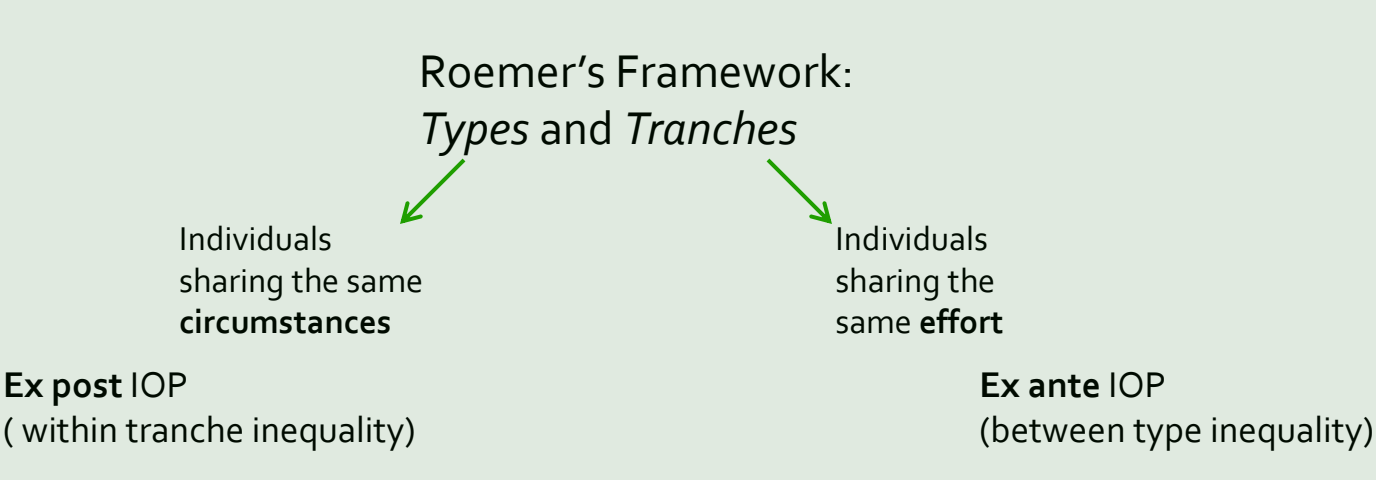
Sources of Income Inequality in China: Individual Effort or Circumstances

Dongjie Wu, Prasada Rao, Kam Ki Tang and Pravin Trivedi
(University of Queensland)

discussion by
Maria C. Lo Bue (University of Verona)

Premise of the Paper

- Rapid income growth in China, accompanied by increasing inequality (more rapid income growth of the rich)
- Drivers of inequality: globalization, migration and private ownership of assets
- Public perception of poor "procedural justice"
- Inequality caused by factors outside personal responsibility ("Inequality of Opportunity")



Research Objective:

- To evaluate IOP in China at both national and regional levels
  - Analyzing the contribution of IOP to overall income inequality over various development stages
  - Including individuals with zero income
  - Considering the effect of each circumstance on the heteroskedasticity

Snapshot of main findings:

- At the national level, circumstances account for around 31% of the income inequality in 2010 and 43% in 2012.
  The figures rise around 25% if heteroskedasticity between types as parts of IOP is included.

- GRP appears to have a negative relationship with income inequality and inequality of effort at the provincial level, but no discernible relationship with the level of IOP.
  → the share of IOP in the overall inequality rises with the increase of GRP.

- The results from the Oaxaca decomposition showed that getting rich does not require better circumstances per se but the bigger influence of circumstances to income. In addition, the shares of IOPs in the overall inequality are similar across regions.

Method: 1) Measure IoP

$y=g(c,e)$

- Set of types: C= {1,…,n}
- Set of tranches E={1,…,m}

$Y=\{Y_1,\ldots,Y_n\}$ with $\mu(Y_i)$ avg. outcome in type *i*

$Y=\{Y^1,\ldots,Y^m\}$ with $\mu(Y^j)$ avg. outcome in tranche *j*

Assumption 1: Function *g* is monotonically increasing in effort *e*

Assumption 2: The conditional distribution of effort *e* is independent of circumstances *c*

→Measure ex ante IOP by computing the ineq. of a counterfactual income distribution (Y_c) in which the contribution of effort has been eliminated or counterfactual income distribution Y_e by ruling out the contribution of circumstances

$$Y_c = \{\mu(Y_1)\mathbf{1}_{N_1},\ldots,\mu(Y_i)\mathbf{1}_{N_i},\ldots,\mu(Y_n)\mathbf{1}_{N_n}\}$$

$\mu(Y_i)$ predicted value when circumstances c_i corresponds to type *i*: $\mu(Y_i)=\bar{y}_i=f(\mathbf{c}_i)$

Method: 1) Measure IoP

Inequality of Opportunity Level (IOL)

$$IOL = I(Y_c)$$

Inequality of Opportunity Ratio (IOR)

$$IOR = \frac{I(Y_c)}{I(Y)}$$

Inequality of Opportunity Ratio (IOE)

$$IOE = I(Y_e)$$

$I : \mathbb{R}_+^N \mapsto \mathbb{R}_+$ is an ineq. Index (i.e. Gini, Variance, Theil)

➡ $I(Y) = I(Y_c) + I(Y_e)$

I () is path independent

Use Shapley Decomposition with the Gini coefficients
→ more flexibility to decompose IOP into each circumstances

Method: 2) Empirical Strategy

2.1. Lognormal Hurdle Model

- Taking into account the zero –income individuals
- Identifying the expected income for each type

2.2. Type Heterogeneity of effort

- The within type income distribution might be not identical between types

(i.e.heteroskedastisticty)

→indirect effects of circumstances on income inequality / Correlation between circumstances and effort

- Use MLE to identify the effects of each circumstance on the mean and the variance

2.3 Shapley Decomposition

- Computing the contribution of each circumstance

Method: 2) Empirical Strategy

2.4 Oaxaca Decomposition

- Analyzing how the effect of circumstances on income differ between groups (i.e. female and male, urban and rural, minority and majority, under-developed and developed region)

Decomposing the between-group difference to three components:

R=EN+CO+INT

EN: the extent to which the difference in income between groups is due to difference in circumstances between groups

$$EN = \{E(\mathbf{c}_A) - E(\mathbf{c}_B)\}'\beta_B$$

CO: the amount of inequality between groups coming from the effect of circumstances

$$CO = E(\mathbf{c}_B)'(\beta_A - \beta_B)$$

INT: interaction

$$INT = \{E(\mathbf{c}_A) - E(\mathbf{c}_B)\}'(\beta_A - \beta_B)$$

Data

China Family Panel Study 2010 and 2012

Table 1: Summary Statistics (Respondents)

Statistic	N	Mean	St. Dev.	Min	Max
Individual income(2010)	19,736	10,575.07	21,520.59	0.00	980,000.00
Individual income(2012)	19,736	14,519.89	30,255.13	0.00	1,804,500.00
Household income per capita(2010)	18,729	9,157.06	15,668.13	1.67	1,000,000.00
Household income per capita(2012)	19,248	12,117.97	16,630.24	0.20	612,700.00
Male	19,736	0.47	0.50	0	1
Minority	19,696	0.08	0.27	0	1
Age	19,736	42.25	10.79	21	60
Urban Hukou at age 12	19,625	0.15	0.35	0	1
Live in Coastal Province at age 12	19,736	0.43	0.50	0	1
Number of Sibling	19,550	3.01	1.88	0	14
Married	19,736	0.90	0.30	0	1
CCP Member in 2010	19,736	0.06	0.24	0	1
CCP Member in 2012	19,736	0.07	0.25	0	1

Table 2: Summary Statistics (Respondents' Parents)

Statistic	N	Mean	St. Dev.	Min	Max
Low Occupation	17,309	0.79	0.41	0	1
Mid Occupation	17,309	0.13	0.33	0	1
High Occupation	17,309	0.09	0.28	0	1
CCP member	19,736	0.16	0.37	0	1
Low Education	19,736	0.64	0.48	0	1
Mid Education	19,736	0.22	0.42	0	1
High Education	19,736	0.13	0.34	0	1

Note: 1. All variables are at respondents' 14 years old.

2. All variables only account the higher value within parents.

Data

Table 5: Income difference in dichotomous data (Part 1)

	(1)		(2)		(3)	
	Rich	Poor	Slow growth	Fast growth	Under-developed	Developed
HHincome(2010)	16055.9	6146.5	9217.6	9056.4	7290.5	11597.3
	(25127.0)	(6877.7)	(15409.0)	(16090.1)	(15940.3)	(14958.6)
INDincome(2010)	28921.4	2632.1	10962.8	10072.9	8211.8	13787.9
	(32154.1)	(2694.1)	(22354.5)	(20062.0)	(18983.2)	(24092.5)
HHincome(2012)	21043.6	7877.8	13759.9	9587.3	9836.1	15887.5
	(23922.6)	(8960.1)	(18589.9)	(12635.5)	(14309.4)	(19298.2)
INDincome(2012)	38667.6	3221.0	17453.0	10411.2	10914.0	20906.9
	(44562.4)	(3753.0)	(34894.7)	(21350.7)	(23083.1)	(39028.2)

Notes:1. HHincome is the household income per capita.

2. INDincome is the individual income per capita.

3. The values in parentheses are the standard error.

Table 6: Income difference in dichotomous data (Part 2)

	(1)		(2)		(3)	
	Female	Male	Majority	Minority	Rural	Urban
HHincome(2010)	9155.3	9236.0	9395.9	6850.0	6305.6	12921.4
	(14198.7)	(17250.9)	(16040.2)	(11279.2)	(11009.9)	(19615.4)
HHincome(2012)	11885.0	12396.2	12478.3	8096.8	8848.5	16358.8
	(15677.6)	(17981.0)	(17197.3)	(10859.4)	(10012.6)	(22046.7)
INDincome(2010)	7472.9	14252.8	11034.3	6682.5	7301.1	15044.2
	(15414.2)	(26359.5)	(22090.6)	(14046.5)	(14777.3)	(27395.6)
INDincome(2012)	10737.5	19139.4	15217.5	8924.9	8968.3	22128.9
	(19151.9)	(39425.7)	(31690.4)	(16118.4)	(15814.9)	(41802.9)

Notes:1. HHincome is the household income per capita

2. INDincome is the individual income per capita

3. The values in parentheses are the standard error.

- Rich/Poor: INDincome above/below avg.
- Slow/Fast growth & Under-dev./Dev.: GRP growth rates and levels (Chinese Stat. Yearbook)

Results: 1. IoP at the national level

Table 7: The Hurdle Model at the National Level

	Dependent variable:			
	2010		2012	
	logistic	*OLS*	*logistic*	*OLS*
	(1)	(2)	(3)	(4)
Male	1.664***	0.681***	1.667***	0.730***
	(0.064)	(0.028)	(0.060)	(0.027)
Minority	0.939	−0.288***	1.416***	−0.518***
	(0.116)	(0.052)	(0.125)	(0.050)
Urban Hukou at age 12	0.598***	0.801***	0.973	1.340***
	(0.083)	(0.043)	(0.085)	(0.042)
Coastal Province at age 12	0.768***	0.336***	0.687***	0.499***
	(0.062)	(0.028)	(0.058)	(0.027)
Mid education(Parents)	0.995	0.224***	0.861**	0.255***
	(0.078)	(0.036)	(0.071)	(0.035)
High education(Parents)	0.917	0.022	1.067	0.029
	(0.104)	(0.049)	(0.103)	(0.048)
Mid occupation(Parents)	1.137	0.200***	0.904	0.298***
	(0.100)	(0.045)	(0.091)	(0.045)
High occupation(Parents)	0.939	0.164***	0.825*	0.275***
	(0.114)	(0.054)	(0.106)	(0.054)
Member of CCP(Parents)	1.066	0.171***	1.124	0.152***
	(0.087)	(0.039)	(0.082)	(0.039)
Number of sibling	1.031*	−0.033***	1.005	−0.061***
	(0.017)	(0.008)	(0.016)	(0.007)
Constant	12.462***	7.663***	11.493***	7.816***
	(0.082)	(0.037)	(0.077)	(0.036)
Observations	17,009	15,852	17,009	15,672

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

The standard error is in the parenthesis.

Table 8: The MLE with Type Heteroskedasticity at the National Level

	Mean		Variance	
	2010	2012	2010	2012
Constant	7.651***	7.831***	1.017***	1.113***
	(0.036)	(0.036)	(0.031)	(0.031)
Male	0.677***	0.668***	0.078***	0.057**
	(0.027)	(0.027)	(0.023)	(0.023)
Minority	−0.251***	−0.499***	−0.203***	0.047
	(0.045)	(0.051)	(0.042)	(0.042)
Urban Hukou at age 12	0.803***	1.377***	0.004	−0.630***
	(0.044)	(0.035)	(0.035)	(0.035)
Coastal Province at age 12	0.318***	0.510***	0.342***	0.065***
	(0.028)	(0.027)	(0.023)	(0.023)
Mid education(Parents)	0.214***	0.205***	0.066**	0.068**
	(0.036)	(0.034)	(0.029)	(0.030)
High education(Parents)	0.032	0.062	−0.017	−0.026
	(0.047)	(0.045)	(0.040)	(0.040)
Mid occupation(Parents)	0.190***	0.284***	0.049	−0.052
	(0.046)	(0.041)	(0.037)	(0.037)
High occupation(Parents)	0.143**	0.261***	0.191***	0.058
	(0.057)	(0.051)	(0.044)	(0.045)
Member of CCP(Parents)	0.172***	0.146***	−0.058*	−0.053*
	(0.038)	(0.036)	(0.032)	(0.032)
Number of sibling	−0.026***	−0.056***	−0.048***	−0.019***
	(0.007)	(0.007)	(0.006)	(0.006)

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

The standard error is in the parenthesis.

Results: 1. IoP at the national level

Table 9: The Shapley decomposition at the National Level

	OLS		Hurdle model 1		Hurdle model 2	
	2010	2012	2010	2012	2010	2012
Gender	10.45	9.72	10.13	9.42	15.19	13.76
Ethnicity	0.98	1.52	0.94	1.40	1.63	1.63
Geographic	12.81	21.99	12.19	20.81	25.65	28.56
Parents' SOE	5.48	6.71	5.19	6.27	10.66	10.67
Sibling_number	1.96	3.55	1.78	3.25	5.52	6.45
Income: +/0			0.24	0.38	0.19	0.37
IOE	68.32	56.51	69.55	58.47	41.15	38.57
IOR	31.68	43.49	30.45	41.53	58.85	61.43

[1] OLS is the regression without zero-income. Hurdle model 1 is the regression using the hurdle model with type homoskedasticity. Hurdle model 2 is the regression using the hurdle model with type heteroskedasticity.
[2] The "Geographic" factor includes individuals' Hukou status when they were 12 years old.
[3] Parents' SOE is the parents' socioeconomic status which include parents' educational level, occupational status and political affiliations.
[4] Income: +/0 is the contribution of probability to have a positive income.
[5] All values are presented in percentage

The inclusion of zero income only slightly changed the Shapley value for each factor
In total IOR decreases by 1%→ those who have advantages in circumstances might be more likely to receive zero income

Gender and geographic characteristics are the two main sources of IOP

Difference in IOR between homoskedasticity and heteroskedasticity→ circumstances also largely affect income ineq. indirectly through effort

Results: 2. IoP at the regional level

Table 11: Inequality of Opportunity at the Regional Level(2010)

	Metropolitan	Mid-North	North	East	Mid-South	South	West	Northern West
GRP	74308	27476	33677	46421	23624	29891	20161	20126
Observed Gini	0.5729	0.6630	0.6305	0.6415	0.6623	0.7119	0.6506	0.6887
Gender	10.40	16.17	13.91	12.34	10.73	8.78	9.28	10.67
Ethnicity	0.02	2.84	0.31	0.03	0.50	0.61	3.57	0.47
Hukou	6.10	5.10	4.43	0.22	3.71	7.66	4.68	12.53
Parents' SOE	7.15	4.04	3.72	8.18	5.78	11.68	4.07	3.03
Sibling_number	3.11	0.15	2.89	3.09	1.67	0.03	1.09	0.72
Income: +/0	0.65	1.06	0.59	0.63	0.45	0.85	0.47	-0.06
IOE	72.57	70.65	74.16	75.51	77.16	70.39	76.84	72.64
IOR	27.43	29.35	25.84	24.49	22.84	29.61	23.16	27.36

[1] Parents' SOE is the parents' socioeconomic status which include parents' educational level, occupational status and political affiliations.
[2] Income: +/0 is the contribution of probability to have a positive income.
[3] All values are presented in percentage

Table 12: Inequality of Opportunity at the Regional Level(2012)

	Metropolitan	Mid-North	North	East	Mid-South	South	West	Northern West
GRP	88674	35106	45259	59063	31916	38509	28174	29137
Observed Gini	0.5132	0.6854	0.6183	0.6608	0.6454	0.6730	0.7013	0.7036
Gender	4.74	16.20	9.56	11.84	12.97	7.61	10.39	12.79
Ethnicity	-0.00	1.74	0.06	0.11	0.41	0.21	4.02	0.74
Hukou	11.25	7.01	16.29	4.87	9.07	14.39	8.55	15.53
Parents' SOE	6.90	7.92	4.05	4.39	8.27	11.49	5.15	5.23
Sibling_number	5.92	0.47	2.89	7.04	3.95	3.20	1.58	0.28
Income: +/0	2.27	0.65	0.91	0.68	0.20	1.00	0.07	0.58
IOE	68.92	66.00	66.23	71.08	65.13	62.11	70.24	64.86
IOR	31.08	34.00	33.77	28.92	34.87	37.89	29.76	35.14

[1] Parents' SOE is the parents' socioeconomic status which include parents' educational level, occupational status and political affiliations.
[2] Income: +/0 is the contribution of probability to have a positive income.
[3] All values except GRP and Gini are presented in percentage

Lower figures than those at the national level (probably because the regional disparity contributes IOP at the national level).

The differences between the highest and the lowest IOR are around 7% in 2010 and 9% in 2012 →regional disparity in inequality of opportunity in China.

Regional disparity exists not only in income inequality but also in its sources.

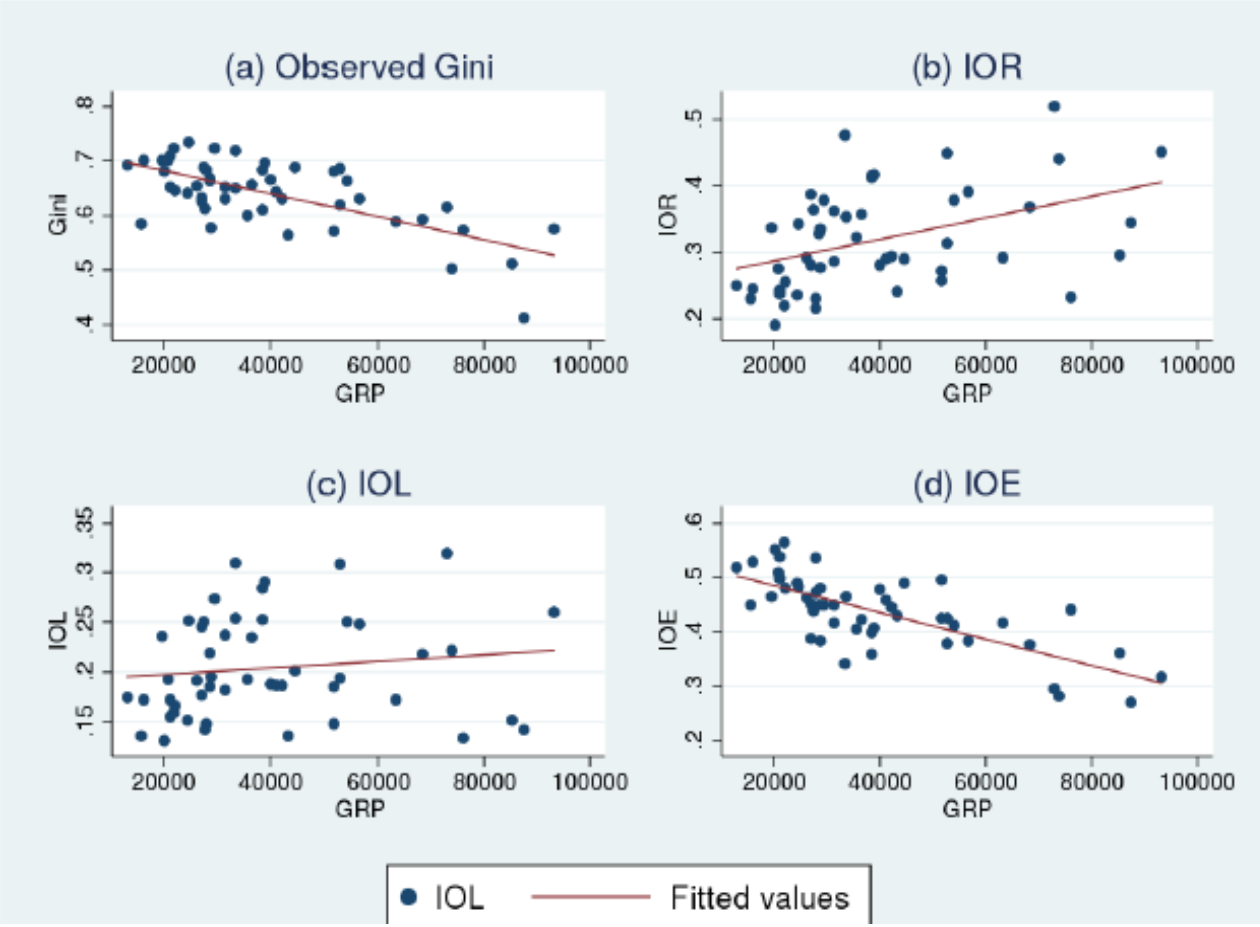
Rich regions: lower level of income inequality but higher IOR;

Poor regions: higher level of income inequality but lower IOR.

The contributions of the three main sources of income ineq. varies remarkably across regions.

Results: 3. Relationship between Provincial IoP and GRP

Figure 4: Provincial Inequality and GRP per capita



IOPs computed using the Shapley decomposition with the hurdle model

Income inequality reduces from about 0.7 to 0.5 when GRP per capita rises from below 20,000 Yuan to more than 90,000 Yuan.

This reduction seems mostly contributed by the decrease in IOE → a poor province has a more diverse distribution of effort or a bigger influence of effort on income inequality.

Results: 4. Oaxaca Decomposition

Table 13: Oaxaca Decomposition(2010)

	(1) Rich	(2) Growth	(3) Developed	(4) Male	(5) Urban	(6) Majority
Differential						
Disadvantage	7.375***	8.200***	8.015***	7.901***	7.981***	7.816***
	(0.0157)	(0.0198)	(0.0185)	(0.0209)	(0.0167)	(0.0481)
Advantage	10.07***	8.276***	8.537***	8.588***	9.133***	8.263***
	(0.00908)	(0.0256)	(0.0271)	(0.0228)	(0.0357)	(0.0165)
Difference	-2.697***	-0.0765**	-0.522***	-0.686***	-1.152***	-0.446***
	(0.0181)	(0.0324)	(0.0328)	(0.0309)	(0.0394)	(0.0509)
Decomposition						
Endowments	-0.139***	0.0404**	-0.0313	-0.0135	-0.354***	-0.152***
	(0.00822)	(0.0154)	(0.121)	(0.00878)	(0.0397)	(0.0177)
Coefficients	-2.668***	-0.167***	-0.0716	-0.682***	-0.762***	-0.339***
	(0.0242)	(0.0329)	(0.0933)	(0.0300)	(0.162)	(0.0538)
Interaction	0.109***	0.0501**	-0.419**	0.00868	-0.0359	0.0442*
	(0.0182)	(0.0170)	(0.150)	(0.00533)	(0.162)	(0.0257)
Observations	12724	12724	12724	12724	12711	12724

Advantage is the predicted income when the dummy variable listed in column equal to 1.

Disadvantage is the predicted income when the dummy variable listed in column equal to 0.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Table 14: Oaxaca Decomposition(2012)

	(1) Rich	(2) Growth	(3) Developed	(4) Male	(5) Urban	(6) Majority
Differential						
Disadvantage	6.734***	7.981***	7.641***	7.382***	7.501***	7.315***
	(0.0270)	(0.0332)	(0.0271)	(0.0331)	(0.0254)	(0.0752)
Advantage	10.37***	7.627***	8.190***	8.350***	9.019***	7.879***
	(0.00842)	(0.0332)	(0.0455)	(0.0331)	(0.0555)	(0.0251)
Difference	-3.639***	0.354***	-0.549***	-0.967***	-1.518***	-0.564***
	(0.0283)	(0.0470)	(0.0530)	(0.0469)	(0.0610)	(0.0793)
Decomposition						
Endowments	-0.148***	0.0888	-0.203	-0.0145	-0.508***	-0.174***
	(0.00785)	(0.0803)	(0.169)	(0.0100)	(0.0617)	(0.0231)
Coefficients	-3.768***	-0.0833	-0.210**	-0.955***	-1.594***	-0.428***
	(0.0441)	(0.0653)	(0.0844)	(0.0461)	(0.245)	(0.0845)
Interaction	0.277***	0.349***	-0.136	0.00170	0.584**	0.0391
	(0.0351)	(0.0924)	(0.182)	(0.00837)	(0.245)	(0.0404)
Observations	13561	13561	13561	13561	13545	13561

Advantage is the predicted income when the dummy variable listed in column equal to 1.

Disadvantage is the predicted income when the dummy variable listed in column equal to 0.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Dataset separated by individuals' income, growth rate at the province level, GRP per capita at the province level, gender , Hokou status and ethnicity

Results: 4. Oaxaca Decomposition

•Predicted income of the advantaged groups is all higher than the disadvantaged except "growth" in 2012.

•The highest difference comes from the comparison between the high-income group and the low-income group.

• Most of them can be explained by the difference in coefficients even though the endowments effect is also significant.

• The rich has slightly better circumstances but their incomes are greatly benefited from their circumstances →better circumstances might help to get ahead in the society but only slightly (those who have got ahead take better advantage of their circumstances)

Discussion: General Questions and Comments

1. Individual Income and Household Income per capita
(Ind. income=labour income +HH's business income/HH size)
HH Income per capita: how is it defined?

2. "…Over 79% individuals report their parents low status of occupation"
(→why not considering further disaggregation?)

3. What about age? Maybe running reg.separately for different cohorts…

4. Individuals with zero income: from table 3 it seems that they have higher socio-ec. Background

5. At p. 24: "Parents' SOE affects the amount of income earned but has less implication in the labour participation" --> Where can I see these results?

6. Income variables were not adjusted for inflation…

Thanks for the nice paper!

Maria C. Lo Bue
mlobue@uni-goettingen.de

Table 3: Zero Income Vs Positive Income (the Independent t-test)

	(1) 2010		(2) 2012	
Male	-0.121***	(-7.95)	-0.123***	(-8.69)
Minority	-0.00141	(-0.17)	-0.0255***	(-3.35)
Coastal province at age 12	0.0729***	(4.82)	0.0989***	(7.00)
Urban Hukou status at age 12	0.0750***	(7.05)	0.0164	(1.64)
Mid education(Parents)	0.0194	(1.51)	0.0405***	(3.36)
High education(Parents)	0.0132	(1.49)	-0.00462	(-0.56)
Mid occupation(Parents)	0.00700	(0.69)	0.0169*	(1.78)
High occupation(Parents)	0.0175**	(2.07)	0.0168**	(2.11)
Member of CCP(Parents)	0.00556	(0.48)	-0.00429	(-0.40)
Number of sibling	-0.165**	(-2.88)	-0.0666	(-1.24)

[1] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[2] The standard error is in the parentheses.

[3] The coefficients represent the mean difference between the zero-income group and the positive-income group.