Decomposition of Differences in Distribution under Sample Selection and the Gender Wage Gap

Santiago Pereda Fernández

*Universidad de Cantabria

EEA-ESEM Congress 2023
29/08/2023

Motivation

- Decomposition methods useful to understand sources of differences between two groups.
- Originally done for mean outcomes in linear models under ignorability assumption: Oaxaca (1973), Blinder(1973).
- Two components: endowments and coefficients.
  - Endowments captures differences in explanatory variables.
  - Coefficients captures differences in returns to explanatory variables (structural): often interpreted as discrimination, differences in preferences, or average treatment effect (Kline, 2011).

Motivation

- Decomposition methods have been extended to cover other outcomes and models (Fortin et al., 2011).
- Decomposition of outcomes other than the mean, most notably the unconditional distributions:
  - Quantile regression: Machado and Mata (2005), Melly (2006), Chernozhukov et al. (2013).
  - Distribution regression: Chernozhukov et al. (2013).
  - Reweighting methods: DiNardo et al. (1996), Firpo et al. (2018).
- Decomposition in models with selection on unobservables:
  - Linear models: Neuman and Oaxaca (2003,2004), Mora (2008), Cukrowska-Torzewska and Lovasz (2016), Huber et al. (2020).
  - Nonlinear models: Maasoumi and Wang (2017,2019), Chernozhukov et al. (2019).

Motivation

- Maasoumi and Wang (2017):
  - Estimation of the distribution of potential outcomes for the entire population (using reweighting methods).
  - Bounds for the decomposed components.
  - Decomposition into two components (same as Oaxaca-Blinder).
- Maasoumi and Wang (2019):
  - Estimation of the distribution of potential outcomes for the entire population (using quantile regression methods).
  - Comparison of the gap (for the mean and several other functions) if everybody participated.
  - Empirical paper on gender wage gap (not theoretical).
- Chernozhukov et al. (2019):
  - Estimation of the distribution of observed outcomes for the selected sample (using distribution regression methods).
  - Decomposition into four components (2 from Oaxaca-Blinder+sorting+"employment structure").
  - Selection modeled using *local* copulas.

This paper

- Model sample selection with quantiles (Arellano and Bonhomme, 2017).
  - Same as in Maasoumi and Wang (2019).
  - Clearer interpretation than Chernozhukov et al. (2019).
  - Provide uniform asymptotic theory.
- Consider two types of decompositions:
  - Distribution of actual outcomes for participants.
  - Distribution of actual outcomes for the entire population.
  - They depend on four components (2 from Oaxaca-Blinder+participation+self-selection).
- Consider two ancillary decompositions:
  - Participation decomposition.
  - Self-selection decomposition.
- Review gender wage gap estimates by Maasoumi and Wang (2019) & decompositions.

Roadmap of the presentation

1. Introduction
2. Model
3. Decompositions
4. Estimation
5. Application
6. Participation
7. Self-Selection
8. Earnings
9. Conclusion

Model and components

- Sample selection model:

$$Y = g_D(X, U)\, S$$
$$S = \mathbf{1}(\pi_D(Z) - V > 0)$$

- $Y \equiv$ outcome (continuous).
- $X \equiv$ predetermined covariates.
- $Z \equiv (Z_1, X)$ instrument for participation+predetermined covariates.
- $S \equiv$ participation indicator.
- $U \equiv$ unobservable r.v. of outcome equation.
- $V \equiv$ unobservable r.v. of participation equation.
- $D \equiv$ group indicator.
- $\pi_D \equiv$ propensity score.
- $g_d \equiv$ structural quantile function.

Model and components

- In a gender wage gap analysis:
- $Y \equiv$ log wage.
- $X \equiv$ education, marital status, age polynomial, race, regional dummies,...
- $Z_1 \equiv$ has any children less than 5 years old.
- $S \equiv 1$ if employed.
- $U \equiv$ unobserved ability.
- $V \equiv$ unobserved propensity to participate.
- $D \equiv$ gender indicator.
- $\pi_D(Z) \equiv$ probability to work for person with characteristics Z.
- $g_d(X, U) \equiv$ potential (latent) wage for a person with characteristics X, U.

Unobservables

- V can be normalized to be uniform (Heckman and Vytlacil, 2005).
- U can also be normalized $\rightarrow$ Skorohod representation: $U|X \sim U(0,1)$
- Can interpret U as conditional (on X) rank or quantile.
- Normalization without loss of generality (Lemma 1).
- Joint distribution independent $\Leftrightarrow$ lack of self-selection on unobservables.
- Model joint distribution with a copula: $C_{d,x}(u,v) \equiv \mathbb{P}(U \leq u, V \leq v | D = d, X = x)$ (definition).
  - Negative correlation of the copula $\Leftrightarrow$ positive selection into participation.
  - Copula informative of potential outcomes for nonparticipants.

Structural components

- Structural Quantile Function: $g_d(x, u)$.
- Distribution of the observables: $F_Z^d(z)$.
- Propensity score: $\pi_d(z)$.
- Conditional (on X=x) copula of the unobservables: $C_{d,x}(u, v)$.
- Conditional (on S=1 and X=x) copula of the unobservables: $G_{d,x}(u, v) \equiv \mathbb{P}(U \leq u | D = d, X = x, V \leq v) = \frac{C_{d,x}(u,v)}{v}$.

Primitives of the decompositions

- Let $h, k, l, m = \{0, 1\}$.
- Mean outcome (for participants and full population):

$$\mathbb{E}\left[Y^{hklm}|S=1\right] \equiv \int_{\mathcal{Z}} \int_0^1 g_k(x,u)\, dG_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

$$\mathbb{E}\left[Y^{hklm}\right] \equiv \int_{\mathcal{Z}} \int_0^1 g_k(x,u)\, dC_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

- Unconditional distribution (for participants and full population):

$$F_{Y|S=1}^{hklm}(y) \equiv \int_{\mathcal{Z}} \int_0^1 \mathbf{1}(g_k(x,u) \leq y)\, dG_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

$$F_Y^{hklm}(y) \equiv \int_{\mathcal{Z}} \left[\int_0^1 \mathbf{1}(g_k(x,u) \leq y)\, dC_{l,x}(u, \pi_m(z)) + (1 - \pi_m(z))\right] dF_Z^h(z)$$

- Unconditional quantile function (for participants and full population):

$$Q_{Y|S=1}^{hklm}(\tau) \equiv \inf\left\{y : \tau \leq F_{Y|S=1}^{hklm}(y)\right\}$$

$$Q_Y^{hklm}(\tau) \equiv \inf\left\{y : \tau \leq F_Y^{hklm}(y)\right\}$$

Particular cases

- This model nests several others considered in the literature.
- Feasible to do same decompositions under simplifying assumptions:
  - Additively separable unobserved term (Appendix).
  - Linear model (Appendix).
  - No self-selection (Appendix).

Mean decomposition for participants

- By definition, $\mathbb{E}[Y|D=1,S=1]=\mathbb{E}\left[Y^{1111}|S=1\right]$ and $\mathbb{E}[Y|D=0,S=1]=\mathbb{E}\left[Y^{0000}|S=1\right]$.
- Decompose the difference between these two as:

$$
\begin{aligned}
\mathbb{E}[Y|D=1,S=1]-\mathbb{E}[Y|D=0,S=1]=&\underbrace{\mathbb{E}\left[Y^{1111}|S=1\right]-\mathbb{E}\left[Y^{0111}|S=1\right]}_{\text{endowments component}}\\
&+\underbrace{\mathbb{E}\left[Y^{0111}|S=1\right]-\mathbb{E}\left[Y^{0011}|S=1\right]}_{\text{coefficients component}}\\
&+\underbrace{\mathbb{E}\left[Y^{0011}|S=1\right]-\mathbb{E}\left[Y^{0001}|S=1\right]}_{\text{selection component}}\\
&+\underbrace{\mathbb{E}\left[Y^{0001}|S=1\right]-\mathbb{E}\left[Y^{0000}|S=1\right]}_{\text{participation component}}
\end{aligned}
$$

Interpretation of the components

- Endowments component: overall gap explained by changes in distribution of observable characteristics.
  - *The group with better average characteristics has a higher mean outcome.*
- Coefficients component: overall gap explained by differences in SQF.
  - *The group with highest returns to characteristics has a higher mean outcome.*
- Selection component: overall gap explained by differences in the amount of self-selection (copula).
  - *The group with higher unobserved ability for participants has a higher mean outcome.*
- Participation component: overall gap explained by differences in structural propensity to participate.
  - *With positive (negative) selection, the group with lowest (highest) participation has a higher mean unobserved ability and outcome.*

Additional decompositions

- The same decomposition can be applied to the entire population by using $\mathbb{E}\left[Y^{hklm}\right]$ instead.
  - Participation component has two effects: as participation increases, less individuals with 0 outcome, so higher mean outcome; in addition, changes in average U.

- Similar decompositions can be performed for the unconditional distribution and functionals of it.
  - Leading case, unconditional quantile distribution:

$$Q_{Y|D=1,S=1}(\tau) - Q_{Y|D=0,S=1}(\tau) = \underbrace{Q^{1111}_{Y|S=1}(\tau) - Q^{0111}_{Y|S=1}(\tau)}_{\text{endowments component}} + \underbrace{Q^{0111}_{Y|S=1}(\tau) - Q^{0011}_{Y|S=1}(\tau)}_{\text{coefficients component}}$$
$$+ \underbrace{Q^{0011}_{Y|S=1}(\tau) - Q^{0001}_{Y|S=1}(\tau)}_{\text{selection component}} + \underbrace{Q^{0001}_{Y|S=1}(\tau) - Q^{0000}_{Y|S=1}(\tau)}_{\text{participation component}}$$

Ancillary decompositions

- Participation decomposition: Oaxaca-Blinder with a (possibly) non-linear propensity score.

- Self-selection decomposition.
  - Average value of U for $h, l, m = \{0, 1\}$:
    $\mathbb{E}\left[U^{hlm}|S=1\right] \equiv \int_{\mathcal{Z}} \int_0^1 u dG_{l,x}\left(u, \pi_m(z)\right) dF_Z^h(z)$.
  - Decomposition:

$$\mathbb{E}\left[U|D=1, S=1\right] - \mathbb{E}\left[U|D=0, S=1\right] = \underbrace{\mathbb{E}\left[U^{111}|S=1\right] - \mathbb{E}\left[U^{011}|S=1\right]}_{\text{endowments component}}$$
$$+ \underbrace{\mathbb{E}\left[U^{011}|S=1\right] - \mathbb{E}\left[U^{001}|S=1\right]}_{\text{selection component}}$$
$$+ \underbrace{\mathbb{E}\left[U^{001}|S=1\right] - \mathbb{E}\left[U^{000}|S=1\right]}_{\text{participation component}}$$

Estimation

- Let $\upsilon_\ell(z,\tau,\eta,f) \equiv \left(g_k(x,\tau), c_{l,x}(\tau,\eta), \pi_m(z), \int_{\mathcal{Z}} f dF_Z^h\right)$, where $\ell \equiv (h,k,l,m)$.

Condition 1

The estimator of the components of the decomposition for the general model, $\hat{\upsilon}_\ell(z,\tau,\eta,f)$, satisfies the law $\sqrt{n}\left(\hat{\upsilon}_\ell(z,\tau,\eta,f) - \upsilon_\ell(z,\tau,\eta,f)\right) \Rightarrow \mathbb{Z}_{\upsilon_\ell}(z,\tau,\eta,f)$ for all ℓ, where $\mathbb{Z}_{\upsilon_\ell}(z,\tau,\eta,f)$ is a zero-mean Gaussian process.

- This is satisfied by the Rotated Quantile Regression estimator (Arellano and Bonhomme, 2017; Theorem 2 & Corollary 1).
  1. The propensity score is calculated using common methods (e.g., probit, logit, linear probability model).
  2. The quantiles are specified linearly, i.e., $g_d(x,\tau) = x'\beta_d(\tau)$.
  3. The copula is modeled parametrically (many choices: Gaussian, t, Frank, Clayton, Gumbel, Bernstein,...).
- Under simplifying assumptions, alternative methods available (e.g., Heckman 2 stage).

Implementation

1. Estimate the propensity score by $\hat{\pi}_d(z_i) \equiv \pi_d(z_i, \hat{\gamma}_d)$.
2. Fix a value of $t \in \Theta$. For $d = 0, 1$ and $\tau \in \mathcal{T}$, compute $\hat{\beta}_d(\tau; t)$ as

$$\hat{\beta}_d(\tau; t) \equiv \arg\min_{b \in \mathcal{B}} \sum_{i=1}^{n} \mathbf{1}(D_i = d) S_i \rho_{G_{d,x}(\tau, \hat{\pi}_d(Z_i); t)} \left(Y_i - X_i' b \right)$$

   where $\rho_u(x) \equiv xu\mathbf{1}(x \geq 0) - (1-u)x\mathbf{1}(x < 0)$ denotes the check function.
3. Estimate the copula parameters for $d = 0, 1$ by minimizing over $t \in \Theta$:

$$\hat{\theta}_d \equiv \arg\min_{t \in \Theta} \left\| \sum_{i=1}^{n} \int_{\varepsilon}^{1-\varepsilon} \mathbf{1}(D_i = d) S_i \varphi(\tau, Z_i) \left[\mathbf{1}\left(Y_i \leq X_i' \hat{\beta}_d(\tau; t) \right) - G_{d,x}(\tau, \hat{\pi}(Z_i); t) \right] d\tau \right\|$$

   where $\varphi(\tau, Z_i)$ is an instrument function (*E.g.*, a polynomial of the propensity score).
4. The slope parameters are obtained by $\hat{\beta}_d(\tau) \equiv \hat{\beta}_d\left(\tau; \hat{\theta}_d\right)$ for $d = 0, 1$.
5. The SQF and the copula are respectively given by $\hat{g}_d(x, \tau) = x' \hat{\beta}_d(\tau)$ and $\hat{C}_{d,x}(\tau, \pi) = C_{d,x}\left(\tau, \pi; \hat{\theta}_d\right)$.

Implementation

- Estimate the counterfactual with covariates from group h, SQF from group k, copula from group l and propensity score from group m ($\ell = (h, k, l, m)$).

- Substitute each function by its estimated counterpart:

- $\hat{\mathbb{E}}\left[Y^{\ell}|S=1\right] =$
$\frac{1}{n_h}\sum_{i=1}^{n}\int_{\varepsilon}^{1-\varepsilon}\hat{g}_k\left(X_i,u\right)d\hat{G}_{l,x}\left(u,\hat{\pi}_m\left(Z_i\right)\right)\mathbf{1}\left(D_i=h\right)$.

- $\hat{F}^{\ell}_{Y|S=1}\left(y\right) =$
$\frac{1}{n_h}\sum_{i=1}^{n}\left[\varepsilon+\int_{\varepsilon}^{1-\varepsilon}\mathbf{1}\left(\hat{g}_k\left(X_i,u\right)\leq y\right)d\hat{G}_{l,x}\left(u,\hat{\pi}_m\left(Z_i\right)\right)\right]\mathbf{1}\left(D_i=h\right)$.

- $\hat{Q}^{\ell}_{Y|S=1}\left(\tau\right)=\inf\left\{y:\tau\leq\hat{F}^{\ell}_{Y|S=1}\left(y\right)\right\}$.

- Similarly for the entire population.

Ancillary decompositions

- Counterfactual mean propensity score:

$$\hat{\mathbb{E}}\left[\pi^{\ell}|S=1\right]=\frac{1}{n_h}\sum_{i=1}^{n}\hat{\pi}_m\left(z_i\right)\mathbf{1}\left(D_i=h\right)$$

- Counterfactual mean value of the unobservables:

$$\hat{\mathbb{E}}\left[U^{\ell}|S=1\right]=\frac{1}{n_h}\sum_{i=1}^{n}\int_{\varepsilon}^{1-\varepsilon}ud\,\hat{G}_{l,x}\left(u,\hat{\pi}_m\left(Z_i\right)\right)\mathbf{1}\left(D_i=h\right)$$

- The different components of the decomposition are computed accordingly.
- Estimators are asymptotically Gaussian.

Inference

- Asymptotic variances are complicated integrals.
- Matrices of the asymptotic variances depend on many density functions.
- Common approach in the literature (Melly, 2006; Chernozhukov et al., 2013; Pereda-Fernández, 2019): bootstrap.
- In this paper, multiplicative bootstrap (Ma and Kosorok, 2005).

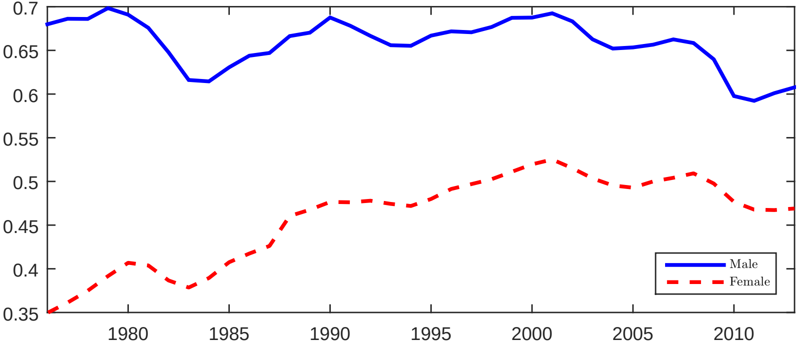
Data

- Current Population Survey (CPS) dataset.
- Same sample as Maasoumi & Wang (2019):
  - 1976-2013 period.
  - Individuals between 18 and 64 years old, who work for wages and salary and do not live in group quarters.
  - Minimum 20 weeks of 25 hours worked in the previous year.

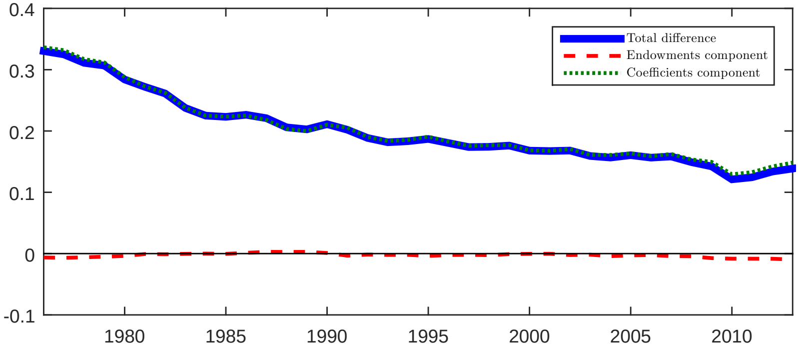
Specifications

- Dependent variable: log hourly wages.
- Regressors: third degree polynomial of age, four levels of education, four regional dummies, marital status, an indicator for white race, and the interactions between age and the other listed covariates, plus another variable they did not use: the number of children.
- Propensity score $\rightarrow$ probit.
- Two parametric copulas: Frank and Gaussian.
- Additional heterogeneous copulas: white vs non-white; college vs less than college; married vs non-married.
- Additionally: Heckman 2-stage estimator (*i.e.*, linear, homogeneous coefficients & fully Gaussian).

Average propensity to work by gender



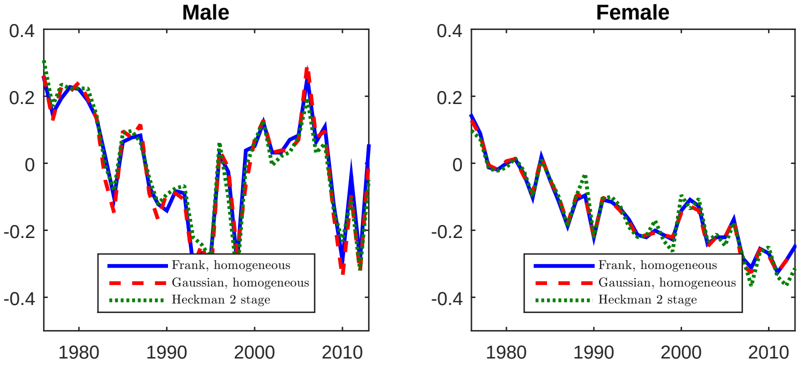
Participation decomposition



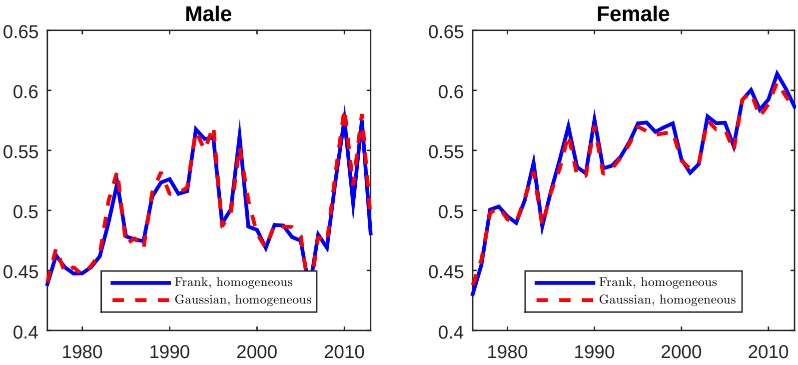
Participation trends

- Male participation stable around 2/3 until financial crisis; drop to 60% afterwards.
- Female participation steadily increased from 35% to around 50%; slight drop after the financial crisis.
- Participation gap more than halved during the period (from 33% to 16%).
- Small gap increase after the financial crisis.
- Participation decomposition shows change driven by coefficients component (structural increase in female participation).

Kendall's τ correlation coefficients

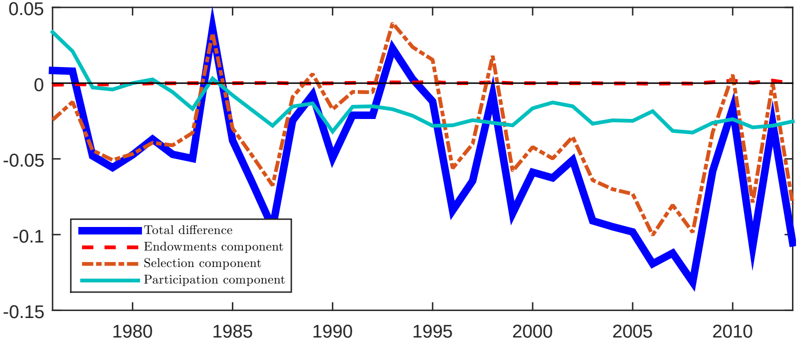


Mean value of u for participants



Moving average estimates

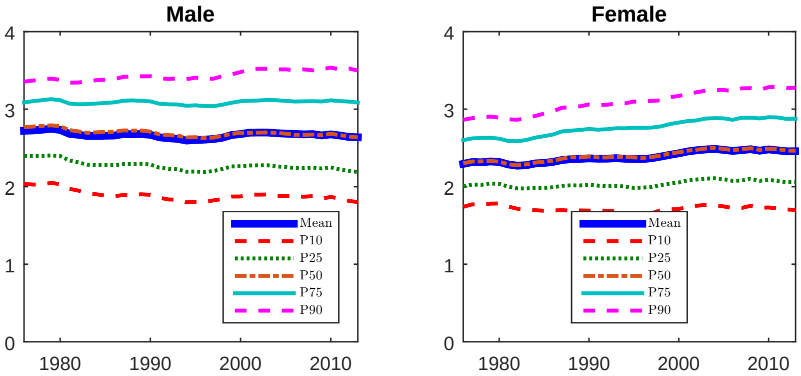
Self-selection decomposition (Frank copula)



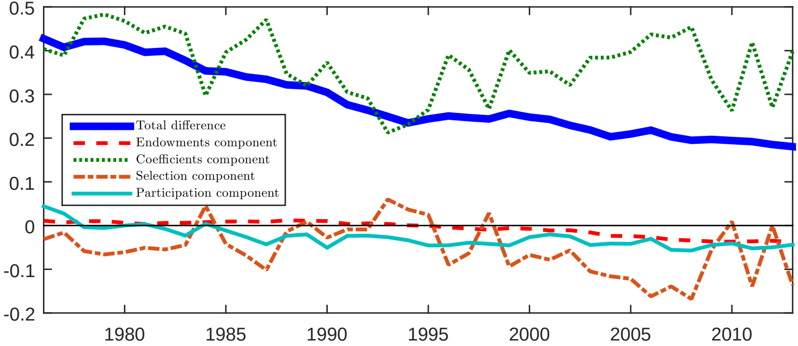
Self-selection trends

- Estimates show some degree of fluctuation, particularly for men.
    - Related to weak instrument for men?
- Steady increase of self-selection for females (positive since the 80s).
- Flatter long-term trend for males.
- Gender gap changed from less than 1 point in favor of males, to over 10 in favor of females.
- Selection and participation components contributed to the reversal.
- Selection component fluctuated a lot.
- Results very similar with both copulas.

Actual earnings distributions for participants by gender

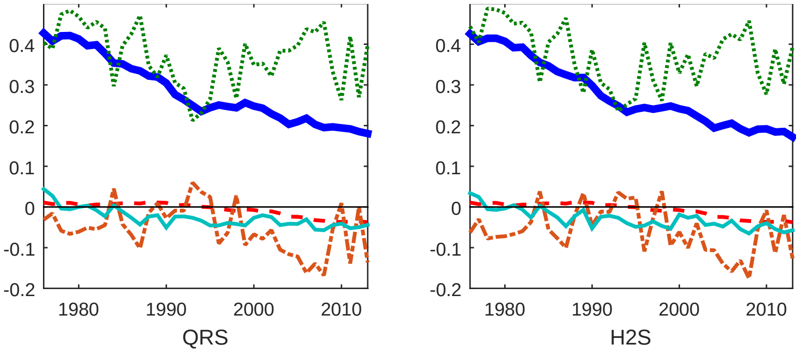


Actual earnings decomposition for participants

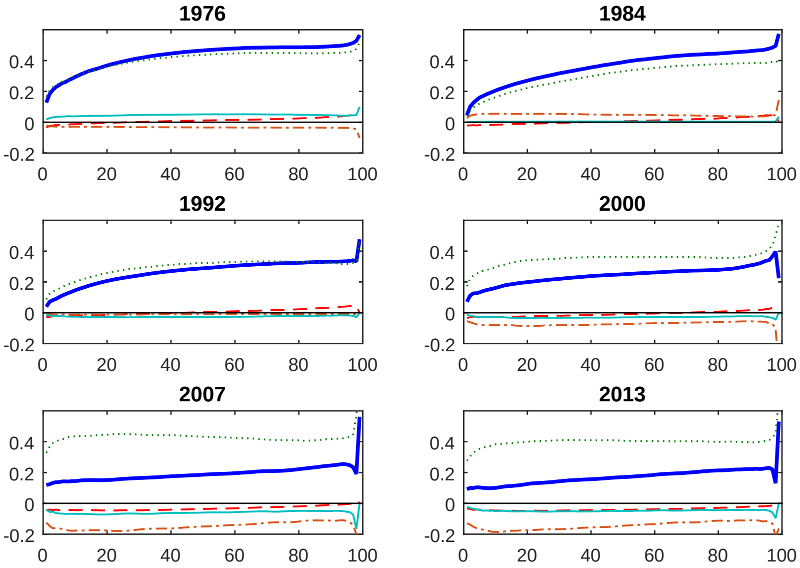


Moving average estimates

Actual earnings decomposition for participants



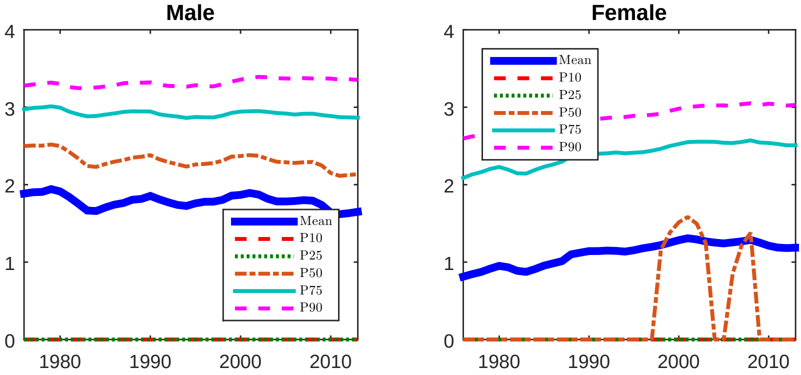
Actual earnings decomposition for participants



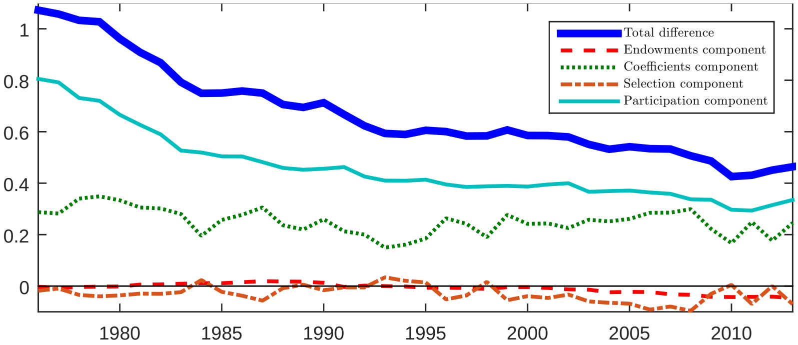
Actual earnings for the entire distribution trends

- Earnings gaps for participants decreased over the period.
  - Largest fall until mid-nineties.
  - Slower catch-up since then.
- Gap larger on the right tail of the distribution.
- Main contributors to the decrease: selection & participation components.
- Smaller contribution by endowments component.
- Selection & coefficients component volatile and mirror each other (weak instrument?).
- Most components quite constant across quantiles.
- Exception: coefficients component.

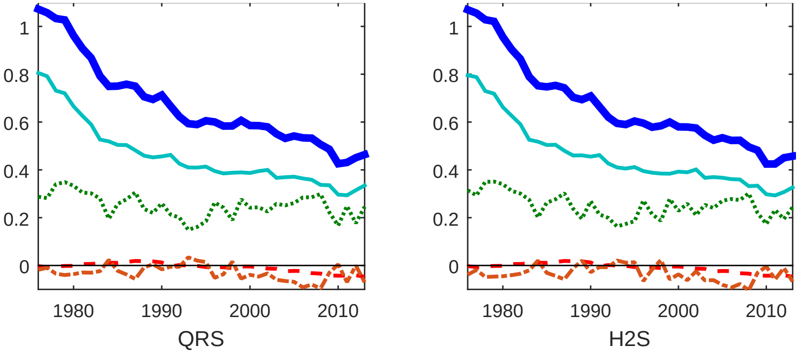
Actual earnings distributions for the entire population by gender



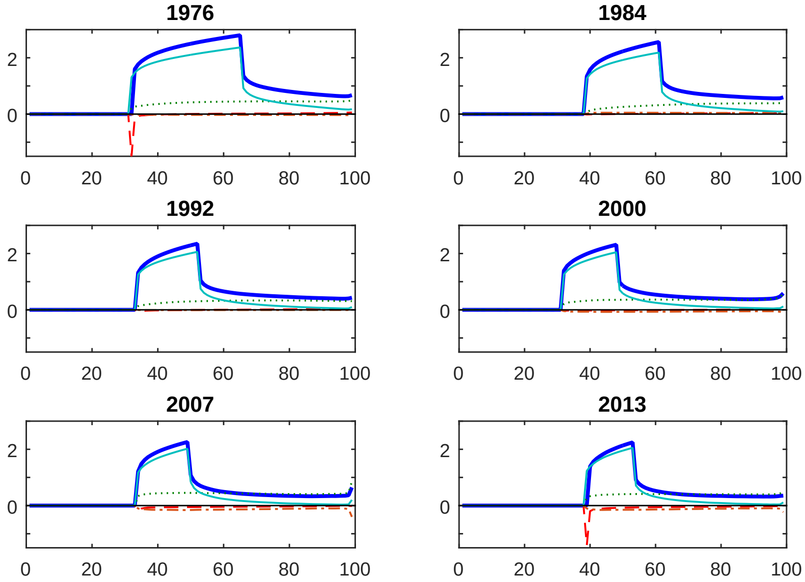
Actual earnings decomposition for the entire population



Actual earnings decomposition for the entire population



Actual earnings decomposition for the entire population



Actual earnings for the entire population trends

- Greater gap & greater decrease than for just participants.
- Participation component becomes main driver of dynamics:
  - Periods with structural increase in female participation close the gap because more women have labor earnings.
  - Also, positive selection implies smaller female participation further reduces the gap.
- Smaller contributions of selection and endowments components to close the gap.
- Small gap increase in the aftermath of the financial crisis → female participation hit harder than male's.

Conclusion (I)

- Presented method to decompose differences between two populations when there is sample selection.
- Differences depend on four components, two of which capture differences in participation and self-selection.
- Ancillary decompositions help understanding sources of main gaps.
- Provided uniform asymptotic theory.
- Existing estimators can be used to estimate the structural functions.
- Multiplier bootstrap to carry on uniformly valid inference.

Conclusion (II)

- Apply methods to evolution of gender pay gap in the US.
- Ancillary decomposition show improvements in unobserved ability more marked for women & larger increases in structural participation.
- Participation and selection components main drivers of gap decreases.
- Considering actual earnings for the entire population reveals larger gaps and larger gaps decrease.
- Structural component remains as main driver of the gap $\rightarrow$ need further research into this (STEM, personality traits, career interruptions, discrimination).

Lemma 1

Let $Y = \tilde{g}_D\left(X, \tilde{U}\right) S$ *and* $D = \mathbf{1}\left(\tilde{\pi}_D(Z) - \tilde{V} > 0\right)$*, where the distribution of the unobservables is given by* $\tilde{F}_{\tilde{U},\tilde{V}|D,X}\left(\tilde{U}, \tilde{V}|D, X\right)$*, with marginal distributions* $\tilde{F}_{\tilde{U}|D,X}$ *and* $\tilde{F}_{\tilde{V}|D,X}$*, that may depend non-trivially on* X*. Then, there exist* g_D*,* π_D *such that the model given by Equations 1-2, where* $U|D, X \sim U[0,1]$ *and* $V|D, X \sim U[0,1]$*, generates the same distribution of* (Y, S, D, Z)*.*

Back to main presentation

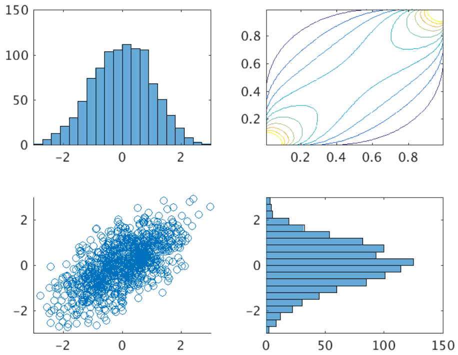
Model the Correlation with Copulas

- X multivariate random variable with marginals F_i and $u_i \equiv F_i(x_i)$ for $i = 1, ..., d$, *i.e.* rank of variable.
- Definition of copula: $C(u_1, ..., u_d) = \mathbb{P}(U_1 \leq u_1, ..., U_d \leq u_d)$.
- Sklar's Theorem (any multivariate cdf can be written in terms of a copula).
- Separately model the correlation from the marginal distribution of the effects.
- Independence copula: $C(u_1, ..., u_d) = \prod_{i=1}^{d} u_i$.
- Main parametric copulas: Archimedean (e.g. Clayton, Gumbel, Frank), Elliptical (e.g. Gaussian, t), Bernstein.

Back to main presentation

Example

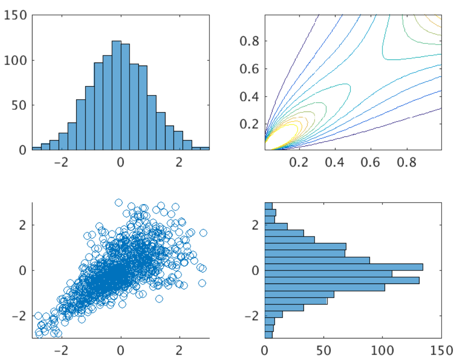
- $X_1 \sim \mathcal{N}(0,1)$, $X_2 \sim \mathcal{N}(0,1)$
- $(F_1(X_1), F_2(X_2)) \sim Gaussian(\rho)$



Back to main presentation

Example

- $X_1 \sim \mathcal{N}(0,1)$, $X_2 \sim \mathcal{N}(0,1)$
- $(F_1(X_1), F_2(X_2)) \sim Clayton(\rho)$



Back to main presentation

Additively separable unobservable term

- $Y = \left(g_D(X) + \tilde{U}\right) S$.
- where $\tilde{U} = Q_{U|D,X}(U)$ denotes the conditional quantile function of the uniformly distributed random variable U.
- Simplification to the expected outcome:

$$\mathbb{E}\left[Y^{hklm}|S=1\right] = \int_{\mathcal{Z}}\int_0^1 \left[g_k(x) + Q_{U|k,x}(u)\right] dG_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

$$= \int_{\mathcal{X}} g_k(x)\, dF_X^h(x) + \int_{\mathcal{Z}}\int_0^1 Q_{U|k,x}(u)\, dG_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

$$\mathbb{E}\left[Y^{hklm}\right] = \int_{\mathcal{X}} g_k(x)\, dF_X^h(x) + \int_{\mathcal{Z}}\int_0^1 Q_{U|k,x}(u)\, dC_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

- First term depends on structural function and distribution of covariates.
- Second term depends on the copula, the propensity score and the distribution of the covariates.
- Can estimate g nonparametrically (e.g., Das et al., 2003).

Back to main presentation

Linear model

- $g_d(x, u) = x'\beta_d + \tilde{u}$.
- where $\tilde{u} = Q_{u|d,x}(u)$ denotes the conditional quantile function of the uniformly distributed random variable U.
- Further simplification from previous case:

$$\mathbb{E}\left[Y^{hklm}|S=1\right] = \mathbb{E}_h(X)'\beta_k + \int_{\mathcal{Z}}\int_0^1 Q_{U|k,x}(u)\, dG_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

$$\mathbb{E}\left[Y^{hklm}\right] = \mathbb{E}_h(X)'\beta_k + \int_{\mathcal{Z}}\int_0^1 Q_{U|k,x}(u)\, dC_{l,x}(u, \pi_m(z))\, dF_Z^h(z)$$

- Can estimate g either parametrically (e.g., Heckman, 1976; Heckman, 1979; Lee, 1983) or semiparametrically (e.g., Newey, 2009).

Back to main presentation

No self-selection

- Independence copula: $C_{d,x}(u,v) = uv$ and $G_{d,x}(u,v) = u$.
- No selection component in the decompositions.
- Participation component vanishes for decompositions for participants.
- Mean and unconditional quantiles for entire population depend on propensity score.
- No need to estimate the copula; use methods without sample selection (Machado and Mata, 2005; Chernozhukov et al., 2013, DiNardo et al. 1996).
- Under linearity, can use the Oaxaca-Blinder decomposition.

Back to main presentation

Assumptions

1. $(Y_i, S_i, D_i, Z_i')'$ are *iid* for $i = 1, ..., n$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and take values in a compact set.

2. The sample size for the d-th group is non-decreasing in n, such that $n/n_d \to p_d \in [0, \infty)\, \forall d$ as $n \to \infty$.

3. $\pi_d(Z) \equiv \pi_d(Z; \gamma_d)$, with $\dim(\gamma_d) < \infty$. $\pi_d(Z; \gamma_d)$ is continuously differentiable with respect to γ_d. Moreover, there exists an asymptotically linear estimator $\hat{\gamma} \equiv (\hat{\gamma}_0', \hat{\gamma}_1')'$ that admits the following representation:
$\hat{\gamma} - \gamma = -B^{-1} \frac{1}{n_d} \sum_{i=1}^{n} r_d(s_i, z_i; \gamma) + o_P\left(\frac{1}{\sqrt{n}}\right)$.

4. Y has conditional density that is bounded from above and away from zero, a.s. on a compact set $\mathcal{Y}$. The density is given by $f_{Y|S=1,D,Z}(y)$ for $D = 0, 1$.

Assumptions

5. $g_d(x,\tau) = x'\beta_d(\tau)$ for $d = 0,1$, where β_d is continuous and such that $g_d(x,\tau)$ is increasing in its last argument.
6. Let $C_{d,x}(u,v) \equiv C_{d,x}(u,v;\theta_d)$, with $\dim(\theta_d) < \infty$ for $d = 0,1$. $C_{d,x}(u,v;\theta_d)$ is uniformly continuous and differentiable with respect to its arguments *a.e.* Its density, $c_{d,x}(u,v;\theta_d)$, is well-defined and finite.
7. Let $\beta(\tau) \equiv \left(\beta_1(\tau)', \beta_0(\tau)'\right)'$ and $\theta \equiv (\theta_1', \theta_0')'$. $\forall \tau \in \mathcal{T}$, $\left(\beta(\tau)', \theta', \gamma'\right)' \in int\mathcal{B} \times \Theta \times \mathcal{G}$, where $\mathcal{B} \times \Theta \times \mathcal{G}$ is compact and convex, and $\mathcal{T} = [\varepsilon, 1-\varepsilon]$, for some constant ε that is used to avoid the estimation of extreme quantiles.
8. Matrices of derivatives of the moments $J_{\beta d}(\tau)$, $\tilde{J}_{\beta d}(\tau)$, $J_{\gamma d}(\tau)$, $\tilde{J}_{\gamma d}(\tau)$, $J_{\theta d}(\tau)$, $\tilde{J}_{\theta d}(\tau)$ for $d = 0,1$, as defined in Appendix A, are continuous and have full rank, uniformly over $\mathcal{B} \times \Theta \times \Gamma \times \mathcal{T}$ and $d = 0,1$.
9. Denote the support of $\pi_d(Z)$ conditional on $X = x$ by $\mathcal{P}_{d,x}$. $\forall x \in \mathcal{X}$ and $d = 0,1$, $\mathcal{P}_{d,x} \in [0,1]$ is an open interval.

Theorem 2

Let $\hat{\vartheta}_d(\tau) \equiv \left(\hat{\beta}_d(\tau)', \hat{\theta}_d', \hat{\gamma}_d'\right)'$. *Under Assumptions 1-9, their joint asymptotic distribution is given by* $\sqrt{n}\left(\hat{\vartheta}_d(\tau) - \vartheta_d(\tau)\right) \Rightarrow \mathbb{Z}_{\vartheta_d}(\tau)$, *where* $\mathbb{Z}_{\vartheta_d}(\tau)$ *is a zero-mean Gaussian process with covariance function* $\Sigma_{\vartheta_d}(\tau, \tau')$, *where:*

$$\Sigma_{\vartheta_d}(\tau, \tau') = \sqrt{p_d p_{d'}} H_d(\tau) \Sigma_{R_d}(\tau, \tau') H_{d'}(\tau')'$$

$$H_d(\tau) = F_d^I(\tau)\left[C_d(\tau) + \left(I - \int_{\varepsilon}^{1-\varepsilon} D_d(u) F_d^I(u)\, du\right)^{-1} \int_{\varepsilon}^{1-\varepsilon} D_d(u) F_d$$

$$\Sigma_{R_d}(\tau, \tau') = \mathbb{E}\left[\mathbb{Z}_{R_d}(\tau) \mathbb{Z}_{R_{d'}}(\tau')'\right]$$

and functions $C_d(\tau)$, $D_d(\tau)$, $F_d^I(\tau)$ *and* $\mathbb{Z}_{R_d}(\tau)$ *are defined in Appendix A.*

Corollary 1

Let $\hat{g}_d(x,\tau) = x'\hat{\beta}_d(\tau)$, $\hat{c}_{d,x}(\tau,\eta) = c_{d,x}\left(\tau,\eta;\hat{\theta}_d\right)$, $\hat{\pi}_d(z) = \pi_d(z;\hat{\gamma}_d)$ *and* $\hat{F}_Z^d(z) = \frac{1}{n_d}\sum_{i=1}^{n}\mathbf{1}(d_i = d)\mathbf{1}(Z_i \leq z)$. *They satisfy Condition 1.*

Back to main presentation

Implementation for the full population

- $\hat{\mathbb{E}}\left[Y^{\ell}\right] = \frac{1}{n_h}\sum_{i=1}^{n}\int_{\varepsilon}^{1-\varepsilon}\hat{g}_k\left(x_i, u\right) d\,\hat{C}_{l,x}\left(u, \hat{\pi}_m\left(z_i\right)\right)\mathbf{1}\left(d_i = h\right).$

- $\hat{F}_Y^{\ell}\left(y\right) =$
  $\frac{1}{n_h}\sum_{i=1}^{n}\left[\left[\varepsilon + \int_{\varepsilon}^{1-\varepsilon}\mathbf{1}\left(\hat{g}_k\left(x_i, u\right) \leq y\right) d\,\hat{C}_{l,x}\left(u, \hat{\pi}_m\left(z_i\right)\right)\right] + \left(1 - \hat{\pi}_m\left(z\right.\right.\right.$

- $\hat{Q}_Y^{\ell}\left(\tau\right) = \inf\left\{y : \tau \leq \hat{F}_Y^{\ell}\left(y\right)\right\}.$

Back to main presentation

Differences relative to Maasoumi and Wang (2019)

- There are some implementation differences:
  1. Include additional covariate (number of children).
  2. Quantile grid used for the estimation: $(0.01, 0.02, ..., 0.99)$; they used $(0.3, 0.4, ..., 0.7)$.
  3. Instrument used: $\varphi(u, z) = \hat{\pi}(z)$; they used $\varphi(u, z) = \sqrt{u(1-u)\hat{\pi}(z)}$.
- As a result, slightly different objective function to estimate θ.
- On top of that, worse performance of Stata relative to Matlab.

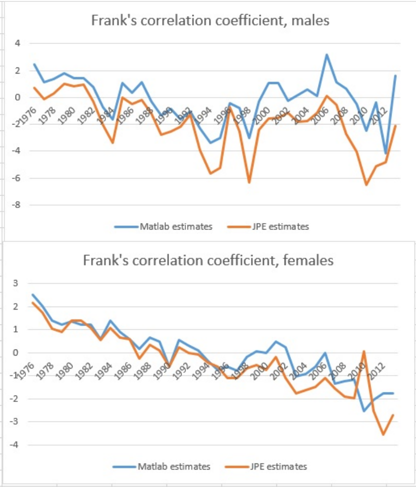
Differences relative to Maasoumi and Wang (2019)

Table: Objective function for θ

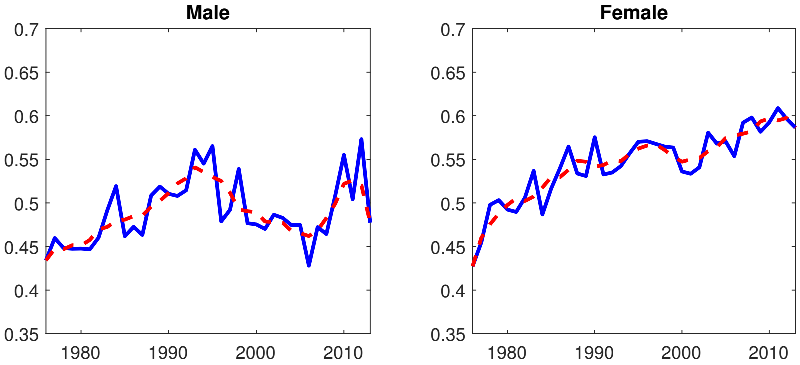
	τ				
	0.3	0.4	0.5	0.6	0.7
Matlab estimates	1216.086	2044.821	2672.171	3051.696	3090.995
Stata estimates	1216.519	2045.354	2673.005	3052.189	3091.509

- Stata estimator of the check function performs worse.
- Small errors sum up in the objective function for θ.
- Slightly biased estimates in Stata.

Differences relative to Maasoumi and Wang (2019)

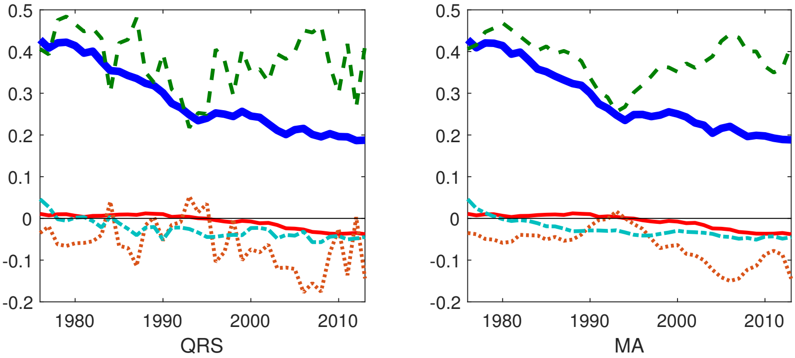


Moving average estimates of average U



Back to main presentation

Moving average estimates of average U



Back to main presentation