Measuring Lifetime Inequality*

John Knowles†
Dept. of Economics
University of Pennsylvania

May 22, 1999

Abstract

Previous studies of intergenerational income mobility (eg Solon, 1992, Zimmerman, 1992), rely on averages of annual income as a measure of lifetime income. However, given that the way in which income evolves with age differs by demographic group and education, it is possible to estimate lifetime income more accurately by estimating the entire age-income profile. This paper shows that correcting for heterogeneity of age-income profiles, as well as aggregate and cohort variation in income, results in measures of inter-generational income correlation that are substantially higher than previous estimates. The distribution of lifetime income is found to be much less skewed than averages of annual income.

*Chapter 1 of my PhD dissertation, prepared for submission to the Economics Dept. of the University of Rochester, under the supervision of Per Krusell. Many thanks to Mark Bils and Petra Todd for helpful discussions. This is a first draft, so comments are particularly welcome, but please check with me before citing my results.

†Email: **jknowles@econ.sas.upenn.edu**

1 Introduction

Recent empirical research suggests that transmission of economic inequality across generations is much more important than was previously believed; this finding has resulted in renewed interest in theories of the distribution of lifetime income. There is a sense however in which our beliefs about the statistics on which these theories rest is largely provisional: there is no publicly available lifetime income data.

The standard response to this difficulty has been to substitute for lifetime income averages taken over several years. One problem with this approach is that it does not use all of the relevant information available in household data sets. In particular, the way in which income evolves with age differs by demographic group and education, so it is possible to construct a more accurate estimate of lifetime income by estimating the entire age-income profile.

The problem of missing data is especially acute for the study of intergenerational dynamics, because data is required on the lifetime incomes of both parent and child, while the length of the series in panel data is around twenty-five years. As a result, the measurement of children's income tends to be taken when the children are younger than the age at which the measurement is taken for the parents. For instance, given the persistence of education differences across generations, and the relative steepness of the age-income profiles of the most educated, the income of the children of the highly educated parent will be low relative to the children's cohort while the parent's is high relative to the parent's cohort. This suggests that by not allowing for heterogeneity of the age-income profiles, the standard procedure results in estimates of intergenerational correlation that are biased down towards zero.

In this paper I show that correcting for heterogeneity of age-income profiles, as well as aggregate and cohort variation in income, results in measures of IG income correlation that are substantially higher than previous estimates. I also apply the new measures of lifetime income to another question related to long-term inequality: estimation of the relationship between fertility and income. This takes two forms: the relationship between income and completed family size , and that between income and the age at which an individual first becomes a parent. Finally, I report the moments of the lifetime-income distribution and apply a formal test of log-normality.

That age-earnings profiles are steeper for college graduates is well-known and has been shown in both cross-sectional (?), Card, 1994) and longitudinal studies (?). More recently, Baker (1997) has tested and rejected an important alternative model of earnings dynamics, the hypothesis that earnings follow a random walk (MaCurdy, 1982). This is relevant here because under the random walk hypothesis current earnings are a sufficient statistic for future earnings; therefore it would not be possible to improve upon averages as a predictor of lifetime income. This would also be true if earnings growth variations over the years were purely transitory, or if everyone's earnings grew at the same rate, but these hypotheses are patently in disagreement with the data. Baker restimates profiles of earnings adjusted for time and experience. His tests reject the hypothesis that the autocorrelation parameter is 1.0 and find substantial heterogeneity in growth rates across demographic groups[1].

The implications of this heterogeneity for intergenerational dynamics are interesting for two reasons. First, the qualitative importance of these dynamics is still disputed. Although a number of recent studies have shown father-son correlations to be quite high relative to older estimates, Couch and Dunn (1997) argue that this may be an artifact, the result of omitting parent-child pairs where one or the other has zero income, and a tendency to ignore income earned between ages 18 and 25. By correcting for these factors, they obtain much lower correlations, on the order of 0.18. The problem is that panels are not long enough to allow good lifetime income measurement for people that young. Second, even taking the qualitative importance as given, predictions of income inequality and their implication for the effectiveness of policy are strongly dependent on the magnitude of this correlation.

This paper estimates age income profiles by race and education from age 30 to age 80 on a representative sample of men drawn from the Panel Study of Income Dynamics. The available data is for the years 1967 to 1991, so income outside of this data window is imputed using the estimated coefficients from an income regression on the pooled observations of the entire sample. The main result is that allowing for dependence of the slope of the age-income profile on observable heterogeneity alone results in an estimate of the father-son correlation of income of 0.73, that is roughly a third higher than the currently received estimates of about 0.53 (Mulligan, 1996). The estimated earnings correlations on the other hand is relatively low, about 0.27.

Some of the increase in the intergenerational correlation is plausibly due

[1]Solon (1992) mentions in a footnote that Baker's work may prove useful for obtaining improved estimates of intergenerational income correlations.

simply to aging of the sample. For instance, Mulligan (1996) obtains a much higher father-son correlation of income than does Solon (1992), simply because he considers the income of children at a later age. This does not weaken the case for the importance of profile heterogeneity in these estimates. What does this show? That profile heterogeneity is at work: high-eduation profiles are approaching the average shape. Stokey (1996) argues that this increase in the apparent correlation is due to the reduction of noise from taking averages over longer periods. Geweke and Keane (1997) show that by age 30, observed income is practically a sufficient statistic for lifetime income, conditional on demographic variables, which is inconsistent with the Stokey interpretation.

The method described in this paper is useful not only because it allows more accurate estimation of the lifetime incomes of a given sample, but it also allows calcualtion of intergenerational correlations over a wider range of age cohorts, and extends the ages of analysis into old age, where income inequality is greatest. The method also allows improved estimation of inequality and poverty in countries where panel data sets do not exist or are much shorter than the PSID.

The procedure for estimation of family income dynamics is outlined in the next section. The results reported there are used in the following section to construct lifetime income profiles, which concludes by describing the distributions of lifetime income and earnings. This is followed by an analysis of intergenerational correlations. The paper concludes with a summary of the results and a discussion of some implications for future research.

2 An Empirical Model of Family Income

The eventual goal of this exercise is to construct lifetime income profiles for people we can observe in panel data only for a relatively short window, twenty-five years at most. To estimate income for the ages that fall outside this window, we will therefore use the income information from other sample members for whom observations for those ages are available. The problem with this strategy is that people from different age cohorts will have different income levels at the same age: income-age profiles are changing over time, and also are affected by influences specific to a cohort, such as events that affected their human capital accumulation (eg World War II). This paper proceeds by decomposing the process driving income dynamics into two components: one that is time-varying and one that is constant over time.

The goal of this section is to estimate these components so that the constant component of the income process can be used in the following section to predict entire income profiles, conditional on information available in the years in which income was actually observed.

Note that to the extent that the estimation is conditioned on personal characteristics, these must be observable at the ages for which we have observations on the people whose profiles we want to estimate. For instance, suppose we want to predict income in old age of people for whom the latest observations we have are in middle age. In this case, we cannot include marital status in old age as an explanatory variable when estimating the income process on old people, because this variable would not be available when we attempt to predict income for the younger cohort. It is better instead to condition the income process on marital status at age 35 say, when marital status is observable for the younger cohort, or the earliest marital status observation available for those people who are older than 35 when the panel begins[2].

There are several possible income concepts to be estimated. Total household income includes income of all current members of the household; likewise, let earnings includes labor income of all households. For the present, both of these income concepts will be modelled identically, and hence referred to simply as "household income". The analysis takes the characteristics of the head of the household as determining not only his own income, but also that of the other household members[3]. The basic assumption is that the log of annual household income y_{it} follows an MA(1) process with annual disturbances ε_{it} drawn from a conditional log-normal distribution. Income depends on a vector x_{it} of polynomial functions of the individual's age. The specification also allows for individual fixed effects α_i and for time-varying effects δ_t of individual characteristics w_i,where D_t represents year dummy variables :

$$y_{it} = \alpha_i + \beta x_{it} + d_t x_{it} + \delta_t D_t w_i + \varepsilon_{it} \tag{1}$$

[2]The PSID actually contains marital history variables that can be used to infer marital status at any age before the panel began. These were not used here; they are probably missing for a large part of the sample.

[3]This is standard procedure (eg Mulligan, 1996). To the extent that the head's characteristics affect family decisions and marital oppprtunities, the effects of these on income will be attributed to the head's characteristics.

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + \upsilon_{it}, \quad \log(\upsilon_{it}) \sim N(0, \sigma) \tag{2}$$

. The characteristics w_i are assumed constant (eg education, race or gender); those that vary deterministically (eg age, age of spouse or children) are included in x_{it}, while those that vary stochastically with time are not inferrable outside of the dataset and hence cannot be used to predict income in the future. The non-time-varying effects of an individual's observable characteristics are accounted for by the individual fixed effect, which will also reflect the effects of unobservable heterogeneity among individuals.

The process for the error ε_{it} is assumed to be independent of time and identical for all individuals. Since we allow for unobserved heterogeneity in (1), this amounts to assuming that the unobserved heterogeneity can affect the levels of income but not the growth rates[4].

The log-normal specification is quite standard for estimation of income dynamics. Lillard and Willis (1978) estimate an MA specification with random individual effects. On the other hand, Geweke and Keane(1997) find a better fit for the log of men's labor income with a mixture of normals distribution; the log-normal apparently understates the persistence of income shocks from year to year[5]. While the Geweke and Keane method may be more precise (they use Bayesian techniques to estimate probability distributions over the paramters of the income process, then for each individual find income in each year by integrating over future income paths), the clarity of this paper is much better served by sticking to the standard procedure[6].

For the purpose of characterizing the income distribution, the grave shortcoming of the log-normal specification is the necessity of excluding observations with zero income. Previous studies of income dynamics either exclude these observations or treat these as missing data and impute positive income. This problem will be discussed below in the discussion of the data.

[4]Baker(1997) argues that the portion of growth-rate due to the unobserved heterogeneity is at least as large as that due to observable factors. If this unobserved portion is also correlated across generations, then IG correlation estimates may be even higher.

[5]The paper actually says the logonormal *overstates* short-run persistence, but this seems inconsistent with claims later in the paper. I should think about this some more (or drop this comment).

[6]The idea is to bring in a standard technique to show that garden-variety econometrics is enough to improve our estimates of IG correlation. Plus I would not have the least idea how to go about implementing Geweke's technique.

2.1 Data

The sample consists of male household heads drawn from the representative cross-section (SRC) dataset of the Panel Study of Income Dynamics. All men whose ages in 1968 were between 10 and 80 are included except those who did not report the number of children they had (this restriction alone reduced the sample size by 800 men). This results in a sample of 1895 men.

The income dynamics sample consists of all annual observations on this sample made while the men were between 30 and 80 years old. The lower age bound was chosen so that virtually all men would have completed their schooling and most would have made lasting marital decisions by the time observations began. The upper bound was chosen so that income in retirement would be part of the sample. Mortality also plays a role; since the rich live longer, their lifetime income is higher. Descriptive statistics are shown in Table 1, which presents sample means by age in 1968. The sample contains roughly equal numbers with age in 1968 between 10 and 17, between 18 and 31, and between 32 and 49. The oldest age group 50-74, accounts for roughly 1/2 of the other groups, though the range of admissible ages is more than three times that of the youngest group.

Household income is taken to be the sum of total money income in the household and the value of food stamps received. Money income is the sum of taxable income and total transfers received by the head, wife and other members of the household. Transfers includes government aid, such as AFDC and SSI, social security, pensions, unemployment pay and help from relatives. Household labor earnings is the sum of the labor income of the head and the spouse. Labor income is the sum of wages income, professional income, bonuses, overtime commissions and an imputed labor portion of each of farm income, garden income and income from room and board. Income and earnings averages by time period are also reported, in 1967 dollars, in Table 1. It can be seen that income is increasing with age until age 50, both within cohorts over time, and across cohorts at any given point in time. The earnings decline in old age is much more rapid than that of income.

Of the personal characteristics that will be used to explain income, the trickiest to measure are work experience and education. In fact, since work experience depends on employment events that cannot be predicted from age 30, true experience is not used; instead experience is taken to be the difference between age and years of schooling, less 6 years of pre-school. Education itself is only measured in years in the PSID; there is no classification by level completed or degrees attained. Table 1 shows that average years of education (MAXGRD) is increasing as the sample gets younger, except that the peak is actually in the 18-31 age range. Since education tendencies have varied quite a bit over the different cohorts of the sample, respondent's education was reported as his percentile rank relative to his cohort. To account for possible non-linearities, these percentiles were converted to dummy variables for the 10th, 25th, 50th, 75th and 90th education years percentile. These variables are denoted P10ed, P25ed, P50ed, P75ed and P90ed.

Because previous studies suggest that marital status of the men plays an important role in predicting future income, particular attention was paid here to establishing the marital status of the men, especially to identify those who were in stable relationships. For the purpose of this study, a man over age 32 in 1968 was considered married if he was married or living with a woman in the 1968 wave of the survey. Younger men were considered married if they reported the same woman as their spouse during a majority of the years between ages 30 and 35.

The race of the men was taken to be that of the household head in 1968, since there seems to be no separate race variable for the children in 1968 households.The percentage of black people in the sample is highest for the young age group, because black men in the sample had more children, though this doesn't explain the high proportion of blacks in the oldest group.

The number of children of the household head may relate to income growth in at least two ways: family size may reduce income growth by reducing the time available for work, and income growth may itself reduce family size by raising the cost of fertility, since children seem to require large inputs of parental time. This variable is given in the PSID as the total number of children ever born to the individual. The average number of children (NUMKIDS) of the respondent is reported in Table 1, showing that fertility has fallen drastically for the youngest group. The oldest group however also seems to have had relatively few children.

Sample attrition is an important issue in the PSID. Although the original sample of households was representative in 1968, Solon (1992) points out that respondents with extremely high or low incomes have been more likely to drop out of the survey. This reduces the income dispersion of the sample, and would certainly affect the results of estimating the income dynamics if the specification relied only on observable effects to explain income. However to the extent that these dropouts differ systematically only in levels of income and not in the shape of the age-income profile (conditional on observables), omission of these people affects only the distribution of the unobservable fixed effects.So this is an issue for estimating the parameters of the lifetime income distribution, and for estimates of intergenerational correlation, but not for estimating the process that maps panel data onto lifetime income.

Not all men in the sample are equally represented in the income dynamics data set. This is due to two types of missing data: observations that are missing because they correspond to years for which data is not available, and observations that are missing either income variables or explanatory variables. The number of observations missing because the PSID started after the respondent's 30th birthday is 14,020. An additional 49,330 observations are missing because the last wave of the survey (1992) occurred before the respondent's 80th birthday[7].

Given these restrictions, the maximum number of observations is 32,273. Of these, only 20 are missing due to non-positive or missing total income, but 3,644, or more than 10% of the maximum sample size, have zero or missing labor income. About 20% of these were missing only one year of labor income, but 65% were missing more than three years. On average, labor income in the preceding year was extremely low, around $700, and total income was about 50% less than the sample average. People with observations in this category tend to be of retirement age (average age 68 years) and married (90%); about 8 % are black, close to the sample average. The chief distinction then of the zero-earnings sub-sample seems to be their age, so it is reasonable to suppose that labor income is zero because they have chosen not to work, or that they work unofficially and have chosen not to report the income because of the high implicit tax rates facing pensioners[8]. Under both of these hypotheses, it would be important to estimate imputed labor income and include this as part of lifetime labor income[9].

The idea behind the inclusion of the year dummy variables D_t in equation (1) is that income depends on economy-wide factors, such as business cycles, evolution of the skill premium, and trends in wage discrimination against

[7]It is the large number of observations in these missing categories that justifies the method of this paper.

[8]I'm not sure if there IS a high tax rate on earnings for these old geezers.

[9]An alternative hypothesis is that labor income fell to zero due to involuntary unemployment or illness. Under this hypothesis, the addition of imputed labor income would overestimate the lifetime income of those who reported zero labor income in old age. This possibility is dealt with later in the paper by estimating the probability of zero labor income in old age.

black people or women. The simplest way to allow for such effects is to include a dummy variable for each year, interacted with the personal characteristics. The problem however is that the experience variable is exactly co-linear with an individual's age and years of schooling, so the year-dummy variables would be co-linear with experience and the individual's fixed effect. Therefore it is necessary to model these effects directly. As a preliminary attempt at this, the US unemployment rate for each year is substituted for D_t. This variable is present as a level effect and in interactions with dummy variables for race, marital status, number of kids and education.

2.2 Estimation Results

The income model (1) was estimated on the income dynamics sample by OLS, using dummy variables for each person to reflect the individual fixed effects. The results are reported in Table 2 for the log of annual labor income and the log of annual total income. Given the significant auto-correlation of the error term expected from such a model, the test statistics should be based on an adjusted variance-covariance matrix[10].This work is in progress; for now, the table reports test statistics based on the standard OLS variance estimates. The explanatory power of the model, given by the R^2 statistic, appears somewhat greater for earnings than for income.

The table shows that the aggregate effects were generally not signficant, except for the fact that earnings varies less strongly with unemployment when the respondent has more children. The slope of the income profile appears to vary much less across demographic groups than does the slope of the earnings profiles. Additional children slow income growth somewhat, as does very high education, but the effects on earnings are much greater. Black people tend to have significantly lower rates of earnings growth.An important category of father-son pairs excluded by this procedure consists of those households where the father was not present in 1968. This was the case for 34 of the 454 sons in the sample. To the extent that this sample selection reflects income-related variables, then this is likely to bias the estimates of intergenerational correlation. In fact marital instability is much more common in low-income households, and children from these households are much more likely to have low income themselves (?). So the omission of these pairs is likely to reduce

[10]The adjusted variance-covariance matrix is a straight-forward application of the Eicker-White adjustment for heteroscedasticity, generalized to allow for serial correlation.

the estimated effect of parental income.

Given the polynomial specification of earnings in the regression equation, it is difficult to interpret the effect of experience. The age-income profiles implied by the estimated coefficients are seen more clearly in Figure 1. Each diagram shows the income or earnings profile of a given demographic group (dark line) compares to the average profiles implied by the experience coefficients alone. These projection do not include any of the aggregate effects, nor do they include the fixed effect; therefore the level itself is not meaningful, only the shape of the profile. What is striking here is that for both income and earnings, the profile for black people falls further below the average profile with age, while for the most educated half of the population, both forms of income start far below the average profile only intersecting the average around age 60. This confirms the idea that motivated this paper, that people with high lifetime incomes are likely to be further below their own average income than the rest of the population while young, reducing the apparent correlation of their income with that of their parents[11].

3 Lifetime Income Distributions

Given parameter estimates for the basic specification, the next step is to construct an income distribution from the sample. Because the sample is composed of cohorts from a wide range of birth years, differences in lifetime income are contaminated by the effects of aggregate growth and other cohort effects, such as variation over the sample period in the returns to education or experience. This is dealt with here by using the estimated time-dependent effects to strip out the cohort effects. Adjusted income $\tilde{y}_{it}$ is defined as the difference between realised income and income predicted from the time-dependent effects alone:

$$\tilde{y}_{it} = y_{it} - d_t x_{it} - \delta_t D_t w_i \tag{3}$$

. This leads naturally to the definition of predicted adjusted income:

$$\tilde{y}_{it} = \hat{\alpha}_i + \widehat{\beta} x_{it} \widehat{\gamma} w_i \tag{4}$$

[11]Note that profile heterogeneity appears stronger for total household income than it does for household earnings; the gap between the baseline and the highly-educated profile, for instance, is much less for earnings than it is for income. This suggests that correction for *observed* heterogeneity will have less impact on the lifetime income distribution.

.It is also necessary to adjust the estimated fixed effects $\hat{\alpha}_i$ for two reasons. First, because lifetime income is likely to depend on the birth-cohort of respondent, due to cohort-specific effects like wars or technological change. Second, because the age interval during which income is observed determines whether the estimated fixed effect is higher or lower than the individual's mean income over the lifecycle. This is done by OLS estimation of the effect of a polynomial A_i in age on the estimated fixed effects $\hat{\alpha}_i$, controlling for other personal characteristics w_i:

$$\hat{\alpha}_i = \zeta_0 + \zeta_1 W_i + \zeta_2 A_i + u_i \tag{5}$$

Letting the estimated effect of age be denoted $\widehat{\zeta}_2$, the adjusted fixed effect is defined as:

$$\widetilde{\hat{\alpha}}_i = \hat{\alpha}_i - \widehat{\zeta}_2 A_i \tag{6}$$

. Incorporating this adjustment to the fixed effect leads to our final adjusted-income concept, time-invariant annual income,

$$\breve{y}_{it} = \widetilde{\hat{\alpha}}_i + \widehat{\beta} x_{it} + \widehat{\gamma} w_i \tag{7}$$

Lifetime income Y_i is defined as the present discounted value of time-invariant annual income over the adult life. The discount rate r is supposed to represent the median rate at which young adults can borrow. This was set to 5% annually, in keeping with the literature on borrowing constraints(cite??).

$$Y_i = \sum_{t=t_0}^{T} \left(\frac{1}{1+r} \right)^t E_{t_0} \left[y_{it} | x_{i,t_0}, w_{i,t_0}, \alpha_i \right]$$

The empirical equivalent $\tilde{Y}_i$ of this expected life time income is obtained by substituting time-invariant annual income $\breve{y}_{it}$ for the expected future income $E_{t_0} \left[y_{it} | x_{i,t_0}, w_{i,t_0}, \alpha_i \right]$:

$$\tilde{Y}_i = \sum_{t=t_0}^{T} \left(\frac{1}{1+r} \right)^t \breve{y}_{it} \tag{8}$$

The results of the regression for adjusting the fixed effects are shown in the Appendix. It turns out that the age effects are not so important in the fixed effects, which is reassuring; this implies less risk of accidentally contaminating the fixed effects while purging the effect of age. The regressions show that the observable heterogeneity explains only about a third of the variation in the fixed effects across the sample, thereby justifying the incorporation of unobservable heterogeneity in the specification.

The distribution of time-invariant annual income was then calculated from the estimated coefficients in Table 2 for both income concepts. Table 4 gives the first three moments of the estimated distributions, and compares these to the mean of annual income over a 5-year period.

From Figure 2, it is evident that both the distribution of lifetime labor earnings and that of total income are symmetric and single-peaked; clearly the distributions are roughly log-normal. The main difference is that very low income levels are more likely for earnings than for total income.While earnings have a higher coefficient of variation than income, the skewness of income is somwhat higher, which results in a higher estimated Gini coefficient for income than for earnings.What is surprising about these results is that although the lifetime quantities have less dispersion than the 5-year averages, lifetime income is substantially more skewed than annual income. Table 4 also gives the average income by percentile rank. From this it can be seen that the ratios of the average income of the top percentiles to those of the lower percentiles, a standard way of comparing inequality across income distributions, are much lower for lifetime income than for the average income, indicating less inequality of lifetime income.

4 Intergenerational Transmission

This section compares the correlation of estimated lifetime income between father and son with those in the previous studies of the PSID, such as Solon(1992), Mulligan (1996) and Couch and Dunn (1997).All of these studies use the technique of representing lifetime income by 5-year averages, adjusted for the ages of the father and the son. These papers therefore implictly assume uniformity of the age-income profiles, because the age adjustments are not conditioned on the observed heterogeneity of the sample members.

As these previous papers, the intergenerational correlation is calculated as the effect ρ of the log of father's income y_i^p on the log of children's income

y_i^c:

$$y_i^c = \alpha + \rho y_i^p + f(x_i^c, x_i^p) + \varepsilon_i \tag{9}$$

In order to account for the effects of measuring parents and children at different ages, the previous studies also included age variables, typically polynomials in the ages of father and child; this is represented here by $f(x_i^c, x_i^p)$. Given the way in which lifetime income was calculated here however, age adjustments should not make much of a difference; all available information has already been incorporated into the income profile. Table 5 shows the results of repeating the Solon regression for lifetime income and earnings, as well as the 5-year averages calculated on the same sample, for the years.

An important category of father-son pairs excluded by this procedure consists of those households where the father was not present in 1968. This was the case for 34 of the 454 sons in the sample. To the extent that this sample selection reflects income-related variables, then this is likely to bias the estimates of intergenerational correlation. In fact marital instability is much more common in low-income households, and children from these households are much more likely to have low income themselves (**?**). So the omission of these pairs is likely to reduce the estimated effect of parental income.

The main result is that the intergenerational correlation of income is much higher than in the previous studies, 0.73, compared to 0.53 for Mulligan (1996); the number obtained here is however quite close to the instrumental variables estimate obtained by Mulligan, using an instriument for parental education. Earnings on the other hand actually appear less correlated than in the previous studies, 0.34 compared to 0.43 in Solon(1992). The standard errors for these estimates are quite small according to the standard calculations, but since a number of regression steps were used to produce the income data, the true standard errors are likely to be larger than the reported statistics suggest.

Given that earnings make up such a large fraction of the total income of households, two questions arise: 1) why did Solon's study find that earnings were more weakly correlated than income across generations?, and why did the lifetime income estimates increase the gap between the correlations of income and earnings? Solon raises the possibility of non-linearity of the effect of parental income, and indeed finds that low-earnings parents are more likely to produce high-earnings kids than vice versa: once families become skilled they are more likely to remain that way in the next generation. This effect was found to be stronger for earnings than income, but was quite small in either case. To test for a similar effect on lifetime income, the parent-child regressions were re-estimated with the square and cube of parental log income in the regression. These results are also shown in Table 5. The explanatory power of the regression for income was barely increased by this addition, but that for earnings more than doubled. The actual coefficients for labor income suggest that additional parental earnings is more effective in raising the earnings of the children of the rich than those of the poor: the coefficient on parental income is negative once higher powers of income are added to the regression. The relationship between the earnings of parent and child is shown in Figure 3. It is apparent that the effect of parental earnings on those of the children is indeed weaker for parents with low earnings, and that the linear specification actually understates the effect of parental earnings for parents with higher labor income.

It is clear that for poor familes, income contains a component that is much more persistent than earnings. While existing theory does predict this as the outcome of the trade-off between human capital investment and savings for bequests (**?**), an alternative explanation is that transfers are negatively correlated with earnings for the poor. One way to explore this in the future is to repeat the entire estimation procedure for other components of income, such as transfers and asset income. The problem is that these tend to be zero for a large portion of the population, so the log specification for income dynamics doesn't work. One way around this is to estimate the correlation of earnings plus each component separately: for instance earnings plus transfers is always positive over the earnings sample. If this sum turns out be more correlated across generations than earnings alone, one could safely infer that it is transfers that account for the fact that income is more highly correlated across generations than earnings.

5 Conclusion

The goal of this paper was to produce improved measures of lifetime inequality, and the intergenerational dynamics thereof. The argument was that existing measures might understate intergenerational persistence because income while young, relative to an individual's lifetime average, tends to be lower for rich people than for poor. The main implication of this argument is that measuring lifetime income requires the estimation of the entire age-income profile, with a specification that allows for the effects of this heterogeneity. This was achieved by estimating a standard model of income dynamics, while allowing the growth rate of income to depend on observable differences among people, such as race and education. The model was estimated for total household income and total household labor earnings, as a function of the characteristics of the household head and aggregate variables.

The results of the income dynamics estimation confirmed previous findings that there are indeed significant differences in income growth across racial and educational groups. Most significantly for the paper's argument, people with high education turned out have substantially lower income relative to their mean until age 40. For labor earnings however, these differences were somewhat smaller. The labor earnings of single black people with low education rose more slowly than those of similar white people.

Individual lifetime profiles were then constructed for all men in the sample by imputing annual income between ages 30 and 80, purged of aggregate and cohort effects. This was achieved by adjusting the estimated individual fixed effects from the dynamics regressions for age, and using this adjusted effect and the estimated coefficients, excluding the aggregate effects, to predict each individual's income for each year that their age was in the specified range. In this way, the income profiles allowed for both observable and non-observable heterogeneity. The present value of lifetime income was then calculated from these imputed income profiles. Table 4 reported the moments of the implied distributions of lifetime inequality and a simple tabulation of these distributions. It turned out that lifetime income had a lower coefficient of variation than labor earnings, but was more highly skewed. The Gini coefficient turned out to be slightly higher for earnings than for total income, but was in both cases much lower than for annual earnings.

With these estimates of lifetime inequality in hand, it was at last possible to estimate the effect of parental lifetime income on that of their children. This was done by linking those sample members who were living with their parents and aged between 10 and 20 in 1968 to their fathers. The effect of the father's income was found from a log-linear regression of the son's lifetime income; it turned out that the total household income is much more persistent across generations than earnings; 0.73 for income, 0.30 for labor earnings. The large discrepancy between the two correlations was largely due to non-linearity in the effect of parental earnings on those of the children: when higher-order terms in the parental earnings were included, the explanatory power of the earnings regression doubled, while the income regression was unaffacted. The non-linear specification for earnings implied that the effect of parental earnings on those of the children was higher for the rich than for the poor. This seems due to the fact that transfer income is negatively correlated with earnings.

The estimates presented here are not the final word in this research program. Not only is the length of the available panels increasing with time, which may increase the precision of estimates, but also the estimates of lifetime income presented here could be improved by using a better specification of the income dynamics process. There are two types of effects that have not been fully specified here. The first is the effect of time; the unemployment rate is currently the only aggregate variable in the specification for income dynamics; this means there is no control for variations in the wage premia for skill or experience, nor variations in the wage differentials for women and racial minorities. These effects are likely to be important over the period under study and will affect the mapping from the observed income of the old people in the panel to the predicted income of the young. The second class of effects that could be treated more thoroughly is that of unobservable individual effects on the shape of the age-income profile, since most of the variation across individuals can not be explained by observable variables (**?**). If these effects are correlated across generations, then the intergenerational income correlation may be even higher than the numbers reported here.

????

???? ??

???? ???

Correction for Missing Earnings

Because many of the earnings observations were missing or zero, they were excluded from the earnings dynamics regression. In the estimated profiles in Figure 6, these observations have effectively been replaced by expected annual income, conditioned on the person's characteristics and their fixed effect. Under the hypothesis that these excluded observations reflect voluntary unemployment or missreporting, the accuracy of the profiles is enhanced by ignoring the fact that these observations were zero. Under the hypothesis that these observations represent incapacity for work or involuntary unemployment however, the earnings profile may exagerate the labor wealth of people with excluded observations.

To correct for this, the following equation for employment was estimated using a probit model, resulting in estimated probabilities of employment by age and observable characteristics[12]:

$$E_{it} = \phi_0 + \phi_1 x_{it} + \phi_{2t} x_{it} + \phi_{3t} D_t w_i + \phi_4 w_i + \nu_{it}$$

In this specification, the explanatory variables are essentially as in the dynamics model (1) in the main body of the paper, except that personal characteristics enter w_i directly, since there are no individual fixed effects in this specification. The results of this regression are given in table XA. The error term ν_{it} here is of course likely to extremely persistent, but OLS is still unbiased. The test statistics are not accurate, as they haven't yet been corrected for this autocorrelation.

The baseline here, as in the main body of the paper, is unmarried non-black population with very little education and no children. The results show that unemployment is less likely for married people but more likely for people with more children, and that unemployment is more likely when the aggregate unemployment rate is high (uneducated black people are slighlty more likely to feel the effects of aggregate unemployment). The interactions with experience suggest the employment probabilities of educated people increase more slowly than the baseline, while those of black people increase more quickly.

[12] A fixed-effects regression would have resulted in the effects of unobservables also being incorporated into the probability estimates, and hence more precise estimates for each individual in the sample. However I couldn't get this specification to work (linear dependence?). Nor is there any correction here for the substantial autocorrelation of the error term.

Figure 3 aids in interpreting the effect of experience by showing the evolution of the employment probability with age for black men and for highly-educated men. It is evident that the employment probability stays very close to one until men attain an age of 50, at which point employment starts to decline for the baseline men and for blacks, while for the most highly educated employment stays close to one until they are nearly 60 years old, and declines much more slowly thereafter.

The adjustment of earnings for unemployment consists of defining adjusted earnings to equal the product estimated time-invariant earnings and the time-invariant probability of employment. The latter is obtained by subtracted the estimated effects of aggregate unemployment from the estimated employment probability. It turns out that this adjustment has very little effect on the results; it mainly affects earnings in old age which are relatively low and highly discounted when calculating labor wealth as of age 30. Table Xa compares the distribution of lifetime labor earnings with and without the employment adjustment. It is apparent from the table that this correction for the possibility of zero earnings makes very little difference to the earnings distribution. It is possible that this might change if we were able to specify the unemployment probability as a function of unobservable heterogeneity.

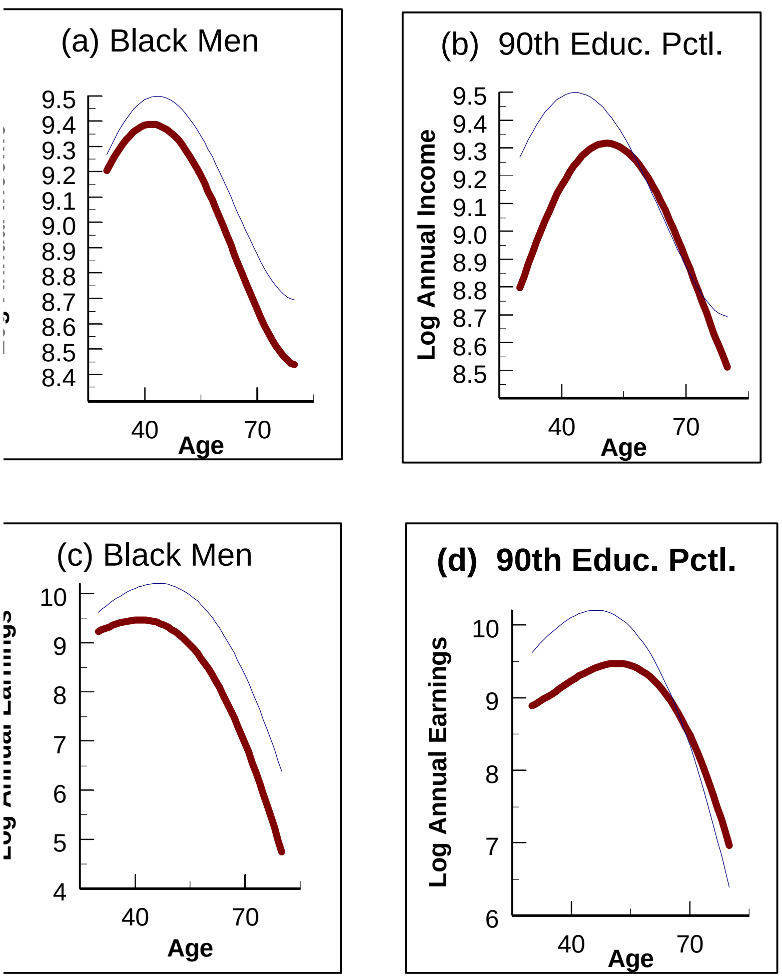
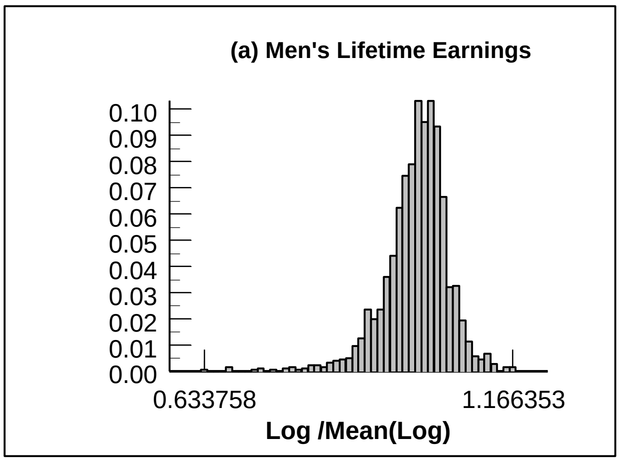

(a) Black Men
9.5
9.4
9.3
9.2
9.1
9.0
8.9
8.8
8.7
8.6
8.5
8.4
40
70
Age
(b) 90th Educ. Pctl.
Log Annual Income
9.5
9.4
9.3
9.2
9.1
9.0
8.9
8.8
8.7
8.6
8.5
40
70
Age
(c) Black Men
10
9
8
7
6
5
4
40
70
Age
(d) 90th Educ. Pctl.
Log Annual Earnings
10
9
8
7
6
40
70
Age

Figure 1:


(a) Men's Lifetime Earnings
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.633758
1.166353
Log /Mean(Log)

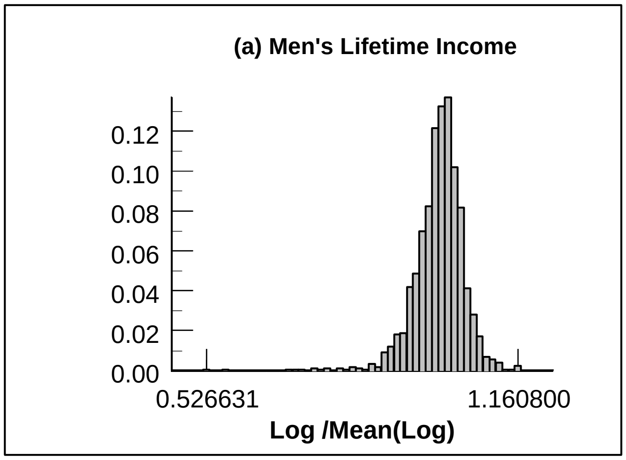

(a) Men's Lifetime Income
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.526631
1.160800
Log /Mean(Log)

Figure 2:

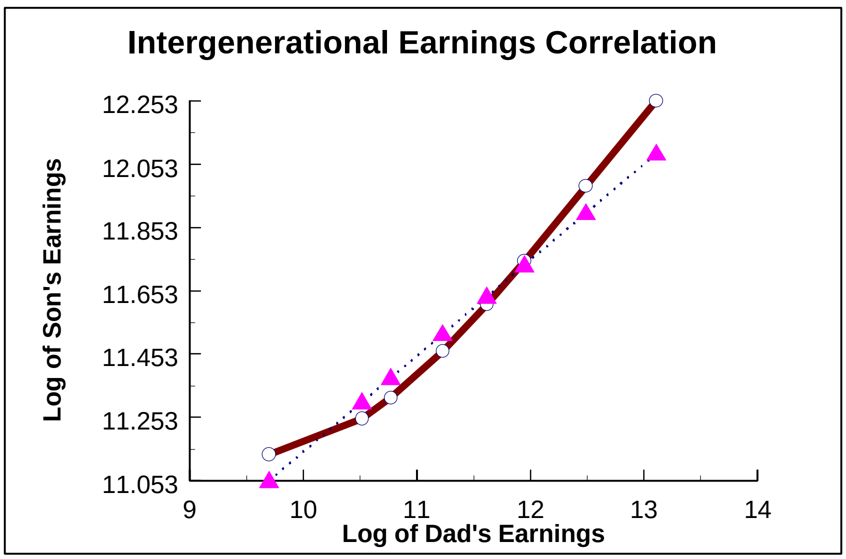
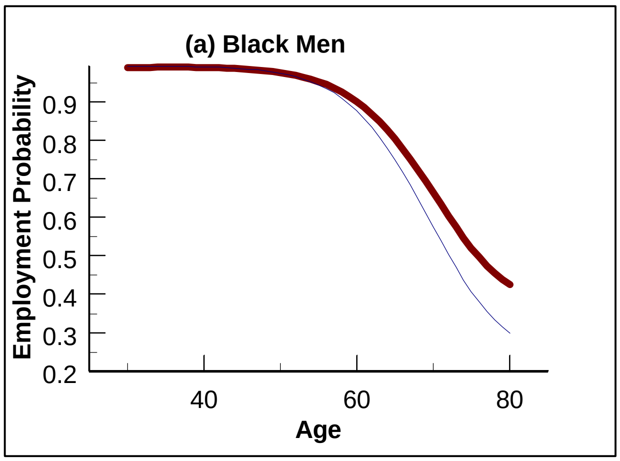

Intergenerational Earnings Correlation
12.253
12.053
11.853
11.653
11.453
11.253
11.053
Log of Son's Earnings
9
10
11
12
13
14
Log of Dad's Earnings

Figure 3:


(a) Black Men
Employment Probability
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
40
60
80
Age

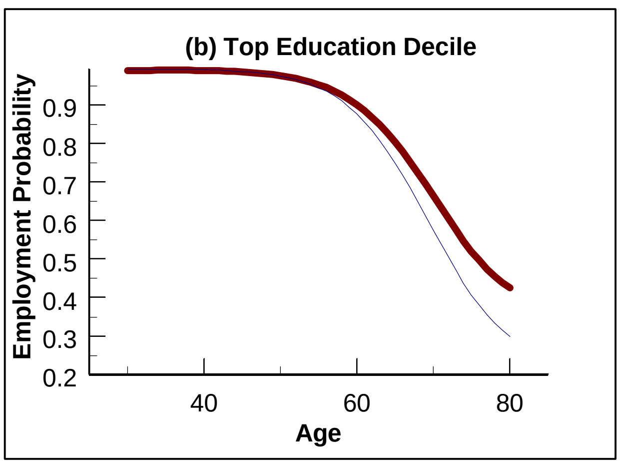

(b) Top Education Decile
Employment Probability
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
40
60
80
Age

Figure 4:

TABLE 1

Means of Variables for Income Dynamics Sample

	1968 Age Range of Subsample					
	10-12	13-17	18-31	32-49	50-62	63-74
N	200	316	535	570	213	61
MAXGRD	13.60	13.50	13.94	11.37	11.57	11.37
BLACK	0.09	0.09	0.05	0.05	0.08	0.08
WHITE	0.88	0.91	0.93	0.93	0.90	0.92
MARRIED	0.59	0.68	0.82	0.96	0.90	0.92
NUMKIDS	1.40	1.66	2.27	2.97	3.65	1.70
Income						
1970-74			$10,334	$11,901	$11,122	$9,344
1975-80			$10,485	$13,171	$11,050	$8,233
1980-84			$11,299	$13,940	$8,898	$7,488
1985-90		$10,178	$12,142	$13,439	$7,506	$11,780
1990-92	$10,657	$11,996	$15,227	$13,393	$7,192	$7,929
Earnings						
1970-74			$9,818	$10,666	$9,133	$4,549
1975-80			$9,887	$11,369	$8,494	$2,618
1980-84			$10,159	$11,225	$4,947	$1,288
1985-90		$9,339	$10,494	$9,984	$1,687	$408
1990-92	$9,957	$10,771	$12,667	$8,222	$635	$21

NOTE: Earnings and Income are as of age 30

Figure 5:

TABLE 2

OLS Log Income and Earnings Regression Estimates

Variable	Income	Earnings	Variable	Income	Earnings
EXPYRS	0.045326	-0.013159	WHITUNEM	-0.029001	-0.012278
	0.0001	0.3144		0.1013	0.5147
EXP2	0.00028	0.005311	UNEMP	0.022619	-0.020533
	0.5632	0.0001		0.2606	0.3448
EXP3	-4.193E-05	-0.000145	P90EDEXP	-0.003349	-0.025147
	0.0001	0.0001		0.183	0.0001
EXP4	4.05E-07	9.17E-07	P75EDEXP	-0.000373	-0.022483
	0.0001	0.0001		0.8713	0.0001
BLACUNEM	-0.023731	-0.032315	P50EDEXP	-0.002143	-0.015805
	0.2333	0.1334		0.3292	0.0001
P90EUNEM	-0.006309	-0.002688	P25EDEXP	0.001725	-0.002011
	0.4958	0.7917		0.5342	0.5577
P75EUNEM	-0.004468	-0.020119	BLEXP	-0.004894	-0.024203
	0.6095	0.0407		0.3359	0.0001
P50EUNEM	-0.005258	0.000372	MAREXP	0.006302	0.002453
	0.5204	0.9674		0.0068	0.3504
P25EUNEM	-0.003634	-0.013804	WHEXP	-0.002615	-0.014995
	0.7331	0.2682		0.5539	0.0023
MARRUNEM	-0.000799	0.017012	**R-squared**	0.55735	0.68674
	0.9214	0.0533	**Observations**	30816	27210

Figure 6:

Table 3
OLS Regression to Adjust Fixed Effects

Variable	Income	Earnings	Transfers
INTERCEP	0.17551797 0.0001	0.19056993 0.0001	8355.9992 0.0001
AGE68	0.01392387 0.3247	0.01533366 0.2175	-57.908517 0.0634
AGESQR	0.00040698 0.648	0.00045279 0.2907	3.674389 0.0001
AGECUB	0.00000357 0.9306	0.00000402 0.342	-0.056509 0.0001
BLACK	0.1126564 0.8101	0.12104817 0.0009	-1931.958 0.0001
P90ED	0.05688168 0.0001	0.06162822 0.0001	-1081.1287 0.0001
P75ED	0.05521199 0.0001	0.05990362 0.0001	-1090.2652 0.0001
P50ED	0.05208806 0.0001	0.05654027 0.0001	-837.07794 0.0001
P25ED	0.06657583 0.0001	0.07202102 0.0046	-36.709793 0.802
MARRIED	0.04120901 0.0001	0.04494759 0.0001	126.286444 0.1671
WHITE	0.09834117 0.1112	0.10547757 0.0054	-1138.2896 0.0001
NUMKIDS	0.00842789 0.0035	0.00912527 0.0001	19.368472 0.2966
Observations	1808	1808	1808
R-squared	0.2742	0.3615	0.354

Figure 7:

TABLE 4

Distributions of Lifetime and Average Incomes

	Total		Labor	
	5-year	Lifetime	5-year	Lifetime
Mean/Med.	1.09	1.11	1.09	1.13
Coef. Var	0.59	0.62	0.66	0.69
Skewness	3.59	3.87	3.69	3.49
Gini	0.28	0.29	0.32	0.32
Percentile				
5	0.38	0.37	0.24	0.30
10	0.49	0.48	0.41	0.41
25	0.70	0.70	0.68	0.67
75	1.33	1.34	1.34	1.38
95	1.97	2.16	2.08	2.35
99	3.41	3.44	3.30	4.15
N	603	1809	603	1809

Figure 8:

TABLE 5

Regression of Log Income on Log of Father's Income

Variable	Household Head		Biological Father	
Total Income				
Father's Income	0.74	0.00	0.67	86.83
	1.00E-04	1.00E+00	1.00E-04	1.41E-01
F.Inc. Squared		0.19		-6.74
		8.55E-01		1.61E-01
F. Inc. Cubed		-0.01		0.17
		7.71E-01		1.80E-01
R-Sq	0.23	0.24	0.15	0.17
N	378	378	300	300
Labor Income				
Father's Earnings	0.34	-0.09	0.33	-0.13
	1.00E-04	2.75E-01	1.00E-04	2.24E-01
F.Earn. Squared		0.09		0.38
		6.18E-02		3.30E-03
F. Earn. Cubed		0.00		-0.02
		0.211		0.01
R-Sq	0.10	0.21	0.07	0.17
N	378	378	300	300

Figure 9:

Table A1
Probit Regression Estimates for Employment Probabilities

Variable	Estimate	Variable	Estimate	Variable	Estimate
INTERCPT	2.587 0.0001	BLEXP	0.0104 0.0019	MARRIED	1.1933 0.0001
EXPYRS	-0.0379 0.4836	MAREXP	-0.0196 0.0001	NUMKIDS	-0.1682 0.0001
EXP2	0.00362 0.1257	KIDEXP	0.00305 0.0001	BLACUNEM	-0.0243 0.4573
EXP3	-0.00012 0.006	UNEMP	-0.1331 0.0001	P90EUNEM	0.0412 0.1994
EXP4	9.211E-07 0.0008	BLACK	-0.3509 0.1373	P75EUNEM	0.0457 0.097
P90EDEXP	-0.0194 0.0001	P90ED	0.3717 0.1489	P50EUNEM	0.0171 0.5287
P75EDEXP	-0.0176 0.0001	P75ED	0.3674 0.1059	P25EUNEM	0.00925 0.7661
P50EDEXP	-0.0135 0.0004	P50ED	0.3716 0.0956	MARRUNEM	0.0155 0.5717
P25EDEXP	-0.0127 0.0127	P25ED	0.7255 0.0118	NUMKUNEM	0.00534 0.1946
N =31730		R-sq = 0.2834			

Figure 10: