Attempts to increase (direct ex-ante lower bound) estimates of Inequality of Opportunity

Andreas Peichl
(ZEW, U Mannheim)

Canazei Winter School 2016

Background

- Typically, share of IOp due to circumstances is surprisingly small
- Low estimates of IOp have led to questions on its policy usefulness (Kanbur / Wagstaff, 2014)
- Identification of circumstances crucial for measuring IOp
  - but not all circumstances observable
  - disagreement about distinction between circumstances and effort
- Previous literature: mostly lower bound estimates of IOp (Bourguignon et al., 2007, Ferreira & Gignoux, 2011)
  - Niehues & Peichl (2014) upper bound estimator
- **Aim of this talk:** *some attempts to increase LB estimates*

Agenda

- Parametric ex-ante approach; $w_i = f(C_i, E_i(C_i), u_i)$

$$w_i = \alpha C_i + \beta E_i + u_i \tag{1}$$

$$E_i = \kappa C_i + v_i \tag{2}$$

- Log-linearization & estimate reduced form via OLS:

$$\ln w_i = \underbrace{(\alpha + \beta\kappa)}_{\psi} C_i + \underbrace{\beta v_i + u_i}_{\eta_i}. \tag{3}$$

  - $\widehat{\psi}$ measures overall effect of observed C_i on w_i
  - lower bound since including any additional C can only increase the share of inequality explained by C_i (intuition like R^2)

- Parametric ex-ante approach; $w_i = f(C_i, E_i(C_i), u_i)$

$$w_i = \alpha C_i + \beta E_i + u_i \quad (1)$$

$$E_i = \kappa C_i + v_i \quad (2)$$

- Log-linearization & estimate reduced form via OLS:

$$\ln w_i = \underbrace{(\alpha + \beta\kappa)}_{\psi} C_i + \underbrace{\beta v_i + u_i}_{\eta_i}. \quad (3)$$

  - $\widehat{\psi}$ measures overall effect of observed C_i on w_i
  - lower bound since including any additional C can only increase the share of inequality explained by C_i (intuition like R^2)

- Parametric prediction of smoothed distribution: $\widetilde{\mu} = \exp[\widehat{\psi} C_i + \sigma^2/2]$
  - Absolute level of IOp: $IOL = I_0(\widetilde{\mu})$
  - Relative share of IOp: $IOR = \frac{I_0(\widetilde{\mu})}{I_0(w_i)}$; usually MLD

Balcazar (2015, EL): LB on IOp and measurement error

Country	Total inequality $MLD(X) \times 100$	Between-type inequality $MLD(\widehat{X}_B) \times 100$	Within-type inequality $[MLD(X) - MLD(\widehat{X}_B)] \times 100$	Relative within-type inequality $IR(\widehat{X}_W)$
Azerbaijan	0.216	0.027	0.190	87.70
Bangladesh	0.184	0.024	0.160	86.89
Bolivia	0.159	0.030	0.129	81.25
Burkina Faso	0.246	0.024	0.221	90.05
Burundi	0.189	0.024	0.164	87.09
Cambodia	0.181	0.024	0.157	86.97
Cameroon	0.235	0.026	0.208	88.76
Chad	0.341	0.023	0.318	93.23
Colombia	0.114	0.028	0.086	75.79
Cote d'Ivoire	0.205	0.026	0.179	87.49
Egypt	0.351	0.028	0.323	92.05
Ethiopia	0.259	0.025	0.234	90.38
Guinea	0.271	0.023	0.248	91.45
Haiti	0.171	0.026	0.145	84.61
Honduras	0.127	0.027	0.100	78.69
Jordon	0.134	0.022	0.112	83.46
Kenya	0.261	0.025	0.236	90.48
Lesotho	0.234	0.022	0.212	90.50
Liberia	0.246	0.026	0.220	89.35
Morocco	0.305	0.029	0.276	90.39
Mozambique	0.266	0.024	0.242	90.91
Niger	0.301	0.022	0.279	92.59
Peru	0.132	0.028	0.104	78.84
Rwanda	0.192	0.024	0.167	87.25
Tanzania	0.200	0.024	0.176	87.77
Turkey	0.162	0.026	0.136	83.78

- outcome: height of toddlers $\rightarrow$ no effort
- substantial variation: interpreted as measurement error

Lara Ibarra & Martinez Cruz (2015, WB WP): Exploring the sources of downward bias in measuring in IOp

Table 4. Difference between true IOO and median estimated IOO in percentage terms: Baseline scenario

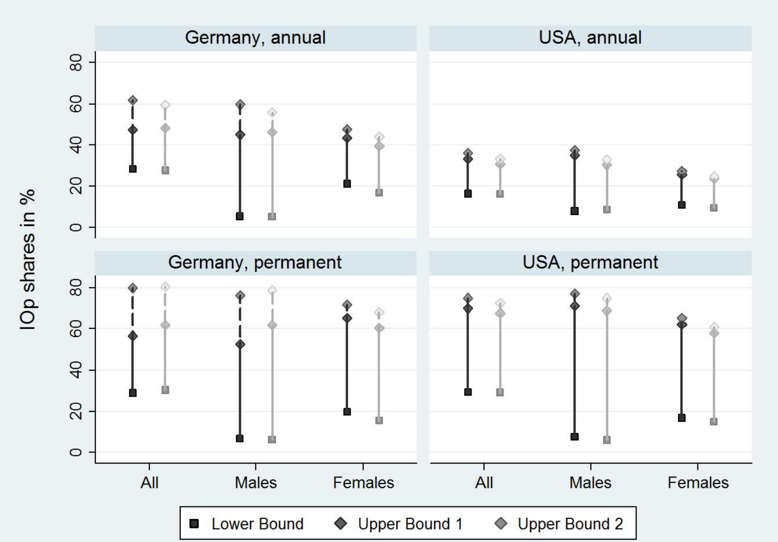
Observed population	Excluded Circumstances					
	None (1)	Gender (2)	Urban (3)	Region (4)	Father's Education (5)	Mother's education (6)
	True IO share = 0.978					
All	0.00	-27.88	-4.14	-39.82	-1.07	-5.35
Top 1% truncated	-0.14	-29.65	-4.56	-40.91	-1.27	-5.82
Top 5% truncated	-0.82	-37.14	-6.41	-42.76	-2.25	-8.02
	True IO share = 0.635					
All	0.00	-27.99	-4.29	-39.83	-1.18	-5.46
Top 1% truncated	-3.25	-32.14	-7.47	-41.66	-4.34	-8.76
Top 5% truncated	-12.81	-45.35	-17.71	-45.07	-14.05	-19.20
	True IO share = 0.468					
All	0.00	-27.95	-4.12	-39.77	-1.10	-5.32
Top 1% truncated	-5.03	-33.58	-9.35	-42.03	-6.24	-10.55
Top 5% truncated	-18.01	-48.01	-22.50	-47.24	-19.14	-23.76

Niehues/Peichl (2014, SCWE): two-stage estimator for upper bound

1. Fixed-effects earnings regression to derive measure of *constant unobserved heterogeneity*
   - individual FE captures all time-invariant variables: circumstances (per definition exogenous) and constant effort
   - = upper bound for the influence of circumstances
2. FE as circumstance measure to quantify maximum amount of IOp

- compare to lower bounds based on rich set of circumstance variables
  - Intuition: How much variance explained by FE vs. observed C?

Niehues / Peichl (2014, SCWE): baseline results



NP extended to dev countries (work in progress)

Year	Country	UB Level	Total Inequality	UB Ratio	Unit of Obs.
2013	Argentina	0.288	0.302	0.954	Individual
2010	China	0.540	0.583	0.926	Individual
2006	Mexico	0.877	1.221	0.718	Individual
2001	Malawi	1.239	1.514	0.818	Individual
2004	South Africa	0.602	0.754	0.799	Household
2009	Ethiopia	0.465	0.740	0.628	Household

What circumstances are we missing?

- Existing LB estimates much lower than UB
- FE indicate that unobserved ability and talent are important circumstances – see also Björklund, Jäntti and Roemer (2012)

What circumstances are we missing?

- Existing LB estimates much lower than UB
- FE indicate that unobserved ability and talent are important circumstances – see also Björklund, Jäntti and Roemer (2012)
- All accomplishments of child before "age of consent" (14 or 16 yrs) should be treated as due to circumstances – both nature and nurture.
- Hufe/Peichl/Roemer/Ungerer (2015): use NLSY & BCS data
  - use measures of (cognitive and non-cognitive) ability at this age and child health as circumstance
  - also more/better information on family background and childhood

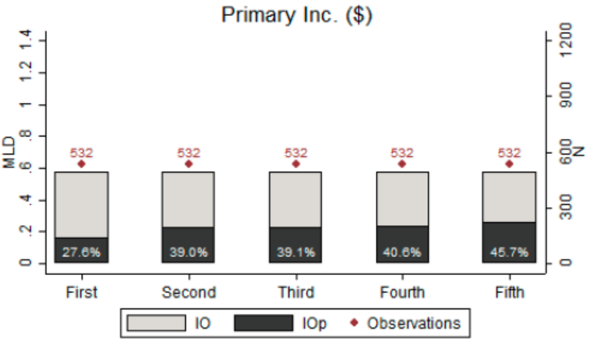
Circumstance sets

Scenario						Circumstance Set	Circumstance Var.
				Second	First	Base	Sex, Country of Birth, Ethnic Affiliation, Cohort, Age, Academic Achievement Mother, Occupation Code Mother, Rural/Urban, Height (16), Family Income
		Fourth	Third			Ability	PIAT Math, PIAT Reading
Sixth	Fifth					Behavioral Problems	Behavioral Problems Index (BPI)
						Child-Parent Relationship	Play/Schoolwork w/ Parents, Perceived Quantity of Time w/ Parents, Parents Split, Parental Income
						Health-Related Behavior	Smoking Habits Mother, Drinking Habits Mother, Health Restrictions Child
						Survey Specifics	Specific to NLSY79 and BCS70. See text for more information.

Table 1: Overview of Circumstance Scenarios

NLSY: baseline

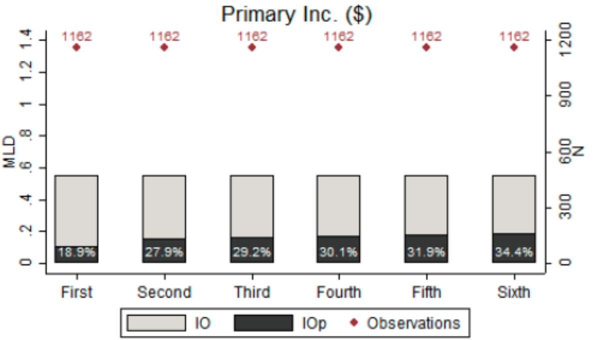
Figure 2: IOp with varying circumstance sets (NLSY79), comparable sample, average income



Note: The overall bar yields the extent of outcome inequality IO. The black colored share of each bar yields inequality attributed to circumstances, i.e. the lower bound absolute measure of inequality of opportunity IOp. The residual gray colored share of each bar can be interpreted as an upper bound measure of inequality attributed to differential efforts. The white labels at the bottom of each bar indicate the share of IOp in IO, i.e. the relative measure of inequality of opportunity r.

NLSY: average income

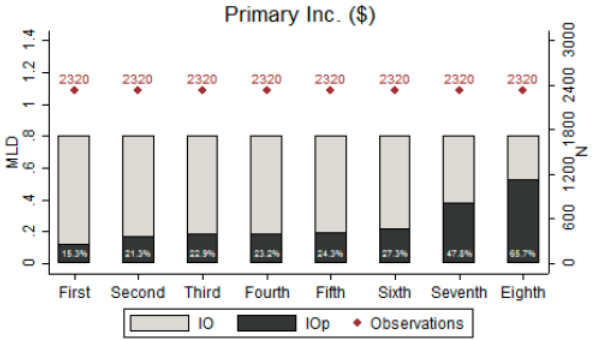
Figure 3: IOp with varying circumstance sets (NLSY79), survey-specific sample, average income



Note: The overall bar yields the extent of outcome inequality IO. The black colored share of each bar yields inequality attributed to circumstances, i.e. the lower bound absolute measure of inequality of opportunity IOp. The residual gray colored share of each bar can be interpreted as an upper bound measure of inequality attributed to differential efforts. The white labels at the bottom of each bar indicate the share of IOp in IO, i.e. the relative measure of inequality of opportunity r.

NLSY: pooled sample

Figure 4: IOp with varying circumstance sets (NLSY79), survey-specific pooled sample



Note: The overall bar yields the extent of outcome inequality IO. The black colored share of each bar yields inequality attributed to circumstances, i.e. the lower bound absolute measure of inequality of opportunity IOp. The residual gray colored share of each bar can be interpreted as an upper bound measure of inequality attributed to differential efforts. The white labels at the bottom of each bar indicate the share of IOp in IO, i.e. the relative measure of inequality of opportunity r.

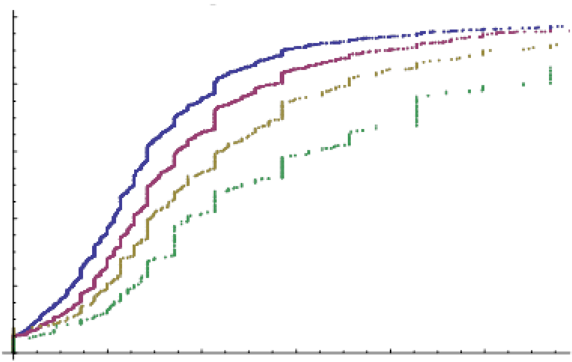
Properties of MLD

- Typically, share of IOp due to circumstances is surprisingly small "due to having information only on few circumstances"

- However: MLD often used to estimate IOp (because of axioms) ... and we are only able to "explain" some maximum amount of total inequality with any given set of C in its decomposition (Ravi Kanbur)
- Roemer (2015): maximum possible amount approx. 65% of total inequality (dep. on assumptions!) $\rightarrow$ IOR +54%:

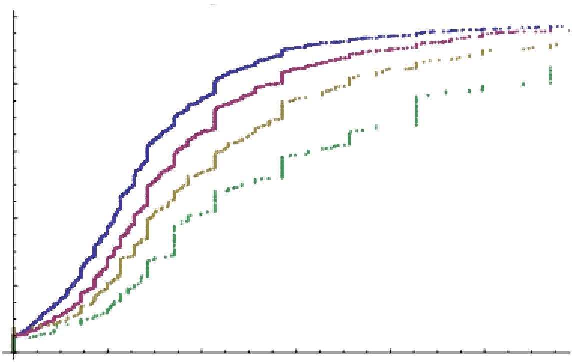
IOR	normalized IOR
10	15.38
20	30.77
30	46.15
40	61.54
50	76.92

IOp in Egypt: Assaad, Krafft and Roemer (2015)



- 4 types according to parental education → stochastic dominance

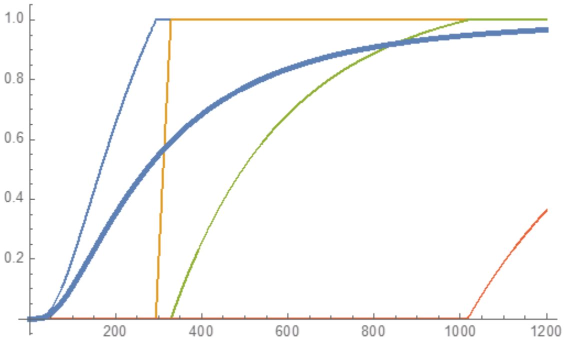
IOp in Egypt: Assaad, Krafft and Roemer (2015)



- 4 types according to parental education → stochastic dominance
- BUT: IOR = 10.3%. Why so low?

- Roemer (2015): what is maximum IOR possible given the data?

- Roemer (2015): what is maximum IOR possible given the data?



- "maximal" decomposition: the supports of the four component distributions are mutually disjoint → IOR = 83.3%
- Figure 1: supports of the four component distributions are essentially identical – very far from being disjoint.

Specification of earnings equation

- Hufe / Peichl (2015): “Lower bounds and the linearity assumption in parametric estimations of IOp”
- Standard approach:
  - Implicit Homogeneity Assumption: Effect of one C independent of other C
  - and: no type-specific effort variance

Specification of earnings equation

- Hufe / Peichl (2015): "Lower bounds and the linearity assumption in parametric estimations of IOp"
- Standard approach:
    - Implicit Homogeneity Assumption: Effect of one C independent of other C
    - and: no type-specific effort variance

	Female	Male
Graduate Mother	Type 1	Type 2
Non-Graduate Mother	Type 3	Type 4

An Implicit Homogeneity Assumption

- The standard approach would proceed as follows:

$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \tilde{\epsilon}_i \tag{4}$$

An Implicit Homogeneity Assumption

- The standard approach would proceed as follows:

$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \tilde{\epsilon}_i \tag{4}$$

- However, is the homogeneity assumption reasonable?

An Implicit Homogeneity Assumption

- The standard approach would proceed as follows:

$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \tilde{\epsilon}_i \tag{4}$$

- However, is the homogeneity assumption reasonable?
- If not, (4) is "biased":

$$\tilde{\epsilon}_i = \beta_4 C_i^{female} \times C_i^{HS} + \epsilon_i \tag{5}$$

- We should estimate instead:

$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \beta_4 C_i^{female} \times C_i^{HS} + \epsilon_i \tag{6}$$

Effort Levels and Effort Variance

The standard approach implicitly nets out type-specific differences in effort levels:

$$y = g(\Omega, \theta(\Omega), \epsilon) \tag{7}$$

Effort Levels and Effort Variance

The standard approach implicitly nets out type-specific differences in effort levels:

$$y = g(\Omega, \theta(\Omega), \epsilon) \tag{7}$$

However, it does not control for differences in type-specific effort variance.

Björklund et al. (2012) suggest the following remedy:

$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \beta_4 C_i^{female} \times C_i^{HS} + \epsilon_i + u_i - u_i \quad (8)$$

Björklund et al. (2012) suggest the following remedy:

$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \beta_4 C_i^{female} \times C_i^{HS} + \epsilon_i + u_i - u_i \tag{8}$$

$$u_i = \epsilon_i \frac{\sigma}{\sigma_{T^k}} \tag{9}$$

Björklund et al. (2012) suggest the following remedy:

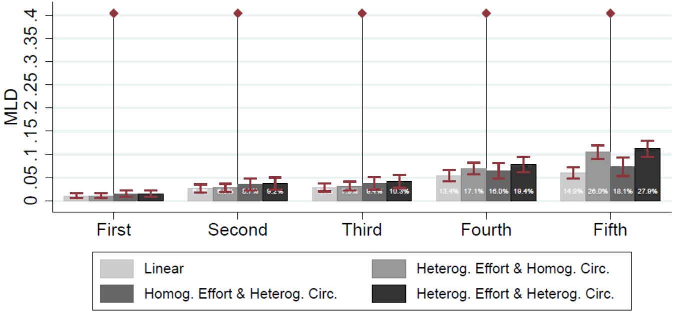
$$\ln y_i = \beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \beta_4 C_i^{female} \times C_i^{HS} + \epsilon_i + u_i - u_i \quad (8)$$

$$u_i = \epsilon_i \frac{\sigma}{\sigma_{T^k}} \quad (9)$$

$$\mu^k(p) = \exp\left[\beta_1 + \beta_2 C_i^{female} + \beta_3 C_i^{HS} + \beta_4 C_i^{female} \times C_i^{HS} + \epsilon_i - \underbrace{\epsilon_i \sigma / \sigma_{T^k}}_{=u_i}\right]$$

Application: NLSY data

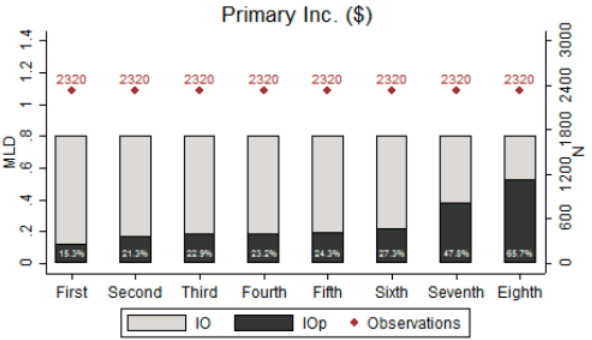
- 5 C vars: gender, race, region of birth, family income, parental education → 192 non-overlapping types



- Estimates of IOp are downward biased by neglecting type-specific heterogeneity in C influence

NLSY: pooled sample

Figure 4: IOp with varying circumstance sets (NLSY79), survey-specific pooled sample



Note: The overall bar yields the extent of outcome inequality IO. The black colored share of each bar yields inequality attributed to circumstances, i.e. the lower bound absolute measure of inequality of opportunity IOp. The residual gray colored share of each bar can be interpreted as an upper bound measure of inequality attributed to differential efforts. The white labels at the bottom of each bar indicate the share of IOp in IO, i.e. the relative measure of inequality of opportunity r.

1. Introduction
2. Conceptual Framework
3. Extensions
   - Upper bounds
   - Childhood characteristics
   - Maximum IOp
   - Interactions
   - Spouses
4. Summary

Peichl / Ungerer (2015): Role of spouses in couples

- Current approach (equation (3)) implicitly assumes full responsibility for partner's circumstance, income and effort variables.
- Peichl / Ungerer (2015): 3 extensions to baseline of *Full resp.*
- *(ii) Responsible for partners' circumstances and effort (unitary model)*:

$$\ln w_i = \psi C_i + \zeta \ln w_i^P + \eta_i. \qquad (10)$$

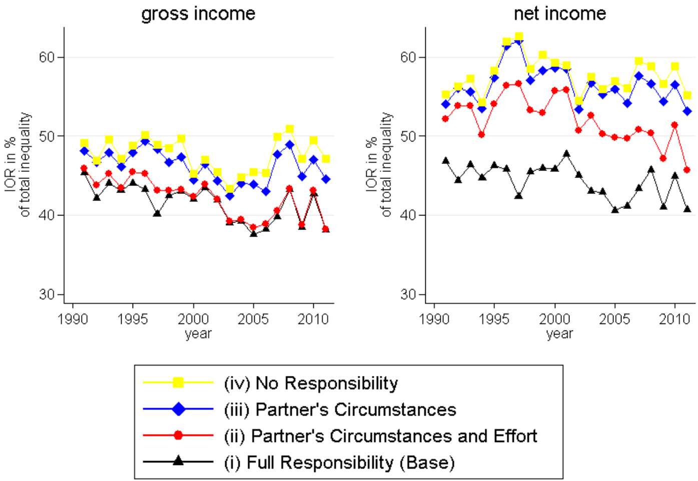
- *(iii) Responsible for partner's circumstances (collective model)*:

$$\ln w_i = \psi C_i + \zeta \ln w_i^P + \lambda E_i^P + \eta_i. \qquad (11)$$

- *(iv) No responsibility*:

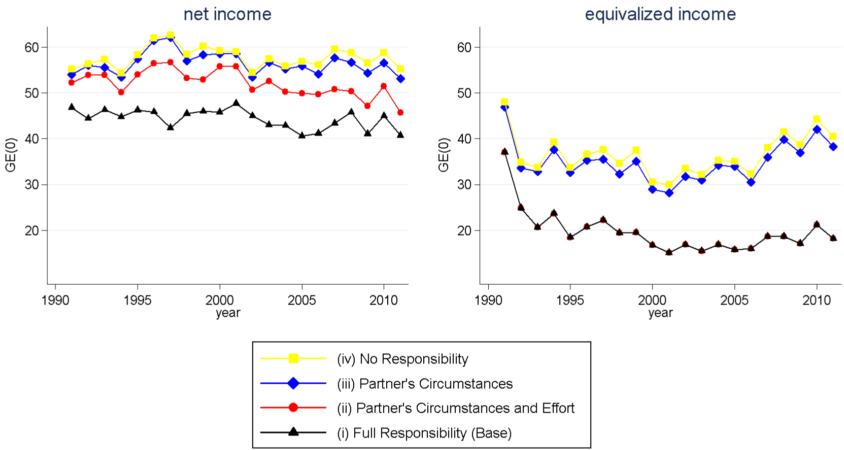
$$\ln w_i = \psi C_i + \zeta \ln w_i^P + \lambda E_i^P + \phi C_i^P + \eta_i. \qquad (12)$$

Accounting for the Spouse when Measuring IOp



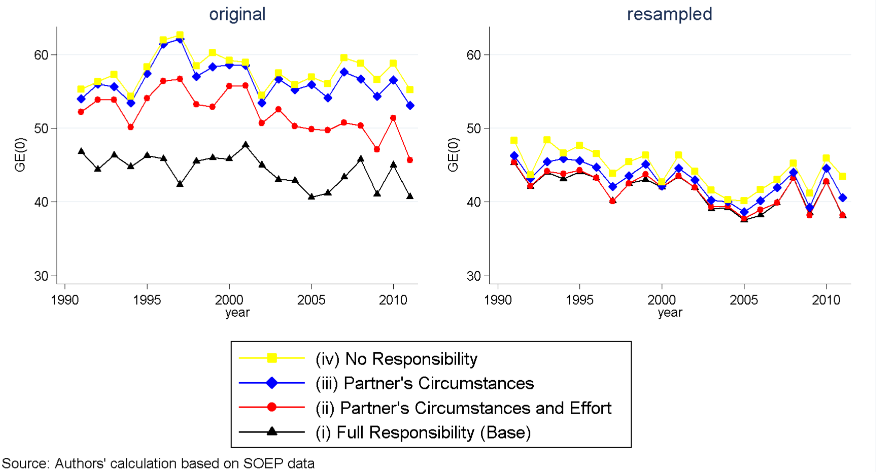
Source: Authors' calculation based on SOEP data

Individual vs. household income



Source: Authors' calculation based on SOEP data

Role of assortative mating?



Summary

- Previous IOp estimates too low ...
- good news: IOp estimates can be improved
- ... but more work needs to be done
  - Hufe & Peichl (2016): use genetic information as C
  - Hufe / Kanbur / Peichl (2016): Extend standard IOp with poverty sensitivity
  - ...

Link to ex-post approach

- Fleurbaey / Peragine / Ramos (2015): Ex Post Inequality of Opportunity Comparisons

		% Overall Inequality				
Class	.274	.324	.354	10	10	8
		[.350]	[.414]			
Type	.243	.294	.320			8
		[.318]	[.374]			
Tranche	.412	.344	.325		10	8
		[.371]	[.338]			
Class	.279	.331	.361	20	10	8
		[.357]	[.420]			
Class	.262	.312	.340	20	20	8
		[.337]	[.397]			
Tranche	.384	.320	.303		20	8
		[.345]	[.355]			

Thank you for your attention!

Comments? Questions?

peichl@zew.de