BIGDATA: Collaborative Research: IA: Population Reproduction of Poverty at Birth from Surveys, Censuses, and Birth Registrations

Michael S. Rendall, Principal Investigator, UMD; Mark Handcock, Co-Principal Investigator, UCLA;
Margaret Weden, Co-Principal Investigator, RAND

Goals of the Study

Domain Goal: collaboration between statisticians and social demographers to generate new forms of causal inference about processes of social mobility and social reproduction across generations in a population

BIGDATA Goals: develop, evaluate, and disseminate methods of estimation and simulation that combine a maximum amount of large- and medium-scale survey data with population-level census and administrative data on the same population

BIGDATA Challenge I: Complexity of Population Models of Social Reproduction

Need to include both mother's and father's characteristics in the modeling of intergenerational *inheritance* and of *assortative mating* on these characteristics.

Patterns of assortative mating and reproduction by various combinations of race, ethnicity, country of birth, and educational attainment in the U.S. has expanded greatly over recent decades

Simple model structures and single data sources are unable to model the increasing complexity of American society and its socioeconomic disparities across racial, ethnic, immigrant and gender divisions

BIGDATA Challenge II: Need for Model Parameter Estimation and Model Validation that Combines Data Across Multiple Sources

Methods for combining data across multiple sources of very large-scale, medium-scale, and smaller-scale population-representative data* sources are essential for estimating the parameters of a sufficiently realistic model of intergenerational social reproduction

Combining survey data sources with different sampling designs, however, raises additional problems for which solutions have only begun to be developed**

* Ridder, G., and R.A. Moffitt (2007) "The econometrics of data combination," pp.5469-5547 in J.J. Heckman and E.E. Leamer (Eds.) Handbook of Econometrics Vol.6b. Amersterdam: North Holland.

**Rendall, M. S., Handcock, M. S., and Jonsson, S. H. (2009) "Bayesian Estimation of Hispanic Fertility Hazards from Survey and Population Data" Demography 46(1):65-83.

BIGDATA Challenge II: The Need for Simulation to Model Outcomes Not Observable in a Single Data Source

There is no source of U.S.-population-representative data on whether a child was born into poverty as well as whether both the individuals' mother *and* father were born into poverty: that is, the poverty-at-birth circumstances of mother-father-child triads

It is possible to observe a random sample of mother-child dyads and of father-child dyads and then simulate mother-father-child triads if one understands consequences of poverty at birth on assortative mating and childbearing patterns

We structure a simulation model to capture consequences of poverty-status at birth on educational attainment and thereby assortative mating and childbearing

Table 1. Population Model

Age a=0	Age a=16	a=17	29 33	a=45
"Parent" generation				
Ethnicity $e_{1(k,j)}$ $e_{1(f,j)}$		$e_{1(k,j)}$ $e_{1(f,j)}$		$e_{1(k,j)}$ $e_{1(f,j)}$
Poverty $o_{1(k,j)}$ $o_{1(f,j)}$				
	Schooling years $c_{1(k,j)}$ $c_{1(f,j)}$	$c_{1(k,j)}$ $c_{1(f,j)}$		$c_{1(k,j)}$ $c_{1(f,j)}$
	Partnership status $p_{1(k,j)}$=0 $p_{1(f,j)}$=0	$p_{1(k,j)}$ $p_{1(f,j)}$		$p_{1(k,j)}$ $p_{1(f,j)}$
	Partner characteristics $q_{1(k,j)}$=0 $q_{1(f,j)}$=0	$q_{1(k,j)}$={$w_{1q(k,j)}$,$a_{1q(k,j)}$,$e_{1q(k,j)}$} $q_{1(f,j)}$= {$w_{1q(f,j)}$,$a_{1q(f,j)}$,$e_{1q(f,j)}$}		$q_{1(k,j)}$={$w_{1q(k,j)}$,$a_{1q(k,j)}$,$e_{1q(k,j)}$} $q_{1(f,j)}$={$w_{1q(f,j)}$,$a_{1q(f,j)}$,$e_{1q(f,j)}$}
	Parity $r_{1(k,j)}$=0 $r_{1(f,j)}$=0	$r_{1(k,j)}$=1 $r_{1(f,j)}$=0		$r_{1(k,j)}$=1 $r_{1(f,j)}$=2
"Child" generation				
		{$a_{2y(k,j,v)}$=0;$o_{2y(k,j,v)}$,$s_{2y(k,j,v)}$,$e_{2y(k,j,v)}$}		{$a_{2y(k,j,v)}$=16; $s_{2y(k,j,v)}$,$c_{2y(k,j,v)}$,$e_{2y(k,j,v)}$}
			{$a_{2z(f,j,v)}$=0; $s_{2z(f,j,v)}$,$c_{2z(f,j,v)}$,$e_{2z(f,j,v)}$}	
				{$a_{2z(f,j,v)}$=4; $s_{2z(f,j,v)}$,$c_{2z(f,j,v)}$,$e_{2z(f,j,v)}$}

Notes: o = poverty; a = age; e = race/ethnicity/nativity; c = schooling; r = parity; p = partnership status (married, cohabiting, unpartnered); q = partner characteristics; s = sex (gender) of child. Subscripts *(k,i) (i,j)* index the mother *i=1,2,…,I* born to generation-0 "grandparent unit" *k* and father *j=1,2,…,J* of generation 1 born to generation-0 "grandparent unit" *l*. Subscripts *(x,v), (y,v),* and *(z,v)* index their children *v=1,2,…,V* of generation 2

Table 2. Data Sources to Estimate Poverty Status at Birth

Data Source	Period	Number of newborns ('000)	Poverty status at own birth?	Poverty status at 0,1,2 parents' births?	Educ., race, ethnicity, nativity of 0,1,2 parents
Birth registration file (NCHS)	1970-2015	175,000	no	0	1 or 2
Census 5% Public Use Microdata Sample (PUMS)	1970, 1980, 1990, 2000	770	yes	0	1 or 2
American Community Survey (ACS)	2005-2015	460	yes	0	1 or 2
Current Population Survey (CPS)	1970-2015	75	yes	0	1 or 2
Survey of Income and Program Participation (SIPP)	1984-2015	43	yes	0	1 or 2
Panel Study of Income Dynamics (PSID)	1968-2015	7.5	yes	1	1 or 2
National Longitudinal Survey of Youth 1979 (NLSY79)	1979-2015	5.3	yes	1	1 or 2
National Longitudinal Survey of Youth 1997 (NLSY97)	1997-2015	1.6	yes	0	2

Data Sources for Simulation Model Components

1) **Poverty at birth**
- see Table 2 for data sources: **note** that mother and father's poverty at births are never *both* observed in relationship to their children's poverty at birth, and therefore need the mother-father-child triad emerges through simulating steps 2) to 4) below:

2) Male and Female **Educational Attainment** given own poverty status at birth:
- PSID, NLSY79, SIPP

3) Assortative **Union Formation and Dissolution** given educational attainment:
- ACS, CPS, SIPP (also: PSID, NLSY79, NLSY97)

4) Couple and Unpartnered Women's **Fertility** given education of both woman and any coresident partner
- NCHS, ACS, CPS, SIPP, PSID, and NLSY79

First subproject: Assortative Union Formation in a "Two-Sex" Model

The single most challenging model component is to **generate partner characteristics q** of Table 1.

Partner characteristics depend on the "two-sex" process of forming marital or cohabiting unions, or unions that produce children outside of a coresidential union.

"Two sex" refers to taking into account both male and female preferences and the distribution of available opposite-sex women and men

We use the model of Logan, Hoff and Newton (2008)*, implemented in the software package RPM (Admiraal and Handcock 2008)**

* Logan J.A., Hoff, P.D., and Newton M.A. (2008) "Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion" Journal of the American Statistical Association 103(482):559-569.

**Admiraal R., and Handcock M.S. (2008) RPM: Estimated revealed preferences from observed matchings. Version 1.0, University of Washington, Seattle.

Utilities V, U of men i and women j depend on preferences α, β and status characteristics; we generate **dyad-specific random utilities** for each female-male dyad in the population

$$V_{i,j} = \beta^T W_{i,j} + \gamma_{i,j}$$

$$U_{j,i} = \alpha^T W_{j,i} + \varepsilon_{j,i}$$

We then find **a stable set of unions** that satisfy for each female-male dyad in the population, the set of unions between woman i, mp(i), and man j, fp(j), that satisfy for all **opportunity sets** O(i) and O(j) of utility-maximizing unions for individuals of the other sex:

$$V_{i,mp(i)} \geq V_{i,j} \forall j \in O(i) \cup \varnothing$$

$$U_{j,fp(j)} \geq U_{j,i} \forall i \in O(j) \cup \varnothing$$

Preliminary Data for Model Estimation and Evaluation, SIPP 2001, 2004, 2008 Panels, **married** men ages 18-59, Own race/ethnicity by Wife's race/ethnicity

Wife:	Hispanic	Black	White	Asian	N (Unweighted)
Self: Hispanic	85.9	0.4	13.2	0.6	4,414
Black	2.5	87.8	8.5	1.2	2,870
White	2.4	0.4	96.1	1.2	28,273
Asian	1.3	0.5	6.7	91.4	1,631

Preliminary Data for Model Estimation and Evaluation, SIPP 2001, 2004, 2008 Panels, **cohabiting** men ages 18-59

Wife:	Hispanic	Black	White	Asian	N (Unweighted)
Self: Hispanic	73.6	1.5	22.8	2.1	564
Black	4.5	78.7	16.0	0.8	479
White	2.9	0.6	94.9	1.7	2,762
Asian	9.9	0.0	34.8	55.3	55

This research is supported by grant NSF IIS-1546259. We thank Lucia Lykke for research assistance