The Sources of Life Chances: Does Education, Class Category, Occupation, or Short-Term Earnings Predict 20-Year Long-Term Earnings?

ChangHwan Kim,[a] Christopher R. Tamborini,[b] Arthur Sakamoto[c]

a) University of Kansas; b) U.S. Social Security Administration; c) Texas A&M University

Abstract: In sociological studies of economic stratification and intergenerational mobility, occupation has long been presumed to reflect lifetime earnings better than do short-term earnings. However, few studies have actually tested this critical assumption. In this study, we investigate the cross-sectional determinants of 20-year accumulated earnings using data that match respondents in the Survey of Income and Program Participation to their longitudinal earnings records based on administrative tax information from 1990 to 2009. Fit statistics of regression models are estimated to assess the predictive power of various proxy variables, including occupation, education, and short-term earnings, on cumulative earnings over the 20-year time period. Contrary to the popular assumption in sociology, our results find that cross-sectional earnings have greater predictive power on long-term earnings than occupation-based class classifications, including three-digit detailed occupations for both men and women. The model based on educational attainment, including field of study, has slightly better fit than models based on one-digit occupation or the Erikson, Goldthorpe, and Portocarero class scheme. We discuss the theoretical implications of these findings for the sociology of stratification and intergenerational mobility.

Keywords: long-term earnings; educational attainment; occupation; class; administrative data

LONG-TERM earnings are a consequential source of socioeconomic well-being and life chances in contemporary societies (Tamborini, Kim, and Sakamoto 2015). Long-term earnings are associated with a range of outcomes, including savings and investment behaviors, wealth accumulation, retirement income, Social Security benefit levels, social class identity, feelings of self-worth, health, life expectancy, overall life satisfaction, and marital stability (Hout 2008; Kawachi et al. 1997; Rainwater 1974; Stronks et al. 1997; Tamborini, Iams, and Reznik 2012; Western et al. 2012). Long-term earnings also affect intergenerational processes, such as the degree to which parents bequeath wealth to their offspring (Becker and Tomes 1979; Mazumder 2005). Understanding long-term earnings inequality is thus consistent with Weber's emphasis on the importance of "life chances" (Weber [1922] 1978).

Despite being critically important for a variety of socioeconomic outcomes, long-term earnings have not been extensively studied in prior studies. The main factor hindering empirical research is the scarcity of longitudinal data for a broad portion of the labor market. Consequently, sociological studies often use various cross-sectional indicators on the presumption that they shed light on an individual's long-term socioeconomic circumstances.

In sociology, occupation is often considered to be the best proxy of an individual's social class (Erikson, Goldthorpe, and Portocarero 1979; Weeden and Grusky

2005) as well as his or her long-term or lifetime "permanent" income (Featherman and Hauser 1978; Hauser and Warren 1997; Hauser 2010). Occupation is assumed to correlate more highly with lifetime earnings than do short-term earnings and to suffer less from measurement error (Hauser and Warren 1997). Asserting that occupational mobility is superior to earnings mobility in studying intergenerational socioeconomic association (e.g., Torche 2015), sociologists tend to use occupational status to study both cross-sectional stratification and intergenerational stratification processes. However, despite its theoretical importance for a wide range of sociological studies, the widespread assumption that occupation is the best proxy for long-term earnings has surprisingly been subjected to very little empirical validation.

This study seeks to provide robust statistical evidence about the empirical associations of various socioeconomic indicators with long-term earnings. Using restricted-use data that links workers from the Survey of Income and Program Participation (SIPP) to their longitudinal tax records, we provide new evidence on the relationships between key cross-sectional socioeconomic variables and an individual's cumulative earnings over a 20-year window. A range of cross-sectional indicators are considered, including occupation; annual earnings; three-year cumulative earnings; the Erikson, Goldthorpe, and Portocarero (EGP) class scheme; Weeden–Grusky microclass; educational attainment; and field of study.

To be clear, our objective is not to engage in a mindless "sociological horse race" (Grusky 2001:29) but to explore empirical evidence about which set of covariates provides the most empirically justified predictors of long-term earnings. Our analysis thereby ascertains whether different proxies are complementing each other by measuring different sources of long-term earnings or whether the proxies are substituting for one another. Establishing the relative strength of cross-sectional indicators with an individual's cumulative earnings is ultimately important for understanding the determinants of long-term socioeconomic well-being, the mechanisms shaping intra-generational earnings trajectories, and the validity of measures used to estimate intergenerational mobility.

Theories on the Association between Variables Observed in Cross Section and Long-Term Earnings

The assumption that cross-sectional variables can be used to describe an individual's long-term socioeconomic circumstance has been articulated most explicitly in the sociological literature. Perspectives on which cross-sectional proxy variables best predict long-term earnings may be classified into three broad approaches. The first is using occupational classifications (which are sometimes supplemented with employment relations). Another approach is focusing on short-term earnings, which has been popular among both sociologists and economists. A third implicit proxy for long-term earnings is education. These three approaches are certainly not mutually exclusive, and each may be associated with long-term earnings. To some extent, the latter may be better indicated by using occupation, short-term earnings,

and education altogether if they complement each other in predicting long-term earnings.

Occupation-Based Approaches to Indicate Long-Term Earnings

One important strand of research suggests that occupation is a reliable and valid (albeit ordinal) measure of long-term earnings and, more generally, life chances (Blau and Duncan 1967; Hauser and Warren 1997; Hout and DiPrete 2006; Wilkinson 1966). Most sociological studies of intergenerational mobility have relied on occupational information (in some cases, combined with measures of employment relations and authority) as their foundation (Breen and Jonsson 2005; Featherman and Hauser 1978). As stated by a prominent figure in the study of intergenerational occupational mobility, "...there is a 'permanent' level of occupational status around which there are temporary fluctuations" (Hauser 2010:6). Thus, "one might regard occupational socioeconomic status as roughly equivalent to permanent income in studies of intergenerational mobility" (Hauser 2010:7). In this way, Hauser (2010) conceptualizes occupation as a useful proxy of lifetime or long-term earnings. Accordingly, many sociologists have asserted that an individual's occupational status is a better indicator of long-term earnings than is current income (e.g., Hauser and Warren 1997; Torche 2015; Wright 2005).

There are a variety of ways in which occupation has been used to indicate long-term socioeconomic standing. One approach is to rely on the occupational codes constructed by the U.S. Census Bureau. Studies in this strand of research classify an individual's occupation with varying levels of detail using one-digit, two-digit, or three-digit occupational codes. Another approach is to employ occupation to develop class-based schemes. A well-known example in this regard is the Weberian class classification developed by Erikson et al. (1979). The latter creates a typology of job categories differentiated by one's generic relationship to the market and type of employment contract (referred to hereafter as EGP). Although the EGP class scheme was not developed specifically to proxy long-term earnings, the typology has more recently been described as providing a good indicator of long-term earnings (Breen 2005; Erikson and Goldthorpe 1992; Goldthorpe 2012).[1]

Drawing from occupation-based measures, another group of scholars propose a related approach that is based on a neo-Durkheimian "microclass" conceptualization (Jonsson et al. 2009; Weeden and Grusky 2005; Weeden et al. 2007). One of the rationales for this microclass scheme is that the aggregate macroclass approach has never been popular outside academia and has "fail[ed] the realist test" (Grusky 2005:51). According to the microclass scheme, three-digit occupations are collapsed into 126 categories representing discrete groupings in the labor market. These in turn are presumed to indicate an individual's long-term socioeconomic standing (Weeden and Grusky 2005).

Education-Based Approaches to Indicate Long-Term Earnings

An alternative to using occupation-based approaches is to focus on education as an enduring resource affecting long-term earnings (Kim, Tamborini, and Sakamoto 2015; Tamborini et al. 2015). Although the importance of education has been well

appreciated in occupational-based approaches (e.g., Ishida, Muller, and Ridge 1995), it has traditionally been seen more as the major mediating factor in determining occupation rather than the direct source of long-term socioeconomic standing per se. For this reason, only a few sociologists have explored education as a central resource directly affecting long-term earnings because the most important proximate determinant of the latter has been presumed to be occupation.

However, a recent study documents how earnings accumulate over an individual's life by educational attainment and result in large gaps in lifetime earnings across educational groups (Tamborini et al. 2015). In describing demographic trends during the twentieth century, Fischer and Hout (2006:247) concurringly state that "the division between the less- and more-educated grew and emerged as a powerful determinant of life chances and lifestyles." Some related studies in labor market sociology have also investigated the direct effects of education on earnings and have characterized these effects as deriving from factors other than occupation per se (e.g., human capital, enhanced skill development, improved trainability, cumulative advantage, and greater bargaining power of highly educated workers). However, these labor market studies investigate data on cross-sectional earnings rather than long-term earnings (DiPrete and Eirich 2006; Hout 2012; Sakamoto and Kim 2014; Sakamoto and Wang 2017; Tomaskovic-Devey, Thomas, and Johnson 2005).

In economics, educational attainment is also a critical variable in human capital models. From the point of view of human capital models, education reflects an individual's actual and potential productive skills, which in turn are thought to determine long-term earnings (Becker and Tomes 1979). Long-term earnings are particularly important because the returns to human capital investment occur later in time. Modest incomes in a cross section may sometimes be attributed to a high level of human capital investment (e.g., graduate school or on-the-job training, such as an internship) that often result in high earnings later in the work career.

Another aspect of education that is important to long-term earnings is the growing significance of its horizontal dimensions (Gerber and Cheung 2008; Sakamoto and Wang 2017). For example, recent research reveals strong effects of field of study in differentiating the long-term earnings of the college educated (Kim et al. 2015; Ma and Savas 2014). Perhaps not surprisingly, undergraduate degrees in science, technology, engineering, and math (STEM); business; and health science have notably above-average long-term earnings. At the graduate level, persons with degrees in STEM, business, medicine, dentistry, and law have very high long-term earnings. Kim et al. (2015) even suggest that "horizontal stratification in education across field of study may now be more consequential for long-term rewards in the labor market than vertical stratification."

Current (or Short-Term) Earnings as a Proxy for Long-Term Earnings

Another approach is to use cross-sectional or short-term earnings to proxy long-term earnings. A recent study utilizing longitudinal administrative tax information finds that the relative rank of annual earnings is fairly stable between the ages of 30 and 60 (Chetty et al. 2014). This finding, however, does not necessarily mean that cross-sectional annual earnings is the best proxy of long-term earnings. The permanent

income hypothesis proposed by Friedman (1957) stated that consumption decisions are made based on long-term income so that current income is a poor determinant of the current consumption patterns (e.g., Bernanke 1984; DeJuan, Seater, and Wirjanto 2006; Flavin 1981). In economics, permanent income refers to the expected long-term average income. From the perspective of the permanent income hypothesis, short-term current income at any one point in an individual's career is subject to yearly fluctuations as well as measurement error, so it is a poor proxy of their long-term income.

For example, an accountant aged 25 may earn less than an accountant aged 45 in a given year, but their long-term incomes might be similar. In this case, occupation might be a better proxy than annual income. Yet, an alternative scenario is also plausible. For example, a non–tenure-track sociology professor in a regional university will earn substantially less than a tenured full professor of sociology in an elite university in a given year. The discrepancy in annual earnings between these two professors may reflect their gap in long-term earnings quite well in contrast to occupation, which is the same in this case.

A related issue is the extent to which time spent out of the labor force influences long-term earnings (Brenner 2010; Cooper 2013; Davis et al. 2011; Sakamoto, Tamborini, and Kim 2018). Researchers usually drop respondents with nonpositive earnings in a cross-sectional data set from their analyses. In the case of occupation, workers who are out of the labor force or who are unemployed are typically assigned the occupation of their last job, which is potentially misleading in regard to their income. The relative predictive power between occupation and short-term earnings depends in part on which variable reflects the likelihood of extended period of unemployment better.

Earnings are also used to measure the intergenerational income elasticity. The standard economics model of intergenerational mobility is a regression with the individual's earnings as the dependent variable and parental earnings as the independent variable (e.g., Becker and Tomes 1979; Black and Devereux 2011; Solon 1999). The regression coefficient for parental long-term earnings (sometimes referred to as the intergenerational elasticity [IGE]) is interpreted as a measure of intergenerational mobility, with a larger value indicating greater inheritance (Solon 1999).[2] Because long-term earnings are not available in most survey data, economists and a growing number of sociologists use annual earnings for intergenerational mobility.

Previous Empirical Studies on a Proxy for Long-Term Earnings

A small handful of studies directly assess the relationship between different proxy variables and long-term earnings. Brady et al. (2017), to our knowledge, is the only sociological study that has addressed the predictive powers of the various proxy variables on long-term income. Although Brady and his colleagues focused on explaining long-term post-tax and post-transfer household income (rather than individual earnings), their findings are informative. Using the U.S. Panel Study of Income Dynamics (PSID) and the German Socio-Economic Panel (SOEP), they compared the *R*-squared in a model using cross-sectional reports of occupation with that using one year of household income. They find that one year of household

income predicts long-term household income better than cross-sectional occupation even at the three-digit level.

Two economic studies, Goldberger (1989) and Zimmerman (1992), are frequently cited to justify the superiority of occupation (e.g., Hauser and Warren 1997; Torche 2015). Goldberger (1989) suggested that occupation may be a better measure of lifetime earnings than one-year annual earnings. However, his conclusion is based on a simulation rather than actual observational data. Analyzing the National Longitudinal Survey, Zimmerman (1992) argued that the occupation-based Duncan socioeconomic index provides more a accurate measure of long-term earnings than short-term (average) earnings or wages. Zimmerman's (1992) conclusion, however, was inferred from the finding that the intergenerational association using the Duncan index varies less across respondents' ages than the association using short-term (average) earnings. His study did not compare the predictive power between occupation and earnings in accounting for long-term earnings explicitly.

Another strand of research, mostly in economics, has examined the association between current earnings and long-term earnings without comparing it with occupation. Using longitudinal Social Security earnings matched to the SIPP, Mazumder (2001) reported that even a five-year average of earnings is a poor measure for mean lifetime earnings, which leads to an estimate of the intergenerational earnings elasticity that is biased down by 30 percent. However, not all short-term earnings are poor proxies of long-term earnings. The association between current earnings and lifetime earnings varies by age (i.e., lifecycle bias). Haider and Solon (2006) measured the association between current and lifetime earnings across age using the Health and Retirement Study (HRS) male participants' linked Social Security earnings. They found that the association is fairly strong and unbiased if the earnings used are in the early thirties and mid-forties. Using Swedish and German administrative data, respectively, Böhlmark and Lindquist (2006) and Brenner (2010) replicated Haider and Solon's study. One of the implications of these findings is that the intergenerational earnings elasticity is sensitive to the age profile of the sample.

These previous studies on proxies for long-term earnings constitute a fairly small literature, which is insufficient given the critical importance of long-term earnings. To our knowledge, there is no prior study (in either sociology or economics) that explicitly compares the predictive power of various proxy variables on long-term earnings. We seek to shed light on this issue by providing evidence about how a critical aspect of life chances is associated with key cross-sectional indicators of socioeconomic characteristics.

Research Methods

Data

We use data from the 1990 SIPP (Wave 2) matched to the Detailed Earnings Record (DER) file at the Social Security Administration (SSA). The SIPP data provide demographic, labor market, and socioeconomic characteristics of a nationally representative sample at the time of the survey. Wave 2 provides detailed information

on respondents' educational histories. The DER file matched to the SIPP is used to measure respondents' annual earnings based on their W-2 tax records from 1990 to 2009. We henceforth refer to this matched longitudinal data set as the SIPP-DER. More detailed descriptions of SSA administrative records and survey matches may be found elsewhere (see McNabb et al. 2009; Tamborini and Iams 2011).

The SIPP-DER data have several advantages for the current study. A key asset is that we can observe substantially more years of earnings for the same individual compared to the SIPP panel alone by using the linked administrative tax records. The linked earnings records also contain less measurement error than self-reported or imputed earnings in surveys (Kim and Tamborini 2014). Furthermore, after being linked to the SIPP, there is no sample attrition in tracking respondents' earnings for 20 years. In addition, the annual earnings from the SIPP-DER are not "top-coded." Those who work in a volatile labor market may have very high earnings for a few years (e.g., as the CEO of a successful start-up company) but could have very low earnings a few years later (e.g., if the company fails).

Approximately 90 percent of SIPP respondents were successfully linked to the administrative data. Even though this is a high match rate, our analyses nonetheless use a SIPP weight that adjusts for nonmatched respondents to maintain the national representation of the sample. Our analysis sample consists of native-born persons born between 1945 and 1965 whose age is between 25 and 45 in 1990, the year of survey. We limit our sample to those who reported positive earnings and occupations in the 1990 SIPP panel (Wave 2). Respondents who subsequently died over the observational period (1990 to 2009) were removed from the sample using death records contained in the administrative data (i.e., the "Numident" file). We also excluded a small number of respondents who ever received a Social Security disability benefit up to the year of the survey using linked benefit records from the SSA (i.e., Master Beneficiary Record). Individuals who never had a W-2 form submitted for them during the observation period would be excluded from the target population. The final sample sizes for our analyses are 6,066 men and 5,543 women.

Dependent Variables

The dependent variable is long-term earnings, defined as 20-year cumulative earnings. More specifically, long-term earnings refers to the sum of an individual's total taxable earnings accumulated over 20 years from all formal employment (adjusted for inflation, consumer price index for all urban consumers) as recorded by the Internal Revenue Service (IRS) for the period of 1990 to 2009. For the youngest respondents in our sample (born in 1965), the 20-year measure accounts for total earnings from age 25 to age 44. For the oldest respondents in our sample (born in 1945), it reflects total earnings from age 45 to age 64. At the median (age 35), 20-year earnings cover prime working ages from 35 to 54.

Explanatory (Proxy) Variables

We assess the predictive power of the following 10 variables on 20-year long-term earnings:

Level of education. Educational attainment is measured using four dichotomous variables: (1) less than high school, (2) some college, (3) bachelor's degree, and (4) graduate education (with high school graduates as the reference group).

Level of education and field of study (EducFoS). We disaggregate educational attainment further by college major of highest degree using the SIPP's education module (Wave 2). Eight dichotomous variables are introduced: (1) less than high school, (2) some college, (3) BA in STEM majors, (4) BA in law or business majors, (5) BA in other majors, (6) graduate degree in STEM, (7) graduate degree in law or business majors, and (8) graduate degree in other majors. High school graduates serve as the reference group.

One-digit occupation. Using the occupation recorded in Wave 2 of the SIPP, we construct six broad occupational categories (using five dichotomous variables): (1) managerial and professional specialty; (2) technical, sales, and administrative support; (3) service; (4) farming, forestry, and fishing; (5) precision production, craft, and repair; and (6) operators, fabricators, and laborers.

Three-digit occupation. We construct a detailed three-digit occupation measure based on 1990 Standard Occupational Classification (SOC) codes. Note that the observed number of detailed occupations in the 1990 SIPP differs between men (406) and women (309).

Occupational education. We construct a continuous variable that quantifies the percentage of people in the respondent's three-digit occupational category who had completed one or more years of college. For this variable, we used the "edscor90" score constructed in the Integrated Public Use Microdata Series (Ruggles et al. 2015).[3]

EGP classes. Following Morgan and Cha (2007), we created 10 EGP classes (using nine dummy variables): (1) Class I includes higher-grade professionals, administrators, and officials; managers in large industrial establishments; and large proprietors. (2) Class II includes lower-grade professionals, administrators, and officials; higher-grade technicians; managers in small industrial establishments; and supervisors of nonmanual employees. (3) Class IIIa includes routine nonmanual employees of a higher-grade (administration and commerce). (4) Class IIIb includes routine nonmanual employees of a lower-grade (sales and service). (5) Class IVa and IVb include small proprietors, artisans, et cetera with employees and small proprietors, artisans, et cetera without employees. (6) Class IVc includes farmers, smallholders, and other self-employed workers in primary production. (7) Class V includes lower-grade technicians and supervisors of manual workers. (8) Class VI includes skilled manual workers. (9) Class VIIa includes semi- and unskilled manual workers (not in agriculture). (10) Class VIIb includes agricultural and other workers in primary production.

Weeden–Grusky (WG) microclass. Following the codes of Weeden and Grusky (2012), we created their "microclass" typology.[4] WG microclass requires the expansion of the number of observations to allocate the 1990 occupation categories to the 1970 occupation categories proportionally. For this reason, the sample sizes for the analyses with the WG micro-class are larger than the original sample sizes. Because of the difference in the reported number of occupations between men and women, the total number of dichotomous variables is 124 for men and 121 for women.

One-year (1990) W-2 earnings. This is a continuous variable indicating an individual's total earnings reported to the IRS in 1990. All income variables are log-transformed and adjusted for inflation.

Three-year (1990 to 1992) W-2 earnings. This continuous variable is the sum of earnings across three years from 1990 to 1992 as reported to the IRS.

One-year (1990) annualized SIPP earnings. The SIPP respondents report their monthly earnings every four months. We combined 12 months of earnings from January 1990 to December 1990 to create self-reported SIPP earnings.

In addition to these covariates, we control for demographic variables in some models. The demographic variables include age, age squared, race and/or ethnicity (three dichotomous variables), number of children (by 1990), ever divorced (by 1990), born in the South, self-employed (in 1990), and region (nine census regions). The total number of demographic variables is 18.

Analytical Strategy

To explore the predictive power of the aforementioned variables on the 20-year long-term earnings, we estimate ordinary least squares (OLS) regression models and compare four fit statistics: R-squared (R^2), adjusted R-squared (adjusted R^2), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). R^2 quantifies the proportion of the variance of the dependent variable (y) explained by independent variables (or 1 minus the proportion of the residual variance compared to the total variation of y) as shown in Equation (1). The model that has the highest R^2 can be said to be the best model in predicting y. A well-known problem of R^2, however, is that it will never decrease when another variable is added to a regression equation. Consequently, some analyses may add many explanatory variables to acquire a high R^2 but at the expense of including covariates that are substantively dubious (Xie 1999).

$$R^2 = 1 - \frac{\sum e^2}{\sum (y - \bar{y})^2} \qquad (1)$$

To address this concern, several different approaches have been suggested. One of them is adjusted R^2, which penalizes the additional explanatory variables as follows:

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \qquad (2)$$

in which k is a number of the explanatory variables. Whether adjusted R^2 rises or falls depends on whether the improvement of R^2 as a result of additional explanatory variables is associated with a change in t-value larger than 1 (Greene 2003:34–35). Although adjusted R^2 is usually preferred to R^2 for assessing the fit of forecasting models, it has been criticized as not penalizing the loss of degrees of freedom heavily enough (Greene 2003:159–160). Adjusted R^2 has the least amount of adjustment for extra explanatory variables compared to other fit statistics (Kennedy 1998:103). Nonetheless, adjusted R^2 can be used to evaluate two models with two different sets of explanatory variables and/or across models with varying

sample sizes (Wooldridge 2016:183). Generally, a model with a higher adjusted R^2 is preferred to a model with a lower one.

Two other measures of goodness of fit that we use include the AIC and BIC. For both of these measures, the smaller the value, the better the model fits the data. As shown in Equations (3) and (4), both the AIC and BIC improve as R^2 becomes larger. In contrast to R^2, however, the values of the AIC and BIC increase as the sample size increases. Thus, the BIC or AIC for two models with different sample sizes cannot be directly compared. This latter feature is not a significant problem in our analysis because the bulk of our models are estimated using the same sample. When sample sizes differ, we compare the predictive power of different proxy variables using adjusted R^2.

$$AIC = s_y^2(1 - R^2)e^{2k/n} = 2k + n \ln \frac{\sum e^2}{n} \tag{3}$$

$$BIC = s_y^2(1 - R^2)n^{k/n} = k \ln(n) + n \ln \frac{\sum e^2}{n} \tag{4}$$

Mathematically, the AIC and BIC differ only by the extent to which the number of explanatory variables (k) is penalized. For the AIC, the measure increases by $2k$. For the BIC, the measure increases by $k\ln(n)$. Despite their mathematical similarities, the underlining theoretical assumptions between the AIC and BIC differ substantially. The BIC assumes that the true model exists among the candidates that are tested, and that the true model's dimension (k) remains fixed regardless of the sample size. The penalty associated with a larger sample size implies that the BIC guarantees the selection of the true model as the sample size grows infinitely (Vrieze 2012). This is known as the consistency property of the BIC. A problem with this property is that when the assumptions are not met, the BIC is not efficient (Burnham and Anderson 2004; Vrieze 2012). In other words, if the number of parameters in the true model increases with a larger sample size or if the true model does not exist among the candidates, then the model selected with a smaller BIC score is not necessarily the best model.

The WG microclass perspective postulates that a particular set of *gemeinschaft* groupings are organized in terms of certain related occupations, which are also said to be proxies for life chances. According to this view, the BIC should be preferred to the AIC. The model of the WG microclass is expected to yield a smaller BIC than other models. Indeed, Weeden and Grusky (2005:162) used BIC statistics in evaluating their models, stating that the BIC assesses whether "the wanton expenditure of degrees of freedom is warranted."

Unlike the BIC, the AIC does not assume that there is the true model among the candidates that are being considered. Independent of the true model, the AIC chooses whichever model minimizes the mean squared error of prediction (Vrieze 2012). Whether or not the true model exists among the candidate models, the AIC finds the optimal specification (Yang 2005). In sum, the AIC is preferred for the prediction of the outcome variables, whereas the BIC is preferred in finding the true model if it is believed to exist among the considered specifications (Vrieze 2012).

In our study, we do not make any a priori assumption that one proxy variable or any combination of proxies is definitely superior to others. Our aim is not to find

the true proxy variable. Instead, we simply seek to ascertain the best fitting model. Given the efficiency of the AIC in finding the optimal model, we mainly rely on the AIC in comparing models.[5]

One issue, however, is that the AIC can perform poorly when there are so many parameters relative to the sample size (Burnham and Anderson 2002:66). To address this problem, an additional bias-correction term may be added in the AIC that is derived from a second-order variant of the AIC. This second-order variant is called AICc and is shown in Equation (5).

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{5}$$

Note that the difference between the AIC and AICc is negligible when the sample size is large. However, if the ratio of the sample size (n) to the number of parameters (k) is small (less than 40), AICc is strongly recommended (Burnham and Anderson 2002; Vrieze 2012). In this study, we therefore rely on AICc as our primary model selection criteria. When appropriate, we also consider other goodness-of-fit statistics in our discussion.

Empirical Results

OLS Regression Models for Men

The descriptive statistics for our sample are presented in Table 1. Table 2 reports the OLS regressions on 20-year long-term (logged) earnings for men. For each model, we report the degrees of freedom, R-squared, adjusted R-squared, the AICc, and the BIC. The first model uses educational levels as the only independent covariate (i.e., 4 dichotomous variables). This model has an R-squared of 0.1884, an AICc statistic of 14,692, and a BIC statistic of 14,725. The second model also uses educational levels but introduces the major area of field of study at the tertiary level (i.e., STEM, law and business, or other). This "EducFoS" model uses eight degrees of freedom for the model. The EducFoS model has an AICc statistic of 14,582. Because the AIC statistic is clearly lower in the EducFoS model compared to the model using traditional educational level (i.e., 14,692 versus 14,582), the model incorporating the field of study is statistically preferable in terms of predicting long-term earnings. Other fitness statistics lead to that same conclusion. We therefore use educational levels differentiated by field of study as our measure of educational attainment.

The additional results are shown under panel I. "One Proxy Variable" in Table 2 refer to specifications using other variables of interest, including occupational variables (one digit or three digit), the EGP class typology, mean years of schooling in the detailed occupation schema (i.e., "Occupational Education"), and short-term earnings from the linked W-2 data (one year [1990] and three years [1990 to 1992]).

Results show that the EducFoS model outperforms all of the occupation-based models. In each occupation-based model, the AICc is higher than the EducFoS model. The lowest AICc score among the occupation models is the three-digit measure (14,625), which is still 33 points higher than in the EducFoS model. Some readers may consider this 33-point gap relatively small compared to the total AICc.

Table 1: Descriptive statistics.

		Male	Female
Sample Size		6,066	5,543
Log 20 Year (1990–2009) Cumulative Earnings	Mean	13.696	13.162
	(Variance)	(0.812)	(0.964)
Earnings: Independent Variables			
— Log 1 Year (1990) W-2 Earnings	Mean	10.594	10.101
	(Variance)	(0.755)	(0.876)
— Log 3 Year (1990–1992) W-2 Earnings	Mean	11.698	11.192
	(Variance)	(0.762)	(0.905)
— Log 1 Year (1990) SIPP Earnings	Mean	10.619	10.086
	(Variance)	(0.471)	(0.778)
Education			
— Less Than High School		8.2	5.8
— High School Graduates		33.7	34.9
— Some College		28.5	31.2
— BA		19.5	19.2
— Advanced Degree		8.7	7.7
Occupation			
— Manager/Professionals		27.2	32.4
— Technical, Service, and Administration Occupation		20.1	43.6
— Service Occupation		8.0	13.5
— Farming, Forestry, and Fishing Occupation		2.7	0.4
— Precision, Craft, and Repair Occupation		20.6	2.0
— Operators, Fabricators, and Laborers		21.5	8.1
Birth Years		1945–1965	1945–1965
Age in 1990		25–45	25–45
Race			
— Non-Hispanic Whites		85.9	83.6
— Non-Hispanic Blacks		7.7	10.8
— Hispanics		5.0	4.5
— Other Races		1.4	1.2
Married in 1990		66.0	63.2
Ever Divorced		21.2	26.3
Number of Children (Mean)		1.2	1.4
Self-Employed		10.3	4.8
Born in the South Region		26.9	29.2

Notes: Number of children is capped at 4.

Table 2: Results for OLS models of log–long-term earnings, men.

Independent Variables	Reduced df	R^2	Adjusted R^2	AICc	BIC
I. One Proxy Variable					
Education (Level of Education)	4	0.1884	0.1879	14,692	14,725
Education Level and Field of Study (EducFoS)	8	0.2040	0.2030	14,582	14,642
One-Digit Occupation	5	0.1411	0.1404	15,037	15,077
Three-Digit Occupation	405	0.3045	0.2548	14,615	17,282
EGP Class	9	0.1744	0.1731	14,806	14,873
Occupational Education	1	0.1032	0.1030	15,291	15,305
One-Year (1990) W-2 Earnings	1	0.4343	0.4342	12,497	12,510
Three-Year (1990–1992) W-2 Earnings	1	0.5823	0.5822	10,657	10,670
II. Two or More Proxy Variables					
EducFoS + 1-Digit Occ	13	0.2351	0.2335	14,350	14,444
EducFoS + 3-Digit Occ	413	0.3538	0.3066	14,188	16,905
EducFoS + EGP Class	17	0.2593	0.2572	14,164	14,284
EducFoS + 1 Year W-2 Earn	9	0.4954	0.4953	11,812	11,879
EducFoS + 1 Year W-2 Earn + 1-Digit Occ	14	0.5017	0.5005	11,753	11,853
EducFoS + 1 Year W-2 Earn + 3-Digit Occ	414	0.5494	0.5163	12,004	14,728
EducFoS + 1 Year W-2 Earn + EGP	18	0.5041	0.5026	11,732	11,859
EducFoS + 3 Year W-2 Earn	9	0.6161	0.6135	10,161	10,228
EducFoS + 3 Year W-2 Earn + 1-Digit Occ	14	0.6191	0.6182	10,123	10,224
EducFoS + 3 Year W-2 Earn + 3-Digit Occ	414	0.6553	0.6300	10,379	13,103
EducFoS + 3 Year W-2 Earn + EGP	18	0.6211	0.6200	10,099	10,226
III. With the Control of Demographic Variables					
Demographic Variables Only	18	0.0914	0.0887	15,405	15,532
Dem + EducFoS	26	0.2619	0.2606	14,145	14,326
Dem + 1-Digit Occupation	23	0.2020	0.1989	14,627	14,788
Dem + 3-Digit Occupation	423	0.3426	0.2933	14,315	17,097
Dem + EGP Class	27	0.2229	0.2195	14,474	14,662
Dem + Occupational Education	19	0.1771	0.1745	14,806	14,940
Dem + EducFoS + 1-Digit Occupation	31	0.2869	0.2833	13,961	14,175
Dem + EducFoS + 3-Digit Occupation	431	0.3883	0.3415	13,897	16,730
Dem + EducFoS + EGP Class	35	0.2998	0.2957	13,859	14,100
Dem + EducFoS + Occupational Education	27	0.2777	0.2745	14,031	14,218
Dem + EducFoS + 1 Yr W-2 Earn	27	0.5178	0.5157	11,579	11,767
Dem + EducFoS + 1 Yr W-2 Earn + 1-Digit Occ	32	0.5227	0.5202	11,527	11,749
Dem + EducFoS + 1 Yr W-2 Earn + 3-Digit Occ	432	0.5684	0.5353	11,783	14,622
Dem + EducFoS + 1 Yr W-2 Earn + EGP	36	0.5247	0.5218	11,511	11,759
Dem + EducFoS + 3 Yr W-2 Earn	27	0.6362	0.6346	9,870	10,058
Dem + EducFoS + 3 Yr W-2 Earn + 1-Digit Occ	32	0.6370	0.6369	9,836	10,057
Dem + EducFoS + 3 Yr W-2 Earn + 3-Digit Occ	432	0.6733	0.6483	10,093	12,933
Dem + EducFoS + 3 Yr W-2 Earn + EGP	36	0.6405	0.6383	9,817	10,065

Notes: Data source: SIPP-DER. Sample size is 6,066 for all analyses. Demographic variables include age, age squared, race (black, Hispanic, other versus white), whether ever divorced as of 1990, and whether married in 1990.

Recall, however, that the absolute size of the AICc statistic does not convey any substantive meaning. We compute the likelihood that the EducFoS model fits better than the three-digit occupation model by using the Akaike weights (Burnham and Anderson 2002:74–81).[6] Among the education- and class-based one-proxy-variable models, the chance that the three-digit occupation model is a better fit than the EducFoS model is extremely low.

The *R*-squared of 0.3045 for the three-digit occupation model is large, but that high value in part derives from the much greater expenditure of degrees of freedom (i.e., 405, as shown in Table 2). These findings imply that the three-digit occupational model is "over-fitting" in the sense of adding an excessive number of independent variables (Raffalovich et al. 2008). The BIC statistic reinforces that conclusion.

The specifications using cross-sectional earnings provide additional insights. The model using one year of administrative earnings (i.e., earnings in 1990 as recorded in the W-2 tax form) has an AICc statistic of 12,497. This value is notably lower than any of the AICc statistics mentioned above, including for the EducFoS model. Thus, just one year of earnings predicts an individual's subsequent 20-year earnings better than do education or class variables. The model using three years of administrative earnings (1990 to 1992) predicts more than half of the variation in a person's long-term earnings (i.e., the *R*-squared is 0.582) and yields an even greater relative drop in the AICc statistic to 10,657.

Results shown under the second heading, "Two or More Proxy Variables," refer to models using various combinations of independent variables. In general, these results show that adding the occupational-based variables to educational attainment (i.e., the EducFoS model) improves the predictive power of the model. Among the occupational-based class models, the EGP typology seems to have the most predictive power in terms of the AICc statistic and is followed by the three-digit occupation model.

The models that include cross-sectional earnings have decidedly lower AICc statistics, suggesting higher predictive power. After adding one year of W-2 earnings to educational attainment, the AICc statistic falls to 11,812. The fit of this model is improved when one-digit occupation or the EGP class is added. However, the addition of the three-digit occupation is associated with a higher AICc statistic, which indicates lower predictive power.

The results reported in the bottom panel of Table 2 (i.e., heading III) refer to models that add demographic controls. The model with demographic characteristics and educational attainment fits quite well, with an AICc statistic of 14,145. This is clearly an improvement relative to the EducFoS-only model (14,582). It also predicts long-term earnings better than models with educational attainment and any of the occupation variables.

Although the demography-educational-attainment model fits the data better than any of the models combining demographic and occupational-based variables, the addition of occupation-based variables on top of the demography-educational-attainment model also improves the fitness substantially. For example, the AICc statistic of the models "Dem + EduFoS + Occupation" and "Dem + EducFoS + EGP Class" are lower than the AICc statistic for the "demography-educational-attainment" model. This implies that occupation accounts for some of the hetero-

geneity in long-term earnings that is not measured by demographic and educational variables.

The model that adds one year of administrative earnings into the demography-educational-attainment model shows a markedly lower AICc (to 11,579). Adding in either one-digit occupation or EGP improves the fit even after controlling for one-year administrative earnings. The model with three-digit occupation does not improve the fit statistic as much as the models with one-digit occupation. That is, once information on earnings is controlled for, the value of detailed occupation as an additional control is small in accounting for long-term earnings. Not surprisingly, the best-fitting models in Table 2 include the variable summing up three years of administrative earnings around the beginning of the observation window.

OLS Regression Models for Women

Table 3 provides the results for women, which parallel the models used in Table 2. In general, the results are fairly similar to those for men. Demographic and educational variables have comparatively high predictive power, whereas models using three-digit occupation have high AICc statistics when included in models with additional covariates. Models with cross-sectional earnings have higher predictive power than models without those variables.

However, there are some noteworthy gender differences. Firstly, demographic variables are somewhat less predictive of long-term earnings for women than for men. Secondly, educational field of study is slightly less predictive of long-term earnings for women than for men. Thirdly, occupational-based variables are less predictive of long-term earnings for women than for men. Lastly, cross-sectional earnings are less predictive of long-term earnings for women than for men.

Interestingly, for women, models with occupation-based variables (such as three-digit occupation or the EGP typology) have better fit (AICc of three-digit occupation = 14,755; AICc of EGP = 14,740) than models with educational attainment (AICc = 14,818). These results suggest that, on average, the long-term earnings of female workers could be more affected by gender-based job segregation or by exogenous factors relating to family and household circumstances that do not affect male workers as much. The gender difference of the predictive power of three proxy variables such as occupation, education, and short-term earnings might relate to women's labor force participation rate for the cohort that we analyze (i.e., born in 1945 to 1965). For this cohort, women's long-term earnings can be better predicted by the endogenous labor market variables, such as occupation, compared to exogenous variables to the labor market, such as demographic characteristics and educational attainment.

Models with Self-Reported Earnings

To recap, the models with one-year and three-year administrative earnings presented in Tables 2 and 3 had higher predictive power than other models. Although the number of studies utilizing administrative W-2 earnings is increasing, these restricted-use data are still fairly difficult to access. The available information for most studies will be self-reported earnings. Furthermore, the high predictive power

Table 3: Results for OLS models of log–long-term earnings, women.

Independent Variables	Reduced df	R^2	Adjusted R^2	AICc	BIC
I. One Proxy Variable					
Education (Level of Education)	4	0.1192	0.1186	14,832	14,865
Education Level and Field of Study (EducFoS)	8	0.1227	0.1214	14,818	14,878
One-Digit Occupation	5	0.1160	0.1152	14,854	14,894
Three-Digit Occupation	308	0.2266	0.1811	14,755	16,765
EGP Class	9	0.1352	0.1338	14,740	14,807
Occupational Education	1	0.0749	0.0747	15,098	15,112
One-Year (1990) W-2 Earnings	1	0.3830	0.3829	12,853	12,866
Three-Year (1990–1992) W-2 Earnings	1	0.5147	0.5146	11,522	11,535
II. Two or More Proxy Variables					
EducFoS + 1-Digit Occ	13	0.1627	0.1608	14,569	14,662
EducFoS + 3-Digit Occ	316	0.2575	0.2126	14,547	16,608
EducFoS + EGP Class	17	0.1816	0.1791	14,451	14,570
EducFoS + 1 Year W-2 Earn	9	0.4126	0.4117	12,596	12,662
EducFoS + 1 Year W-2 Earn + 1-Digit Occ	14	0.4172	0.4158	12,563	12,662
EducFoS + 1 Year W-2 Earn + 3-Digit Occ	317	0.4570	0.4241	12,815	14,882
EducFoS + 1 Year W-2 Earn + EGP	18	0.4178	0.4159	12,565	12,691
EducFoS + 3 Year W-2 Earn	9	0.5293	0.5282	11,369	11,435
EducFoS + 3 Year W-2 Earn + 1-Digit Occ	14	0.5326	0.5314	11,340	11,439
EducFoS + 3 Year W-2 Earn + 3-Digit Occ	317	0.5635	0.5370	11,605	13,672
EducFoS + 3 Year W-2 Earn + EGP	18	0.5322	0.5307	11,352	11,478
III. With the Control of Demographic Variables					
Demographic Variables Only	18	0.0655	0.0624	15,188	15,314
Dem + EducFoS	26	0.1632	0.1592	14,592	14,771
Dem + 1-Digit Occupation	23	0.1595	0.1560	14,611	14,770
Dem + 3-Digit Occupation	326	0.2564	0.2099	14,578	16,702
Dem + EGP Class	27	0.1689	0.1648	14,556	14,741
Dem + Occupational Education	19	0.1268	0.1238	14,814	14,947
Dem + EducFoS + 1-Digit Occupation	31	0.1994	0.1949	14,357	14,569
Dem + EducFoS + 3-Digit Occupation	334	0.2827	0.2367	14,397	16,572
Dem + EducFoS + EGP Class	35	0.2059	0.2009	14,320	14,558
Dem + EducFoS + Occupational Education	27	0.1775	0.1735	14,498	14,683
Dem + EducFoS + 1 Yr W-2 Earn	27	0.4239	0.4211	12,525	12,710
Dem + EducFoS + 1 Yr W-2 Earn + 1-Digit Occ	32	0.4292	0.4259	12,483	12,702
Dem + EducFoS + 1 Yr W-2 Earn + 3-Digit Occ	335	0.4696	0.4355	12,726	14,907
Dem + EducFoS + 1 Yr W-2 Earn + EGP	36	0.4302	0.4265	12,482	12,726
Dem + EducFoS + 3 Yr W-2 Earn	27	0.5421	0.5398	11,252	11,438
Dem + EducFoS + 3 Yr W-2 Earn + 1-Digit Occ	32	0.5458	0.5432	11,217	11,435
Dem + EducFoS + 3 Yr W-2 Earn + 3-Digit Occ	335	0.5771	0.5499	11,470	13,651
Dem + EducFoS + 3 Yr W-2 Earn + EGP	36	0.5458	0.5429	11,224	11,469

Notes: Data source: SIPP-DER. Sample size is 5,543 for all analyses. Demographic variables include age, age squared, race (black, Hispanic, other versus white), whether ever divorced as of 1990, and whether married in 1990.

Table 4: Results for OLS models of log–long-term earnings using samples limited to those who reported SIPP self-reported earnings.

Independent Variables	Reduced df	R^2	Adjusted R^2	AICc	BIC
I. Male					
One-Year (1990) Annualized SIPP Earnings	1	0.3327	0.3326	13,792	13,806
Three-Digit Occupation	402	0.2993	0.2466	14,935	17,557
EducFoS + 1 Year SIPP Earn	9	0.3801	0.3791	13,385	13,451
EducFoS + 1 Year SIPP Earn + 1-Digit Occ	14	0.3877	0.3862	13,325	13,424
EducFoS + 1 Year SIPP Earn + 3-Digit Occ	411	0.4641	0.4228	13,416	16,094
EducFoS + 1 Year SIPP Earn + EGP	18	0.4044	0.4025	13,173	13,300
Dem + 1 Year SIPP Earn	19	0.3740	0.3719	13,461	13,594
Dem + 3-Digit Occ	420	0.3513	0.3001	14,534	17,270
Dem + EducFoS + 1 Yr SIPP Earn	27	0.4195	0.4168	13,043	13,230
Dem + EducFoS + 1 Yr SIPP Earn + 1-Digit Occ	32	0.4259	0.4227	12,990	13,210
Dem + EducFoS + 1 Yr SIPP Earn + 3-Digit Occ	344	0.4641	0.4228	13,352	16,094
Dem + EducFoS + 1 Yr SIPP Earn + EGP	36	0.4307	0.4271	12,950	13,197
II. Female					
One-Year (1990) Annualized SIPP Earnings	1	0.3087	0.3086	13,500	13,513
Three-Digit Occupation	307	0.2340	0.1870	14,694	16,682
EducFoS + 1 Year SIPP Earn	9	0.3344	0.3333	13,314	13,380
EducFoS + 1 Year SIPP Earn + 1-Digit Occ	14	0.3414	0.3397	13,268	13,367
EducFoS + 1 Year SIPP Earn + 3-Digit Occ	316	0.3889	0.3502	13,515	15,560
EducFoS + 1 Year SIPP Earn + EGP	18	0.3500	0.3478	13,206	13,331
Dem + 1 Year SIPP Earn	19	0.3273	0.3249	13,391	13,522
Dem + 3-Digit Occ	325	0.2634	0.2153	14,527	16,628
Dem + EducFoS + 1 Yr SIPP Earn	27	0.3542	0.3509	13,190	13,374
Dem + EducFoS + 1 Yr SIPP Earn + 1-Digit Occ	32	0.3610	0.3572	13,143	13,361
Dem + EducFoS + 1 Yr SIPP Earn + 3-Digit Occ	334	0.4046	0.3646	13,417	15,576
Dem + EducFoS + 1 Yr SIPP Earn + EGP	36	0.3631	0.3587	13,135	13,378

Notes: Data source: SIPP-DER. Sample size for men is 5,745 and for women is 5,308. Demographic variables include age, age squared, race (black, Hispanic, other versus white), whether ever divorced as of 1990, and whether married in 1990.

of the one-year and three-year administrative earnings is at least partially driven by the endogenous nature of these variables with the 20-year cumulative earnings.

Consequently, we explore the predictive power of self-reported annual SIPP earnings in Table 4. Note that the sample sizes of Table 4 differ from those of Tables 2 and 3 because we needed to limit the sample for this analysis to respondents who reported earnings over the entire calendar year of 1990 (encompassing multiple waves of the SIPP panel). Also recall that the AICc statistic is not comparable when the sample sizes differ. Thus, the results in Table 4 cannot be directly compared with those in Tables 2 and 3.

Overall, our results demonstrate that self-reported annual earnings in 1990 are a much stronger proxy for subsequent 20-year long-term earnings than three-digit

occupation in 1990. This is consistent with our findings using one-year administrative earnings in 1990. All occupational codes except three-digit occupation improve the model fit when they are added to the self-reported earnings. For both men and women, the three-digit occupation code worsens the model fit when it is additionally controlled for on top of self-reported annual earnings. Among all occupational codes, the EGP code improves the model fit the most when it is added to self-reported annual earnings and other covariates.

The Predictive Power of Weeden–Grusky Microclasses

The final portion of our analysis considers the WG microclass scheme, which collapses three-digit occupations into more than 100 microclasses. This process requires imputation and results in an artificial increase in sample size. Thus, the AICc of the WG microclass cannot be compared with other models in Tables 2, 3, and 4.

The main question here is whether the WG microclass scheme accounts for long-term earnings better than three-digit occupation or one-year self-reported SIPP earnings. Our results demonstrate that the WG microclass scheme has smaller predictive power than three-digit occupation or one-year self-reported earnings in accounting for 20-year long-term earnings. All four fit statistics reported in Table 5 indicate that the remaining errors after the introduction of WG microclass in a model are larger than after the introduction of one-year SIPP earnings. Comparing the WG microclass scheme to three-digit occupation, R^2, adjusted R^2, and the AICc statistic shows that three-digit occupation is a better proxy of long-term earnings than WG microclass. The result does not vary by gender.

Robustness Checks

Some may wonder whether the results we report here hold across age range changes. To address this concern about life cycle bias (Böhlmark and Lindquist 2006; Brenner 2010; Haider and Solon 2006), we ran the same models in Tables 2 and 3 using age-stratified samples (25 to 34 and 35 to 45 in 1990).[7] We find that the relative predictive power of the proxy variables remain the same in both groups regardless of gender in terms of the AICc fitness statistic. Interestingly, the predictive power of three-digit occupation on long-term earnings seems to be remarkably similar between age 25 to 34 men (adjusted $R^2 = 0.260$) and age 35 to 45 men (adjusted $R^2 = 0.261$). The same consistency is evident among women. We also limit our sample to age 30 to 39 so that their 20-year long-term earnings cover earnings from age 30 to 50 for the youngest and age 39 to 59 for the oldest. Our main results are not altered. An implication of this finding is that the transitory fluctuation of the detailed occupation over age in accounting for lifetime earnings is smaller than for the short-term earnings.

In another robustness check, we limit our base sample to the full-time workers in the 1990 SIPP survey (i.e., those who work 35 or more hours per week). This sensitivity test yields estimates that are consistent with our main results. In another robustness check, we exclude the top 1 percent and the bottom 1 percent from our sample and re-estimated our models to consider whether our findings are driven

Table 5: Results for OLS models of log–long-term earnings using the imputed sample to create Weeden–Grusky microclasses.

Independent Variables	Reduced df	R^2	Adjusted R^2	AICc	BIC
I. Male					
WG Microclass	123	0.2380	0.2342	58,282	59,288
One-Digit Occupation	5	0.1452	0.1450	60,890	60,939
Three-Digit Occupation	397	0.3135	0.3024	56,256	59,473
One-Year (1990) W-2 Earnings	1	0.4354	0.4354	50,610	50,626
Three-Year (1990–1992) W-2 Earnings	1	0.5795	0.5794	43,314	43,330
One-Year (1990) Annualized SIPP Earnings	1	0.3784	0.3783	52,994	53,010
II. Female					
WG Microclass	120	0.1655	0.1589	40,250	41,174
One-Digit Occupation	5	0.1136	0.1133	40,947	40,993
Three-Digit Occupation	303	0.2306	0.2152	39,378	41,688
One-Year (1990) W-2 Earnings	1	0.3951	0.3950	35,066	35,081
Three-Year (1990-1992) W-2 Earnings	1	0.5285	0.5285	31,235	31,250
One-Year (1990) Annualized SIPP Earnings	1	0.3244	0.3244	36,764	36,779

Notes: Data source: SIPP-DER. Samples are limited to those who reported positive earnings for both the SIPP and their W-2. Sample size for men is 24,768 and for women is 15,372. Demographic variables include age, age squared, race (black, Hispanic, other versus white), whether ever divorced as of 1990, and whether married in 1990.

by a small number of the extremely high long-term earnings. Again, the results are consistent.

Discussion and Conclusion

This study investigated the predictive power of a set of cross-sectional predictors on 20-year cumulative long-term earnings using data that link national survey data in 1990 (SIPP) with longitudinal W-2 earnings records. Overall, the findings advance knowledge about the determinants of long-term earnings and provide insights into the reliability of measures often used to measure important concepts such as intergenerational mobility. Our results indicate that even one year of cross-sectional earnings is more predictive of long-term earnings than demography, education, or occupation-based variables. Reinforcing this view, the fit statistics of self-reported annual earnings also suggest relatively high predictive power compared to other models. Understanding the sources of cross-sectional earnings inequality may therefore be more indicative of the sources of life chances than the occupation-based class variables that are commonly considered in the sociological literature.

The findings help clarify the relevance of occupation observed in a single year for subsequent long-term earnings. Contrary to common assumptions in the literature, the occupation-based independent variables observed in a cross section have less notable net effects on long-term earnings than other variables we examined. Despite the large expenditures in degrees of freedom, the predictive powers of three-digit

occupation and the WG typology are not higher than those of education and short-term earnings variables. Compared to the Weeden–Grusky microclass, three-digit occupation performs better for both genders regardless of which fit statistics are considered.[8] Adding occupational variables to annual earnings sometimes slightly improves the model fitness.

At the same time, class-based classifications and occupational codes, including one-digit occupation, three-digit occupation, and the EGP class, improve model fitness when they are added to demography and educational attainment variables. This implies that the broad occupational classification explains the additional dimension of long-term earnings that is not captured by annual earnings. In particular, the employment relation of the EGP class seems to be a valuable additional dimension. Among broader class-based typology, the EGP class seems better than one-digit occupation and occupational education in terms of the predictive power of long-term earnings.

The predictive power of occupation on long-term earnings varies substantially by gender in contrast to the predictive power of annual earnings, which is fairly consistent by gender. This implies that within-occupational inequality in accounting for long-term earnings is much larger for women than for men. For women, labor supply issues are likely to be more important than for men with regard to long-term earnings. To some extent, this finding is consistent with Hauser and Warren's (1997) critique of occupational socioeconomic indexes. On average, women's occupational standing does not have the same implication as do men's. Annual earnings are therefore a more reliable and consistent proxy of long-term earnings than detailed occupation for each gender.

Our AICc statistics imply that using hundreds of dichotomous variables to indicate detailed occupation as observed in a cross-section may be "over-fitting" the model at least in regard to predicting long-term earnings. The problem of overfitting can be attenuated if sample size is large enough. As shown in Table 5, when we used the artificially augmented sample, detailed occupational codes start to perform better than one-digit occupational codes in terms of the AICc. A practical implication of this finding is that when the annual earnings variable is missing, detailed occupation can be a decent proxy of lifetime earnings as long as the sample size is fairly large. The rising nonresponse rate of self-reported earnings in surveys (Mouw and Kalleberg 2010) is probably another reason why detailed occupation is still practical when annual earnings information is missing.

Some may argue that the smaller recall error for occupation than for earnings is another reason why occupation may be preferred to earnings. It is true that there may be nontrivial measurement error in self-reported earnings (Kim and Tamborini 2012; Kim and Tamborini 2014). However, researchers should be aware that occupational coding is not error free either (Fisher and Houseworth 2013). In most national surveys conducted by government agencies, interviewers ask open-ended questions about job duties, and professional coders recode them in the data editing process. Previous studies have consistently uncovered nontrivial occupational measurement errors (e.g., Belloni et al. 2016; Mathiowetz 1992; Mellow and Sider 1983). In a recent study, Belloni et al. (2016) suggest that the disagreement rate between coders is as high as 40 percent even for one-digit occupation. According to one estimate (Speer

2016), incorrect occupational coding can lead to about 90 percent overestimation of intergenerational mobility.

Our study also underscores the significance of educational attainment for an individual's long-term earnings. Both the highest level attained and field of study (among the college educated) matter for a person's long-term earnings. Our statistics show that educational attainment for men performs slightly better than does aggregate occupation when horizontal stratification in higher education is accounted for by using field of study. This result is consistent with recent studies that emphasize the growing significance of the horizontal stratification within the same level of education in determining life chances (Gerber and Cheung 2008; Kim et al. 2015; Ma and Savas 2014).

The study has several limitations worth noting. First, estimates of relationships based on different birth cohorts may vary from those presented here. The examination of more recent cohorts, as they age over the life course, would also be of interest. Second, although it covers a substantially long period of time, 20-year cumulative earnings is not necessarily equal to lifetime earnings. We cannot rule out a possibility that the predictive power of proxy variables might change in regard to 40-year lifetime earnings. Third, the covariates examined in this study were largely measured at the time of the SIPP survey and thus do not capture changes over time.

Notwithstanding these and other limitations, our findings provide new evidence on the association between proxy variables and lifetime earnings by elucidating the relationships between an individual's demographic, educational, and labor market characteristics and his or her cumulative earnings over a 20-year time period. We emphasize that our results should not be taken as evidence that sociologists should lessen their interest in occupational structure or intergenerational occupational mobility. Instead, what our results clearly call into question is the claim that occupational mobility is superior to earnings mobility in measuring intergenerational socioeconomic mobility because occupation is a better proxy of long-term earnings than short-term earnings. Studies on intergenerational occupational mobility or on the effects of occupational structure in accounting for rising income inequality are still valuable not because occupations (or occupation-based class schemes) measure life chances better than short-term earnings but because occupation reflects important dimensions of life chances that are not captured by short-term earnings. Our investigation also demonstrates that more refined empirical analyses of labor market outcomes are possible as more administrative data become available.

Notes

1 In a somewhat similar vein, Wright's (2005) typology seeks to consider both life chances and location within production relations. His approach nonetheless ends up relying heavily upon occupation to classify workers and employers into different classes.

2 Estimates of the IGE tend to be larger when based on estimates of long-term earnings (for either fathers or their sons) that use data for earnings in the age range of 35 to 45 rather than in the 20s or late 50s (Black and Devereux 2011).

3 The data are available at https://usa.ipums.org/usa/.

4 The code is available at
   http://www.kimweeden.com/wp-content/uploads/work/gw2codes.zip.

5 Some readers may prefer the BIC to the AIC because the BIC is derived from the Bayesian framework. However, Burnham and Anderson (2002) demonstrate that the AIC can also be derived from the Bayesian framework. They argue that the AIC has theoretical advantages over the BIC in terms of the principle of information and a priori assumptions.

6 The Akaike weights (w_i) can be computed as $\frac{e^{-\frac{1}{2}\Delta AIC_c}}{\sum_{r=1}^{R} e^{-\frac{1}{2}\Delta AIC_{c_r}}}$ where r refers to model r. The ratio of the Akaike weights is computed as w_i/w_j (Burnham and Anderson 2002:74–81).

7 The results of the robustness checks are not reported here but may be obtained from the authors upon request.

8 Occupational licensing has been a major theoretical justification for this approach, but recent research casts doubt on the assumption that licensing has significant effects on either wages or employment (Redbird 2017).

References

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy* 87:1153–89. https://www.jstor.org/stable/1833328.

Belloni, Michele, Agar Brugiavini, Elena Meschi, and Kea Tijdens. 2016. "Measuring and Detecting Errors in Occupational Coding: an Analysis of SHARE Data." *Journal of Official Statistics* 32:917–45. https://www.degruyter.com/view/j/jos.2016.32.issue-4/jos-2016-0049/jos-2016-0049.xml.

Bernanke, Ben S. 1984. "Permanent Income, Liquidity, and Expenditure on Automobiles: Evidence From Panel Data." *The Quarterly Journal of Economics* 99:587–614. https://academic.oup.com/qje/article-lookup/doi/10.2307/1885966.

Black, Sandra E., and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics* 4:1487–1541. https://www.sciencedirect.com/science/article/pii/S0169721811024142.

Blau, Peter, and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York, NY: The Free Press.

Böhlmark, Anders, and Matthew J. Lindquist. 2006. "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden." *Journal of Labor Economics* 24:879–96. http://www.journals.uchicago.edu/doi/10.1086/506489.

Brady, David, Marco Giesselmann, Ulrich Kohler, and Anke Radenacker. 2017. "How to Measure and Proxy Permanent Income: Evidence from Germany and the U.S." *The Journal of Economic Inequality,* first published on December 16, 2017 as doi:10.1007/s10888-017-9363-9. https://doi.org/10.1007/s10888-017-9363-9.

Breen, Richard. 2005. "Foundations of a Neo-Weberian Class Analysis." In *Approaches to Class Analysis*, edited by E. O. Wright, pp. 31–50, New York, NY. Cambridge University Press.

Breen, Richard, and Jan O. Jonsson. 2005. "Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility." *Annual*

Review of Sociology 31:223–43. http://www.annualreviews.org/doi/10.1146/annurev.soc.31.041304.122232.

Brenner, Jan. 2010. "Life-cycle Variations in the Association between Current and Lifetime Earnings: Evidence for German Natives and Guest Workers." *Labour Economics* 17:392–406. https://www.sciencedirect.com/science/article/pii/S0927537109000608.

Burnham, Kenneth P., and David Raymond Anderson. 2002. *Model Selection and Multimodel Inference : A Practical Information-Theoretic Approach*. New York, NY: Springer-Verlag.

Burnham, Kenneth P., and David R. Anderson. 2004. "Multimodel Inference." *Sociological Methods & Research* 33:261–304. http://journals.sagepub.com/doi/10.1177/0049124104268644.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129:1553–623. https://academic.oup.com/qje/article-lookup/doi/10.1093/qje/qju022.

Cooper, Daniel. 2013. "The Effect of Unemployment Duration on Future Earnings and Other Outcomes." *Working Paper No. 13-8, FRB of Boston.* Retrieved January 5, 2018. http://www.ssrn.com/abstract=2366542.

Davis, Steven J., Till von Wachter, Robert E. Hall, Richard Rogerson, and Till von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity* pp. 1–72. https://www.brookings.edu/bpea-articles/recessions-and-the-costs-of-job-loss/https://www.jstor.org/stable/41473597.

DeJuan, Joseph P., John J. Seater, and Tony S. Wirjanto. 2006. "Testing the Permanent-Income Hypothesis: New Evidence from West-German States (Länder)." *Empirical Economics* 31:613–29. http://link.springer.com/10.1007/s00181-005-0035-4.

DiPrete, Thomas A., and Gregory M. Eirich. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32:271–97. https://doi.org/10.1146/annurev.soc.32.061604.123127.

Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford, United Kingdom: Oxford University Press.

Erikson, Robert, John H. Goldthorpe, and Lucienne Portocarero. 1979. "Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden." *The British Journal of Sociology* 30:415. http://www.jstor.org/stable/589632?origin=crossref.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York, NY: Academic Press.

Fischer, Claude S., and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York, NY: Russell Sage Foundation.

Fisher, Jonathan D., and Christina A. Houseworth. 2013. "Occupation Inflation in the Current Population Survey." *Journal of Economic and Social Measurement* 38:243–61. https://content.iospress.com/articles/journal-of-economic-and-social-measurement/jem00377.

Flavin, Marjorie A. 1981. "The Adjustment of Consumption to Changing Expectations About Future Income." *Journal of Political Economy* 89:974–1009. https://www.jstor.org/stable/1830816.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.

Gerber, Theodore P., and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education: Forms, Explanations, and Implications." *Annual Review of Sociology* 34:299–318. https://doi.org/10.1146/annurev.soc.34.040507.134604.

Goldberger, Arthur S. 1989. "Economic and Mechanical Models of Intergenerational Transmission." *The American Economic Review* 79:504–13. http://www.jstor.org/stable/1806859.

Goldthorpe, John H. 2012. "Back to Class and Status: Or Why a Sociological View of Social Inequality Should Be Reasserted." *Revista Española de Investigaciones Sociológicas* 137:201–16. https://doi.org/10.5477/cis/reis.137.201.

Greene, William H. 2003. *Econometric Analysis*. Upper Saddle River, NJ: Pearson Education.

Grusky, David B. 2001. "The Past, Present, and Future of Social Inequality." In *Sociological Stratification in Sociological Perspective*, edited by D. B. Grusky, pp. 3–51. Boulder, CO: Westview Press.

Grusky, David B. 2005. "Foundations of a Neo-Durkheimian Class Analysis." In *Approaches to Class Analysis*, edited by E. O. Wright, pp. 51–81. New York, NY: Cambridge University Press.

Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96:1308–20. https://doi.org/10.1257/aer.96.4.1308.

Hauser, Robert M. 2010. "Intergenerational Economic Mobility in the United States Measures, Differentials, and Trends." *Euramerica* 40:635–81.

Hauser, Robert M., and John Robert Warren. 1997. "Socioeconomic Indexes for Occupations: A Review, Update, and Critique." *Sociological Methodology* 27:177–298. https://doi.org/10.1111/1467-9531.271028.

Hout, Michael. 2008. "How Class Works in Popular Conception: Most Americans Identify with the Class Their Income, Occupation, and Education Implies for Them." In *Social Class: How Does It Work?*, edited by A. Lareau and D. Conlye, pp. 25–64. New York, NY: Russell Sage Foundation.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38:379–400. https://doi.org/10.1146/annurev.soc.012809.102503.

Hout, Michael, and Thomas DiPrete. 2006. "What Have We Learned? RC28's Contributions to Knowledge about Social Stratification." *Research in Social Stratification and Mobility* 24:1–20. https://doi.org/10.1016/j.rssm.2005.10.001.

Ishida, Hiroshi, Walter Muller, and John M. Ridge. 1995. "Class Origin, Class Destination, and Education: A Cross-National Study of Ten Industrial Nations." *American Journal of Sociology* 101:145–93. https://www.jstor.org/stable/2782508.

Jonsson, Jan O., David B. Grusky, Matthew Di Carlo, Reinhard Pollak, and Mary C. Brinton. 2009. "Microclass Mobility: Social Reproduction in Four Countries." *American Journal of Sociology* 114:977–1036. http://www.journals.uchicago.edu/doi/10.1086/596566.

Kawachi, Ichiro, Bruce P. Kennedy, Kimberly Lochner, and Deborah Prothrow-Stith. 1997. "Social Capital, Income Inequality, and Mortality." *American Journal of Public Health* 87:1491–98. https://doi.org/10.2105/AJPH.87.9.1491.

Kennedy, Peter. 1998. *A Guide to Econometrics*. Boston, MA: MIT Press.

Kim, ChangHwan, and Christopher R. Tamborini. 2012. "Do Survey Data Estimate Earnings Inequality Correctly? Measurement Errors Among Black and White Male Workers." *Social Forces* 90:1157–81. http://sf.oxfordjournals.org/cgi/doi/10.1093/sf/sor042.

Kim, ChangHwan and Christopher R Tamborini. 2014. "Response Error in Earnings: An Analysis of the Survey of Income and Program Participation Matched With Administrative Data." *Sociological Methods & Research* 43:39–72. http://smr.sagepub.com/cgi/content/abstract/43/1/39.

Kim, ChangHwan, Christopher R. Tamborini, and Arthur Sakamoto. 2015. "Field of Study in College and Lifetime Earnings in the United States." *Sociology of Education* 88:320–39. http://soe.sagepub.com/cgi/doi/10.1177/0038040715602132.

Ma, Yingyi, and Gokhan Savas. 2014. "Which Is More Consequential: Fields of Study or Institutional Selectivity?" *The Review of Higher Education* 37:221–47. https://muse.jhu.edu/article/525235.

Mathiowetz, Nancy A. 1992. "Errors in Reports of Occupation." *The Public Opinion Quarterly* 56:352–55. https://www.jstor.org/stable/2749157.

Mazumder, Bhashkar. 2001. "The Mis-Measurement of Permanent Earnings: New Evidence from Social Security Earnings Data." *Working Paper 2001-2024, Federal Reserve Bank of Chicago.* Retrieved October 28, 2017. http://www.ssrn.com/abstract=296131.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *Review of Economics and Statistics* 87:235–55. http://www.mitpressjournals.org/doi/10.1162/0034653053970249.

McNabb, Jennifer, David Timmons, Jae Song, and Carolyn Puckett. 2009. "Uses of Administrative Data at the Social Security Administration." *Social Security Bulletin* 69:75–84.

Mellow, Wesley, and Hal Sider. 1983. "Accuracy of Response in Labor Market Surveys: Evidence and Implications." *Journal of Labor Economics* 1:331–44. https://www.jstor.org/stable/2534858.

Morgan, Stephen L., and Youngjoo Cha. 2007. "Rent and the Evolution of Inequality in Late Industrial United States." *American Behavioral Scientist* 50:677–701. http://journals.sagepub.com/doi/10.1177/0002764206295018.

Mouw, Ted, and Arne L. Kalleberg. 2010. "Occupations and the Structure of Wage Inequality in the United States, 1980s to 2000s." *American Sociological Review* 75:402–431. http://journals.sagepub.com/doi/10.1177/0003122410363564.

Raffalovich, Lawrence E., Glenn D. Deane, David Armstrong, and Hui-Shien Tsao. 2008. "Model Selection Procedures in Social Research: Monte-Carlo Simulation Results." *Journal of Applied Statistics* 35:1093–114. https://www.tandfonline.com/doi/full/10.1080/03081070802203959.

Rainwater, Lee. 1974. *What Money Buys: Inequality and the Social Meanings of Income*. New York, NY: Basic Books.

Redbird, Beth. 2017. "The New Closed Shop? The Economic and Structural Effects of Occupational Licensure." *American Sociological Review* 82:600–624. `http://journals.sagepub.com/doi/10.1177/0003122417706463`.

Ruggles, Steve, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0 [dataset]*. Minneapolis, MN: University of Minnesota.

Sakamoto, Arthur, and Changhwan Kim. 2014. "Bringing Productivity Back In: Rising Inequality and Economic Rents in the U.S. Manufacturing Sector, 1971 to 2001." *Sociological Quarterly* 55:282–314. `http://doi.wiley.com/10.1111/tsq.12055`.

Sakamoto, Arthur, Christopher R. Tamborini, and ChangHwan Kim. 2018. "Long-Term Earnings Differentials Between African American and White Men by Educational Level." *Population Research and Policy Review* 37:91–116. `http://link.springer.com/10.1007/s11113-017-9453-1`.

Sakamoto, Arthur, and Sharron Xuanren Wang. 2017. "Occupational and Organizational Effects on Wages among College-educated Workers in 2003 and 2010." *Social Currents* 4:175–95. `http://journals.sagepub.com/doi/10.1177/2329496516651643`.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." *Handbook of Labor Economics* 3:1761–800. `https://www.sciencedirect.com/science/article/pii/S1573446399030102`.

Speer, Jamin D. 2016. "How Bad is Occupational Coding Error? A Task-Based Approach." *Economics Letters* 141:166–68. `https://www.sciencedirect.com/science/article/pii/S0165176516300544`.

Stronks, Karien, H., Van De Mheen, Johannes van den Bos, and J. P. Mackenbach. 1997. "The Interrelationship between Income, Health and Employment Status." *International Journal of Epidemiology* 26:592–600. `https://doi.org/10.1093/ije/26.3.592`.

Tamborini, Christopher R., and Howard M. Iams. 2011. "Are Generation X'ers Different than Late Boomers? Family and Earnings Trends among Recent Cohorts of Women at Young Adulthood." *Population Research and Policy Review* 30:59–79. `http://link.springer.com/10.1007/s11113-010-9178-x`.

Tamborini, Christopher R., Howard M. Iams, and Gayle L. Reznik. 2012. "Women's Earnings Before and After Marital Dissolution: Evidence from Longitudinal Earnings Records Matched to Survey Data." *Journal of Family and Economic Issues* 33:69–82. `http://link.springer.com/10.1007/s10834-011-9264-1`.

Tamborini, Christopher R., ChangHwan Kim, and Arthur Sakamoto. 2015. "Education and Lifetime Earnings in the United States." *Demography* 52:1383–407. `http://link.springer.com/10.1007/s13524-015-0407-0`.

Tomaskovic-Devey, Donald, Melvin Thomas, and Kecia Johnson. 2005. "Race and the Accumulation of Human Capital across the Career: A Theoretical Model and Fixed-Effects Application." *American Journal of Sociology* 111:58–89. `https://doi.org/10.1086/431779`.

Torche, Florencia. 2015. "Analyses of Intergenerational Mobility." *The ANNALS of the American Academy of Political and Social Science* 657:37–62. http://journals.sagepub.com/doi/10.1177/0002716214547476.

Vrieze, Scott I. 2012. "Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)." *Psychological Methods* 17:228–43. http://doi.apa.org/getdoi.cfm?doi=10.1037/a0027127.

Weber, Max. 1978. *Economy and Society*. Berkeley, CA: University of California Press.

Weeden, Kim A., and David B. Grusky. 2005. "The Case for a New Class Map." *American Journal of Sociology* 111:141–212. https://doi.org/10.1086/428815.

Weeden, Kim A., and David B. Grusky. 2012. "The Three Worlds of Inequality." *American Journal of Sociology* 117:1723–85. http://www.journals.uchicago.edu/doi/10.1086/665035.

Weeden, Kim A.,, Young-Mi Kim, Matthew Di Carlo, and David B. Grusky. 2007. "Social Class and Earnings Inequality." *American Behavioral Scientist* 50:702–36. https://doi.org/10.1177/0002764206295015.

Western, Bruce, Deirdre Bloome, Benjamin Sosnaud, and Laura Tach. 2012. "Economic Insecurity and Social Stratification." *Annual Review of Sociology* 38:341–59. http://www.annualreviews.org/doi/10.1146/annurev-soc-071811-145434.

Wilkinson, Bruce W. 1966. "Present Values of Lifetime Earnings for Different Occupations." *Journal of Political Economy* 74:556–72. https://doi.org/10.1086/259220.

Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach*. Mason, OH: South-Western College Publishing.

Wright, Erik Olin. 2005. "Foundations of a Neo-Marxist Class Analysis." In *Approaches to Class Analysis*, edited by E. O. Wright, pp. 4–30, Cambridge, United Kingdom. Cambridge University Press.

Xie, Yu. 1999. "The Tension between Generality and Accuracy." *Sociological Methods & Research* 27:428–35. http://journals.sagepub.com/doi/10.1177/0049124199027003006.

Yang, Yuhong. 2005. "Can the Strengths of AIC and BIC Be Shared? A Conflict between Model Indentification and Regression Estimation." *Biometrika* 92:937–50. http://academic.oup.com/biomet/article/92/4/937/389439/Can-the-strengths-of-AIC-and-BIC-be-shared-A.

Zimmerman, David J. 1992. "Regression Toward Mediocrity in Economic Stature." *The American Economic Review* 82:409–29. http://www.jstor.org/stable/2117313.

ChangHwan Kim: Department of Sociology, University of Kansas.
E-mail: chkim@ku.edu.

Christopher R. Tamborini: Office of Retirement Policy, U.S. Social Security Administration.
E-mail: chris.tamborini@ssa.gov.

Arthur Sakamoto: Department of Sociology, Texas A&M University.
E-mail: asakamoto@tamu.edu.