 1.89 | 3.05 | 0.19 | 0.24 | 13.12 | 18.74 | 82,510 | 2.81 | 0.14 | 0.24 | 13.26 | 18.96 | 79,873 | |
| 1996 | 1.97 | 3.04 | 0.19 | 0.26 | 13.12 | 18.98 | 73,034 | 2.81 | 0.14 | 0.25 | 13.30 | 19.08 | 71,468 | |
| 1997 | 2.03 | 3.05 | 0.18 | 0.27 | 13.10 | 19.09 | 74,576 | 2.83 | 0.14 | 0.26 | 13.30 | 19.27 | 72,763 | |

 A2015 Constant Dollars

^BUnion status of workers was not collected in the outgoing rotation group supplements from 1980 to 1982. However, using the May pension supplement it may be possible estimate this summary statistic for a subsample of the population

 $^{C}\operatorname{Potential}$ experience is computed as age - years of education - 5

Summary Statistics

| | | | | | Men | | Women | | | | | | |
|------|--|------------------|--------------------|----------|-----------|------------------|----------|------------------|--------------------|----------|-----------|------------------|----------|
| Year | Log Real Minimum Wage ^A | Log Real Wage | Union ^B | Nonwhite | Education | $Experience^{C}$ | No. Obs. | Log Real Wage | Union ^B | Nonwhite | Education | $Experience^{C}$ | No. Obs. |
| 1998 | 2.01 | 3.09 | 0.18 | 0.27 | 13.13 | 19.22 | 75,589 | 2.86 | 0.13 | 0.27 | 13.32 | 19.34 | 73,450 |
| 1999 | 1.99 | 3.12 | 0.18 | 0.27 | 13.18 | 19.34 | 76,746 | 2.88 | 0.13 | 0.27 | 13.35 | 19.45 | 74,432 |
| 2000 | 1.96 | 3.13 | 0.17 | 0.28 | 13.18 | 19.46 | 77,712 | 2.89 | 0.13 | 0.28 | 13.36 | 19.60 | 75,166 |
| 2001 | 1.93 | 3.14 | 0.17 | 0.28 | 13.23 | 19.73 | 82,348 | 2.92 | 0.13 | 0.28 | 13.41 | 19.85 | 80,038 |
| 2002 | 1.91 | 3.16 | 0.16 | 0.28 | 13.26 | 20.01 | 87,798 | 2.94 | 0.13 | 0.28 | 13.46 | 20.09 | 86,464 |
| 2003 | 1.89 | 3.15 | 0.16 | 0.31 | 13.23 | 20.16 | 85,502 | 2.95 | 0.13 | 0.30 | 13.50 | 20.46 | 85,298 |
| 2004 | 1.87 | 3.15 | 0.15 | 0.32 | 13.25 | 20.25 | 84,260 | 2.94 | 0.13 | 0.30 | 13.54 | 20.56 | 83,181 |
| 2005 | 1.83 | 3.14 | 0.15 | 0.32 | 13.24 | 20.47 | 84,803 | 2.94 | 0.13 | 0.31 | 13.58 | 20.69 | 83,569 |
| 2006 | 1.80 | 3.14 | 0.14 | 0.33 | 13.26 | 20.52 | 84,987 | 2.95 | 0.12 | 0.31 | 13.61 | 20.79 | 82,819 |
| 2007 | 1.90 | 3.14 | 0.14 | 0.33 | 13.31 | 20.60 | 83,717 | 2.95 | 0.13 | 0.32 | 13.67 | 20.87 | 82,302 |
| 2008 | 1.98 | 3.14 | 0.15 | 0.33 | 13.40 | 20.81 | 82,286 | 2.96 | 0.13 | 0.32 | 13.75 | 21.04 | 81,588 |
| 2009 | 2.08 | 3.17 | 0.15 | 0.33 | 13.46 | 21.15 | 78,864 | 2.98 | 0.13 | 0.32 | 13.80 | 21.31 | 79,945 |
| 2010 | 2.06 | 3.16 | 0.14 | 0.33 | 13.49 | 21.27 | 78,083 | 2.98 | 0.13 | 0.32 | 13.85 | 21.44 | 78,894 |
| 2011 | 2.03 | 3.13 | 0.14 | 0.34 | 13.52 | 21.19 | 77,769 | 2.97 | 0.13 | 0.32 | 13.89 | 21.50 | 77,611 |
| 2012 | 2.01 | 3.14 | 0.13 | 0.35 | 13.56 | 21.35 | 78,070 | 2.96 | 0.12 | 0.35 | 13.92 | 21.53 | 76,725 |
| 2013 | 2.00 | 3.13 | 0.13 | 0.36 | 13.58 | 21.35 | 78,030 | 2.97 | 0.12 | 0.35 | 13.99 | 21.45 | 76,301 |
| 2014 | 1.98 | 3.13 | 0.13 | 0.37 | 13.59 | 21.35 | 78,714 | 2.97 | 0.12 | 0.36 | 14.02 | 21.35 | 76,464 |
| 2015 | 1.98 | 3.15 | 0.13 | 0.37 | 13.63 | 21.30 | 77,740 | 2.99 | 0.12 | 0.37 | 14.06 | 21.29 | 75,537 |

 A2015 Constant Dollars

^BUnion status of workers was not collected in the outgoing rotation group supplements from 1980 to 1982. However, using the May pension supplement it may be possible estimate this summary statistic for a subsample of the population

 $^{C}\operatorname{Potential}$ experience is computed as age - years of education - 5

Summary Statistics

On the Agenda

Introduction

2 Hourly Wage Series From the CPS

3 Additive Decomposition of the Gini Coefficient

- 4 Estimation Procedure
- 5 Empirical Application

6 Conclusions

The Lorenz Curve and the Gini Coefficient

▶ The relation between the index and curve: Gini coefficient is twice the area between the line of equality (45°) and the Lorenz curve



Cumulative share of people from lowest to highest income or wages 100%

C. Hurtado (UIUC - Economics)

Sources of Inequality

The Lorenz Curve and the Gini Coefficient

▶ Following [Koenker, 2005], the Lorentz curve can be express as

$$L(\tau) = \frac{\int_{0}^{\tau} Q_{Y}(t) dt}{\int_{0}^{1} Q_{Y}(t) dt} = \frac{1}{\mu} \int_{0}^{\tau} Q_{Y}(t) dt$$

where, Y is a positive random variable, quantile function $Q_Y(t)$, and mean $0 < \mu < \infty$.

▶ Let $h(\cdot)$ be a monotone transformation such that, $h(Y) \ge 0$ and $0 < \mu_h < \infty$, with $\mu_h = E[h(y)]$. Then,

$$L_h(\tau) = \frac{1}{\mu_h} \int_0^\tau Q_{h(Y)}(t) dt \tag{1}$$

The Lorenz Curve and the Gini Coefficient

A researcher may model the conditional quantile function, given a vector x ∈ ℝ^P of covariates, as

$$Q_{h(Y)}(t|x) = x^{T}\beta(t) = \sum_{j=1}^{P} x_{j}\beta_{j}(t), \qquad (2)$$

where each $\beta_j(t)$ is the coefficient corresponding to the characteristic j at the *t*-th quantile, and $t \in (0, 1)$.

> The conditional Lorenz curve of the transformed variable is reduced to

$$L_{h}(\tau|x) = \frac{1}{\mu_{h}} \int_{0}^{\tau} Q_{h(Y)}(t|x) dt = \frac{1}{\mu_{h}} \sum_{j=1}^{P} x_{j} \int_{0}^{\tau} \beta_{j}(t) dt.$$
(3)

The Gini Coefficient and the Quantile Regression

► The Gini index given a vector of covariates, x ∈ ℝ^P, can be derived by replacing the conditional Lorenz curve into the definition of the Gini coefficient, to obtain

$$G_{h}(x) = 1 - 2 \int_{0}^{1} L_{h}(\tau | x) d\tau$$

= $1 - \frac{1}{\mu_{h}} \sum_{j=1}^{P} x_{j} \int_{0}^{1} \int_{0}^{\tau} 2\beta_{j}(t) dt d\tau.$ (4)

► Equation (4) is an additive decomposition of the Gini index. We can use it to investigate the evolution of changes in the distribution of h(Y) as a function of the initial endowments and sociodemographic characteristics, x_j , as well as the returns (prices or weights) of these endowments and characteristics, $\frac{1}{\mu_h} \int_0^1 \int_0^\tau 2\beta_j(t) dt d\tau$.

Impact of Individual Characteristics on the Gini Index

Caveats

- ► The linear decomposition of the Gini index is computed for the variable h(Y).
- Possible conflict between the statistical objective and the economic objective of study.
- An interesting extension is to understand the impact of factors on the untransformed variable.
- We have a link between quantile regression and the Gini coefficient, but a natural extension of the analysis is to perform counterfactual scenarios using a hypothesized distribution for the covariates for different years, similar to [Machado and Mata, 2005].

Estimates of the Gini Index

Denote by

$$\frac{\prod_j}{\mu_h} \equiv \frac{1}{\mu_h} \int_0^1 \int_0^\tau 2\beta_j(t) dt d\tau$$

An estimate of the Gini index can be computed as

$$\hat{G}_h = 1 - \sum_{j=1}^P \bar{X}_j \frac{\hat{\Pi}_j}{\hat{\mu}_h}$$

where \bar{X}_j is the weighted average of covariate j, and $\hat{\mu}_h$ and $\hat{\Pi}_j$ are estimates for μ_h and $\int_0^1 \int_0^\tau 2\beta_j(t) dt d\tau$ respectively.

On the Agenda

Introduction

- 2 Hourly Wage Series From the CPS
- 3 Additive Decomposition of the Gini Coefficient
- 4 Estimation Procedure
- 5 Empirical Application

6 Conclusions

▶ To implement the procedure we would like to estimate

$$\Pi_j = \int\limits_0^1 \int\limits_0^ au 2eta_j(t) dt d au;$$

To implement the procedure we would like to estimate

$$\Pi_j = \int\limits_0^1 \int\limits_0^ au 2eta_j(t) dt d au;$$



C. Hurtado (UIUC - Economics)

Sources of Inequality

To implement the procedure we would like to estimate

$${\sf \Pi}_j = \int\limits_0^1 \int\limits_0^ au 2eta_j(t) dt d au;$$



C. Hurtado (UIUC - Economics)

$$\hat{\Pi}_j = \int\limits_0^1 \int\limits_0^\tau 2\hat{eta}_j(t) dt d au;$$

- A polynomial approximation for β_j(t) provides a close form for its antiderivative.
- A value for $\hat{\Pi}_j$ can be easily obtain from that approximation.

Bootstrap:

Let N be the sample size and R the number of repetitions for the bootstrap.

- 1. In each iteration re-sample N observations with replacement.
- 2. Using the re-sample estimate the quantile regression coefficients.
- 3. Use polynomial approximation to calculate $\hat{\Pi}_j$.
- 4. To compute the point estimate use the average of the $\hat{\Pi}_j$.

- Is this methodology accurate?
 - Simulate data with known quantile process, $\beta(t)$, and estimate the known impact of the *j*-th covariate, $\prod_i \land Accurate?$
- Degree of polynomial approximation?
 - No problem, as long as the polynomial approximation is inside the confidence band.
 Degree?
- Do we learn something new?
 - Let me spend the remainder of this presentation contrasting the results from this methodology and those from [Machado and Mata, 2005]

On the Agenda

- 2 Hourly Wage Series From the CPS
- 3 Additive Decomposition of the Gini Coefficient
- 4 Estimation Procedure
- 5 Empirical Application
- 6 Conclusions

- What do we learn from the US data?
- set t = 0 to be 1986 and t = 1 to be 1995.
- Get a subsample of size 80,000 for each year.
- Define the model as:

$$Q_t(\log(w)|x;t) = x'\beta(u_i)$$

where x is a vector that contains individual characteristics on unionization status, potential experience and its square, classification of schooling, non-white dummy, female dummy, part time dummy, marital status and index variables for 3 regions and 19 industries.

















- ► ABC of [Machado and Mata, 2005]:
 - If U is a uniform random variable on [0, 1], then $F^{-1}(U)$ has distribution F.
 - 1. Random sample of size *m* form $U[0,1]: u_1, \cdots, u_m$.
 - Estimate Q_{ui}(w|x; t), the conditional quantile model of log wages, yielding to m estimates of the QR coefficients β^t(u_i) = (β^t₁(u_i), · · · , β^t_p(u_i)).
 - Generate a random sample of size m with replacement from the covariates matrix X(t), denoted by {x_i^{*}(t)}_{i=1}^m.
 - 4. Finally, generate a random sample of log wages that is consistent with the conditional distribution defined by the model: $\{w_i^*(t) \equiv x_i^*(t)' \hat{\beta}^t(u_i)\}_{i=1}^m$

▶ ABC of [Machado and Mata, 2005]

- Counterfactual Densities:

To generate a random sample from the marginal wage distribution that would have prevailed in t = 1 if all covariates had been distributed as in t = 0, just follow the algorithm but drawing the bootstrap sample of the third step from the rows of X(0).

- Here they are assuming that workers had been paid according to the t = 1 schedule, but characteristics are as in t = 0.

- ► ABC of [Machado and Mata, 2005]
 - Counterfactual Densities: effect of an individual covariate ?
 - * Let y(t) denote one particular covariate of interest at time t.
 - * Create a partition of y(t) into J classes, $C_1(t), \cdots, C_J(t)$.
 - * Denote by $f_j(t)$, for $j = 1, \dots, J$, the relative frequency of class $C_j(t)$.
 - Follow steps 1 to 4 as before to generate {w_i^{*}(1)}_{i=1}^m, a random sample of size m of the wage density at t = 1.
 - 2. Take one class, say $C_1(1)$
 - Let *I*₁ = {*i* = 1, · · · , *m*|*y_i*(1) ∈ *C*₁(1)}; select the subset of the random sample generated in step 1 corresponding to *I*₁.
 - Generate a random sample of size $m \times f_1(0)$ with replacement from $\{w_i^*(1)\}_{i \in I_1}$.
 - 3. Repeat step 2 for $j = 2 \cdots, J$.

- ► ABC of [Machado and Mata, 2005]
 - Decomposing the Changes in the Wage Density:
 - Observed sample $\{w_i(t)\}$: Denote by f(w(t)) an estimator of the marginal density.
 - Generated sample $\{w_i^*(t)\}$: Denote by $f^*(w(t))$ an estimator of the marginal density implied by the model.
 - Counterfactual density: Denoted by $f^*(w(1); X(0))$. This is if all covariates had their t = 0 distribution.
 - Counterfactual density: Denoted by $f^*(w(1); y(0))$. If only one covariate had the distribution as in t = 0.

- ► ABC of [Machado and Mata, 2005]
 - Decomposing the Changes in the Wage Density:
 - Let $\alpha(\cdot)$ denote a summary statistic.

$$\alpha(f(w(1))) - \alpha(f(w(0))) = \underbrace{\alpha(f^*(w(1); X(0))) - \alpha(f^*(w(0)))}_{\text{coefficients}} + \underbrace{\alpha(f^*(w(1))) - \alpha(f^*(w(1); X(0)))}_{\text{covariates}}$$

• The contribution of an individual covariate

$$\alpha(f^*(w(1))) - \alpha(f^*(w(1); y(0)))$$

individual covariate

Results using [Machado and Mata, 2005]



(log) wage densities for 1988 and 1995 (\$2015)

C. Hurtado (UIUC - Economics)

Sources of Inequality

Results using [Machado and Mata, 2005]



Coefficients: (log) wage densities for 1988 and counterfactual 1995 (\$2015)

Results using [Machado and Mata, 2005]



Covariates: (log) wage densities for 1988 and counterfactual 1995 (\$2015)

| | | Margii | nals | Aggreg | gate contributio | ns |
|--------------|------------|------------|---------------|---------------|------------------|---------|
| | 1986^{A} | 1995^{A} | Change | Cova | Coef | Res |
| 1st quant. | 1.469 | 1.498 | 0.029 | -0.005 | 0.044 | -0.0100 |
| | | | -0.322; 0.362 | -0.240; 0.188 | -0.147; 0.191 | |
| | | | | -0.167 | 1.509 | -0.342 |
| 10th quant. | 2.065 | 2.051 | -0.014 | 0.011 | -0.037 | 0.0122 |
| | | | -0.067; 0.028 | -0.056; 0.072 | -0.094; 0.031 | |
| | | | | -0.793 | 2.662 | -0.869 |
| 25th quant. | 2.380 | 2.351 | -0.029 | 0.006 | -0.062 | 0.0273 |
| | | | -0.092; 0.008 | -0.044; 0.059 | -0.117; -0.008 | |
| | | | | -0.209 | 2.164 | -0.955 |
| Median | 2.799 | 2.744 | -0.054 | 0.018 | -0.072 | 0.0002 |
| | | | -0.115; 0.008 | -0.031; 0.064 | -0.128; -0.028 | |
| | | | | -0.324 | 1.328 | -0.004 |
| 75th quant. | 3.221 | 3.175 | -0.047 | 0.038 | -0.086 | 0.0009 |
| | | | -0.109; 0.029 | -0.017; 0.094 | -0.134; -0.031 | |
| | | | | -0.809 | 1.827 | -0.019 |
| 90th quant. | 3.543 | 3.533 | -0.010 | 0.051 | -0.081 | 0.0194 |
| | | | -0.088; 0.091 | -0.020; 0.122 | -0.159; -0.008 | |
| | | | | -4.953 | 7.846 | -1.893 |
| 99th quant. | 3.988 | 4.094 | 0.106 | 0.026 | 0.009 | 0.0708 |
| | | | 0.02; 0.314 | -0.127; 0.196 | -0.105; 0.173 | |
| | | | | 0.250 | 0.082 | 0.668 |
| Gini of logW | 11.381 | 11.737 | 0.356 | 0.241 | -0.094 | 0.2087 |
| | | | -0.392; 1.061 | -0.382; 0.835 | -0.719; 0.493 | |
| | | | | 0.677 | -0.264 | 0.586 |
| | | | | | | |

^A2015 Constant Dollars

C. Hurtado (UIUC - Economics)

| | Marginals | | | Individual Covariates | | | | | | |
|--------------|------------|-------------------|------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|---|----------------------------------|----------------------------------|
| | 1986^{A} | 1995 ^A | Change | Gender | Union | Race | High School | Some Coll. | College | Exper. |
| 1st quant. | 1.469 | 1.498 | 0.029 -0.322;0.362 | 0.002 -0.251;0.251 | -0.055 -0.288;0.195 | 0.082 -0.185;0.350 | -0.005 -0.242;0.224 | 0.002 -0.257;0.232 | -0.014 -0.270;0.193 | 0.044 -0.218;0.269 |
| 10th quant. | 2.065 | 2.051 | -0.014 -0.067;0.028 | 0.065 -0.027 -0.102;0.052 | -1.875 -0.025 -0.095;0.060 | 2.804 0.001 -0.079;0.085 | -0.164 0.005 -0.071;0.082 | 0.067 0.003 -0.074;0.077 | -0.482 -0.004 -0.096;0.070 | 1.504 0.024 -0.057;0.101 |
| 25th quant. | 2.380 | 2.351 | -0.029 -0.092;0.008 | -0.022 -0.082;0.042 0.783 | -0.013 -0.077;0.055 0.438 | -0.016 -0.067;0.056 0.553 | -0.007 -0.065;0.058 0.252 | -0.131 0.005 -0.051;0.071 -0.173 | -0.013 -0.072;0.053 0.438 | 0.006 -0.044;0.072 -0.196 |
| Median | 2.799 | 2.744 | -0.054 -0.115;0.008 | -0.016 -0.081;0.047 0.296 | -0.016 -0.077;0.047 0.292 | -0.004 -0.074;0.057 0.075 | -0.006 -0.068;0.052 0.106 | -0.009 -0.076;0.053 0.166 | 0.007 -0.058;0.067 -0.132 | 0.023 -0.036;0.084 -0.420 |
| 75th quant. | 3.221 | 3.175 | -0.047 -0.109;0.029 | -0.019 -0.094;0.047 0.398 | -0.019 -0.092;0.047 0.409 | -0.012 -0.074;0.053 0.261 | -0.004 -0.074;0.064 0.095 | -0.029 -0.099;0.041 0.622 | -0.006 -0.067;0.073 0.119 | 0.014 -0.051;0.094 -0.303 |
| 90th quant. | 3.543 | 3.533 | -0.010 -0.088;0.091 | -0.032 -0.120;0.060 3.070 | -0.016 -0.096;0.067 1.561 | 0.004 -0.077;0.096 -0.425 | -0.017 -0.087;0.067 1.695 | -0.033 -0.115;0.062 3.203 | 0.004 -0.085;0.095 -0.415 | 0.018 -0.068;0.105 -1.715 |
| 99th quant. | 3.988 | 4.094 | $0.106 \\ 0.02; 0.314$ | -0.024 -0.208;0.179 -0.226 | -0.068 -0.283;0.138 -0.643 | -0.032 -0.239;0.164 -0.300 | -0.053 -0.250;0.133 -0.499 | -0.032 -0.223;0.173 -0.300 | -0.053 -0.282;0.163 -0.500 | -0.024 -0.188;0.189 -0.224 |
| Gini of logW | 11.381 | 11.737 | 0.356 -0.392;1.061 | -0.012 -0.742;0.667 -0.033 | 0.015 -0.678;0.716 0.043 | -0.061 -0.767;0.590 -0.171 | -0.114 -0.802;0.589 -0.320 | -0.255 -0.965;0.479 -0.715 | 0.115 -0.627;0.844 0.323 | -0.114 -0.861;0.550 -0.319 |

 $^{A}2015$ Constant Dollars

Note 1: The first entry in each cell is the point estimated in the change in the attribute of the density, explained by the indicated factor

Note 2: The second entry is the 95% confidence interval for the change

Note 3: The third entry is the proportion of the total change explained by the indicated factor

 We can investigate the *changes in the wage density* following the framework of [Machado and Mata, 2005]

$$Gini(w(1)) - Gini(w(0)) = \underbrace{Gini(w(1); X(0)) - Gini(w(0))}_{\text{coefficients}} + \underbrace{Gini(w(1)) - Gini(w(1); X(0))}_{\text{covariates}} + \text{residual}$$

► The contribution of an individual covariate

$$\hat{Gini}(w(1)) - \hat{Gini}(w(1); y(0))$$

| | | Margir | nals | Aggreg | Aggregate contributions | | | | |
|--------------|--------|--------|---------------|---------------|-------------------------|---------|--|--|--|
| | 1986 | 1995 | Change | Cova | Coef | Res | | | |
| 1st Quartile | 1.882 | 1.862 | -0.0200 | -0.0015 | 0.0810 | -0.0995 | | | |
| | | | -0.169; 0.119 | -0.071; 0.064 | 0.069; 0.094 | | | | |
| | | | | 0.076 | -4.057 | 4.980 | | | |
| 2nd Quartile | 3.815 | 3.857 | 0.0421 | -0.0473 | 0.2021 | -0.1127 | | | |
| | | | -0.201; 0.273 | -0.267; 0.161 | 0.169; 0.234 | | | | |
| | | | | -1.125 | 4.806 | -2.680 | | | |
| 3rd Quartile | 3.799 | 3.966 | 0.1662 | -0.1430 | 0.3177 | -0.0084 | | | |
| | | | -0.094; 0.428 | -0.528; 0.224 | 0.266; 0.368 | | | | |
| | | | | -0.861 | 1.911 | -0.051 | | | |
| 4th Quartile | 1.884 | 2.051 | 0.1675 | -0.2917 | 0.4127 | 0.0465 | | | |
| | | | 0.017; 0.306 | -0.850; 0.264 | 0.341; 0.481 | | | | |
| | | | | -1.742 | 2.464 | 0.278 | | | |
| Total | 11.381 | 11.737 | 0.356 | -0.4836 | 1.0135 | -0.1741 | | | |
| | | | -0.392; 1.061 | -1.706; 0.697 | 0.848; 1.173 | | | | |
| | | | | -1.359 | 2.848 | -0.489 | | | |

<u>Note 1:</u> The first entry in each cell is the point estimated in the change in the attribute of the density, explained by the indicated factor

Note 2: The second entry is the 95% confidence interval for the change

Note 3: The third entry is the proportion of the total change explained by the indicated factor

| | | Margin | nals | | | Inc | dividual Covari | iates | | |
|--------------|------------|------------|---------------|---------------|---------------|---------------|-----------------|----------------|---------------|---------------|
| | 1986^{A} | 1995^{A} | Change | Gender | Union | Race | High School | Some Coll. | College | Exper. |
| 1st Quartile | 1.882 | 1.862 | -0.0200 | 0.0036 | 0.0136 | 0.0086 | 0.0212 | -0.0323 | -0.0094 | -0.0399 |
| | | | -0.169; 0.119 | -0.010; 0.018 | -0.002; 0.030 | -0.001; 0.019 | 0.008; 0.036 | -0.051; -0.014 | -0.049; 0.028 | -0.089; 0.012 |
| | | | | -0.183 | -0.683 | -0.432 | -1.062 | 1.618 | 0.473 | 1.997 |
| 2nd Quartile | 3.815 | 3.857 | 0.0421 | 0.0117 | 0.0421 | 0.0288 | 0.0693 | -0.1117 | -0.0354 | -0.1484 |
| | | | -0.201; 0.273 | -0.032; 0.057 | -0.007; 0.092 | -0.003; 0.063 | 0.025; 0.118 | -0.176; -0.049 | -0.183; 0.104 | -0.330; 0.046 |
| | | | | 0.279 | 1.001 | 0.685 | 1.647 | -2.657 | -0.842 | -3.530 |
| 3rd Quartile | 3.799 | 3.966 | 0.1662 | 0.0210 | 0.0669 | 0.0498 | 0.1270 | -0.2123 | -0.0677 | -0.2901 |
| | | | -0.094; 0.428 | -0.058; 0.102 | -0.011; 0.146 | -0.006; 0.108 | 0.046; 0.217 | -0.334; -0.094 | -0.350; 0.199 | -0.645; 0.089 |
| | | | | 0.126 | 0.403 | 0.299 | 0.764 | -1.277 | -0.407 | -1.745 |
| 4th Quartile | 1.884 | 2.051 | 0.1675 | 0.0311 | 0.0848 | 0.0688 | 0.1949 | -0.3316 | -0.1040 | -0.4609 |
| | | | 0.017; 0.306 | -0.086; 0.151 | -0.014; 0.185 | -0.008; 0.149 | 0.071; 0.333 | -0.522; -0.146 | -0.539; 0.305 | -1.025; 0.141 |
| | | | | 0.186 | 0.507 | 0.411 | 1.163 | -1.980 | -0.621 | -2.752 |
| Total | 11.381 | 11.737 | 0.356 | 0.0675 | 0.2075 | 0.1559 | 0.4123 | -0.6880 | -0.2166 | -0.9393 |
| | | | -0.392; 1.061 | -0.186; 0.327 | -0.034; 0.452 | -0.019; 0.339 | 0.150; 0.705 | -1.082; -0.303 | -1.121;0.635 | -2.088; 0.288 |
| | | | | 0.190 | 0.583 | 0.438 | 1.159 | -1.934 | -0.609 | -2.640 |

 $^{A}2015$ Constant Dollars

Note 1: The first entry in each cell is the point estimated in the change in the attribute of the density, explained by the indicated factor

Note 2: The second entry is the 95% confidence interval for the change

Note 3: The third entry is the proportion of the total change explained by the indicated factor

On the Agenda

- 2 Hourly Wage Series From the CPS
- 3 Additive Decomposition of the Gini Coefficient
- 4 Estimation Procedure
- 5 Empirical Application



Conclusions

- Using the proposed method it is possible identify by quartiles the individual characteristics that contribute most in order to increase or decrease the inequality of the distribution (log) wages.
- The method complements other approaches that have been developed to answer the same question.
- Returns to schooling change in the US between 1986 and 1995, specially on the top of the distribution.
- ► The wage gap between man and woman has been reduced in the US between 1986 and 1995, except on the top of the distribution
- The change in the proportion of High School graduates between 1986 and 1995 has increased the inequality in the distribution of wages whereas the change in the proportion of people with Thechnical Careers has reduced the inequality of wages.

References I



Blau, F. and Kahn, L. (1996).

International differences in male wage inequality: Institutions versus market forces. *Journal of Political Economy*, 104(4):791–837.



Blinder, A. (1973).

Wage discrimination: Reduced form and structural estimates.

Journal of Human Resources, 8:436-455.



Bound, J. and Johnson, G. (1992).

Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations.

American Economic Review, 82(3):371-92.



Can falling supply explain the rising return to college for younger men? a cohort-based analysis.

The Quarterly Journal of Economics, 116(2):705–746.

References II

| 14 | |
|----|---|
| | _ |
| | |
| | |
| | |
| | |
| | |
| | |

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996).

Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach.

Econometrica, 64:1001-1044.



Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101:410–442.

Katz, L. F. (1999).

Changes in the wage structure and earnings inequality. *Handbook of labor economics*, 3:1463–1555.

References III



Koenker, R. (2005).

Inequality measures and their decomposition.

In Quantile regression. Cambridge University Press, Cambridge.



Levy, F. and Murnane, R. J. (1992).

Us earnings levels and earnings inequality: A review of recent trends and proposed explanations.

Journal of economic literature, 30(3):1333–1381.



Machado, J. A. and Mata, J. (2005).

Counterfactual decomposition of changes in wage distributions using quantile regression.

Journal of Applied Econometrics, 20:445-465.



Oaxaca, R. (1973).

Male-female wage differentials in urban labor markets.

International Economic Review, 14:693–709.

$$Q_{\log(w)}(t|x) = eta_1(t)$$
age + $eta_2(t)$ age²

$$eta_1(t) = 0.2t + 0.05t^2$$

 $eta_2(t) = -0.0023t - 0.0003t^2$

$$\Pi_j = \int_0^1 \int_0^\tau 2\beta_j(t) dt d\tau$$

 $\Pi_1 = 0.075 \\ \Pi_2 = -0.00082$



Bootstrap: Let N = 8,000 be the sample size and R = 1,000 the number of repetitions for the bootstrap.

- 1. In each iteration re-sample N observations with replacement.
- 2. Using the re-sample estimate the quantile regression coefficients.
- 3. Use polynomial approximation of degree 6 to calculate $\hat{\Pi}_i$.
- 4. To compute the point estimate use the average of the $\hat{\Pi}_j$ and use the variance to calculate the confidence intervals.

| | Π_j | P_6 | Q_1 | Q_2 | Q_3 | Q_4 |
|------------------|----------|-------------|-----------|------------|-------------|-------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| age | 0.075 | 0.079*** | 0.003 | 0.032*** | 0.094*** | 0.185*** |
| | | (0.017) | (0.002) | (0.011) | (0.023) | (0.033) |
| age ² | -0.00082 | -0.00084*** | -0.00003 | -0.00034** | -0.00101*** | -0.00197*** |
| | | (0.00024) | (0.00003) | (0.00015) | (0.00033) | (0.00047) |

<u>Note 1:</u> In all the estimations the number of observations used is 8,000. The bootstrap procedure to estimate columns (2) to (6) uses resample with replacement. All estimates in columns (2) to (6) are computed at the same time.

Note 2: The impact estimates in columns (2) to (6) present standard errors in parenthesis.

<u>Note 3:</u> * p < 0.1; ** p < 0.05; *** p < 0.01



▶ What do we learn from the US data?



▶ What do we learn from the US data?



▶ What do we learn from the US data?



▶ What do we learn from the US data?



▶ What do we learn from the US data?





▶ What do we learn from the US data?

