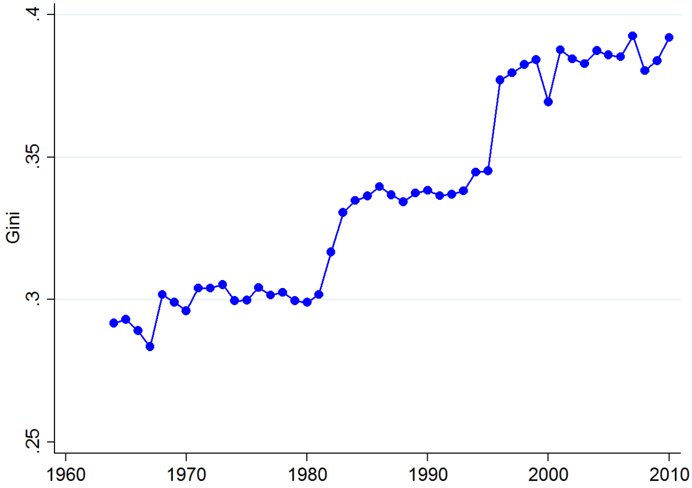
Partnership patterns and long-term trends in US family earnings inequality

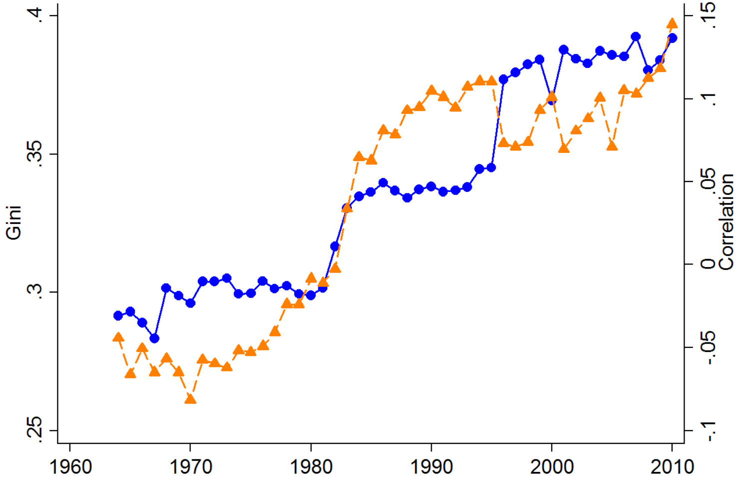
Koen Decancq, CSB, University of Antwerp
Andreas Peichl, ZEW Mannheim
Philippe Van Kerm, CEPS/INSTEAD Luxembourg

CASE Welfare Policy and Analysis Seminar
London School of Economics and Political Science, May 14 2014

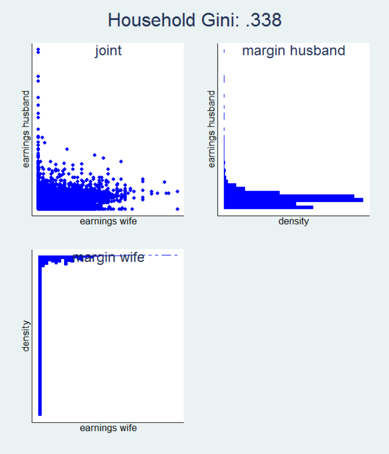
Increasing income inequality (in the US)



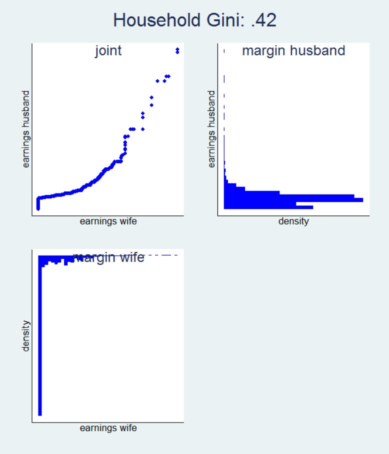
Increased correlation between earnings of spouses *and* increased household earnings inequality (Martin 2006 ; Burtless 2009 ; Schwartz 2010)



Observed vs. perfect pairing illustration



Observed vs. perfect pairing illustration



Questions

- How much of long-term increase in inequality accounted for by increase in partner earnings correlation ?

Questions

- How much of long-term increase in inequality accounted for by increase in partner earnings correlation ?
- How does increasing correlation come about and what aspect does account for increasing inequality ?

Questions

- How much of long-term increase in inequality accounted for by increase in partner earnings correlation ?
- How does increasing correlation come about and what aspect does account for increasing inequality ?
  - Assortative mating (wrt education, race) ?
    - share of couples with same education increased by 17%

Questions

- How much of long-term increase in inequality accounted for by increase in partner earnings correlation ?
- How does increasing correlation come about and what aspect does account for increasing inequality ?
  - Assortative mating (wrt education, race) ?
    - share of couples with same education increased by 17%
  - Joint labour market participation—rise of dual earners
    - share of two earner couples increased by 57%

Questions

- How much of long-term increase in inequality accounted for by increase in partner earnings correlation ?
- How does increasing correlation come about and what aspect does account for increasing inequality ?
  - Assortative mating (wrt education, race) ?
    - share of couples with same education increased by 17%
  - Joint labour market participation—rise of dual earners
    - share of two earner couples increased by 57%
  - Earnings correlation conditional on working ?
    - overall earnings correlation increased from negative to positive

Questions

- How much of long-term increase in inequality accounted for by increase in partner earnings correlation ?
- How does increasing correlation come about and what aspect does account for increasing inequality ?
  - Assortative mating (wrt education, race) ?
    - share of couples with same education increased by 17%
  - Joint labour market participation—rise of dual earners
    - share of two earner couples increased by 57%
  - Earnings correlation conditional on working ?
    - overall earnings correlation increased from negative to positive

Key difficulty : the marginal distributions change too (expansion of college graduates, of women labour force participation, increasing returns to education) (Greenwood et al. 2014)

Usual methods

Two typical approaches :

- Factor decomposition of inequality indices (CV, Gini) (male & female earnings) (Karoly & Burtless 1995 ; Cancian & Reed 1998, 1999 ; Reed & Cancian 2001)
- Simulation : ranking and (re-)matching methods (Burtless 1999 ; Fournier 2001 ; Reed and Cancian 2004 ; Larrimore 2009 ; Greenwood et al. 2014)

Drawback : difficult to point out assortativeness channels (and often ad-hoc and difficult to compare across studies)

Our approach

- Construct and estimate a (‘statistical’) model for the distribution of earnings with explicit ‘assortativeness parameters’ (distinct from marginal distribution parameters)
- ‘Parameter swaps’ simulations (static ! no general equilibrium magnitudes)
- Allows for separation of different channels
  1. Changed assortativeness *in marital selection (education, race)*
  2. Changed assortativeness *in joint (extensive) labour supply decision*
  3. Changed assortativeness *in earnings* in two-earner couples

A model of family earnings

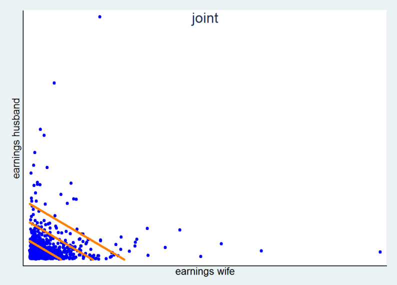
A model of family earnings

Total family (couple) earnings y : sum of earnings of wife (y^w) and husband (y^h).

Probability density function (pdf) of total earnings :

$$g(y) = \int_0^y f(y^w, \underbrace{y - y^w}_{y^h}) dy^w$$

$f(y^w, y^h)$ is the joint pdf of earnings of both partners



A model of family earnings

Conditioning on employment, education and demographic factors

Familiar chain of conditional probabilities :

$$f(\mathbf{y}) = \sum_{\mathbf{l}\in L} \sum_{\mathbf{e}\in E} \sum_{\mathbf{r}\in R} \sum_{\mathbf{z}\in Z} f(\mathbf{y}|\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z}) \underbrace{p(\mathbf{l}|\mathbf{e},\mathbf{r},\mathbf{z})\, p(\mathbf{e}|\mathbf{r},\mathbf{z})\, p(\mathbf{r}|\mathbf{z})\, p(\mathbf{z})}_{p(\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z})}$$

- $\mathbf{y} = (y^w, y^h)$ is *pair* of husband-wife earnings
- $\mathbf{l} = (l^w, l^h)$ is *pair* of husband-wife employment status
- and similarly for **e** (college education), **r** (race), **z** (other factors)

A model of family earnings

Conditioning on employment, education and demographic factors

Familiar chain of conditional probabilities :

$$f(\mathbf{y}) = \sum_{\mathbf{l} \in L} \sum_{\mathbf{e} \in E} \sum_{\mathbf{r} \in R} \sum_{\mathbf{z} \in Z} f(\mathbf{y}|\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z}) \underbrace{p(\mathbf{l}|\mathbf{e},\mathbf{r},\mathbf{z})\, p(\mathbf{e}|\mathbf{r},\mathbf{z})\, p(\mathbf{r}|\mathbf{z})\, p(\mathbf{z})}_{p(\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z})}$$

- $\mathbf{y} = (y^w, y^h)$ is *pair* of husband-wife earnings
- $\mathbf{l} = (l^w, l^h)$ is *pair* of husband-wife employment status
- and similarly for **e** (college education), **r** (race), **z** (other factors)

A model of family earnings

Conditioning on employment, education and demographic factors

Familiar chain of conditional probabilities :

$$f(\mathbf{y}) = \sum_{\mathbf{l} \in L} \sum_{\mathbf{e} \in E} \sum_{\mathbf{r} \in R} \sum_{\mathbf{z} \in Z} f(\mathbf{y}|\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z}) \underbrace{p(\mathbf{l}|\mathbf{e},\mathbf{r},\mathbf{z})\, p(\mathbf{e}|\mathbf{r},\mathbf{z})\, p(\mathbf{r}|\mathbf{z})\, p(\mathbf{z})}_{p(\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z})}$$

- $\mathbf{y} = (y^w, y^h)$ is *pair* of husband-wife earnings
- $\mathbf{l} = (l^w, l^h)$ is *pair* of husband-wife employment status
- and similarly for **e** (college education), **r** (race), **z** (other factors)

A model of family earnings

Conditioning on employment, education and demographic factors

Familiar chain of conditional probabilities :

$$f(\mathbf{y}) = \sum_{\mathbf{l}\in L} \sum_{\mathbf{e}\in E} \sum_{\mathbf{r}\in R} \sum_{\mathbf{z}\in Z} f(\mathbf{y}|\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z}) \underbrace{p(\mathbf{l}|\mathbf{e},\mathbf{r},\mathbf{z})\, p(\mathbf{e}|\mathbf{r},\mathbf{z})\, p(\mathbf{r}|\mathbf{z})\, p(\mathbf{z})}_{p(\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z})}$$

- $\mathbf{y} = (y^w, y^h)$ is *pair* of husband-wife earnings
- $\mathbf{l} = (l^w, l^h)$ is *pair* of husband-wife employment status
- and similarly for **e** (college education), **r** (race), **z** (other factors)

A model of family earnings

Conditioning on employment, education and demographic factors

Familiar chain of conditional probabilities :

$$f(\mathbf{y}) = \sum_{\mathbf{l}\in L}\sum_{\mathbf{e}\in E}\sum_{\mathbf{r}\in R}\sum_{\mathbf{z}\in Z} f(\mathbf{y}|\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z})\underbrace{p(\mathbf{l}|\mathbf{e},\mathbf{r},\mathbf{z})\,p(\mathbf{e}|\mathbf{r},\mathbf{z})\,p(\mathbf{r}|\mathbf{z})\,p(\mathbf{z})}_{p(\mathbf{l},\mathbf{e},\mathbf{r},\mathbf{z})}$$

- $\mathbf{y} = (y^w, y^h)$ is *pair* of husband-wife earnings
- $\mathbf{l} = (l^w, l^h)$ is *pair* of husband-wife employment status
- and similarly for **e** (college education), **r** (race), **z** (other factors)

Dissecting joint probabilities

Dissecting joint probabilities

Discrete (binary) data : Joint probabilities can be expressed as a (complicated) function Ψ of marginal probabilities and of cross-product ratio (Mosteller 1968)

$$p(x^w, x^h) = \Psi(p(x^w), p(x^h), \theta^{h,w})$$

$$\theta^{h,w} = \frac{p(0,0) \times p(1,1)}{p(1,0) \times p(0,1)}$$

Dissecting joint probabilities

Discrete (binary) data : Joint probabilities can be expressed as a (complicated) function Ψ of marginal probabilities and of cross-product ratio (Mosteller 1968)

$$p(x^w, x^h) = \Psi(p(x^w), p(x^h), \theta^{h,w})$$

$$\theta^{h,w} = \frac{p(0,0) \times p(1,1)}{p(1,0) \times p(0,1)}$$

$\Longrightarrow$ counterfactual probabilities by changing $\theta^{h,w}$!

Dissecting joint probabilities

Discrete (binary) data : Joint probabilities can be expressed as a (complicated) function Ψ of marginal probabilities and of cross-product ratio (Mosteller 1968)

$$p(x^w, x^h) = \Psi(p(x^w), p(x^h), \theta^{h,w})$$

$$\theta^{h,w} = \frac{p(0,0) \times p(1,1)}{p(1,0) \times p(0,1)}$$

$\Longrightarrow$ counterfactual probabilities by changing $\theta^{h,w}$!

Continuous data : Joint density can be expressed as product of marginal densities and copula (Sklar 1959)

$$f(y^w, y^h) = f^w(y^w) \times f^h(y^h) \times c(F^w(y^w), F^h(y^h))$$

$\Longrightarrow$ counterfactual probabilities by changing c !

Estimation

Fully parametric model (but flexible)

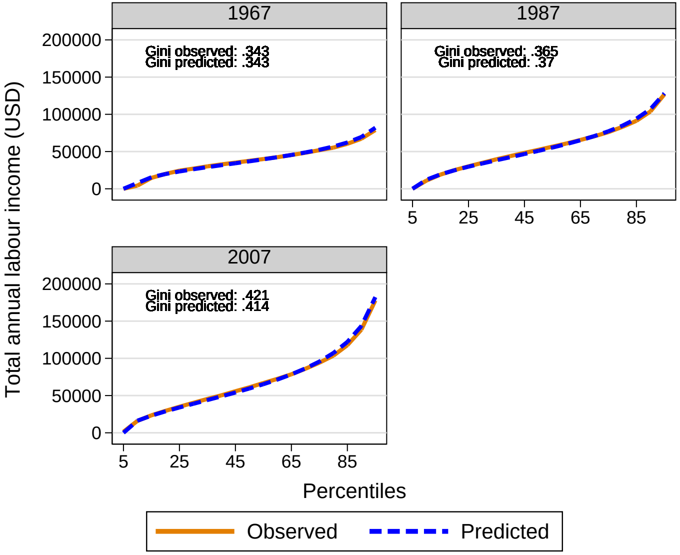
- Conditional earnings distributions assumed Singh-Maddala distributed separately for husband and wives (all three parameters vary with covariates ; ▸ S-M ; ▸ Controls)
- Copula function assumed Plackett (with dependence parameter varying with covariates ; ▸ Plackett copula)
- Joint employment, joint college education and joint race are estimated as bivariate probit
- (Maximum likelihood estimation)

$\Longrightarrow$ Counterfactual distributions (and Gini coefficients) based on (Monte Carlo) simulations with 'parameter swaps' between 2007 and 1967

Data

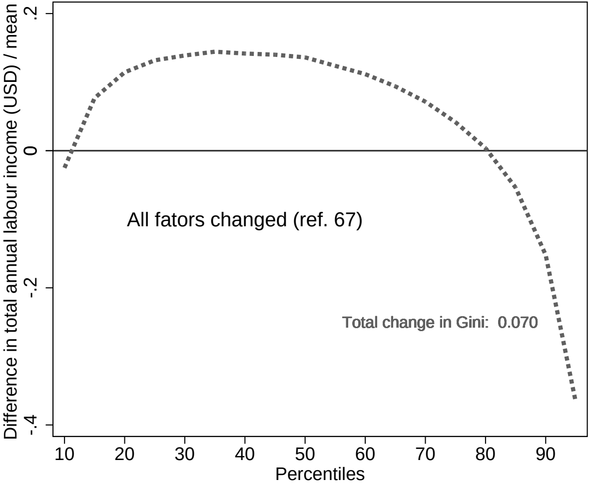
- Current Population Survey (CPS), 1964–2010
  - March Annual Social and Economic Supplement (ASEC)
  - years : 1967 ('target'), (1977, 1987, 1997), 2007 ('reference') (five years pooled)
- Sample selection (follow Cancian & Reed 1998 and Schwartz 2010) :
  - couples (married and cohabiting)
  - working age (21-55)
- *Pre-tax income* : market income from employment and self-employment
  - Topcoding (downward bias of inequality estimates) : Estimate censored Singh-Maddala (S-M) specification
  - Ignore non-labour incomes

	1967	1977	1987	1997	2007
Same age	0.769	0.781	0.753	0.750	0.765
Same educ	0.473	0.479	0.517	0.534	0.555
Same race	0.997	0.989	0.982	0.971	0.959
Both work	0.454	0.579	0.710	0.753	0.712
Husband works	0.981	0.962	0.957	0.955	0.940
Wife works	0.463	0.599	0.736	0.780	0.751
Correlation	–0.065	–0.041	0.078	0.071	0.103
N (couples)	20019	24042	22542	18706	21419

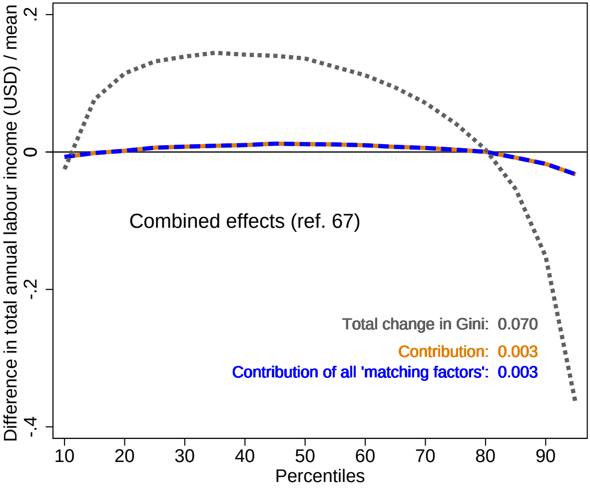

1967
Gini observed: .343
Gini predicted: .343
1987
Gini observed: .365
Gini predicted: .37
2007
Gini observed: .421
Gini predicted: .414
Total annual labour income (USD)
200000
150000
100000
50000
0
5
25
45
65
85
Percentiles
Observed
Predicted

What about 2007 if all cross-product ratios and copulas had remained as in 1967 ?

What about 2007 if all cross-product ratios and copulas had remained as in 1967 ?

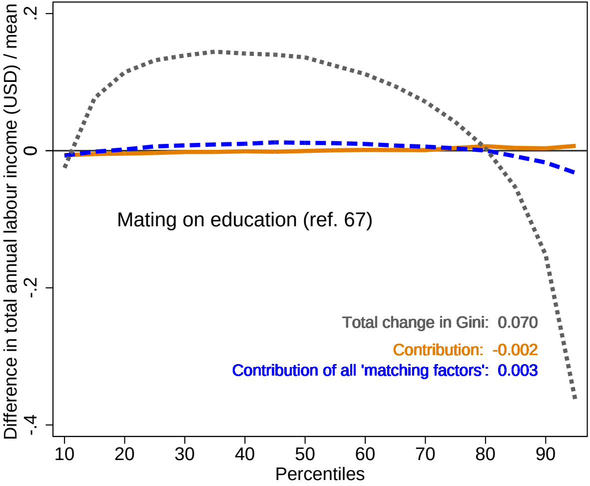


What about 2007 if all cross-product ratios and copulas had remained as in 1967 ?



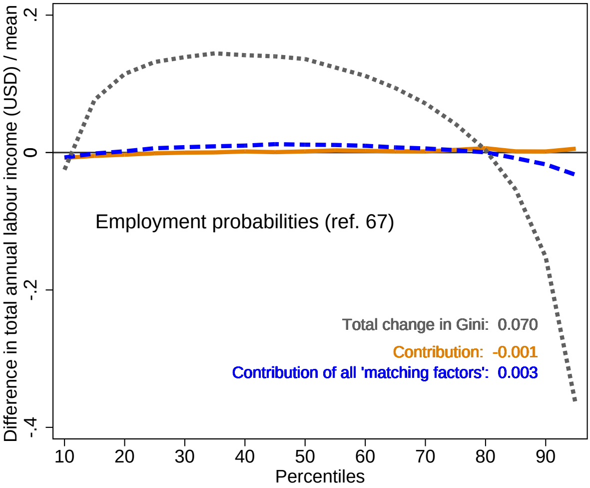
What about 2007 if mating on education (cross-product ratio) had remained as in 1967 ?

What about 2007 if mating on education (cross-product ratio) had remained as in 1967 ?



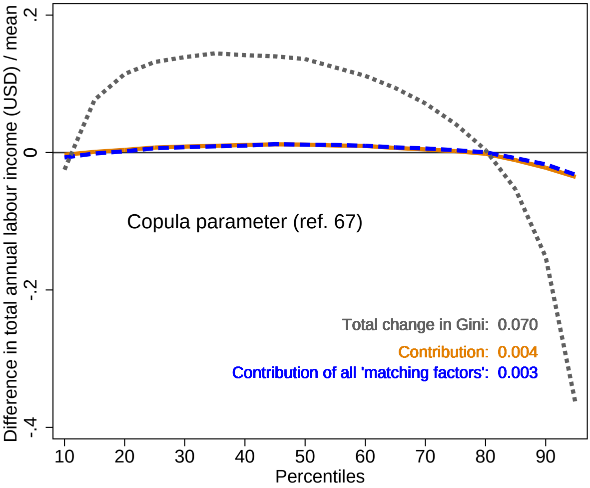
What about 2007 if cross-product ratio on joint employment probabilities had remained as in 1967 ?

What about 2007 if cross-product ratio on joint employment probabilities had remained as in 1967 ?



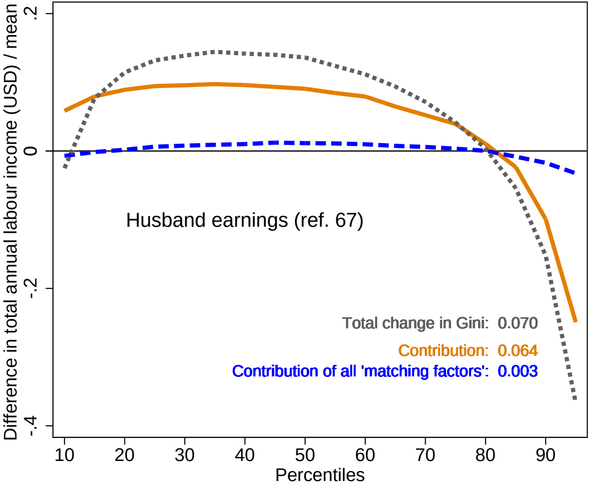
What about 2007 if joint earnings rank association (copula parameters) had remained as in 1967 ?

What about 2007 if joint earnings rank association (copula parameters) had remained as in 1967 ?



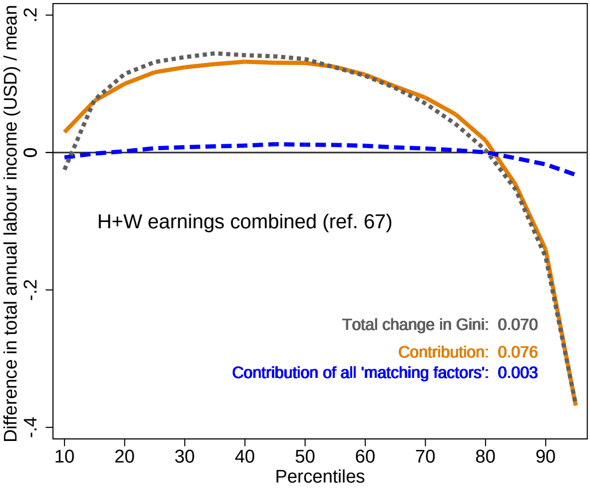
What about 2007 if men earnings distribution parameters had remained as in 1967 ?

What about 2007 if men earnings distribution parameters had remained as in 1967 ?



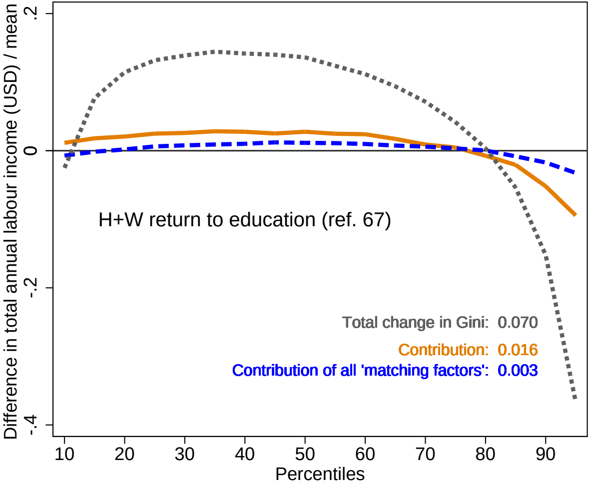
What about 2007 if men and women earnings distribution parameters had remained as in 1967 ?

What about 2007 if men and women earnings distribution parameters had remained as in 1967 ?



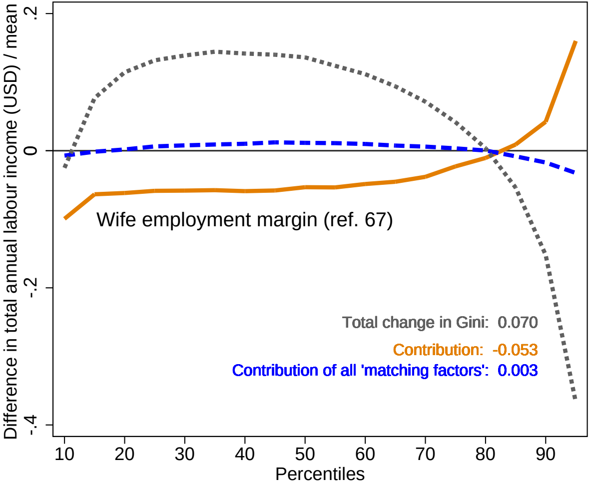
What about 2007 if return to education on earnings had remained as in 1967 ?

What about 2007 if return to education on earnings had remained as in 1967 ?



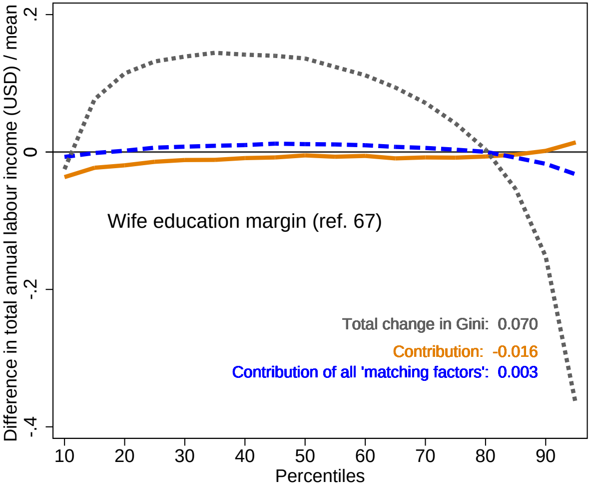
What about 2007 if women employment probabilities had remained as in 1967 ?

What about 2007 if women employment probabilities had remained as in 1967 ?



What about 2007 if women college education probability had remained as in 1967 ?

What about 2007 if women college education probability had remained as in 1967 ?



Summary

- Parametric ‘copula-based’ decomposition of family earnings distributions is implementable (and ‘clean’)
- Main findings (1967 vs. 2007) :
  - Inequality in couple market earnings increased by about one third (according to Gini coefficient)
  - Independently of margin changes, assortative mating accounts for very little of this change (less than 5%)
  - ... and mainly through correlation in earnings among dual earner couples (delayed partnering age ?)
  - These effect are dwarfed by impacts of changes in earnings and (women) employment margins

More to come

- Extensions :
  - Finer definitions (intensive labour supply, education)
  - Single-person households
  - Non-labour income
  - Tax-benefit policies
  - behavioural responses ?

"*Perhaps the greatest challenge is to develop more comprehensive models of the household income distribution, incorporating not only models of labour market earnings but also reflecting income from other sources including social benefits and investment income, and the demographic factors affecting who lives with whom*" (Jenkins and Micklewright 2007 : 19).

Singh-Maddala distribution

- Cumulative distribution function :

$$F(y) = \mathrm{SM}(y; a, b, q) = 1 - \left[1 + \left(\frac{y}{b}\right)^{a}\right]^{-q}.$$

  where :
  - q is a shape parameter (‘upper tail’)
  - a is a shape parameter (‘spread’)
  - b is a scale parameter
- To be precise, we estimate a censored Singh-Maddala distribution (to take account of topcoding)

Plackett copula

- Plackett copula :

$$C(u, v) = \frac{\left((1 + (\theta - 1)(u + v)) - \sqrt{(1 + (\theta - 1)(u + v))^2) - 4uv}\right.}{2(\theta - 1)}$$

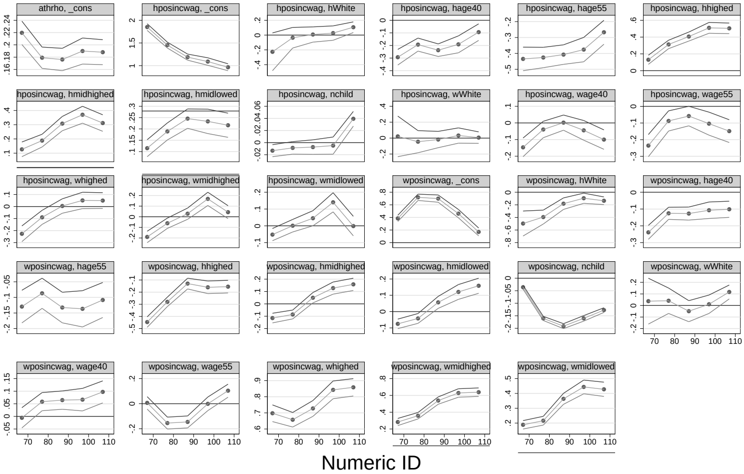
where :

  - θ is a dependence parameter

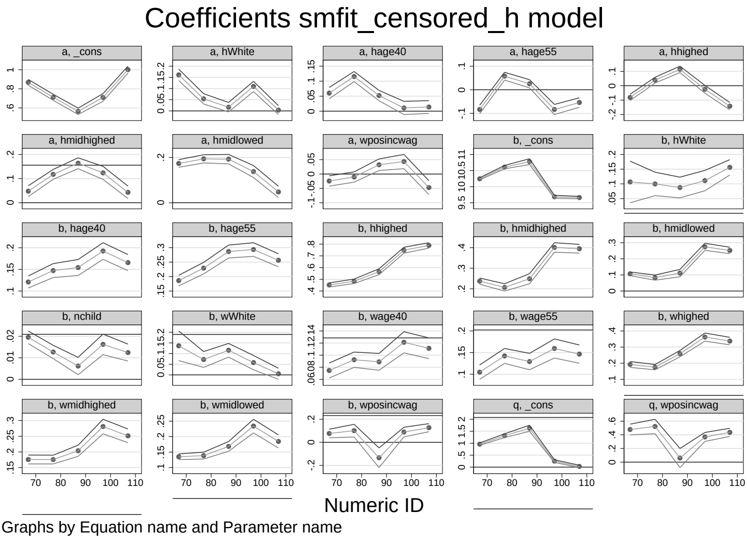
Conditioning Variables

	DFL	Probit	SM^h			SM^w			Plackett
			a	b	q	a	b	q	
# child	X	X					X		X
Age^h	X	X	X	X					X
Age^w	X	X		X		X	X		X
$Age^{h,w}$				X					X
$Age^{h,h}$		X	X	X					X
$Age^{w,w}$		X					X		X
$Race^h$	X	X	X	X					X
$Race^w$	X	X				X	X		X
$Race^{h,w}$	X								
$Educ^h$		X	X	X			X		X
$Educ^w$		X		X		X	X		X
$Educ^{h,h}$		X		X					
$Educ^{w,w}$		X							
cte	X	X	X	X	X	X	X	X	X
partner w			X	X	X	X	X	X	

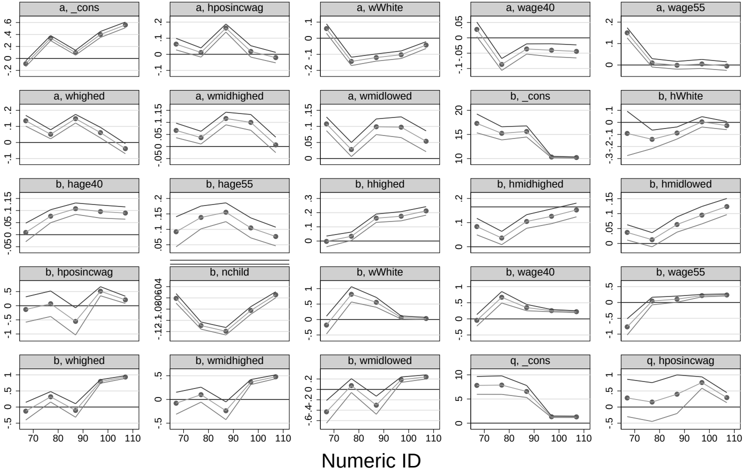
Coefficients biprobit model



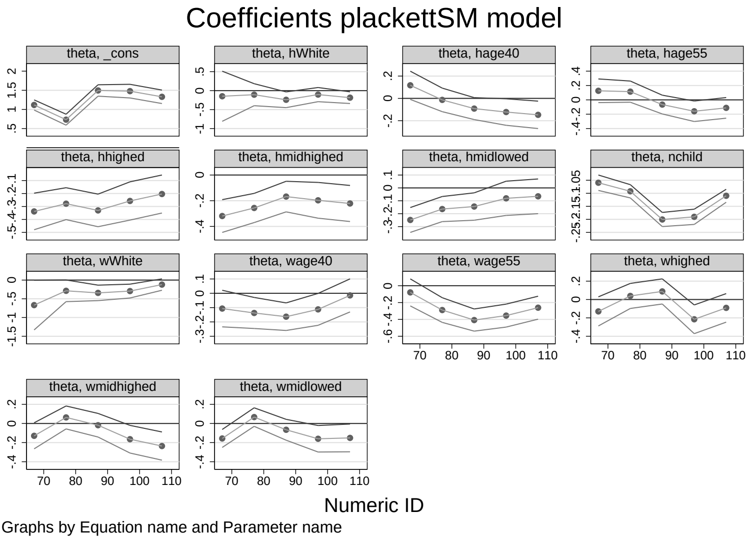
Graphs by Equation name and Parameter name


Coefficients smfit_censored_h model
a, _cons
a, hWhite
a, hage40
a, hage55
a, hhighed
a, hmidhighed
a, hmidlowed
a, wposincwag
b, _cons
b, hWhite
b, hage40
b, hage55
b, hhighed
b, hmidhighed
b, hmidlowed
b, nchild
b, wWhite
b, wage40
b, wage55
b, whighed
b, wmidhighed
b, wmidlowed
b, wposincwag
q, _cons
q, wposincwag
70 80 90 100 110
Numeric ID
Graphs by Equation name and Parameter name

Coefficients smfit_censored_w model



Graphs by Equation name and Parameter name


Coefficients plackettSM model
theta, _cons
theta, hWhite
theta, hage40
theta, hage55
theta, hhighed
theta, hmidhighed
theta, hmidlowed
theta, nchild
theta, wWhite
theta, wage40
theta, wage55
theta, whighed
theta, wmidhighed
theta, wmidlowed
70 80 90 100 110
Numeric ID
Graphs by Equation name and Parameter name