

UCLA

UCLA Electronic Theses and Dissertations

Title

Prospective Versus Retrospective Approaches to the Study of Intergenerational Social Mobility

Permalink

<https://escholarship.org/uc/item/4z26g9zk>

Author

Song, Xi

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

**Prospective Versus Retrospective Approaches to
the Study of Intergenerational Social Mobility**

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Xi Song

2013

© Copyright by
Xi Song
2013

ABSTRACT OF THE THESIS

Prospective Versus Retrospective Approaches to the Study of Intergenerational Social Mobility

by

Xi Song

Master of Science in Statistics

University of California, Los Angeles, 2013

Professor Yingnian Wu, Chair

Most intergenerational social mobility studies are based upon retrospective data, in which samples of individuals report socioeconomic information about their parents. Such an approach suffers from a retrospective reporting bias, because early generations can be recalled only if they have offspring, and those with more offspring are more likely to be recalled. This thesis discusses the conceptual and practical advantages of using an alternative, prospective approach, which examines intergenerational mobility by following a sample of respondents and their progeny. This prospective approach is especially useful for multigenerational mobility analysis that is rising in the field of mobility studies in recent years. We also propose an adjustment method that corrects the retrospective reporting bias in retrospective data, and thus permits them to be used in prospective intergenerational analyses that incorporate both demographic and mobility effects. We illustrate the adjustment method using both two-generation and multigenerational models based on simulated data and empirical data from the Panel Study of Income Dynamics. The results suggest that our adjustment retrospective method removes more than 95% of the bias in the prospective analysis based on retrospective data.

The thesis of Xi Song is approved.

Mark Stephen Handcock

Robert Denis Mare

Yingnian Wu, Committee Chair

University of California, Los Angeles

2013

*To Peter Zhengpeng Zhang,
for the many ways, in which he enriches my life.*

TABLE OF CONTENTS

1	Introduction	1
2	Retrospective and Prospective Mobility Table	5
3	The Demography of Social Mobility	9
3.1	Beyond the Mobility Table: the Intervening Effect of Fertility	9
3.2	A Joint Mobility and Demography Model	11
3.3	Estimation of the Model Parameters	13
3.4	An Adjustment Method for Retrospective Data	17
4	Simulated Example	20
4.1	Data Generating Process	20
4.2	Simulation Results	23
5	Empirical Example	26
5.1	Data	26
5.2	Empirical Results	28
6	Conclusion	37
7	Appendix: Simulation Details	40
	References	45

LIST OF TABLES

4.1	Table 1. Two-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models based on Monte Carlo Simulation	24
4.2	Table 2. Three-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models based on Monte Carlo Simulation	25
5.1	Table 3. Three-Generation Fertility and Mobility for <i>Men</i> Born Between 1930 and 1950 and Their Offspring, <i>PSID</i>	33
5.2	Table 3.[cont.] Three-Generation Fertility and Mobility for <i>Men</i> Born Between 1930 and 1950 and Their Offspring	34
5.3	Table 4 Three-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models, <i>PSID</i>	35
5.4	Table 4.[cont.] Three-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models	36

ACKNOWLEDGMENTS

Getting through graduate school is not an easy process. I am very lucky to have been surrounded by supportive advisors and friends, who have made this process less dull and more pleasant. I owe many thanks to my advisor in statistics, Ying Nian Wu, for his influence on my intellectual development. It was his encouragement in my first year of graduate school that made me aware of the potential benefits of pursuing a master's degree in statistics. Over the past two years, when I sought help with statistical questions, he was very generous with his time, and always provided me with satisfactory answers.

I would also like to thank Mark S. Handcock for playing an important role in preparing me for my career in academia. Mark is gifted in his ability to solve sociological puzzles by applying rigorous statistical methods. Our conversations have influenced my thinking and the direction of my work.

Special thanks also goes to my advisor in sociology, Robert Mare. Rob has mastered the art of mentoring. He always encourages me to try new things, and is more optimistic than I am about my potential to successfully pursue ambitious and difficult research projects. It has been a joy to exchange thoughts with him, whether it is about big research ideas or trivial matters of everyday life.

My academic life would have been much less meaningful without the friendship provided by a large network of friends in the UCLA Statistics and the Sociology departments: Albert Wong, Benjamin Jarvis, Dwight Davis, Hoshun Yang, Pengcheng Li, Sung Park, William Rosales, Yangyi He, Yao Shi, Yi Yi, Yang Zhou, Yuxi Zhou, Yuyan Hu, and Zhongnan Fang. They have provided me with a constant source of support during my academic journey and welcome distractions from the grinding work within academia.

Finally, I would like to thank my family for everything they have given me. Without their enthusiasm, praise, and support, I would not have gotten this far.

This thesis is a modified version of

Song, Xi and Robert D. Mare. 2013. “Retrospective Versus Prospective Approaches to the Study of Intergenerational Social Mobility” Working Paper.

This research was supported by the National Science Foundation (SES-1260456). The authors benefited from facilities and resources provided by the California Center for Population Research at UCLA (CCPR), which receives core support (R24-HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

CHAPTER 1

Introduction

This thesis illustrates the advantages of using a prospective approach in two-generation and three-generation social mobility studies, and provides an adjustment method for conducting prospective analyses when only retrospective data are available. Traditional social mobility studies typically examine the parent-offspring association in socioeconomic status by asking a group of individuals about the characteristics of their parents retrospectively. Such an approach has been the basis of a large and successful literature on intergenerational social mobility (e.g., Beller 2009; Blau and Duncan 1967; Breen 2004; Erikson and Goldthorpe 1992; Featherman and Hauser 1978; Hauser et al. 1975; Hout 1983, 1988). Modern longitudinal studies, however, which follow a sample of adults through the birth and growth to adulthood of their offspring and descendants, afford the possibility of studying social mobility prospectively, especially over multiple generations (Mare 2011).

In Duncan's (1966) classic article on methodology of social mobility studies, he concluded: "Although data in the typical mobility study are collected retrospectively (by questioning the respondent about the past), this is only a convenience in data collection. While it introduces problems of data reliability and validity, it does not commit the analyst to a backward-looking conceptual framework." Whereas this conclusion is, strictly speaking, correct, it leaves unsaid much about the relationship between these two approaches. Important statistical and conceptual issues arise when one considers the possibilities and limitations of these two

approaches.

From a statistical point of view, mobility estimates, such as odds ratios, are not necessarily the same in the two approaches. Retrospective reporting bias may distort data based on retrospective surveys because parents' generation can be recalled only if they have offspring, or more strictly speaking, if they have surviving offspring who could be sampled by the surveys (Mare 2011). Moreover, parents with more offspring are over-represented in exact proportion to their fertility in the retrospective data (Glass 1954; Duncan 1966; Allan and Bytheway 1973). Thus, the respondents are representative of the offspring generation, but their parents are not representative of people who have ever lived in the parent generation. In contrast, for prospective data, in which the survey asks individuals to provide information about themselves and their offspring, the respondents are representative of people in the parents' generation. However, their offspring may not be, either because the respondents are asked about only a selected child rather than all children, or because the respondents have not finished having children. Prospective data provide representative samples for both the parent and the offspring generation only when parents provide information about all their children. Retrospective and prospective estimates agree when family size is unrelated to the intergenerational transmission of socioeconomic characteristics. One of the contributions of this thesis is to provide a method of reconciling mobility estimates from the two approaches.

Although prospective mobility data, like retrospective data, may be used for the statistical analysis of associations between the socioeconomic statuses of parents and offspring, prospective data also afford a wider range of analytic possibilities. For example, the question of intergenerational inequality is not only about intergenerational transmission of status from parents to offspring, but about how intergenerational effects come about—that is, does improving the socioeconomic attainment of an individual increase the probability of eventually having a child

who will also have higher socioeconomic attainment? From a prospective perspective, demography and social inequality are closely intertwined, as the transmission of intergenerational inequality involves not only the inequalities among those who have offspring, but also the inequality between those who have offspring and those who do not. The prospective approach calls attention to the interdependence of demographic processes and the associations of socioeconomic characteristics between members of successive generations. The traditional mobility table approach focuses on the associations between parents' and their offspring's characteristics conditional on the existence of the offspring (Mare and Maralani 2006). But a more complete understanding should account for the interdependence of mobility and differentials in timing and levels of marriage, fertility and mortality. Following the prospective logic of intergenerational inequality, a handful of studies have developed joint demographic and mobility models to investigate research questions such as effects of intergenerational reproduction of education and occupation (Maralani 2013; Mare and Maralani 2006; Matras 1961, 1967), causal aspects of intergenerational effects (Lawrence and Breen 2012), and changes in population composition by generation (Preston 1974; Preston and Campbell 1993).

Mare (2011) calls for current social and demographic research to go beyond traditional two-generation paradigm and understand the roles of grandparents, nonresident kin and remote ancestors in influencing the social mobility of families under different circumstances. Longitudinal surveys that follow families over three or more generations permit mobility analyses to incorporate multigenerational effects beyond the scope of nuclear families. Examples of multigenerational, longitudinal data include the Panel Study of Income Dynamics (PSID), and the Wisconsin Longitudinal Study (WLS). An important impediment to multigenerational studies of mobility is that few longitudinal surveys include intergenerational information beyond two generations, as this requires that surveys follow families more than 50 years. In the absence of prospective data spanning three

or more generations, retrospective, multigenerational data such as those collected by Treiman and colleagues provide potentially useful data sources for the study of intergenerational social mobility (Szelenyi and Treiman 1994; Treiman, Moeno and Schlemmer 1996; Treiman and Walder 1998).

In this thesis, we show how to extend the analysis of retrospective social mobility data to address the broader set of analytic issues that can be studied when prospective data are available. We illustrate our methods using both simulated data and empirical data from the Panel Study of Income Dynamics. Our results suggest that the adjustment method removes more than 95% of the bias in the retrospective joint demographic and mobility effect. The adjustment methods proposed in this thesis are well-suited not only for two-generation social mobility but also multigenerational mobility analysis that is rising in the field of mobility studies in recent years.

We divide the remainder of the thesis into five sections. The first section discusses the statistical relationship between mobility tables constructed from retrospective and prospective data. In the second section, we propose a prospective approach to study social mobility: a joint model of social mobility and demographic reproduction. We then describe our adjustment method, explaining how to conduct the prospective analyses when only retrospective data are available. In the fourth section, we illustrate the method using simulated two-generation and multigenerational data sets. We then use the Panel Study of Income Dynamics as an example to assess the effectiveness of the adjustment method for real data. The conclusion section reviews the capabilities and limitations of our method.

CHAPTER 2

Retrospective and Prospective Mobility Table

Most mobility studies use tables based on retrospective surveys that ask respondents about their own and their parents' socioeconomic characteristics. The mobility table shows the parent-offspring association based on a tabulation of a socioeconomic characteristic of an individual and his or her parents (Hout 1983). Some widely used cross-sectional, retrospective mobility data include the Occupational Changes in a Generation surveys (Blau and Duncan 1967; Featherman and Hauser 1978), the General Social Survey (GSS) (Hout 1988; Beller 2009) and the Comparative Analysis of Social Mobility in Industrial Nations Project (e.g., Erikson, Goldthorpe and Portocarero 1979; Erikson and Goldthorpe 1992). Based on the retrospective data, the mobility table is constructed from the perspective of adult individuals, who are a representative sample of their own generation. These individuals report on their parents, but because this design overrepresents parents who have more offspring and fails to include any members of the parental generation who are childless, the resulting data do not provide a representative sample of the parental generation.

As an alternative, it is possible to construct the mobility table based on prospective data. There are two kinds of prospective mobility data. The first type of data asks a group of respondents about their own socioeconomic status and that of a selected child. For example, the Wisconsin Longitudinal Study (WLS) adopts such a design for collecting intergenerational occupational information. In the sample, the respondents are representative of their generation

regardless of their number of sons, whereas sons of high fertility respondents are underrepresented in their generation. The second type of data asks respondents about all their offspring, or repeatedly follows up with respondents and all their offspring. Examples of this prospective design include the collection of the educational information for the Wisconsin Longitudinal Study, the Panel Study of Income Dynamics, and newer parallel panel studies in other countries such as the German Socioeconomic Panel (started in 1984), the British Household Panel Survey (1991), the Canadian Survey of Labor and Income Dynamics (1993), the Korean Labor and Income Panel Study (1998), the Swiss Household Panel (1999), the Australian Household, Income and Labor Dynamics (2001), and the Chinese Family Panel Studies (2010). These prospective data include a parent sample and an offspring sample, both of which are representative of their own generations.

If we use the first type of prospective data to construct mobility tables, the offspring are not representative of their generation, because offspring from large families are underrepresented. The algebraic relationship between the first type of prospective data and retrospective data is illustrated below. For a traditional retrospective mobility table, the row F represents fathers' occupations and the column S represents sons occupations, with I and J categories respectively (typically, $I = J$). For example, if we use row i' and column j' as reference (i.e., fathers in occupation i' and respondents in occupation j'), the odds ratios (OR_r) are defined as $\frac{n_{ij}n_{i'j'}}{n_{ij'}n_{i'j}}$, which is the relative odds of being in occupation j rather than j' , given one's father is in occupation i , to the odds of being in occupation j rather than j' , given one's father is in occupation i' . For the first type of prospective data, we can estimate retrospective odds ratios by adjusting for the fertility of fathers. If the k^{th} father who is in occupation i and has $r_{ij}^{(k)}$ sons in occupation j and \bar{r}_{ij} refers to the average number of sons in occupation j for fathers in occupations i , that is $\bar{r}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} r_{ij}^{(k)}$, the relationship between the retrospective and prospective odds ratios is:

$$OR_r = \frac{\sum_{k=1}^{n_{ij}} r_{ij}^{(k)} \sum_{k=1}^{n_{i'j'}} r_{i'j'}^{(k)}}{\sum_{k=1}^{n_{ij'}} r_{ij'}^{(k)} \sum_{k=1}^{n_{i'j}} r_{i'j}^{(k)}} = \frac{(n_{ij}\bar{r}_{ij})(n_{i'j'}\bar{r}_{i'j'})}{(n_{ij'}\bar{r}_{ij'})(n_{i'j}\bar{r}_{i'j})} = OR_p \cdot \frac{\bar{r}_{ij}\bar{r}_{i'j'}}{\bar{r}_{ij'}\bar{r}_{i'j}} \quad (2.1)$$

Equation 2.1 shows that to construct retrospective odds ratio based on prospective data, we can simply weight the prospective data so that fathers with more sons are over-represented in exact proportion to their fertility. More specifically, we weight each father by his fertility, or equivalently weight each son by one plus his number of siblings.

When the weighting ratio $\frac{\bar{r}_{ij}\bar{r}_{i'j'}}{\bar{r}_{ij'}\bar{r}_{i'j}} = 1$, then the retrospective and prospective odds yield the same conclusions about mobility. If the father's fertility is only associated with the father's occupation or the son's occupation, then the products in the denominator and the numerator cancel, meaning that the prospective and the retrospective results are the same. Only when there is three-way interaction among father's occupation, son's occupation and fathers fertility (or son's number of siblings) do the retrospective and prospective odds ratios disagree. For example, when the fertility of the fathers of immobile sons is greater than the fertility of the fathers of mobile sons (i.e., $\frac{\bar{r}_{ij}\bar{r}_{i'j'}}{\bar{r}_{ij'}\bar{r}_{i'j}} > 1$), the estimate of OR_p is smaller than OR_r , and vice versa. While there is abundant evidence about the two-way interactions between father's socioeconomic status and fertility (e.g., Blake 1981), as well as between father's fertility and sons status (e.g., Mare and Chen 1986), few studies examine the three-way interaction among father's occupation, son's occupation and father's fertility.

For the second type of prospective data, we can construct mobility tables based on the offspring sample, and link offspring with their parents in the parent sample. If we use data from all sons, the resulting mobility table has the same structure as a traditional mobility table constructed from retrospective data. In this case, the prospective and retrospective odds ratios are equal and no weighting adjustment

(as shown in equation 2.1) is required.

CHAPTER 3

The Demography of Social Mobility

3.1 Beyond the Mobility Table: the Intervening Effect of Fertility

As we show above, prospective data can be used to construct traditional mobility tables after a proper weighting of the data. But this is not the main reason for using prospective mobility data. The real advantage of prospective data is they are suitable for analyses beyond a simple intergenerational correlation of socioeconomic status. Specifically they permit analyses that incorporate demographic pathways into our understanding of intergenerational transmission of inequality. Most mobility research focuses on the associations between parents and their offspring's characteristics conditional on the existence of the offspring. However, a more complete understanding of intergenerational influence should treat as problematic the degree to which offspring will come into existence as well as the effects of parents on their children (Mare and Maralani 2006). The mobility table itself is inadequate for showing how a socioeconomic distribution persists or changes because the mobility process is interdependent with the differentials in timing and levels of fertility, mortality, and migration (Duncan 1966). In the discussion below, we focus on the role of fertility in a one-sex model for men. More comprehensive versions of these models take account of other demographic processes, including marriage, divorce, remarriage, parental and child survival, adoption, migration, and timing of these events, both for women and for two-sex populations

(For related discussions, see Mare 1997; Mare and Schwartz 2006; Maralani and Mare 2005).

The intervening role of fertility in the mobility process operates at both the individual and population levels. At the individual level, a man's socioeconomic characteristics are closely related to his fertility decisions, such as whether, with whom, and when he will have children and how many children he will have. These decisions, in turn, influence the family context in which a child grows up, affecting factors such as sibship size, parental investment in childrearing, and the parent-child relationship, which are linked to a child's chances for upward social mobility (Blau and Duncan 1967; Sobel 1985). Most research examines either the effect of socioeconomic characteristics on fertility or the effect of number of siblings on socioeconomic outcomes, instead of evaluating them jointly for two or more generations. At the population level, changes in socioeconomic distributions by generation are affected by the association between parents' and offspring's socioeconomic characteristics, weighted by the total number of offspring from each socioeconomic group. Families that produce more offspring more successfully replace themselves in the population, and their offspring will be overrepresented in the next generation, whereas families with no offspring cannot pass on their advantages or disadvantages to the next generation. Thus, even if the intergenerational mobility pattern is fixed over time, changes in the relationship between socioeconomic status and fertility could result in changes in the social makeup of the population. Given that changes in fertility and family structure have not occurred uniformly across socioeconomic groups, differential demographic behaviors may magnify or diminish the macro-level implications of the changing mobility patterns of families.

3.2 A Joint Mobility and Demography Model

Based on population projection models, a handful of studies extend the standard mobility table to estimate the joint effect of a man's socioeconomic characteristics on his fertility and the characteristics of whatever sons he has (Bartholomew 1982; Mare 1997; Matras 1961, 1967; Musick and Mare 2004; Preston 1974; Preston and Campbell 1993). Our discussion below relies on the framework of the joint demographic and mobility model in Mare and Maralani (2006). The model specifies the effect of a man's socioeconomic position in one generation (compared to other positions of men in that generation) on the expected number of sons in a given socioeconomic position in the next generation. This model shows how social mobility and fertility contribute to transformations of the socioeconomic distribution of a population. The model can be written as

$$S_{Y_2|Y_1} = F_{Y_1} \cdot f_{Y_1} \cdot p_{Y_2|Y_1} \quad (3.1)$$

where $S_{Y_2|Y_1}$ denotes the number of men in the offspring generation who are in position Y_2 and have fathers in position Y_1 ; F_{Y_1} denotes the number of men in the paternal generation who are in position Y_1 ; f_{Y_1} denotes the expected number of sons born to a man in position Y_1 and who survive to adulthood; $p_{Y_2|Y_1}$ denotes the probability that a son born to a man in position Y_1 will enter position Y_2 . This elementary form only includes fertility mobility effects, but various forms of this model also allow the presence of marriage, mortality and age effects (Lam 1986; Mare and Maralani 2006; Maralani and Mare 2005; Matras 1961, 1967; Preston 1974). To compare this approach to the standard mobility table, we only consider the father-son association, although related models for the female and two-sex populations have also been developed (Preston and Campbell 1993; Mare and Maralani 2006; Maralani 2013).

Based on the model, we define the joint demographic and mobility effect as

$$E(f_{Y_1} p_{Y_2|Y_1}) - E(f_{Y'_1} p_{Y_2|Y'_1}) \quad (3.2)$$

This effect suggests the expected relative advantages of each man in position Y_1 over a man in position Y'_1 in reproducing sons in position Y_2 .

The model can further incorporate potential influences from grandparents, great grandparents, and earlier generations of ancestors in both the fertility and the mobility components. The multigenerational form of the model specifies that:

$$S_{Y_t|\bar{Y}_{t-1}} = F_{\bar{Y}_{t-1}} \cdot f_{\bar{Y}_{t-1}} \cdot p_{Y_t|\bar{Y}_{t-1}} \quad (3.3)$$

where \bar{Y}_{t-1} is a vector that denotes the family history of positions and t denotes the generation sequence; $S_{Y_t|\bar{Y}_{t-1}}$ denotes the number of men in generation t who are in position Y_t and have fathers in position Y_{t-1} , grandfathers in position Y_{t-2} and so forth; $F_{\bar{Y}_{t-1}}$ denotes the number of men in generation $t-1$ and those in prior generations in the position history \bar{Y}_{t-1} ; $f_{\bar{Y}_{t-1}}$ denotes the expected number of sons of men in generation $t-1$, given that positions of early generations are \bar{Y}_{t-1} ; and $p_{Y_t|\bar{Y}_{t-1}}$ denotes the probability that a son born to a family in positional history \bar{Y}_{t-1} achieve position Y_t . Accordingly, the joint demographic and mobility effect in the multigenerational form is:

$$E(f_{\bar{Y}_{t-1}} \cdot p_{Y_t|\bar{Y}_{t-1}}) - E(f_{\bar{Y}'_{t-1}} \cdot p_{Y_t|\bar{Y}'_{t-1}}) \quad (3.4)$$

which is the difference of the expected number of individuals in position Y_t in a population from families with position history \bar{Y}_{t-1} compared to \bar{Y}'_{t-1} . Standard errors of the joint effect, for either the two-generational or multigenerational model, can be estimated by the delta method, based on the mean estimates of f and p , and the variance-covariance matrix of these estimates.

3.3 Estimation of the Model Parameters

To estimate the fertility and mobility parameters above, we rely on regression-based methods. For the fertility component, we assume that the number of offspring for the k^{th} individual in the population is denoted as f_k , which follows a Poisson distribution with parameter $\mu > 0$

$$P(f_k|\mu_k, \mathbf{X}_k) = \frac{\exp(-\mu_k)\mu_k^{f_k}}{f_k!} \quad (3.5)$$

where μ_k is the mean number of offspring. The Poisson distribution assumes that the mean number of offspring equals the variances of the number of offspring. When this assumption is violated, we can choose a negative binomial distribution instead, in which the variance in the number of offspring is assumed to follow a gamma distribution with a parameter that is estimated separately.

$$P(f_k|\mu_k, \theta_k, \mathbf{X}_k) = \frac{\Gamma(f_k + \theta_k)}{\Gamma(\theta_k) \cdot \Gamma(f_k + 1)} \cdot \frac{\mu_k^{f_k} \cdot \theta_k^{\theta_k}}{(\mu_k + \theta_k)^{f_k + \theta_k}} \quad (3.6)$$

where $\Gamma(\cdot)$ is the gamma function. The mean of the negative binomial distribution is μ_k (like the Poisson distribution), but with variance of $\mu_k + (\mu_k^2)/\theta_k$, where θ_k is the dispersion parameter.

In both the Poisson and the negative binomial regressions, we model the expected number of offspring as a function of the socioeconomic characteristics of the fathers, grandfathers and, potentially, earlier generations, and other control variables.

$$\log^?(\mu_k) = \log^?(E(f_k|\mathbf{X}_k)) = \mathbf{X}'_k\boldsymbol{\beta} \quad (3.7)$$

The Poisson model (and the negative binomial model) assumes that the proportion of observed counts at each level of fertility, including zero children, in the

empirical data matches the proportions predicted by the Poisson (and the negative binomial) distribution. This assumption may be particularly problematic for the distinction between childless individuals and those with children, because different mechanisms may account for the influence of individual's status on the probability of having no offspring, and conditional on having at least one offspring, the probabilities of having different numbers of offspring. For example, in most developed societies individuals with high socioeconomic status tend to have fewer children than those low in status, whereas socioeconomic status may have positive or negative associations with childlessness (e.g., Heaton, Jacobson and Holland 1999; Abma and Martinez 2006).

To allow parents' and grandparents' characteristics to have separate effects on the probabilities of being childless, and conditional on having at least one offspring, the total number of offspring, we introduce a mixture Poisson or negative binomial distribution that models processes of having any offspring and having positive number of offspring jointly (Johnson, Kemp and Kotz 2005). Suppose that π and $1 - \pi$ are the probabilities of failure and success for overcoming a 'hurdle' that conditions success at reproduction; or, in other words, the probability of avoiding childlessness. The model specifies that for the k^{th} individual,

$$\begin{aligned} P[f_k = 0 | \mathbf{Z}_k] &= \pi \\ P[f_k = n | \mathbf{X}_k, f_k > 0] &= \frac{(1 - \pi)p_n}{1 - p_0} \end{aligned} \tag{3.8}$$

where $P[f_k = n]$ is the probability that the number of offspring for the k^{th} individual is n ; \mathbf{Z} is the set of covariates to explain having no offspring and \mathbf{X} is the set of covariates to explain a positive number of offspring; p_n (and p_0) is the probability of having a given number of offspring in the Poisson (or negative binomial) distribution.

We use a logit model to predict $P[f_k = 0]$ and assume that nonzero fertility

$P[f_k = n | f_k > 0]$ follows a truncated Poisson or negative binomial distribution. Thus, we can model them jointly by the mixture Poisson model:

$$\begin{aligned}
 P[f_k = 0 | \mathbf{Z}_k] &= \frac{1}{1 + \exp(\mathbf{Z}'_k \boldsymbol{\gamma})} \\
 P[f_k = n | \mathbf{X}_k, f_k > 0] &= \frac{(1 - P[f_k = 0 | \mathbf{Z}_k]) \exp(-\mu_k) \mu_k^{f_k}}{f_k! [1 - \exp(-\mu_k)]}
 \end{aligned} \tag{3.9}$$

In particular, when all the zeros in the fertility distribution are generated by the same Poisson process as all the positive numbers, $P[f_k = 0 | \mathbf{X}_k]$ in the Poisson model is the same as that estimated from the logit model, which equals $\exp(-\mu_k)$ according to equation 3.5. Therefore, $P[f_k = n | \mathbf{X}_k, f_k > 0]$ in the truncated Poisson model becomes $\frac{\exp(-\mu_k) \mu_k^{f_k}}{f_k!}$, which follows the same form as equation 3.5.

The mixture logit and Poisson model has two advantages over the simple Poisson or negative binomial model. First, it allows us to examine whether the mechanisms that determine individuals' decisions about whether to have offspring are the same ones that determine how many offspring they have. Second, the separation of the zero fertility from the positive fertility can increase the precision of the adjustment method we propose for retrospective data. We illustrate the latter point in the next section.

We estimate the mobility probabilities from a multinomial logit model or an ordered logit model, depending on whether the socioeconomic outcome is purely categorical or ordered. In the multinomial model, the categories of the socioeconomic characteristic are $i \in \{1, 2, \dots, n\}$ with 1 being the reference category. The probability of achieving position i for the k^{th} individual is

$$P(y_k = i | \mathbf{X}_k) = \frac{\exp(\mathbf{X}'_k \boldsymbol{\beta}_i)}{1 + \sum_{i=2}^n \exp(\mathbf{X}'_k \boldsymbol{\beta}_i)} \tag{3.10}$$

Similarly, \mathbf{X} includes a set of determinants in the social mobility process, such as the positions of an individual's parent, grandparent, and potentially, earlier

generations, and other control variables. In the ordered logit model, which we use for the analysis of educational attainment shown later in this thesis, we estimate the probability of attaining the i^{th} level of schooling for the k^{th} individual as

$$P(y_k = i | \mathbf{X}_k) = \begin{cases} P(\epsilon_k < \tau_1 - \mathbf{X}'_k \boldsymbol{\beta}) & i = 1 \\ P(\epsilon_k < \tau_i - \mathbf{X}'_k \boldsymbol{\beta}) - P(\epsilon_k < \tau_{i-1} - \mathbf{X}'_k \boldsymbol{\beta}) & 1 < i \leq I - 1 \\ 1 - P(\epsilon_k < \tau_{I-1} - \mathbf{X}'_k \boldsymbol{\beta}) & i = I \end{cases} \quad (3.11)$$

$$= \begin{cases} \frac{1}{1 + \exp(-\tau_1 + \mathbf{X}'_k \boldsymbol{\beta})} & i = 1 \\ \frac{1}{1 + \exp(-\tau_i + \mathbf{X}'_k \boldsymbol{\beta})} - \frac{1}{1 + \exp(-\tau_{i-1} + \mathbf{X}'_k \boldsymbol{\beta})} & 1 < i \leq I - 1 \\ 1 - \frac{1}{1 + \exp(-\tau_{I-1} + \mathbf{X}'_k \boldsymbol{\beta})} & i = I \end{cases} \quad (3.12)$$

where ϵ refers to the error term, which is assumed to follow a logistic distribution with a mean of 0 and a variance of $\pi^2/3$; and τ denotes the cutpoints of the latent variable corresponding to the observed y . The probability that an individual's position is between two consecutive values is the difference between the cumulative distribution function evaluated at these values.

The mobility estimates are based on estimated probabilities of attainment for a single son. The joint demographic and mobility model specifies the number of men in a given position in the fathers' generation, the expected number of sons born to each man in that position, and the probability that a son with a father in that position will attain a specific position. Thus, the mobility probabilities in the model are estimated by giving an equal weight to each man in the sons' generation. This implies that retrospective data would yield the same results as prospective data, if the prospective data include fathers and all their sons. In that case, there is no need to adjust the retrospective mobility estimates. When

the prospective data include only a single randomly selected son for each father, however, we need to adjust the mobility estimates by weighting for differential fertility.

3.4 An Adjustment Method for Retrospective Data

As discussed above, the joint demographic and mobility model estimated from retrospective data may suffer from retrospective reporting bias. This means that (1) the information from men in the fathers' generation who have multiple offspring will be overrepresented in the retrospective samples reported by the respondents, and (2) the information from men in the fathers' generation who have no offspring will be omitted from retrospective samples. The first source of the bias can be corrected by the inverse probability weighting method (Horvitz and Thompson 1952), namely, weighting each respondent by the inverse of the number of siblings of respondents plus themselves, that is, for the k^{th} individual, $w_k = 1/(sibs_k + 1)$. The relationship between the expected value of a variable measured in the weighted retrospective sample, X' , and the same variable measured in the original retrospective sample, X , is as follows:

$$E(X'_k) = \sum_{k=1}^m \frac{1}{m} \cdot X'_k = \frac{\sum_{k=1}^n w_k \cdot X_k}{\sum_{k=1}^n w_k} = \frac{\sum_{k=1}^n \frac{1}{(sibs_k + 1)} \cdot X_k}{\sum_{k=1}^n \frac{1}{(sibs_k + 1)}} \quad (3.13)$$

where n refers to the original sample size of the retrospective sample, and m is the weighted sample size. This new estimator X' is also known as the Hájek estimator (Hájek 1971).

For the second problem, we can estimate childlessness probabilities from ancillary data sources. For populations in which fertility patterns are approximately stable from one generation to the next, it may be possible to use childlessness probabilities of the *offspring generation*, provided that they report their fertili-

ty and a sufficient number of them have completed their childbearing. To see why the two steps of adjustment (for differential fertility and for childlessness in the parents' generation) can yield an unbiased estimate of fertility in the fathers' generation, we express the expected fertility for the k^{th} individual in the fathers' generation as:

$$E(f_k|\mathbf{X}_k) = P(f_k = 0|\mathbf{X}_k) \cdot 0 + (1 - P(f_k = 0|\mathbf{X}_k)) \cdot E(f_k|\mathbf{X}_k, f_k > 0) \quad (3.14)$$

After weighting the data based on equation 3.13, we can estimate $E(f_k|\mathbf{X}_k, f_k > 0)$ from a truncated Poisson (or negative binomial) model for retrospective data, that is,

$$E(f_k|\mathbf{X}_k) = \frac{\mu_k}{P(f_k > 0|\mathbf{X}_k)} = \frac{\mu_k}{1 - \exp(-\mu_k)} \quad (3.15)$$

where $\mu_k = \exp(\mathbf{X}'_k\boldsymbol{\beta})$. In principle, this adjusted retrospective estimate of $f_k(> 0)$ should be the same as the prospective estimate from the truncated part in the mixture logit and Poisson (or negative binomial) model in equation 3.9 (Long 1997).

Now, we only need to approximate $P(f_k = 0|\mathbf{X}_k)$ for the fathers' generation using $P(f'_k = 0|\mathbf{Z}_k)$ from an ancillary source or from the sons' generation, assuming that the distribution of childlessness probabilities does not change substantially across generations. We rely on a logit model to estimate this probability in equation 3.17 below, but other solutions, such as probit or linear probability models, are also possible. When this approximation is accurate, the retrospective estimate of $P(f'_k = 0|\mathbf{Z}_k)$ should be close to the prospective estimate from the logit part of the mixture Poisson (or negative binomial) model in equation 3.9.

Overall, the adjusted retrospective estimate of fertility in the fathers generation

in equation 3.14 can be expressed as

$$E(f_k|\mathbf{X}_k) = (1 - P(f_k = 0|\mathbf{X}_k)) \cdot \frac{\mu_k}{1 - \exp(-\mu_k)} \quad (3.16)$$

$$= \left(1 - \frac{1}{1 + \exp(\mathbf{Z}'_k \boldsymbol{\gamma})}\right) \cdot \frac{\exp(\mathbf{X}'_k \boldsymbol{\beta})}{1 - \exp(-\exp(\mathbf{X}'_k \boldsymbol{\beta}))} \quad (3.17)$$

To evaluate the performance of the adjustment method, we define the bias in the retrospective estimates of the joint demographic and mobility effect as

$$B = \left(E(f_{\bar{\mathbf{Y}}_{t-1}} \cdot p_{Y_t|\bar{\mathbf{Y}}_{t-1}}) - E(f_{\bar{\mathbf{Y}}'_{t-1}} \cdot p_{Y'_t|\bar{\mathbf{Y}}'_{t-1}})\right)_p - \left(E(f_{\bar{\mathbf{Y}}_{t-1}} \cdot p_{Y_t|\bar{\mathbf{Y}}_{t-1}}) - E(f_{\bar{\mathbf{Y}}'_{t-1}} \cdot p_{Y'_t|\bar{\mathbf{Y}}'_{t-1}})\right)_r \quad (3.18)$$

Then adjusted bias and the percent reduction in bias Δ are defined accordingly

$$B = \left(E(f_{\bar{\mathbf{Y}}_{t-1}} \cdot p_{Y_t|\bar{\mathbf{Y}}_{t-1}}) - E(f_{\bar{\mathbf{Y}}'_{t-1}} \cdot p_{Y'_t|\bar{\mathbf{Y}}'_{t-1}})\right)_p - \left(E(f_{\bar{\mathbf{Y}}_{t-1}} \cdot p_{Y_t|\bar{\mathbf{Y}}_{t-1}}) - E(f_{\bar{\mathbf{Y}}'_{t-1}} \cdot p_{Y'_t|\bar{\mathbf{Y}}'_{t-1}})\right)_{adj} \quad (3.19)$$

$$\Delta = 100(1 - |\gamma|)\% \text{ where } B_{adj} = \gamma B \quad (3.20)$$

CHAPTER 4

Simulated Example

In this section, we simulate several prospective data sets under different fertility and mobility assumptions using Monte Carlo methods. We generate samples of data generated by a joint fertility-mobility model with known parameters. Given these simulated data we first obtain the prospective estimates of the joint demography and mobility model. Then we treat the data sets retrospectively, assuming that information from the childless group is missing, and that individuals with different numbers of offspring are disproportionately represented in the sample. This procedure shows the extent of the retrospective reporting bias in the joint demographic and mobility effect estimated from retrospective data. Our illustrations include a two-generation model which only focuses on father-son associations, and a three-generation model which takes both fathers and grandfathers into account. Multi-generational models with four or more generations can be simulated in a similar fashion.

4.1 Data Generating Process

We generate three variables F (number of sons), Y (socioeconomic position), and U (a random variable that summarizes personal attributes, such as ability, genetic endowment, and experiences) for each of 10,000 subjects in the initial generation. We first generate the personal attributes variable U for the initial generation, which is drawn from a standard normal distribution. We assume the socioeconomic position is a dichotomous variable with two categories {1=low,

2=high}, and for each subject, we draw the variable from a Bernoulli distribution with the mean conditional on the exogenous variable (U). Then we draw fertility (F) from a Poisson distribution with the mean parameter to be determined by the socioeconomic position variable (Y) of the current generation. We then generate a dichotomous variable D , indicating whether the fertility is zero ($D = 0$) or positive ($D = 1$). Since we do not count daughters, we assume that the number of men without any offspring is number of men without sons. Once the variables for the initial generation are generated, all the subsequent generations can be generated by fertility and mobility rules specified in the equations below.

We first simulate a two-generation data set, in which we assume a man's fertility at the t^{th} generation (F_t) depends on his socioeconomic position (Y_t , equation 4.1), and his socioeconomic position (Y_t) depends only on his father's position (Y_{t-1} , equation 4.2), given that his father has at least one son ($D_{t-1} = 1$). We then simulate a three-generation data set, in which multigenerational effects are allowed. Specifically, we assume that a man's fertility (F_t) depends on the socioeconomic positions and fertility of all prior generations ($Y_{t-1} \dots Y_1$ and $F_{t-1} \dots F_1$) as well as his own socioeconomic position (Y_t , equation 4.3). Also, a man's socioeconomic position (Y_t) depends on the socioeconomic positions of all prior generations ($Y_{t-1} \dots Y_1$, equation 4.4) and his father's fertility (F_{t-1}), given that his father has at least one son ($D_{t-1} = 1$). The Appendix gives a detailed description of the simulation procedures.

Two-generation model:

$$E(F_2|Y_2, D_2 = 1) = \exp(\beta_0 + \beta_1(Y_2 - \bar{Y}_2)) = \exp(\log(1.1) + 0.6 \cdot (Y_2 - \bar{Y}_2)) \quad (4.1)$$

$$\begin{aligned} \text{logit}(P[Y_2 = 2|U_2, Y_1, D_1 = 1]) &= \delta_0 + \delta_1 \cdot U_2 + \delta_2 \cdot Y_1 & (4.2) \\ &= \log\left(\frac{0.2}{0.8}\right) + \log(2) \cdot U_2 + \log(2.5) \cdot Y_1 \end{aligned}$$

Three generation model:

$$\begin{aligned}
E(F_3|Y_3, Y_2, Y_1, F_2, F_1, D_1 = 1) &= \exp(\zeta_0 + \zeta_1 \cdot (Y_3 - \bar{Y}_3) + \zeta_2 \cdot (Y_2 - \bar{Y}_2) \\
&\quad + \zeta_3 \cdot (Y_1 - \bar{Y}_1) + \zeta_4 \cdot (F_2 - \bar{F}_2) + \zeta_5 \cdot (F_1 - \bar{F}_1)) \\
&= \exp(\log(1.1) + 0.36 \cdot (Y_3 - \bar{Y}_3) + 0.20 \cdot (Y_2 - \bar{Y}_2) + 0.10 \cdot (Y_1 - \bar{Y}_1) \\
&\quad + 0.10 \cdot (F_2 - \bar{F}_2) + 0.03 \cdot (F_1 - \bar{F}_1))
\end{aligned} \tag{4.3}$$

$$\begin{aligned}
\text{logit}(P[Y_3 = 2|U_3, Y_2, U_2, F_2, Y_1, U_1, D_2 = 1]) &= \lambda_0 + \lambda_1 \cdot U_3 + \lambda_2 \cdot Y_2 + \\
&\quad \lambda_3 \cdot U_2 + \lambda_4 \cdot F_2 + \lambda_5 \cdot Y_1 + \lambda_6 \cdot U_1 \\
&= \log\left(\frac{0.15}{0.85}\right) + \log(1.8)U_3 + \log(2.0)Y_2 + \log(1.3)U_2 + \log(1.1)F_2 \\
&\quad + \log(1.5)Y_1 + \log(1.1)U_1
\end{aligned} \tag{4.4}$$

In the two-generation prospective sample, all the variables $F_1, F_2, D_1, D_2, Y_1, Y_2$ are observed, whereas in the retrospective sample, we only know $F_1 > 0$ (i.e., given $D_1 = 1$), F_2, D_2, Y_1 (given $D_1 = 1$), and Y_2 . We use the proportion of childless adults in the sons' generation ($D_2 = 0$) to approximate that of the fathers' generation ($D_1 = 0$) in the adjusted retrospective method. Likewise, in the three-generation prospective sample, all the variables $F_1, F_2, F_3, D_1, D_2, D_3, Y_1, Y_2$, and Y_3 are observed, whereas in the retrospective sample, we only know $F_1 > 0$ (given $D_1 = 1$ and $D_2 = 1$), $F_2 > 0$ (given $D_2 = 1$), F_3, D_3, Y_1 (given $D_1 = 1$ and $D_2 = 1$), Y_2 (given $D_2 = 1$), and Y_3 . We need to use the proportion of childless adults in the sons' generation ($D_3 = 0$) to approximate that of the fathers' generation ($D_2 = 0$) in the adjusted retrospective method. While it is possible to approximate $D_1 = 0$ based on some compound estimates of D_2, D_3, F_2 , and F_3 , we omit the discussion here, since D_1 is not useful for estimating the joint demographic and mobility model.

We randomly generate 1,000 data sets and obtain the unadjusted prospective and the retrospective Monte Carlo estimates for the joint demographic and mobility effect. We then show the adjusted results for the retrospective results by

weighting the overrepresented fathers and approximating the number of childless adults in the fathers' generation.

4.2 Simulation Results

We first simulate a two-generation prospective data set for men only. The first column in Table 1 presents the prospective fertility results of men in the first generation and the predicted mobility probability of their sons based on the correct model that was used to generate the data. On average, a man in position 2 produces 0.54 more sons who are in position 2 than a man in position 1 does. Next we treat the data retrospectively by analyzing men in the second generation and their reported fathers' positions and fertility. The results presented in model 1.2 show that the retrospective method largely overestimates the fertility of the fathers' generation, for men in both high and low socioeconomic positions. As a result, the joint effect increases from 0.54 to 0.77: the retrospective model implies that not only do high status fathers produce more high status sons than low status fathers do, but also that the degree of their advantage is greater than in the correct model. In model 1.3, we adjust for the overrepresentation of sons in the retrospective models; that is, we weight each son by the inverse of his father's fertility. In model 1.4, we approximate the proportion of childless men in the fathers' generation by the proportion of childless men in the sons' generation. The mobility estimates are the same across the four columns, because, as discussed above, mobility probabilities estimated from retrospective and prospective data are the same when prospective data include all the sons of each father.

After weighting the fertility of fathers who have multiple sons, the retrospective fertility estimates decrease from 2.66 to 2.05, and the joint effect decreases from 0.77 to 0.59, only a slight upward bias. The weighting step shown in model 1.3 eliminates 80.7% of the bias in the retrospective estimate of the joint effect.

Furthermore, after we adjust for the proportion of men without any sons in the fathers' generation, the estimates from model 1.4 become almost equal to the prospective estimates.

Table 1. Two-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models based on Monte Carlo Simulation

	1.1 Prospective Method	1.2 Retrospective Method	1.3 Adjusted Retrospective Method (weighting only)	1.4 Adjusted Retrospective Method (weighting + zero fertility)
Fertility f_{G1}				
f_1	0.910 (0.012)	1.909 (0.021)	1.523 (0.012)	0.911 (0.011)
f_2	1.656 (0.024)	2.655 (0.034)	2.047 (0.023)	1.658 (0.023)
Mobility $p_{G2 G1}$				
$p_{2 1}$	0.207 (0.005)			
$p_{2 2}$	0.440 (0.007)			
The joint effect $f_2 p_{2 2} - f_1 p_{2 1}$	0.541 (0.016)	0.774 (0.026)	0.586 (0.019)	0.541 (0.016)
Bias (ref. model 1.1)	-	0.233	0.045	0.000

Note: Figures in the parentheses are standard errors.

* The mobility estimates are the same across all the models, because the prospective and the retrospective data yield the same results.

Next, we simulate a three-generation data set with lagged effects of grandfathers on father's fertility and son's mobility. The results in Table 2 suggest similar patterns to those shown in Table 1, except that the joint demographic and mobility effects are greater than in the two-generation model. The weighting method eliminates the bias in the retrospective estimates of the joint effect by 72.1%. The adjustment for the childless population accounts for the remaining 27.9% of the

bias. On average, a father in position 2 has 0.75 more sons who achieve position 2 than a father in position 1. The retrospective estimates are again very close to the prospective estimates after we adjust for the overrepresentation of high fertility fathers and the omission of men without sons.

Table 2. Three-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models based on Monte Carlo Simulation

	2.1 Prospective Method	2.2 Retrospective Method	2.3 Adjusted Retrospective Method (weighting only)	2.4 Adjusted Retrospective Method (weighting + zero fertility)
Fertility $f_{\{G2,G1\}}$				
$f_{\{1,1\}}$	0.844 (0.014)	1.851 (0.025)	1.484 (0.015)	0.845 (0.014)
$f_{\{2,2\}}$	1.680 (0.026)	2.711 (0.040)	2.077 (0.025)	1.677 (0.026)
Mobility $p_{G3 \{G2,G1\}}$				
$p_{2 \{1,1\}}$	0.162 (0.006)			
$p_{2 \{2,2\}}$	0.530 (0.008)			
The joint effect $f_{\{2,2\}}p_{2 \{2,2\}} - f_{\{1,1\}}p_{2 \{1,1\}}$	0.754 (0.021)	1.138 (0.034)	0.861 (0.024)	0.753 (0.021)
Bias (ref. model 2.1)	-	0.384	0.107	-0.001

Note: Figures in the parentheses are standard errors.

* The mobility estimates are the same across all the models, because the prospective and the retrospective data yield the same results.

CHAPTER 5

Empirical Example

5.1 Data

The simulated example shows the theoretical performance of the adjustment method, when the underlying probabilities of positive fertility and childlessness are fixed over generations. To show the effectiveness of the adjustment method for real data when the true parameters are unknown, we apply it to a retrospective sample constructed from a prospective data set—the Panel Study of Income Dynamics (PSID 1968-2009). The PSID began in 1968 with a household sample of more than 18,000 Americans from roughly 5,000 families. Original panel members have been followed prospectively each year through 1997 and then biannually. The study follows targeted respondents according to a genealogical design. All heads of households recruited into the PSID in 1968 carry the PSID “gene” and are targeted for collection of detailed socioeconomic information. Members of new households created by the offspring of original targeted household heads retain the PSID “gene” themselves and become permanent PSID respondents. Original panel members are asked questions about the social and economic circumstances of their families of origin. As those original panel members’ children grow older, the PSID also includes information about the social and economic circumstances of multiple generations within families. The data have been widely used in intergenerational mobility studies (Corcoran et al. 1992; Smeeding, Jantti and Erikson 2011; Solon 1992; Torche 2011). The design of the PSID is similar to the second type of the prospective data that we discussed above, which include information

on respondents and all their offspring.

We construct our multigenerational sample through the PSID Family Identification Mapping System (FIMS). The FIMS sample links the PSID respondents with their parents and grandparents who are also PSID sample members. We then merge the person ID in the FIMS sample with the yearly individual files, and keep only the latest available fertility and educational information for all the individuals. We restrict our sample to men in the fathers' generation who were born between 1930 and 1950, so that we can get reasonable retrospective estimates for the sons, since they have reached adulthood by the last wave of the survey in 2009.

We estimate the joint demographic and mobility model with respect to educational attainment, which is transformed from the variable "years of education" into an ordinal variable with four levels: 1 (0-11 years of schooling), 2 (12 years of schooling), 3 (13-15 years of schooling), 4 (16+ years of schooling). We rely on the question about an individual's number of live births to estimate his fertility. Because the question does not identify the sex of births, we estimate the proportion of men without any offspring, rather than men without sons. We obtain the predicted number of sons in the joint demographic and mobility model by dividing the predicted number of offspring by 2, assuming that the sex ratio is 1. In the three-generation model, we assume a man's education depends on both his fathers and grandfather's education, while his fertility depends on his own and his fathers education.

We first estimate the prospective joint demographic and mobility effect from the PSID sample. Then we treat the data retrospectively, which provides information about the sons' education, his number of siblings and offspring, and his father's and grandfather's education. As we discussed earlier, since the retrospective sample omits the childless proportion of men in the fathers' generation, we rely on either the proportion of childless adults in the sons' generation to ap-

proximate that for the fathers' generation, or the childless information from other reliable sources.

The first solution, however, may encounter two interrelated problems. First, many adults in the sons' generation in the PSID sample are still too young to finish bearing children. To estimate the childlessness probability of these adults at the end of their reproductive span, we need to adjust for individuals' exposure time during reproductive ages. In the models estimated in the next section, we adjust for the exposure time by controlling for the respondent's age. Similar methods have been used in Kalbfleisch and Prentice (2002: 334). Second, even if we adjust for the exposure time, we find that only a few sons in the PSID are above age 60, which we assume is the age that men finish their reproduction. Therefore, we use the childlessness estimates from the whole PSID data set, not only from the sample of the sons' generation.

5.2 Empirical Results

Table 3 reports results from the prospective and retrospective fertility and mobility models. In model 3.1, we estimate a negative binomial model for the fertility of men in the fathers' generation. A test that is not reported here suggests that the negative binomial model is preferred to a Poisson model, because of overdispersion of the fertility distribution. The education coefficients of the sons and fathers show a clear educational gradient in the level of fertility; that is, highly educated sons, especially those whose fathers are also highly educated, tend to have fewer offspring.

In model 3.2, we differentiate between the childless group and the group of men with at least one child and estimate the level of fertility with a mixture logit and negative binomial regression model. Positive coefficients from the logit regression imply high odds of having at least one child against being childless.

Coefficients from the truncated negative binomial regression represent effects on the total number of offspring, conditional on having at least one child. A man's own and his father's education both have a weaker effect on whether he has any offspring than on how many offspring he has, given that he has at least one child. Specifically, a man's own education has no impact on whether he has offspring or not, but it strongly reduces his total number of offspring. By contrast, his father's education affects both his chance of having any offspring and his total number of offspring. To compare the mixture model to the regular negative binomial model, we rely on the Vuong likelihood ratio test (Vuong 1989). The alpha parameter in the test, however, shows that differences between the two models are statistically insignificant ($\alpha = 0.04$), which means that the effects of the covariates that predict zeros in the negative binomial models are not different from those in the mixture model.

In the retrospective sample, we approximate the fertility of men in the fathers' generation by the number of siblings of men in the sons' generation. In model 3.3 we show the results both with and without weighting men in the sons' generation by the inverse of one plus their number of siblings. Comparing the coefficients with those in the prospective, truncated negative binomial results in model 3.2, we see that the weighted coefficients are much closer than the unweighted coefficients to the coefficients in model 3.2, implying that the bias is lower when accounting for overrepresentation of high fertility fathers in the retrospective reports from sons.

Next, we approximate the probability of childlessness in mixture model 3.2 with a logit regression that makes use of fertility information from all PSID respondents. In the simulated example, we approximate this probability by the proportion of zero fertility in the sons' generation. As we mentioned earlier, since many PSID sample members in the sons' generation are too young to have finished their reproduction, we rely on fertility information from men in the full

PSID sample, not only those in the offspring sample. In the logit model, we control individuals' age group to adjust for exposure time to the total reproductive span. With this adjustment, we can estimate the childlessness probability at the end of each individual's reproductive span.

In model 3.5, we estimate the educational mobility for men in the sons' generation by their fathers' and grandfathers' education, and their age group using an ordered logit model. The results show that, for our sample of PSID families, a son's level of education does not depend on his grandfather's education when his father's education is taken into account.

Based on the model estimates in Table 3, we calculate the level of total fertility for men (from which we estimate the number of sons), their probability of having any offspring, and the mobility probability for their sons. The calculations are reported in Table 4. The first panel shows the average number of sons by education of men and their fathers across different models. The second panel shows the mobility probability, which is the same across all the models. The third panel presents the probability of childlessness based on the mixture model for the prospective data and the binary logit model for the retrospective data. The fertility estimates from the negative binomial model are very close to those from the mixture model. For example, in both models, the average number of sons is roughly 1.7 for men in the first educational group whose fathers are also in this group. The childlessness proportion for these men is roughly 8%. The retrospective estimates are shown in the last three columns. Model estimates from the truncated negative binomial model in the third column present naive estimates without any adjustment. The last column presents the results from our final adjustment method. The unadjusted estimate suggests that the average number of sons is roughly 2.4 for men who along with their fathers are both in the first educational group. This estimate declines to 1.8 after we adjust for the overrepresentation of fathers with more offspring, and further declines to 1.7 after we adjust for the omission of the childless group.

We report the joint demographic and mobility effect in the fourth panel, and compare our preferred model in column 2 with alternative models in columns 1, 3, 4, and 5. Here we present the joint effect based on one pair of father's and grandfather's characteristicsnamely, both in the top educational group or both in the bottom groupout of many possibilities. The results for the preferred model show that families with both the father and grandfather in the top educational group produce 0.39 more offspring in the top educational group than families in which both the father and grandfather are in the bottom educational group,

despite the lower fertility of more educated fathers and grandfathers. The effect is 0.52 for the simple retrospective model, 0.41 for the weighted retrospective model without adjusting for the childlessness probability, and 0.38 for our final adjustment model. Compared to the bias in the unadjusted retrospective model in column 3, our weighting adjustment model in column 4 eliminates 82.1% of the bias, and the final model in column 5 eliminates 95.5% of the bias based on the formula in equation 3.20.

Table 3. Three-Generation Fertility and Mobility for *Men Born Between 1930 and 1950 and Their Offspring, PSID*

Dependent variable	Prospective Fertility			Retrospective Fertility			Mobility
	3.1 Number of children (NegBin)	3.2 Mixture NegBin model	3.3 Offspring's number of siblings + 1	3.4 Zero fertility	3.5 Grandfather-father-son	Ordered Logit regression	
Individuals'							
schooling							
0-8/9-11	-	-	-	-	-	-	-
12	-	-	-	-	-0.183***	-	-
					(0.064)		
13-15	-	-	-	-	-0.429***	-	-
					(0.071)		
16+	-	-	-	-	-0.461***	-	-
					(0.078)		
Schooling of men born 1930-1950							
0-8/9-11	-	-	-	-	-	-	-
12	-0.169***	0.140	-0.201***	-0.223***	0.028	0.865***	(0.107)
	(0.034)	(0.188)	(0.019)	(0.025)	(0.054)	(0.135)	1.596***
13-15	-0.200***	0.010	-0.294***	-0.303***	-0.034	2.603***	(0.137)
	(0.040)	(0.209)	(0.024)	(0.030)	(0.073)	(0.137)	
16+	-0.335***	-0.170	-0.486***	-0.409***	-0.172*		
	(0.042)	(0.203)	(0.026)	(0.029)	(0.072)		
Schooling of the fathers							
0-8/9-11	-	-	-	-	-	-	-
12	-0.188***	-0.660***	-0.114***	-0.144***	-	0.147	(0.113)
	(0.037)	(0.160)	(0.023)	(0.029)	-	0.132	(0.198)
13-15	-0.275***	-0.502***	-0.283***	-0.229***	-	0.148	(0.178)
	(0.067)	(0.267)	(0.045)	(0.044)	-		
16+	-0.178***	-0.638**	-0.062	-0.104*	-		
	(0.060)	(0.237)	(0.037)	(0.411)	-		

Table 3. [cont.] Three-Generation Fertility and Mobility for *Men* Born Between 1930 and 1950 and Their Offspring

Dependent variable	Prospective Fertility			Retrospective Fertility		Mobility
	3.1 Number of children (NegBin)	3.2 Mixture NegBin model	3.3 Offspring's number of siblings + 1	3.4 Zero fertility	3.5 Grandfather-father-son	
	Logit regression ($f_F \neq 0$ vs. $= 0$)	Truncated NB ($f_F > 0$)	Truncated NB ($f_F > 0$)	Logit regression ($f_F \neq 0$ vs. $= 0$)	Ordered Logit regression	
Age group						
20-29	-	-	-	-	-	-
30-39	-	-	-	1.045***	0.035	0.035
				(0.066)	(0.155)	(0.155)
40-49	-	-	-	1.639***	-0.038	-0.038
				(0.066)	(0.155)	(0.155)
50-59	-	-	-	2.273***	0.223	0.223
				(0.073)	(0.185)	(0.185)
60+	-	-	-	2.272***	-	-
				(0.080)	-	-
Intercept	1.218*** (0.024)	1.271*** (0.246)	1.554*** (0.012)	1.270*** (0.018)	-0.412*** (0.076)	-
Cut Point 1	-	-	-	-	-	-1.044
						(0.168)
Cut Point 2	-	-	-	-	-	1.390
						(0.169)
Cut Point 3	-	-	-	-	-	2.881
						(0.177)
N	2,526	2,526	5,201	5,201	13,092	2,141
Log-likelihood	-4874.82	-4858.85	-10225.22	-3053.73	-6791.63	-2514.11
log(theta)	2.610	3.244	3.978	4.239		

Data Source: The Panel Study of Income Dynamics, 1968-2009.

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.005$. Theta refers to the dispersion parameter in the negative binomial distribution. Only three cases in the mobility sample are aged above 60, so we exclude them from the analysis.

Table 4 Three-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models, *PSID*

Fertility $f_{\{G2,G1\}}$	Prospective Approach			Retrospective Approach		
	Negative Binomial	Mixture Negative Binomial	Truncated NB	Truncated NB	Weighted Truncated NB	Weighted Truncated NB + Zero Fertility Adjusted
$f_{\{1,1\}}$	1.691 (1.012)	1.692 (0.036)	2.366 (0.006)	1.781 (0.009)	1.657 (0.043)	1.657 (0.043)
$f_{\{1,4\}}$	1.415 (0.032)	1.400 (0.080)	2.224 (0.019)	1.605 (0.022)	1.482 (0.098)	1.482 (0.098)
$f_{\{4,1\}}$	1.209 (0.018)	1.211 (0.037)	1.455 (0.012)	1.183 (0.012)	1.107 (0.039)	1.107 (0.039)
$f_{\{4,4\}}$	1.012 (0.028)	1.008 (0.043)	1.368 (0.019)	1.067 (0.019)	0.990 (0.052)	0.990 (0.052)

Mobility $p_{G3\{G2,G1\}}$	Ordered Logit
$p_{4\{1,1\}}$	0.054 (0.005)
$p_{4\{1,4\}}$	0.063 (0.012)
$p_{4\{4,1\}}$	0.437 (0.026)
$p_{4\{4,4\}}$	0.473 (0.043)

Probability of being childless $\pi_{\{G2,G1\}}$	Binary Logit from Mixture NB	Binary Logit
$\pi_{\{1,1\}}$	0.084 (0.138)	0.090 (0.069)
$\pi_{\{1,4\}}$	0.149 (0.266)	0.105 (0.101)
$\pi_{\{4,1\}}$	0.098 (0.161)	0.135 (0.077)
$\pi_{\{4,4\}}$	0.171 (0.206)	0.156 (0.083)

Table 4.[cont.] Three-Generation Prospective Models and Unadjusted and Adjusted Retrospective Models

Fertility $f_{\{G2,G1\}}$	Prospective Approach		Retrospective Approach	
	Negative Binomial	Mixture Negative Binomial	Truncated NB	Weighted Truncated NB + Zero Fertility Adjusted
The joint demographic and mobility effect	0.387	0.385	0.519	0.409
$f_{\{4,4\}}p_{4 \{4,4\}} - f_{\{1,1\}}p_{4 \{1,1\}}$	(0.073)	(0.611)	(0.068)	0.091
Bias (ref.model 4.2)	0.002	-	0.134	0.024
				-0.006

Note: The bias equals the difference of the joint effect between the mixture prospective estimate in model 4.2 and the estimate from an alternative model. The predicted probability of mobility is estimated for the mean age group. The predicted childlessness probability is estimated for the age group 60+. Standard errors of the predicted fertility, the mobility probabilities, and the joint demographic and mobility effect are estimated by the delta method.

* The mobility estimates from the ordered logit regression are the same across all the models, because the prospective and the retrospective data yield the same results.

CHAPTER 6

Conclusion

We have discussed prospective and retrospective approaches to the study of intergenerational social mobility. In general, prospective, longitudinal data are superior to retrospective, cross-sectional data, because they include complete fertility information for the early generations of interviewed families, which allows researchers to examine the mediating role of demography in the intergenerational transmission of socioeconomic status. Although a few more recent studies provide prospective data for more than two generations and allow analyses of multigenerational social mobility, such data remain scarce. Thus it remains desirable to exploit traditional retrospective survey data for estimating joint models of social mobility and demographic processes. However, analyses based on retrospective data may suffer from the retrospective reporting bias problem, because of the omission of childless population in early generations, as well as the overrepresentation of early generations with more offspring and descendants. In this thesis, we have examined the algebraic relationship between traditional mobility tables constructed from retrospective and prospective data. We show that the differences between these two approaches with respect to the mobility odds ratios depend on how the prospective data are collected and the interaction between fertility and social mobility. When prospective data are obtained for all offspring for each individual, these data can be easily used to construct traditional mobility tables based on the offspring sample without any statistical adjustment. However, if the data ask individuals about a randomly chosen offspring, the odds ratios estimated

from retrospective and prospective data may differ. We show that the inconsistency between the two approaches depends on the three-way association among parents' fertility, parents' status, and offspring's status. In the presence of the three-way interaction, we propose a weighting method shown in equation 2.1 to convert the prospective odds ratio to the traditional, retrospective odds ratio.

Retrospective reporting bias becomes a more salient problem when researchers are interested in not only the mobility table itself, but also the persistence or changes in socioeconomic distributions across generations. Demographic pathways, such as marriage, assortative mating, geographic mobility, and fertility, modify the extent to which inequality in one generation is reproduced in the next. Traditional mobility tables focus on the inequality between parents and offspring restricted to those who have offspring, but the process of producing offspring itself also involves social inequality associated with parents' socioeconomic status. Based on the framework of the joint demographic and mobility model proposed by Mare and Maralani (2006), we show how to obtain the prospective estimates of the joint demographic and mobility effect in two-generation and multigenerational mobility examples. We also propose an adjustment method for obtaining the prospective estimates using retrospective data.

A Monte Carlo study comparing the prospective approach with the adjusted and unadjusted retrospective approach shows that the adjustment method removes almost all the difference between the prospective and the biased retrospective estimates. Specifically, the weighting method removes more than 70 percent of the bias, while the remaining bias is eliminated by accounting for childlessness.

Our illustrative analyses of the PSID show how to adjust retrospective mobility data using prospective data and to estimate the joint demographic and mobility model using a mixture logit and negative binomial model. The results suggest that overall the adjustment method removes more than 95% of the bias in the retrospective estimates. The methods proposed in this thesis are potentially ap-

plicable to a wide range of models that include a broader variety of demographic processes and socioeconomic outcomes than those presented here. Compared to the retrospective approach in traditional social mobility studies, the prospective approach provides a broader view of intergenerational inequality. We show in this thesis what kind of new knowledge a forward-looking conceptual framework can offer to social mobility studies and how such a prospective approach can be achieved with limited information from retrospective data.

CHAPTER 7

Appendix: Simulation Details

This appendix provides the details for the simulation examples. For the two-generation model, we assume that a man's fertility depends on his own socioeconomic position, and his socioeconomic position depends on only his father's position. There is no lagged effect from the grandfather in both the fertility and mobility equations. The data are generated in the following order according to the specified probability models:

1.1 The exogenous variable U_1 for the fathers' generation is drawn from a standard normal distribution $U_1 \sim N(0, 1)$.

1.2. We then generate the father's position Y_1 (1 vs. 2) for each of the 10,000 subjects:

$$\text{logit}(P[Y_1 = 2|U_1]) = \alpha_0 + \alpha_1 \cdot U_1 = \log\left(\frac{0.3}{0.7}\right) + \log(2) \cdot U_1 \quad (7.1)$$

1.3. The conditional distribution of a father's fertility given Y_1 follows a Poisson distribution with the mean of the fertility satisfies equation below.

$$E(F_1|Y_1) = \exp(\beta_0 + \beta_1 \cdot (Y_1 - \bar{Y}_1)) = \exp(\log(1.1) + 0.6 \cdot (Y_1 - \bar{Y}_1)) \quad (7.2)$$

We then generate a dichotomous variable D_1 based on F_1 , where $D_1 = 1$ if $F_1 > 0$ and $D_1 = 0$ if $F_1 = 0$.

1.4. The conditional distribution of a son's variable U_2 given U_1 and D_1 is drawn from a normal distribution, where the standard deviation is fixed at 1 and

the mean satisfies the equation below.

$$E(U_2|U_1, D_1 = 1) = \gamma_1 \cdot (U_1 - \bar{U}_1) = 0.8 \cdot (U_1 - \bar{U}_1) \quad (7.3)$$

1.5. The conditional distribution of a son's socioeconomic position Y_2 given U_2 , D_1 and Y_1 follows a Bernoulli distribution.

$$\begin{aligned} \text{logit}(P[Y_2 = 2|U_2, Y_1, D_1 = 1]) &= \delta_0 + \delta_1 \cdot U_2 + \delta_2 \cdot Y_1 \\ &= \log\left(\frac{0.2}{0.8}\right) + \log(2) \cdot U_2 + \log(2.5) \cdot Y_1 \end{aligned} \quad (7.4)$$

1.6. The conditional distribution of a son's fertility F_2 given Y_2 , and D_1 follows a Poisson distribution with the mean of the fertility satisfies the equation below.

$$E(F_2|Y_2, D_1 = 1) = \exp(\beta_0 + \beta_1 \cdot (Y_2 - \bar{Y}_2)) = \exp(\log(1.1) + 0.6 \cdot (Y_2 - \bar{Y}_2)) \quad (7.5)$$

We then generate a dichotomous variable D_2 based on F_2 , where $D_2 = 1$ if $F_2 > 0$ and $D_2 = 0$ if $F_2 = 0$.

In the prospective sample, all the variables $F_1, F_2, D_1, D_2, Y_1, Y_2$ are observed, while in the retrospective sample, we only know $F_1 > 0$ (i.e., $D_1 = 1$), F_2, D_2, Y_1, Y_2 . We need to use the proportion of childless adults in the sons' generation ($D_2 = 1$) to approximate that of the fathers' generation ($D_1 = 1$) in the adjusted retrospective method.

For the three-generation model, we assume that a man's fertility depends on the socioeconomic positions and fertility of all prior generations, as well as his own socioeconomic position. In addition, we assume a man's socioeconomic position depends on the socioeconomic positions of all prior generations. We generate the

data by the following steps.

2.1. The exogenous variable U_1 for the grandfathers' generation follows a standard normal distribution $U_1 \sim N(0, 1)$.

2.2. The conditional distribution of a grandfather's socioeconomic position Y_1 (1 vs. 2) given U_1 , follows a Bernoulli distribution.

$$\text{logit}(P[Y_1 = 2|U_1]) = \alpha_0 + \alpha_1 \cdot U_1 = \log\left(\frac{0.3}{0.7}\right) + \log(2) \cdot U_1 \quad (7.6)$$

2.3. The conditional distribution of a grandfather's fertility F_1 given Y_1 follows a Poisson distribution with the mean of the fertility satisfies the equation below.

$$E(F_1|Y_1) = \exp(\beta_0 + \beta_1 \cdot (Y_1 - \bar{Y}_1)) = \exp(\log(1.1) + 0.6 \cdot (Y_1 - \bar{Y}_1)) \quad (7.7)$$

We generate a dichotomous variable D_1 based on F_1 , where $D_1 = 1$ if $F_1 > 0$ and $D_1 = 0$ if $F_1 = 0$.

2.4. The conditional distribution of a father's variable U_2 given U_1 and D_1 follows a normal distribution, where the standard deviation is fixed at 1 and the mean satisfies the equation below.

$$E(U_2|U_1, D_1 = 1) = \gamma_1 \cdot (U_1 - \bar{U}_1) = 0.8 \cdot (U_1 - \bar{U}_1) \quad (7.8)$$

2.5. The conditional distribution of a father's position Y_2 given U_2 , F_1 , D_1 and Y_1 follows a Bernoulli distribution.

$$\begin{aligned} \text{logit}(P[Y_2 = 2|U_2, Y_1, D_1 = 1]) &= \delta_0 + \delta_1 \cdot U_2 + \delta_2 \cdot Y_1 + \delta_3 \cdot F_1 \quad (7.9) \\ &= \log\left(\frac{0.2}{0.8}\right) + \log(2) \cdot U_2 + \log(2.5) \cdot Y_1 + \log(1.1) \cdot F_1 \end{aligned}$$

2.6. The conditional distribution of a father's fertility F_2 given Y_2 , Y_1 , F_1 and D_1 follows a Poisson distribution with the mean of the fertility satisfies the equation below.

$$\begin{aligned} E(F_2|Y_2, Y_1, F_1, D_1 = 1) &= \exp(\theta_0 + \theta_1 \cdot (Y_2 - \bar{Y}_2) + \theta_2 \cdot (Y_1 - \bar{Y}_1) + \theta_3 \cdot (F_1 - \bar{F}_1)) \\ &= \exp(\log(1.1) + 0.4 \cdot (Y_2 - \bar{Y}_2) + 0.2 \cdot (Y_1 - \bar{Y}_1) + 0.1 \cdot (F_1 - \bar{F}_1)) \end{aligned} \quad (7.10)$$

We generate a dichotomous variable D_2 based on F_2 , where $D_2 = 1$ if $F_2 > 0$ and $D_2 = 0$ if $F_2 = 0$.

2.7. The conditional distribution of a son's variable U_3 given U_2 , U_1 and D_2 follows a normal distribution, where the standard deviation is fixed at 1 and the mean satisfies the equation below.

$$\begin{aligned} E(U_3|U_2, U_1, D_2 = 1) &= \pi_1 \cdot (U_2 - \bar{U}_2) + \pi_2 \cdot (U_1 - \bar{U}_1) \\ &= 0.6 \cdot (U_2 - \bar{U}_2) + 0.2 \cdot (U_1 - \bar{U}_1) \end{aligned} \quad (7.11)$$

Note that when $D_2 = 1$, we must have $D_1 = 1$.

2.8. The conditional distribution of a son's position Y_3 given U_3 , Y_2 , U_2 , F_2 , Y_1 , U_1 , and D_2 follows a Bernoulli distribution.

$$\begin{aligned} \text{logit}(P[Y_3 = 2|U_3, Y_2, U_2, F_2, Y_1, U_1, D_2 = 1]) &= \lambda_0 + \lambda_1 \cdot U_3 + \\ &\quad \lambda_2 \cdot Y_2 + \lambda_3 \cdot U_2 + \lambda_4 \cdot F_2 + \lambda_5 \cdot Y_1 + \lambda_6 \cdot U_1 \\ &= \log\left(\frac{0.15}{0.85}\right) + \log(1.8) \cdot U_3 + \log(2.0) \cdot Y_2 + \log(1.3) \cdot U_2 + \\ &\quad \log(1.1) \cdot F_2 + \log(1.5) \cdot Y_1 + \log(1.1) \cdot U_1 \end{aligned} \quad (7.12)$$

2.9. The conditional distribution of a son's fertility F_3 given Y_3, Y_2, Y_1, F_2, F_1 and D_2 follows a Poisson distribution with the mean of the fertility satisfies the equation below.

$$\begin{aligned}
E(F_3|Y_3, Y_2, Y_1, F_2, F_1, D_1 = 1) &= \exp(\zeta_0 + \zeta_1 \cdot (Y_3 - \bar{Y}_3) + \zeta_2 \cdot (Y_2 - \bar{Y}_2) + \\
&\quad \zeta_3 \cdot (Y_1 - \bar{Y}_1) + \zeta_4 \cdot (F_2 - \bar{F}_2) + \zeta_5 \cdot (F_1 - \bar{F}_1)) \quad (7.13) \\
&= \exp(\log(1.1) + 0.36 \cdot (Y_3 - \bar{Y}_3) + 0.20 \cdot (Y_2 - \bar{Y}_2) + 0.10 \cdot (Y_1 - \bar{Y}_1) + \\
&\quad 0.10 \cdot (F_2 - \bar{F}_2) + 0.03 \cdot (F_1 - \bar{F}_1))
\end{aligned}$$

REFERENCES

- [1] Abma, Joyce C. and Gladys M. Martinez. 2006. "Childlessness among Older Women in the United States: Trends and Profiles." *Journal of Marriage and Family* 68(4): 1045-56.
- [2] Allan, Boris, and Bill Bytheway. 1973. "The Effects of Differential Fertility on Sampling in Studies of Intergenerational Social Mobility." *Sociology* 7(2): 273-76.
- [3] Bartholomew, David J. 1982. *Stochastic Models for Social Processes (3rd edition)*. New York: Wiley.
- [4] Beller, Emily. 2009. "Bringing Intergenerational Social Mobility Research into the Twenty-First Century: Why Mothers Matter." *American Sociological Review* 74(4): 507-28.
- [5] Blake, Judith. 1981. "Family Size and the Quality of Children." *Demography* 18(4): 421-42.
- [6] Blau, Peter M., and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Wiley.
- [7] Breen, Richard (ed.). 2004. *Social Mobility in Europe*. Oxford: Oxford University Press.
- [8] British Household Panel Survey (BHPS) User Manual Codebook. 1997. Institute for Social & Economic Research.
- [9] Chinese Family Panel Studies (CFPS). 2010. Chinese Family Panel Studies: A PSC Research Project (<http://www.psc.isr.umich.edu/research/project-detail/34795>).
- [10] Corcoran, Mary, Roger Gordon, Deborah Laren, and Gary Solon. 1992. "The Association between Mens Economic Status and Their Family and Community Origins." *The Journal of Human Resources* 27(4): 575-601.
- [11] Duncan, Otis Dudley. 1966. "Methodological Issues in the Analysis of Social Mobility." Pp. 51-97 in N. J. Smelser and S. M. Lipset (eds.), *Social Structure and Mobility in Economic Development*. Chicago: Aldine.
- [12] Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford University Press, USA.
- [13] Erikson, Robert, John H. Goldthorpe and Lucienne Portocarero. 1979. "Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden." *British Journal of Sociology* 30(4): 415-41

- [14] Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.
- [15] Glass, David V. (ed.). 1954. *Social Mobility in Britain*. London: Routledge and Kegan Paul.
- [16] Hájek, Jaroslav. 1971. "Comment on An Essay on the Logical Foundations of Survey Sampling, Part One," in *The Foundations of Survey Sampling*, Godambe, V.P. and Sprott, D.A. eds., 236, Holt, Rinehart, and Winston.
- [17] Hauser, Robert M., John N Koffel, Harry P. Travis, and Peter J. Dickinson. 1975. "Temporal Change in Occupational Mobility: Evidence for Men in the United States." *American Sociological Review* 40 (3): 279-97.
- [18] Heaton, Tim B., Cardell K. Jacobson, and Kimberlee Holland. 1999. "Persistence and Change in Decisions to Remain Childless." *Journal of Marriage and the Family* 61:531-39.
- [19] Horvitz, Daniel G., and Donovan J. Thompson. 1952. "A Generalization of Sampling without Replacement From A Finite Universe." *Journal of the American Statistical Association* 47 (260): 663-85.
- [20] Household, Income and Labour Dynamics in Australia (HILDA). 2002. "The Household, Income and Labour Dynamics in Australia Survey: Wave 1." *American Economic Review* 35 (3): 339-348
- [21] Hout, Michael. 1983. *Mobility Tables*. Beverly Hills: Sage.
- [22] Hout, Michael. 1988. "More Universalism, Less Structural Mobility: The American Occupational Structure in the 1980s." *American Journal of sociology* 93(6): 1358-1400.
- [23] Johnson, Norman L., Adrienne W. Kemp, and Samuel Kotz. 2005. *Univariate Discrete Distributions (3rd edition)*. New Jersey: John Wiley & Sons, Inc..
- [24] Kalbfleisch, John D., and Ross L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons, Inc.,
- [25] Korean Labor & Income Panel Study (KLIPS). 2012. An Overview of the KLIPS. See (<http://www.kli.re.kr/klips/en/about/introduce.jsp>).
- [26] Lam, David. 1986. "The Dynamics of Population Growth, Differential Fertility, and Inequality." *American Economic Review* 76: 1103-16.
- [27] Lawrence, Matthew, and Richard Breen. 2012. "And Their Children After Them? Estimating the Effect of College on Educational Reproduction." Working paper.

- [28] Long, Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications
- [29] Maralani, Vida. 2013. "The Demography of Social Mobility: Black-White Differences in the Process of Educational Reproduction." *American Journal of Sociology* 118(6): 1509-58.
- [30] Maralani, Vida, and Robert D. Mare. 2005. "Demographic Pathways of Intergenerational Effects: Fertility, Mortality, Marriage and Women's Schooling In Indonesia." CCPR Working Paper
- [31] Mare, Robert D. 1997. "Differential Fertility, Intergenerational Educational Mobility, and Racial Inequality." *Social Science Research* 26 (3): 263-91.
- [32] Mare, Robert D. 2011. "A Multigenerational View of Inequality." *Demography* 48 (1): 1-23.
- [33] Mare, Robert D., and Christine R. Schwartz. 2006. "Educational Assortative Mating and the Family Background of the Next Generation: A Formal Analysis." *Riron to Hoho (Sociological Theory and Methods)* 21: 253C77.
- [34] Mare, Robert D., and Meichu D. Chen. 1986. "Further Evidence on Number of Siblings and Educational Stratification." *American Sociological Review* 51: 403-12.
- [35] Mare, Robert D. and Vida Maralani. 2006. "The Intergenerational Effects Of Changes In Women's Educational Attainments." *American Sociological Review* 71 (4): 542-64.
- [36] Matras, Judah. 1961. "Differential Fertility, Intergenerational Occupational Mobility, and Change in Occupational Distribution: Some Elementary Interrelationships." *Population Studies* 15: 187-97.
- [37] Matras, Judah. 1967. "Social Mobility and Social Structure: Some Insights from the Linear Model." *American Sociological Review* 32: 608-14.
- [38] Musick, Kelly, and Robert D. Mare. 2004. "Family Structure, Intergenerational Mobility, and the Reproduction of Poverty: Evidence for Increasing Polarization?" *Demography* 41(4): 629-48.
- [39] Preston, Samuel H. 1974. "Differential Fertility, Unwanted Fertility, and Racial Trends in Occupational Achievement." *American Sociological Review* 39: 492-506.
- [40] Preston, Samuel H. and Cameron Campbell. 1993. "Differential Fertility and the Distribution of Traits: The Case of IQ." *American Journal of Sociology*, 98: 997C1019.

- [41] PSID Main Interview User Manual: Release 2012.1. Institute for Social Research, University of Michigan, January 23, 2012.
- [42] Smeeding, Timothy M., Markus Jantii, and Robert Erikson, eds. 2011. *Persistence, Privilege, and Parenting: The Comparative Study of Intergenerational Mobility*. Russell Sage Foundation.
- [43] Sobel, Michael E. 1985. "Social Mobility and Fertility Revisited: Some New Models for the Analysis of the Mobility Effects Hypothesis." *American Sociological Review* 50: 699-712.
- [44] Socio-Economic Panel (SOEP) Group. 2000. "The German Socio Economic Panel After More than 15-Years Overview." *Vierteljahresheft zur Wirtschaftsforschung* 70: 7C14.
- [45] Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82 (3): 393-408.
- [46] Survey of Labour and Income Dynamics (SLID, Canada). 2012. A Survey Overview (<http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=75F0011X>).
- [47] Swiss Household Panel (SHP). 2012. Swiss Household Panel Documentations (<http://www.swisspanel.ch/index.php?lang=en>).
- [48] Szelenyi, Ivan and Donald J. Treiman. 1994. *Social Stratification in Eastern Europe after 1989: General Population Survey Provisional codebook*. Unpublished manuscript. Department of Sociology, University of California, Los Angeles.
- [49] Torche, Florencia. 2011. "Is a College Degree Still the Great Equalizer? Intergenerational Mobility across Levels of Schooling in the United States." *American Journal of Sociology* 117(3): 763-807.
- [50] Treiman, Donald. J, and Andrew. G. Walder. 1996. *Life Histories and Social Change in Contemporary China*. Distributed by the UCLA Social Science Data Archive [<http://www.sscnet.ucla.edu/issr/da/>].
- [51] Treiman, Donald. J., Sylvia Moeno, and Lawrence Schlemmer. 1996. *Survey of Socioeconomic Opportunity and Achievement in South Africa Codebook*. Unpublished manuscript. Department of Sociology, University of California Los Angeles.
- [52] Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57(2): 307-33.

- [53] Wisconsin Longitudinal Study (WLS) Handbook. 2006. Wisconsin Longitudinal Study: Tracking the Life Course (http://www.ssc.wisc.edu/wlsresearch/documentation/handbook/WLS_Handbook.pdf).