Economic Mobility in Rural South Asia

Andrew Foster

Brown University

June 10, 2016

THE POPULATION STUDIES AND TRAINING CENTER

BROWN

Introduction

- Repeated cross-sections can be used to study how inequality changes over time.
- But changes the distribution of inequality is an aggregation of responses of individual people to economic circumstances.
- Thus helpful to look at transitions over time–economic mobility.
- Especially important for looking at interactions between household behavior and policies.
    - Does initial advantage at the household level persist?
    - How does village level advantage affect household economic mobility?
    - Does basic primary care and education lead to greater economic mobility and reduce inequality?

Kuznets

- The units for which incomes are recorded and grouped should be family-expenditure units.
- We should be able to trace secular income levels not only through a single generation but at least through two-connecting the incomes of a given generation with those of its immediate descendants.
- Even in this simple initial sketch, findings in the field of demography were used...Uncomfortable as are such ventures into unfamiliar and perhaps treacherous fields, they can not and should not be avoided.

What population are we interested in? I

- Population of individuals
    - Population is clear
    - Limited use over long periods of time
    - Not suitable for looking at changes in age-specific behavior like schooling
    - Individual well-being and its measurement affected by household coresidence

What population are we interested in? II

- Population of biological descendants
  - Clearly defined in principle
  - Biological descendants are often only visible if there is coresidence
  - We do not typically have data on all relevant biological antecedents (e.g., spouse's origin)
  - Generations may be asyncronous (e.g., set of grandchildren born over a large span of time)
  - How do you account for household spillovers?

What population are we interested in? III

- Population defined by households in region
  - Reasonably well-defined at each point in time
  - Captures importance of household in allocation of resources
  - Political and administrative units are spatially defined
  - Symmetry between antecedant and descendent populations
  - Can be implemented in a regionally defined surveilance system
  - Misses consequences of programs, for example, for outmigrants

India

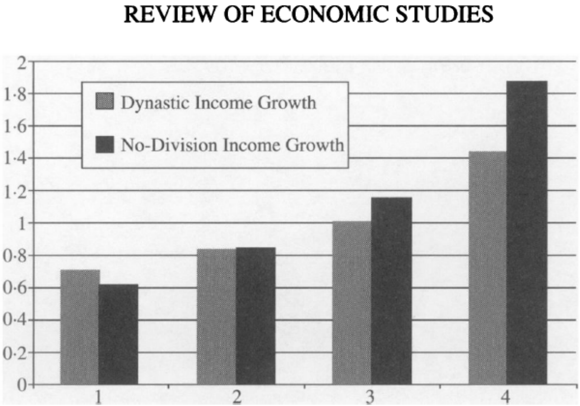
- NCAER REDS/ARIS/SEPRI Data
- Roughly 4000 households in 1967
- Listing of all households and village questionaire for 240 villages
- Resurvey 1969, 1969, 1982, 1999, 2006, 2016
- Some proxy information on migrants and "'refresh"' samples
- Links within village for follow-up households

1970-1982

- Foster and Rosenzweig 2002
- Complicated by follow-up only of households that have same head or did not split off.
- Developed a model that captured tradeoff between educational externality within household and the centripetal forces of heterogeneity in schooling.

Effects of Division

Figure 1: Division and inequality 70-82



Expenditure

Figure 2: Expenditure 82-99

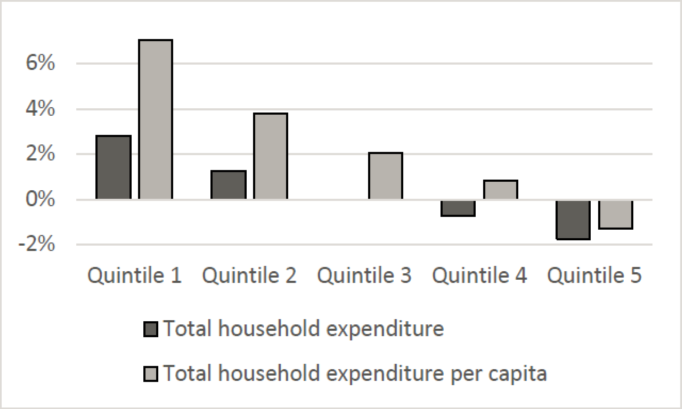


Figure 2A. Mean percentage change per annum between 1982 and 1999 in total household

Expenditure

Figure 3: Expenditure 99-06

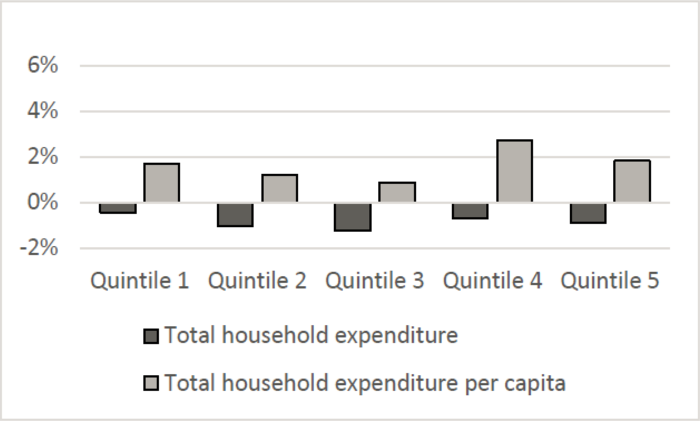


Figure 2B. Mean percentage change per annum

Size

Figure 4: Size 82-99

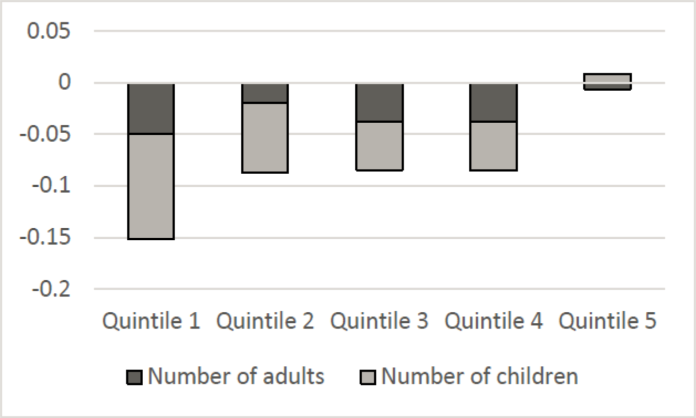


Figure 4A. Mean change in number of household

Size

Figure 5: Size 99-06

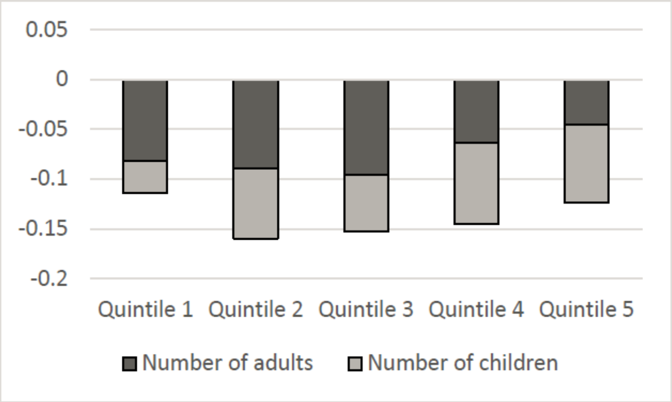


Figure 4B. Mean change in number of household

Split

Figure 6: Split 82-99

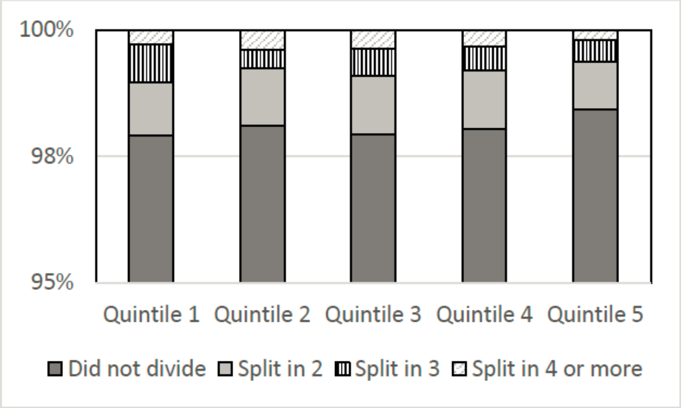


Figure 6A. Percentage of households who divided

Split

Figure 7: Split 99-06

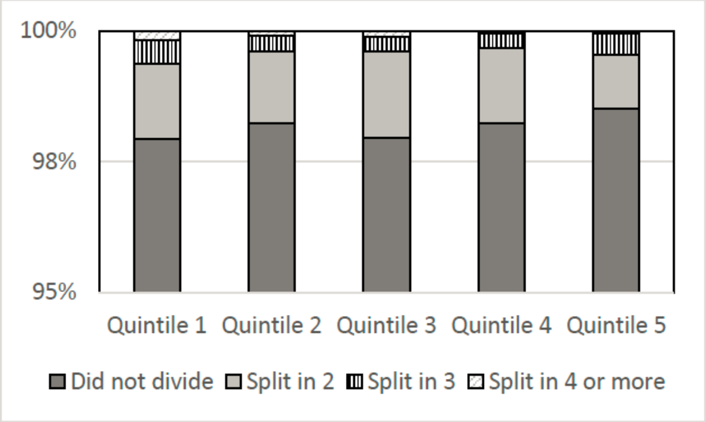


Figure 6B. Percentage of households who divided

Overview

- Demographic change importantly linked to upward economic mobility
- Some inequality reduction in early period linked to reduction in number of children but persistent afterward.
- Other work showing that entry of schools of heterogeneous quality and cost raised schooling among poor but did not translate into higher occupational mobility or migration.
- Human capital formation may effect process of household division and thus given scale economies possible differences in inequality.
- Collection of data follows split-offs but not antecedents–is this a problem?

Recombination

- Data are typically collected from samples of households due to importance of the household in the allocation of resources
- Households are not fixed over time–they "recombine".
- Probability of observing particular individuals/household depends on endogenous behaviors inclusive of recombination
- Therefore to construct weights one needs to keep track of multiple antecedent and desdendants
- With an ex-post household sample one needs to account for selective recruitment of descdendants

Matlab

- Geographically contiguous area of rural Bangladesh with population of 180,000 in 1974.
- There is an ongoing Demographic Surveillance System
- Site of a Health and Family Planning Program in 1978
- Long term follow up in 2014 designed in part to look at longer run consequences of program (not the focuse of this study)

Data

- Periodic censuses with basic data 1974, 1982, 1996
- Vital registration and household residence 1974-2014
- Comprehensive survey of sample 1996 (MHSS), 2014 (MHSS2)
- Most measures of economic well-being, including education for children post 1982, is only available from surveys

Original Sampling Methodology

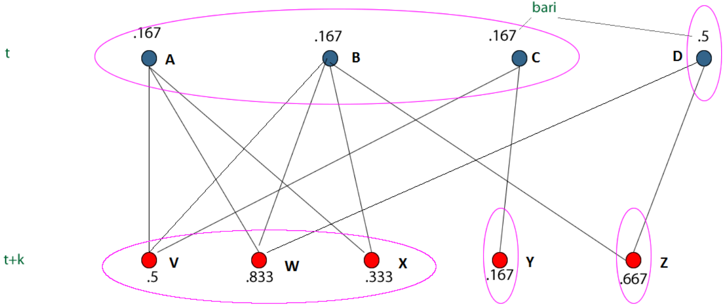
- 1996 MHSS sample based on 1993 census of Matlab population
- Households grouped into baris
  - 7,440 baris in 1993
  - Bari = cluster of households in close physical proximity linked often in a kin network
  - Bari size varies from 1 household to 10, 20 or more
- For the 1996 MHSS, 2,883 baris were randomly picked and one household from each bari was randomly selected
- 2014 MHSS2 followed all descendants of 1996 MHSS, including migrants

Cross-sectional weights I

- Use panel data derived from a random sample of a population in a particular area to estimate population parameters in the same area at a different point in time
- Sample is constructed from households containing descendants of original households augmented with a sample af fully immigrant households in the area.
- Household recombination has not received much attention either from a theoretical or sampling perspective.
  - given typical desent rules the probability of including a t+k household depends on the joint distribution over all period t antecedents,
  - not just the particular antecedents in the sample that caused that household to be selected.

Cross-sectional Weights II

Figure 8: Sampling Probabilities with Household Recombination



Definition of Descendant Household

- Someone in the 1993/2014 household also lived in the 1974 household (Zero Order Link)
- Someone in the 1993/2014 household has lived with a member of the 1974 household at any point between 1974 and 1993/2014 (First Order Lived with Link)
- Someone in the 1993/2014 household has lived with a person who lived with a member of the 1974 household prior to living with the 1993/2014 person (Second Order Forward Lived with Link)

Household Links Construction I

- H_1 : Partitions people (a, b, c, d) into households (α, β) at time 1
- H_2 : Partitions people (b, c, d, e, f) into households $(\delta, \gamma, \epsilon)$ at time 2
- P_1 : Maps people from period 1 to period 2

$$H_1 \quad \begin{array}{c c} & \begin{array}{cc} \alpha & \beta \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \end{array} \qquad H_2 \quad \begin{array}{c c} & \begin{array}{ccc} \delta & \gamma & \epsilon \end{array} \\ \begin{array}{c} b \\ c \\ d \\ e \\ f \end{array} & \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array} \qquad P_1 \quad \begin{array}{c c} & \begin{array}{ccccc} b & c & d & e & f \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

Household Links Construction II

- $C_t = H_t * H_t^T$: Who is coresident with whom at time t
- $LI_1 = C_1 * P_1 * C_2$: Links of people at time 1 to people at time 2 based on coresidence in each period
- $LH_1 = H_1^T * LI_1 * LI_2 * ... * LI_{t-1} * H_t$: Links of households at time 1 to households at time t

$$C_1 \quad \begin{array}{c c} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array} \qquad LI_1 \quad \begin{array}{c c} & \begin{array}{ccccc} b & c & d & e & f \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \end{array} \qquad LH_1 \quad \begin{array}{c c} & \begin{array}{ccc} \delta & \gamma & \epsilon \end{array} \\ \begin{array}{c} \alpha \\ \beta \end{array} & \begin{pmatrix} 2 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} \end{array}$$

Links in Practice

Table 1: Number of Links Between Selected 1974 and 1996 Households

	1996 Households										
	A11*	B11	C11*	D11	D12	D13	E11	C12	F11	B12	B13
1974 Households											
A01	3	4	0	0	0	0	0	0	0	0	0
D01	3	4	7	5	4	4	7	4	5	1	4
G02	3	4	0	0	0	0	0	0	0	0	0
D02	0	0	7	5	4	0	0	4	5	3	4
D03	0	0	1	3	3	0	0	0	5	0	0
C01	0	0	7	0	0	0	0	4	0	0	0
C02	0	0	7	0	0	0	0	4	0	0	0
E01	0	0	0	0	0	1	7	0	0	0	0
H01	0	0	0	0	0	0	0	0	0	1	2
J01	0	0	0	0	0	0	0	0	0	3	0
K01	0	0	0	0	0	0	0	0	0	3	4
G01	0	0	0	0	0	0	0	0	0	3	4
L01	0	0	0	0	0	0	0	0	0	3	0
B01	0	0	0	0	0	0	0	0	0	1	2

*Household was in the 1996 MHSS

Formal Description of Cross-Sectional Weights I

- I_t is the set of households i at time t
- I_{tx} is the set of households i at time t of type x
- $A : I_{t+1} \Rightarrow I_t$ where $A(K)$ is the set of households in I_t that contain antecedents of members of household $K \subset I_{t+1}$
- J_{t+1x} is the set of households j at time $t+1$ such that $A(j) \subset I_{tx}$
- N_{tx} is the number of households i at time t of type x
- S_t is the sample drawn at time t

Formal Description of Cross-Sectional Weights II

- Then for some outcome c_{it} among a population with characteristics x, if $\bar{c}_{tN} = \frac{1}{N_{tx}} \sum_{i \in I_{tx}} c_{it}$
- and $\hat{c}_{tN} = \frac{1}{N_{tx}} \sum_{i \in I_{tx}} c_{it} \frac{\mathbb{1}(i \in A(S_{t+1}))}{\mathbb{E}(\mathbb{1}(i \in A(S_{t+1})))}$
- and $\hat{i}_{tN} = \frac{1}{N_{tx}} \sum_{i \in I_{tx}} \frac{\mathbb{1}(i \in A(S_{t+1}))}{\mathbb{E}(\mathbb{1}(i \in A(S_{t+1})))}$
- then $\operatorname{plim}_{N \to \infty} \frac{\hat{c}_{tN}}{\hat{i}_{tN}} - \bar{c}_{tN} = 0$.
- $\mathbb{E}(\mathbb{1}(i \in A(S_{t+1})))$ can be constructed by simulation using the sampling procedure, the frame from which the sample was drawn, and all antecedent links for these households.
- Note that weights are affected by behavior but any deviation of estimated from population is only sampling error

Table 2: Different 1974 Sample Weights Compared to the Full 1974 Population

	(1)	(2)	(3)	(4)
	Mean Values for 1974 Population and Weighted Samples			
	Full 1974 Population	Our Resampling Weights	No Weights	Propensity Score Weights
Highest Edu	4.136	4.215	4.254	4.180
Number of Cows	1.158	1.179	1.355	1.173
Edu of Head	2.272	2.270	2.130	2.285
Age of Head	45.73	45.63	46.67	45.85
Household Size	6.071	6.090	6.777	6.164
Num of Rooms	1.219	1.210	1.281	1.214
Observations	24,788	5,319	5,319	5,309
Weights	24,788	24,029	5,319	24,594
	P Values for Difference between Full Population and Sample			
		Our Resampling Weights	No Weights	Propensity Score Weights
Highest Edu		0.349	0.036	0.432
Number of Cows		0.528	0.000	0.504
Edu of Head		0.978	0.002	0.787
Age of Head		0.784	0.000	0.549
Household Size		0.718	0.000	0.012
Num of Rooms		0.407	0.000	0.477

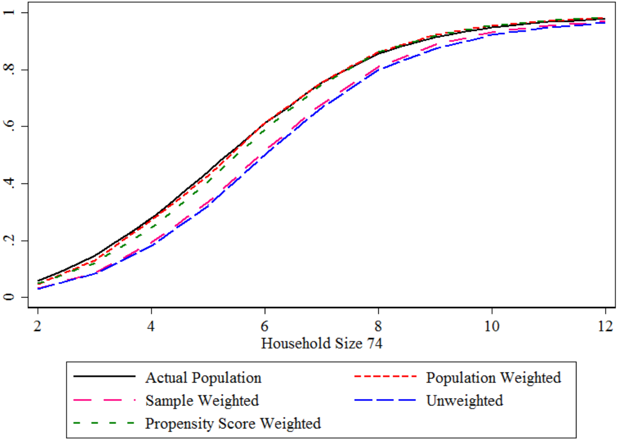
Table 3: Propensity Score Regression

VARIABLES	Dep Var: Dummy=1 if Linked to MHSS1
Highest Edu in Household	-0.0106
	(0.00654)
Household Size	0.130***
	(0.00691)
Articles Owned	-0.000998
	(0.00180)
Number of Cows	0.0165
	(0.0110)
Number of Boats	0.00637
	(0.0289)
Edu of Head of Household	-0.0256***
	(0.00704)
Age of Head of Household	-0.00124
	(0.00126)
Constant	-1.949***
	(0.0615)
Observations	24,757

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Importance of data from unsampled households

Figure 9: Estimated Distribution of 74 Household Size with Various Weights



Family Size 2012 I

- Standard approach carries 1996 weights forward to descendants
- But this is problematic if multiple antecedants are in the sample
  - An issue here because of high sampling probability
  - 1-(1-p1)*(1-p2)
  - But not independent!
- Even if only one antecedants in sample there is a problem of other antecedants not in sample.
- Account for thce fact that descent rules for sampling and estimation may not be the same. (links versus children/spouses/grandchildren).

Family Size 2012 II

Table 4: Estimates of Family Size in 2012 by Weighting Scheme

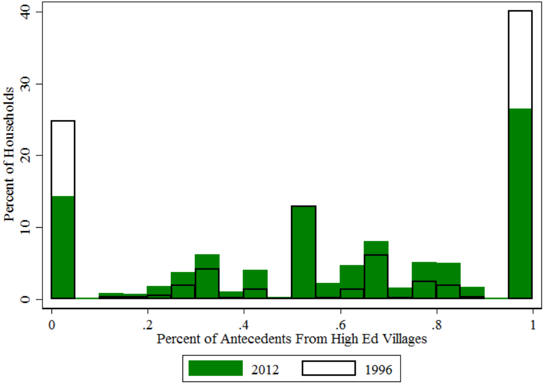
	Full Population	Random Smpl All Desc.	Random Smpl One Desc per 1974 HH	MHSS1 Linked Smpl 1996 Wts	MHSS1 Linked Smpl 2012 Wts
Avg Family Size 2012	4.394795	4.443986	4.542926	4.575445	4.398142
Obs	49,988	25,783	4971	4671	4681

Measurement of Education

- Village Level: Dummy for High vs Low Education village based on percent literate in the village (>40%=High)
- Household Level: Z-score created based on age dependent average education in 1996 for children ages 6 to 16

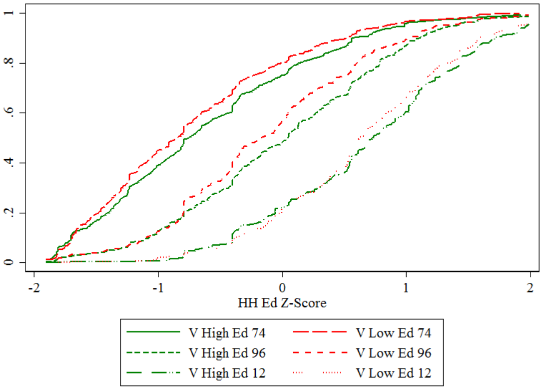
Mixing

Figure 10: Distribution of Fraction of Antecedent Households from 1974 High Ed Villages Among All Households in 1996 and 2012



Cross-sectional change

Figure 11: Distribution of HH Ed Z-score by Year and Village Ed



Regression Estimates 1996

Table 5: Population Estimates 1996 by HH Child Z-score and 74 Village Ed

VARIABLES	(1) Change in Household Size	(2) Change in Household Size	(3) Descdant HHs	(4) Descdant HHs	(5) Consumption Growth	(6) Consumption Growth
Ed Low	-0.0523	0.0223	0.197**	0.115**	-0.0318***	-0.0347***
	(0.0529)	(0.0344)	(0.0779)	(0.0520)	(0.00442)	(0.00295)
Ed High	-0.225***	-0.110**	-0.582***	-0.762***	0.0338***	0.0247***
	(0.0714)	(0.0433)	(0.0913)	(0.0556)	(0.00566)	(0.00351)
H Size Low	2.015***	2.046***	-1.234***	-1.186***	-0.0236***	-0.0252***
	(0.0521)	(0.0335)	(0.0737)	(0.0467)	(0.00469)	(0.00312)
H Size High	-3.337***	-3.411***	1.635***	1.625***	0.0582***	0.0634***
	(0.0823)	(0.0526)	(0.117)	(0.0754)	(0.00559)	(0.00353)
Cons Low	-0.577***	-0.597***	0.219***	0.363***	0.166***	0.168***
	(0.0600)	(0.0391)	(0.0847)	(0.0554)	(0.00448)	(0.00296)
Cons High	0.290***	0.303***	-0.318***	-0.311***	-0.225***	-0.232***
	(0.0628)	(0.0394)	(0.0854)	(0.0517)	(0.00564)	(0.00358)
VH x Ed Low	0.132*		-0.141		-0.00449	
	(0.0697)		(0.106)		(0.00590)	
VH x Ed High	0.184**		-0.286**		-0.0143**	
	(0.0899)		(0.113)		(0.00721)	
VH x H Size Low	0.0526		0.0829		-0.00288	
	(0.0676)		(0.0925)		(0.00628)	
VH x H Size High	-0.122		-0.0232		0.00841	
	(0.108)		(0.150)		(0.00726)	
VH x Cons Low	-0.0394		0.259**		0.00382	
	(0.0788)		(0.114)		(0.00590)	
VH x Cons High	0.0184		0.0151		-0.0112	
	(0.0803)		(0.108)		(0.00726)	
V High Ed	-0.260***	-0.181***	0.00461	0.0127	0.0230***	0.0167***
	(0.0740)	(0.0320)	(0.109)	(0.0466)	(0.00585)	(0.00266)
Constant	-0.755***	-0.803***	4.222***	4.218***	0.0802***	0.0841***
	(0.0579)	(0.0418)	(0.0878)	(0.0645)	(0.00455)	(0.00333)
Observations	19,820	19,820	19,822	19,822	19,313	19,313
R-squared	0.461	0.461	0.120	0.120	0.463	0.462

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression Estimates 2012

Table 6: Population Estimates 2012 by HH Child Z-score and 74 Village Ed

VARIABLES	(1) Change in Household Size	(2) Change in Household Size	(3) Descdant HHs	(4) Descdant HHs
Ed Low	0.207***	0.250***	0.0778	-0.0486
	(0.0387)	(0.0256)	(0.161)	(0.108)
Ed High	-0.399***	-0.316***	-1.185***	-1.437***
	(0.0563)	(0.0349)	(0.195)	(0.120)
H Size Low	2.324***	2.342***	-2.646***	-2.455***
	(0.0382)	(0.0246)	(0.152)	(0.0975)
H Size High	-3.606***	-3.673***	3.322***	3.454***
	(0.0668)	(0.0440)	(0.249)	(0.159)
Cons Low	-0.421***	-0.421***	0.321*	0.612***
	(0.0441)	(0.0294)	(0.179)	(0.119)
Cons High	0.270***	0.264***	-0.588***	-0.623***
	(0.0482)	(0.0300)	(0.177)	(0.110)
VH x Ed Low	0.0757		-0.218	
	(0.0514)		(0.216)	
VH x Ed High	0.134*		-0.398	
	(0.0718)		(0.247)	
VH x H Size Low	0.0302		0.327*	
	(0.0498)		(0.196)	
VH x H Size High	-0.113		0.212	
	(0.0876)		(0.322)	
VH x Cons Low	-0.00300		0.516**	
	(0.0589)		(0.241)	
VH x Cons High	-0.0126		-0.0511	
	(0.0614)		(0.228)	
V High Ed	-0.112**	-0.0667***	0.00159	0.162*
	(0.0533)	(0.0245)	(0.229)	(0.0943)
Constant	-2.079***	-2.107***	8.524***	8.428***
	(0.0421)	(0.0307)	(0.179)	(0.130)
Observations	19,175	19,175	19,177	19,177
R-squared	0.633	0.633	0.115	0.114

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Nonrandom sampling of descendants

- In a regular longitudinal sample one would follow all or a random subset of descendants
- But in retrospective evaluation, “followed” households are a random subsample of population but not necessarily of descendants.
- Given bari-sampling descendant households from the same 1974 antecedent in the same bari will never both show up in the sample
- Related problem arises if following individuals because coresident members are selected together.

Example I

- Suppose two descendant households. If we sample both average income is $\frac{y_1+y_2}{2}$
- If independent sampling then define $\bar{y} = y_i$ if only y_i is observed and $\bar{y} = \frac{y_1+y_2}{2}$ if both are observed, so $E\bar{y} = \frac{y_1+y_2}{2}$ is unbiased.
- But suppose state $B = \{y_1, y_2\}$ is observed with probability 1/2 and state $C = \{y_2\}$ with probability 1/2.
- Taking expectations across states $E\bar{y} = 1/4y_1 + 3/4y_2$ is biased.
- Two options: (a) throw out y_2 if state B (b) throw out observation if state C
- Both seem wasteful and how does this generalize if $\{y_1, y_2\}$ are observed with probability 1/3?

Alternate Measures I

For a 1974 hhold j with 2 descendants of which at least one is picked:

- option A=hhold 1 picked
- option B=hholds 1 and 2 picked
- option C=hhold 2 picked

Find weights w_a, w_b, w_{b1}, w_{b2}, and w_c such that:

$$\mathbb{E}(y) = p_a w_a y_1 + p_b w_b (w_{b1} y_1 + w_{b2} y_2) + p_c w_c y_2 = \frac{1}{2} y_1 + \frac{1}{2} y_2 = \bar{y} \quad (1)$$

Alternate Measures II

We also want to minimize the effect that the variation in the fraction of households in each sample has on y
We therefore want to minimize:

$$Z = [var(p_a)(w_a^2 y_1^2) + var(p_b)(w_b^2)(w_{b1} y_1 + w_{b2} y_2)^2 \quad (2)$$
$$+var(p_c)(w_c^2 y_2^2) - 2cov(p_a, p_b)(w_a y_1)(w_{b1} y_1 + w_{b2} y_2)$$
$$-2cov(p_a, p_c)(w_a y_1)(w_c y_2) - 2cov(p_b, p_c)(w_{b1} y_1 + w_{b2} y_2)(w_c y_2)]$$

Minimization Problem Continued

Taking derivatives of (1), we get the following two conditions:

$$p_a w_a + p_b w_b w_{b1} = \frac{1}{2} \qquad (3)$$
$$p_b w_b w_{b2} + p_c w_c = \frac{1}{2}$$

The criterion function we then use is the sum of the two second derivatives of equation (2):

$$\min_{w_a, w_b, w_{b1}, w_{b2}, w_c} \frac{d^2 Z}{dy_1^2} + \frac{d^2 Z}{dy_2^2} \qquad (4)$$

To find the best possible weights without knowing income, we minimize equation (4) subject to equations (3)

Solution to Minimization Problem

Solving the above minimization problem we find that the weights which minimize the variance are based on the probability of sampling a 1996 hhold

In our two household example, we get the following weights:

$$w_a = w_b w_{b1} = \frac{\Pr(j)}{2 * \Pr(1)} = w_1$$

$$w_c = w_b w_{b2} = \frac{\Pr(j)}{2 * \Pr(2)} = w_2$$

We can generalize this result to assign a weight to every descendant i of a 1974 household j with N descendants:

$$w_i = \frac{\Pr(j)}{N * \Pr(i)}$$

Formal Definition of Mobility Weights

- We wish to estimate $\Delta \bar{c}_{tN} = \frac{1}{N_{tx}} \sum_{i \in I_{tx}} \frac{1}{|A^{-1}(i)|} \sum_{j \in A^{-1}(i)} \left(c_{jt+1} - c_{it} \right)$
- and $\Delta \hat{c}_{tN} = \frac{1}{N_{tx}} \sum_{i \in I_{tx}} \frac{1}{|A^{-1}(i)|} \sum_{j \in A^{-1}(i)} \left(c_{jt+1} - c_{it} \right) \frac{\mathbb{1}(j \in S_{t+1})}{\mathbb{E}(\mathbb{1}(j \in S_{t+1}))}$
- then $\operatorname{plim}_{N \to \infty} \frac{\Delta \hat{c}_{tN}}{\hat{i}_{tN}} - \Delta \bar{c}_{tN} = 0.$

Applying Formalism

- First apply weights to descendants
- Weighted values are averaged across descendants
- 1974 cross-sectional weights are then applied

Simulation i

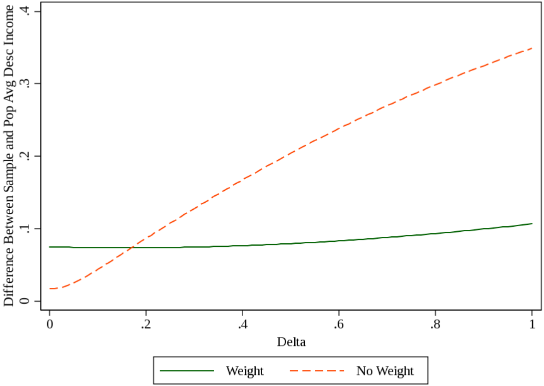
- Simulation exercise with 3 states of the world $\{y_1, y_2, \{y_1, y_2\}\}$ and the probability of each state of the world is correlated with the average y of those households observed.
- Parameter δ measures strength of the correlation

Simulation II

$$\begin{aligned} Pr(1) &= \frac{e^{\delta * y_1}}{e^{\delta * y_1} + e^{\delta * y_2} + e^{\delta * \bar{y}}} \\ Pr(2) &= \frac{e^{\delta * y_2}}{e^{\delta * y_1} + e^{\delta * y_2} + e^{\delta * \bar{y}}} \\ Pr(1\&2) &= \frac{e^{\delta * \bar{y}}}{e^{\delta * y_1} + e^{\delta * y_2} + e^{\delta * \bar{y}}} \end{aligned} \quad (5)$$

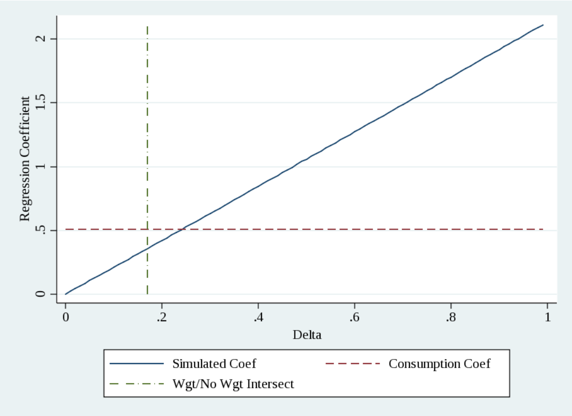
Simulation III

Figure 12: Average absolute difference between sample and actual descendant income means for different levels of correlation



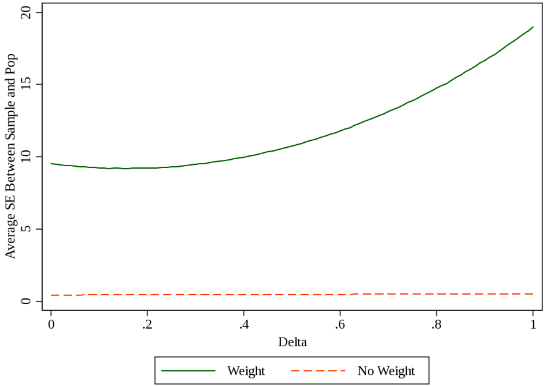
Simulation IV

Figure 13: Coefficients of Probability Regressed on Income for the Simulated Data and the 1996 Census Data



Simulation V

Figure 14: Average squared difference between sample and actual descendant incomes



How well does it work?

Table 7: Household Size Change by 74 Conditions and 74 Village Ed

VARIABLES	(1) Population	(2) Formal	(3) Predicted	(4) 74 Weights	(5) 74/96 Weights	(6) 96 Weights	(7) No Weights
Ed Low	0.0223	0.018	0.071	0.022	0.024	0.003	-0.002
	(0.0344)	(0.155)	(0.347)	(0.140)	(0.140)	(0.088)	(0.086)
Ed High	-0.110	-0.111	0.345	-0.101	-0.098	-0.167	-0.172
	(0.0433)	(0.184)	(0.381)	(0.167)	(0.168)	(0.112)	(0.110)
H Size Low	2.046	2.057	3.439	1.996	2.005	2.058	2.043
	(0.0335)	(0.148)	(0.352)	(0.138)	(0.138)	(0.085)	(0.084)
H Size High	-3.411	-3.414	-4.669	-3.374	-3.384	-3.608	-3.594
	(0.0526)	(0.229)	(0.457)	(0.187)	(0.187)	(0.105)	(0.103)
Cons Low	-0.597	-0.593	-0.727	-0.590	-0.593	-0.683	-0.679
	(0.0391)	(0.173)	(0.347)	(0.157)	(0.157)	(0.091)	(0.089)
Cons High	0.303	0.305	0.485	0.307	0.309	0.314	0.308
	(0.0394)	(0.176)	(0.391)	(0.164)	(0.164)	(0.098)	(0.096)
V High Ed	-0.181	-0.178	-0.209	-0.175	-0.174	-0.156	-0.156
	(0.0320)	(0.142)	(0.298)	(0.130)	(0.130)	(0.081)	(0.080)
Constant	-0.803	-0.811	-1.901	-0.683	-0.701	-0.731	-0.692
	(0.0418)	(0.184)	(0.401)	(0.169)	(0.170)	(0.119)	(0.118)
Observations	19,820	4,690	4,688	4,690	4,690	4,690	4,690
R-squared	0.310	0.261	0.159	0.325	0.328	0.339	0.335

Standard deviation in parentheses for columns 2-7
Standard error in parentheses for column 1

1996 Educational Mobility Graph

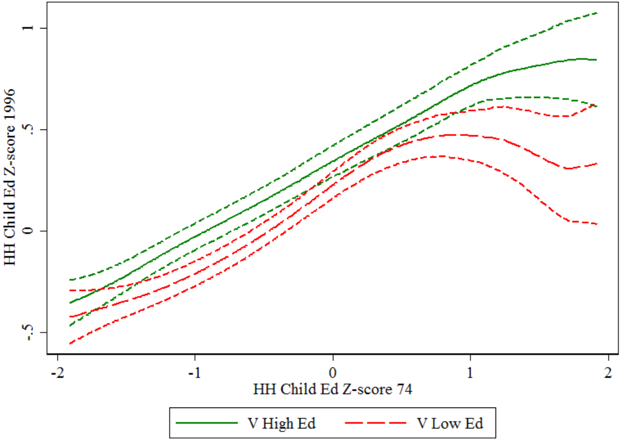


Figure 15: HH Education by 74 HH Education and 74 Village Ed

1996 Educational Mobility Table

Table 8: HH Ed Z-score 96 by 74 Conditions and 74 Village Ed

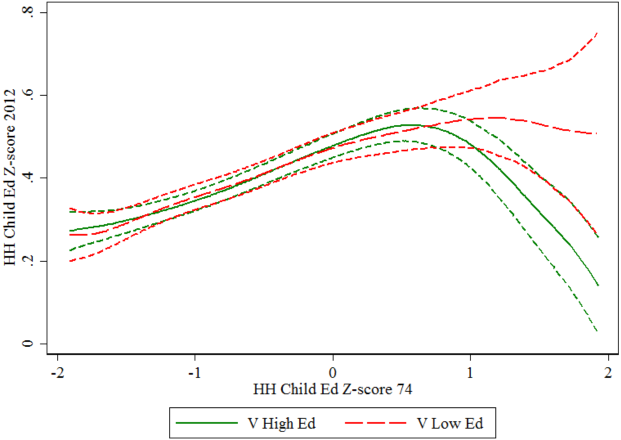
VARIABLES	(1) Formal	(2) Predicted	(3) 74 Weights	(4) 74/96 Weights	(5) 96 Weights	(6) No Weights
Ed Low	-0.226***	-0.270***	-0.244***	-0.234***	-0.244***	-0.265***
	(0.0601)	(0.0592)	(0.0558)	(0.0555)	(0.0314)	(0.0327)
Ed High	0.394***	0.298***	0.395***	0.398***	0.293***	0.288***
	(0.0831)	(0.0833)	(0.0690)	(0.0690)	(0.0386)	(0.0392)
H Size Low	-0.144**	-0.0716	-0.183***	-0.179***	-0.130***	-0.143***
	(0.0615)	(0.0582)	(0.0581)	(0.0580)	(0.0365)	(0.0370)
H Size High	0.0391	0.117*	-0.00203	-0.00748	0.0515	0.0596*
	(0.0701)	(0.0646)	(0.0615)	(0.0611)	(0.0329)	(0.0342)
Cons Low	-0.175***	-0.149***	-0.179***	-0.176***	-0.153***	-0.163***
	(0.0602)	(0.0527)	(0.0582)	(0.0579)	(0.0337)	(0.0350)
Cons High	0.192***	0.224***	0.166**	0.162**	0.170***	0.178***
	(0.0712)	(0.0663)	(0.0664)	(0.0662)	(0.0395)	(0.0408)
V High Ed	0.144***	0.157***	0.130**	0.125**	0.104***	0.114***
	(0.0546)	(0.0522)	(0.0504)	(0.0503)	(0.0300)	(0.0307)
Constant	-0.0428	-0.130*	0.0164	0.0136	-0.00424	0.00312
	(0.0700)	(0.0685)	(0.0667)	(0.0665)	(0.0435)	(0.0445)
Observations	3,469	3,467	3,469	3,469	3,469	3,469
R-squared	0.064	0.028	0.128	0.127	0.103	0.108

Robust standard errors in parentheses

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

2012 Educational Mobility Graph

Figure 16: HH Education by 74 HH Education and 74 Village Ed



2012 Educational Mobility Table

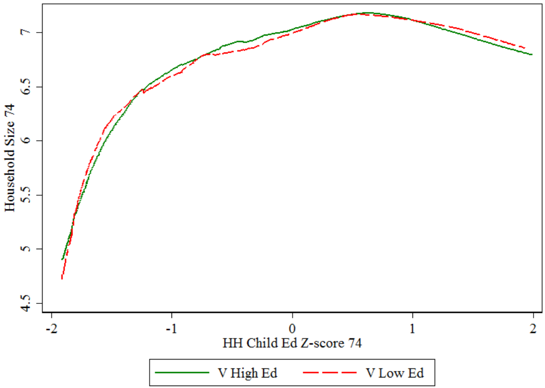
Table 9: HH Ed Z-score 2012 by 74 Conditions and 74 Village Ed

VARIABLES	(1) Formal	(2) 74 Weights	(3) 74/96 Weights	(4) 96 Weights	(5) No Weights
Ed Low	-0.0845***	-0.118**	-0.111**	-0.0981***	-0.104***
	(0.0299)	(0.0459)	(0.0464)	(0.0292)	(0.0285)
Ed High	0.100***	0.201***	0.197***	0.198***	0.198***
	(0.0365)	(0.0540)	(0.0544)	(0.0347)	(0.0338)
H Size Low	-0.0434	-0.0692	-0.0614	-0.0213	-0.0259
	(0.0320)	(0.0512)	(0.0521)	(0.0310)	(0.0305)
H Size High	0.000106	-0.00181	-0.00830	0.0501	0.0532
	(0.0337)	(0.0520)	(0.0528)	(0.0385)	(0.0381)
Cons Low	-0.0410	-0.0674	-0.0684	-0.0561*	-0.0566*
	(0.0316)	(0.0464)	(0.0461)	(0.0323)	(0.0317)
Cons High	0.0383	0.105**	0.101*	0.0265	0.0355
	(0.0346)	(0.0529)	(0.0552)	(0.0342)	(0.0336)
V High Ed	-0.00538	-0.0249	-0.0167	0.0289	0.0264
	(0.0272)	(0.0428)	(0.0433)	(0.0282)	(0.0277)
Constant	0.407***	0.733***	0.733***	0.693***	0.689***
	(0.0385)	(0.0546)	(0.0552)	(0.0376)	(0.0371)
Observations	3,646	3,646	3,646	3,646	3,646
R-squared	0.023	0.042	0.038	0.027	0.030

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

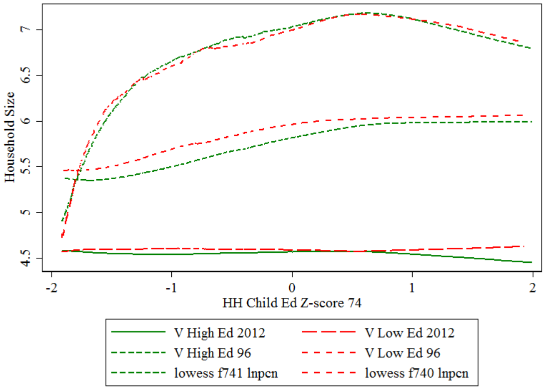
Household Size 74

Figure 17: 74 Household Size by 74 HH Ed Z-score and 74 Village Ed



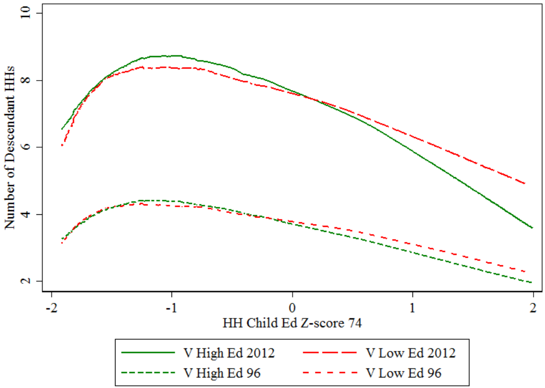
Household Size 1996/2012

Figure 18: Household Size by 74 HH Ed Z-score and 74 Village Ed



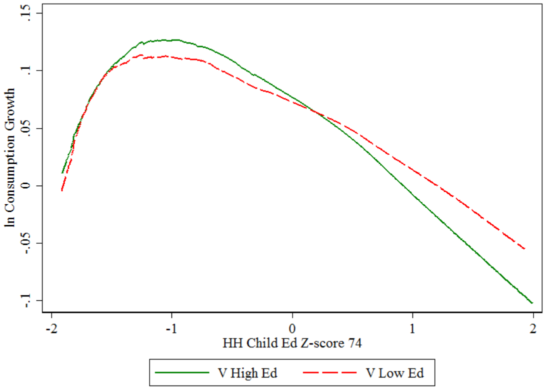
Descendant HHs

Figure 19: Descendant HHs by 74 HH Ed Z-score and 74 Village Ed



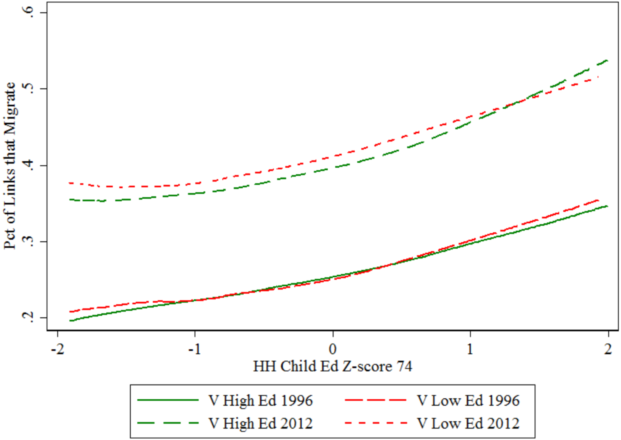
Consumption Growth

Figure 20: Consumption Growth by 74 HH Ed Z-score and 74 Village Ed



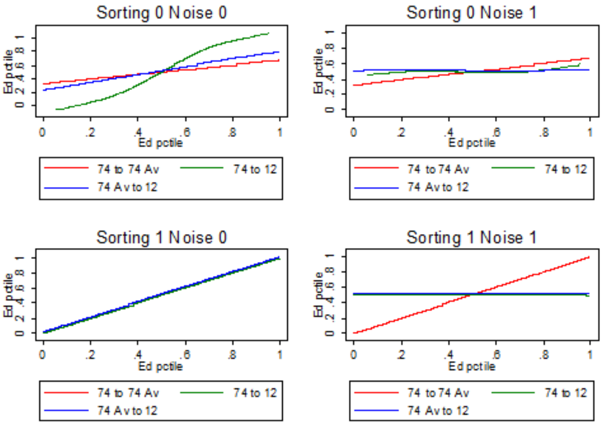
Migration

Figure 21: Percent of Links that Migrate by 74 Village Ed



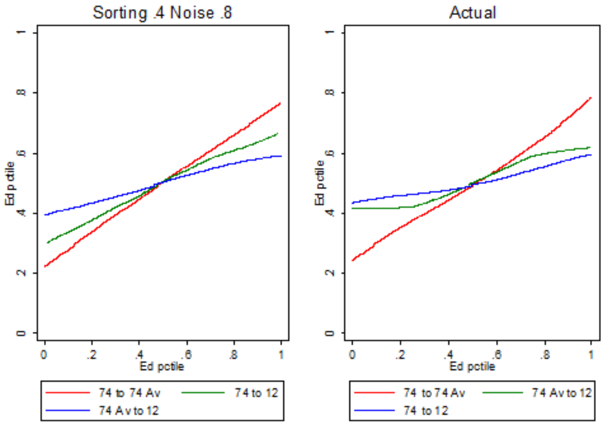
Decomposition I

Figure 22: Decomposition of Mobility Simulation



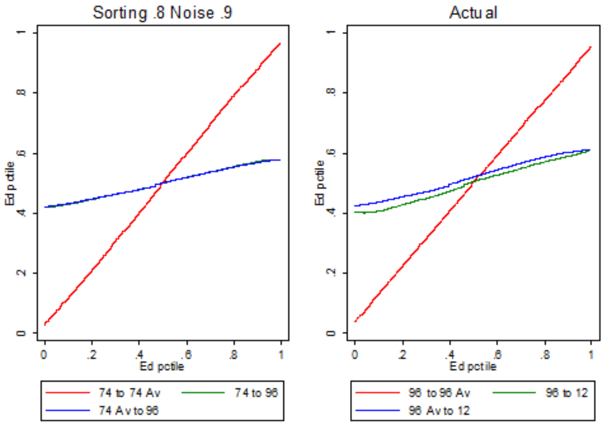
Decomposition II

Figure 23: Decomposition of Mobility 74-12



Decomposition III

Figure 24: Decomposition of Mobility 96-12



Total human capital

Table 10: Effects of 1996 characteristics on 2012 education by weighting scheme

VARIABLES	(1) No weights	(2) 2012 weights	(3) 2012 and 1996 weights
Ed x Fam Size 96	0.119	0.0254	0.173***
	(0.0803)	(0.114)	(0.0482)
Ed 96	0.170***	0.206***	0.0217
	(0.0448)	(0.0633)	(0.0158)
Fam Size 96	-0.0934*	-0.0437	-0.103***
	(0.0479)	(0.0681)	(0.0199)
Constant	0.428***	0.413***	0.499***
	(0.0269)	(0.0386)	(0.0115)
Observations	2,107	2,107	2,107
R-squared	0.047	0.044	0.019

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Conclusions

- Demographic change through family size and household division has importantly influenced economic mobility in South Asia over the last four decades
- Measurement of economic mobility is affected by antecedent as well as descendant processes.
- Endogenity of residence does not make sampling weights endogenous
- Sampling weights require data that is not generally available except in a DSS.
- Matching across economic strata plays some role in increasing mobility
- Basic primary health care leads to increased mobility but not necessarily reduced inequality.
- Some evidence that stock of human capital contributes to economic mobility–which reduces consequences of quality/quantity tradeoff for economic mobility.