An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics

John Fitzgerald
Bowdoin College

Peter Gottschalk
Boston College

Robert Moffitt
Johns Hopkins University

Revised November 97

An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics

The Michigan Panel Study of Income Dynamics (PSID) is unique among large-scale, representative socioeconomic panel data sets in the U.S. in following the descendants of original sample members. Beginning in 1968 with a sample of approximately 5000 families, interviews have been conducted annually with the original sample members and families formed by members of the original households, usually children, who have left.[1] As the panel has aged these children have made up an increasingly large proportion of the relevant sample for the analysis of many adult outcomes such as welfare dynamics and early labor market transitions. This process will continue as parents die and are replaced in the sample by their children. But many of the original children in the PSID attrited either when their parents attrited or after they had set up their own households. The cumulative effects of even fairly low yearly attrition rates applied over a sufficiently long panel has led to the loss of roughly half of the children of original sample members by 1989.

In this paper we explore the impact of this attrition on estimates based on data from the second generation. The study complements our companion paper on attrition among respondents who were adults in 1968[2]. We focus our attention on two different types of questions about the second generation. The first set of questions explores the impact

[1] See Hill (1992) for a description of the PSID.
[2] Throughout this paper we use the term "companion paper" to refer to Fitzgerald, Gottschalk and Moffitt (1997).

of attrition on the mean characteristics (or, more generally, the marginal distribution of the characteristics)of the non-attriting second generation. To answer these questions we rely primarily on a comparison of the 1989 characteristics of surviving children in the PSID (who were 20 to 38 by 1989) to a similar sample drawn from the 1989 Current Population Survey.

The second set of questions that we address focuses on the relationship between adult outcomes of the second generation and their parents. For example, does attrition bias estimates of the intergenerational correlation in earnings, education or welfare participation? The availability of data in the PSID spanning more than one generation has spawned numerous studies examining intergenerational correlations in income (Behrman and Taubman, 1990; Corcoran, Gordon, Laren and Solon, 1992; Couch and Dunn, 1996; Solon, 1992), welfare (Antel, 1992; Duncan, Hill and Hoffman, 1988; Gottschalk, 1995) and economic status (Solon, Corcoran, Gordon, and Laren, 1991). These studies use direct observations not only on the parents' outcomes but also the children's outcome when they become adults.

It has been argued that the comparative advantage of the PSID has become the analysis of intergenerational relationships (Altonji, 1994). Its length and reliance on contemporaneous (rather than retrospective) data from separate interviews of parents and children when each were adults are important strengths in analyzing intergenerational issues. The length of the panel is also, however, a potential weakness if the cumulative attrition is non-random with respect to outcomes of interest.

The paper is divided into three sections. We start by presenting a brief review of the statistical models of attrition that underlie our empirical work. This is followed by a brief discussion of the extent of second-generation attrition in the PSID and a more extensive analysis of the impact of attrition on the distribution of characteristics of the second generation (our first question). The following section focuses on the effects of attrition on estimates of regression coefficients in intergenerational analysis (our second question). The final section draws conclusions based on our analysis.

I. Statistical Issues

General Issue. The statistical framework we use is similar to that developed at length in our companion paper. In this section we briefly review that framework and extend it to the issues with which we are concerned here.

We pose the conventional parametric model of selection as applied to the attrition problem[3]:

(1) $Y_{ct} = X_{ct}\beta + X_{p\tau}\alpha + \varepsilon_t$, observed if $A_t=0$

(2) $A_t^* = Z_t\delta + \nu_t$

(3) $A_t = 1$ if $A_t^* > 0$

$= 0$ if not

with the assumption

(4) $E(\varepsilon_t | X_{ct}, X_{p\tau}) = 0$

[3] In our companion paper we show how this parametric model is a special case of a more general selection model of attrition.

where Y_{ct} is an outcome variable of interest for child c at time t; X_{ct} is a vector of the child's observed characteristics at time t (when the child is an adult), $X_{p\tau}$ is a vector of parental characteristics at some prior time $\tau<t$, and ε_t is a vector of unobservables[4]. A_t is an indicator variable equal to 1 if the child attrites by time t and zero if not, and A_t^* is its latent index.[5] Z_t is a vector of observable characteristics (including X_{ct} and $X_{p\tau}$) that are not necessarily independent of ε_t.[6]

As in our companion paper, we make the important distinction between selection on observables and unobservables. Selection bias in the estimation of (1) on the non-attriting ($A_t=0$) subsample occurs if

Z_t and ε_t are independent but ε_t and ν_t are not (selection on unobservables)

ε_t and ν_t are independent but ε_t and Z_t are not (selection on observables) [7]

The case of selection on unobservables is well known in the econometrics literature. Identification rests either on non-

[4] To focus attention on attrition we assume $X_{p\tau}$ is exogenous. See Gottschalk (1995) and Antel (1992) for discussions of intergenerational correlation in unobservables which would make $X_{p\tau}$ endogenous.

[5] When we say the "child" has attrited, we include the case where the entire parental family attrites before the child has left the household.

[6] The additional assumption that Z_t is mean independent of ν_t is necessary to insure consistent estimates of δ. However, as our companion paper shows, our correction for selection on observables does not require this assumption.

[7] Note that the selection on observable problem cannot be "solved" by entering variables that affect selection but not Y_{ct} in the estimation of (1) (see our companion paper).

linearities in $E(\varepsilon_t | X_{ct}, X_{p\tau}, A_t=0)$ or on an exclusionary restriction (requiring that at least one element of Z_t not appear in X_{ct} or $X_{p\tau}$ and that its δ be non-zero.)

The case of selection on observables is discussed less frequently in the econometrics literature.[8] A selection problem occurs in this case because observables that affect attrition are not independent of ε_t. Thus, while Z_t is not structurally related to Y_{ct} (conditional on X_{ct} and $X_{p\tau}$), they do covary as a result of the selection mechanism. In our companion paper we show that one solution to this selection on observables problem is to first estimate (2), use the resulting estimated coefficients to form weights given by[9]

$$(5)\quad W = \left[\frac{\Pr(A_t = 0 \mid X_{ct}, X_{p\tau}, Z_t)}{\Pr(A_t = 0 \mid X_{ct}, X_{p\tau})}\right]^{-1}$$

and then estimate (1) by WLS[10]. In that paper we show that while selection on Z_t, and hence on ε_t, alters the distribution of ε_t, a consistent estimate of the original density can be obtained by reweighting on the basis of the

[8] The statistical literature has given more attention to selection on observables. See the references in Fitzgerald, Gottschalk and Moffitt (1997).

[9] The literature on choice based sampling cited in our companion paper also makes the point that weighted least squares can be used to eliminate the bias when selection is only on observables. We note that the PSID has constructed "universal" (non-model specific) sample weights as a function of lagged variables, but they have made no systematic effort to include characteristics of both children and their parents in the attrition equations used to construct weights. As we point out in our companion paper, model specific weights are a superior solution.

[10] Note that if Z_t does not affect attrition then all the weights are equal to one so OLS is consistent (because selection on unobservables is assumed not to exist).

observable Z_t's.

The critical variable in the case of selection on observables is, therefore, Z_t. The advantage of panel data is that variables observed in the initial wave of the survey are potential elements of Z_t. As long as these lagged variables are not in the structural model but do covary with the unobservables that affect Y_{ct}, they can be used to account for some of the heterogeneity between attritors and non-attritors. We use characteristics of the child or his parents in the initial interview as elements of Z_t because these lagged values are likely to covary with unobservables in the structural relationship being estimated.

Representativeness of Unconditional Means. Our first question is whether the unconditional means of variables for the second-generation are representative. In the context of the model above, a distinction is made between outcome variables and independent variables, but this distinction depends on the specific model. There is no loss in generality in considering all variables to be potentially outcome variables.

Representativeness of an outcome variable Y can be affected in two ways. First, if $E(\varepsilon_t | X_{ct}, X_{p\tau}, A_t = 0)$ is non-zero then estimates of the unconditional mean of Y_{ct} in the non-attriting sample will, in general, be biased. Second, even if $E(\varepsilon_t | X_{ct}, X_{p\tau}, A_t = 0)$ is zero but selection occurs on one or more of the independent variables, then the unconditional mean of Y_{ct} will again be affected.

Comparisons of second-generation unconditional means in the PSID to those of a corresponding nationally representative sample such as the Current Population Survey (CPS) can partially answer the question of whether the second generation is representative[11]. Note that this comparison tests for selection on unobservables as well as observables.

Bias of Intergenerational Coefficients. Our second question is whether attrition leads to biased estimates of the intergenerational coefficient, α, which will occur if ε_{ct} and $X_{p\tau}$ are not independent in the selected sample. For example, if children with levels of education similar to their parents are more likely to attrite, then estimates of α will be biased toward zero.

The CPS cannot be used to determine the extent of bias in intergenerational coefficients because it does not have information on parental variables. Nor are there other longitudinal data sets which can be used to benchmark the PSID because these data sets also potentially suffer from attrition bias or recall bias. We conduct tests for bias in coefficients with the PSID alone.

As we note in our companion paper, testing for selection on unobservables with the PSID alone can be conducted, absent parametric restrictions on functional forms, with an exogenous variable, or "instrument" for attrition (a Z_t independent of ε_t which is not in X_{ct} or $X_{p\tau}$). Since most of the variables that affect attrition are likely to affect

[11] We recognize that non-response rates in the CPS may also be biasing but we take the close correspondence between the 1990 CPS and the 1990 Census as an indication for the representativeness of the CPS.

behavior, and hence Y_{ct}, there do not appear to be any credible instruments in the PSID for attrition on unobservables. Therefore, we do not test for selection on unobservables in our analysis.

Tests for selection on observables rest on two conditions, either of which are sufficient for the absence of attrition bias on observables: (a) the weights equal one (i.e., Z_t does not affect attrition) or (b) Z_t is independent of Y_{ct} conditional on X_{ct} and X_{pt}.[12] These results are important in the panel-data case because the observable Z_t in question can be lagged exogenous and endogenous variables. For example, we might write:

(6) $Z_t^{*}= f(Y_{c,t-1},Y_{c,t-2},\ldots,Y_{c1},X_{p,t-1},X_{p,t-2},\ldots,X_{p1})$

suppressing lagged values of X_{ct}. If ε_{ct} is not independent of the variables in the function f -- which seems almost certain for the lagged values of Y_{ct} -- bias will result in estimation of equation (1) on the non-attriting sample. But the variables in (6), which are arguments in equation 2, are observed in the data. Equation (2) can, therefore, be estimated directly and used to construct the weights in (5). These can then be used to form WLS estimates.

We use two methods to test whether Z_t affects attrition. The first method is to estimate equations in the form of (2), using values of Y_{c1} and X_{p1} from the first year of the panel, when all individuals were present, to predict later

[12] The test we carry out in this paper are based on condition (a).

attrition:[13]

$$A_t^* = \delta_1 Y_{cl} + \delta_2 X_{pl} + \upsilon_t \tag{7}$$

Significant coefficients on Y_{c1} indicate biasing selection.

The second method of testing for selection on observables is to determine whether the conditional mean of Y_{c1} is different for children who later attrite and for those who do not. This test is based on differences in intercepts (and possibly slope coefficients) of a regression of Y_{c1} on X_{p1} and X_{c1} and A_t.[14] In our companion paper we show that this regression can be derived by inverting equation 7. But the test of whether children who latter attrite have a different conditional mean of Y_{c1} gives a more direct measure of the impact of attrition on outcomes of interest.

Note that the appropriate test is on the difference between the coefficients estimated from the full sample and from the non-attriting sample, rather than between the coefficients estimated from the attriting and non-attriting samples. The latter would be inappropriate because attrition may bias coefficients estimated on both subsamples.[15]

Specifically, we would like to estimate

$$Y_{cl} = \alpha_0 + \alpha_1 Y_{p1} + \alpha_2 X_{cl} + \alpha_3 X_{p1} + \varepsilon_l \tag{8}$$

---

[13] We do not include values of Y_{c2} or Y_{p2} or any other years after the first year since later observations would be affected by attrition between year 1 and t. In ongoing research we are exploring conditions under which values of Y_{cj} and Y_{pj} $(1 < j < t)$ could be included as regressors in equation (7).

[14] This is similar to tests in Becketti et al. (1988) which are referred to as BGLW tests in our companion paper.

[15] We thank a referee for noting that because estimates on both the attriting and non-attriting samples may be biased, a comparison between them is inappropriate.

for the full sample and

(8′) $$Y_{cl} = \alpha_0' + \alpha'_1 Y_{p1} + \alpha'_2 X_{cl} + \alpha_3 X_{p1} + \varepsilon_l'$$

for $A_t=0$

With estimates of these two equations we could test whether estimates of α_1, the intergenerational coefficient, change when the sample is restricted to persons who do not later attrite.

The major difficulty in estimating either (7) or (8) and (8′), is that this procedure requires that we observe the adult outcomes of children before they attrite. Equation (7) tests whether Y_{c1} (e.g. earnings in the first year of the panel) predicts later attrition. Equation (8) tests whether the coefficients estimated on a sample of persons who do not later attrite are different from the coefficients estimated on the full sample. But the adult outcomes of children--such as earnings, education, and welfare participation--cannot be observed in the first year of the panel (1968) because the members of the second generation are by definition less than 18 in that year. Few children (persons under 18) had completed their education or were participating in AFDC before they were 18; nor had they experienced their adult earnings or marital status. These outcomes were only observed when the second generation reached early adulthood. But by that time considerable attrition had already occurred, especially for children who were young in 1968.[16]

To conduct tests using this approach necessarily

[16] This problem does not arise in our companion paper which focuses on outcomes of a single generation. In that paper we can, therefore, examine whether 1968 earnings of adults predict later attrition.

requires additional assumptions. We pick a time point s sufficiently far into the PSID that we can observe adult outcomes of the children. The period during which the second generation is reaching adulthood (when t<s) we call the "pre" period. The period after s we call the "post period". We use the sample of children who did not attrite in the pre-period to estimate the counterparts to equations (8) and (8'). Let A_S equal zero if the respondent has not attrited by s and let A_T equal zero if the respondent has not attrited by T, the end of the panel[17]. The equations we estimate are given by:

(9) $Y_{cs} = \alpha + \alpha_1 Y_{p1} + \alpha_2 X_{cs} + \alpha_3 X_{p1} + \varepsilon_s$

for $A_S = 0$

(9') $Y_{cs} = \alpha' + \alpha'_1 Y_{p1} + \alpha'_2 X_{cs} + \alpha'_3 X_{p1} + \varepsilon'_s$

for $A_S = 0$ and $A_T = 0$

where Y_{cs} and X_{cs} are the values of Y_c and X_c for period s.

The obvious limitation of this strategy is that some attrition will already have occurred in the pre-period (prior to s) and, therefore, estimates based on this sample may already be biased. We must invoke further identifying assumptions in order to test whether attrition is biasing. A sufficient assumption is that attrition biases coefficients in the same direction before and after s. With this assumption any difference between the coefficients estimated on the sample that survived to s and the sample that survived

[17] The corresponding attrition equation, which we also estimate, has parent's and children's adult outcomes (Y_{cs} and X_{p1}) on the right hand side of a binary choice equation estimated on the sample for which A_S is equal to zero and A_T is the indicator variable.

to T, adds to any biasing attrition that occurred in the pre-period. Since, under this assumption, it is not possible for bias in the pre and post-period to offset each other, a finding of attrition bias during the post-period implies bias in the sample that survived through both the pre and post-periods. A finding of no attrition bias in the post-period, however, does not rule out attrition bias during the pre-period.

The model presented in this section can also be used to make two more general points that are sometimes overlooked in the literature. First, it should be clear that the answer to the question of whether the PSID suffers from attrition bias is case specific. The key covariances may be zero for some outcomes but not others. No global statement is possible about attrition bias in a data set. It is up to each researcher to address the question of attrition in the context of the question being addressed. Second, selection on a right-hand-side variable in (1) does not lead to attrition bias.[18] For example, if X_{p1} is in Z_t (i.e., a parental characteristic affects both the child’s outcome and the likelihood that the child attrites), then this selection by itself causes no bias; bias only occurs if Y_{ct} differs for attritors and non-attritors, holding X_{p1} fixed.

II. Extent of Attrition

Our primary sample includes all children 18 years and younger in 1968 (22 to 39 in 1989) living in SEO and SRC households in 1968. Chart 1 shows the proportion of the

[18] This point is made in Solon (1992) and Menchik (1979).

original sample of children remaining in the PSID in each year[19]. The sharp drop in the proportion responding in 1969 shows that more than 10 percent of the sample attrited in the first year of the panel. After this drop, the proportion responding continues to fall at a fairly constant rate reflecting attrition rates of around 3 to 4 percent between 1969 and 1989. While yearly attrition rates are modest, the steady erosion of the sample over a twenty year period has a substantial impact. By 1989, only 52 percent of the children in the original sample were still in the PSID.

Chart 2 shows similar data for children broken down by race. Between 1968 and 1975, blacks and whites had very similar patterns of attrition. However, starting in 1975 blacks attrited at substantially higher rates than whites, leading to 49 percent of the initial sample of black children still in the sample by 1989. In contrast, 59 percent of whites were still in the sample. The pattern for children of all other races differs in two important ways. First, there was substantially more attrition in the first year of the panel but within the next two years there was less attrition, with the result that in 1971 the proportion remaining was only slightly lower for this group than for either blacks or whites. However, after 1971 the attrition rates for children of "other" races were considerably higher than attrition rates for either whites or blacks, with the result that only a third of the original sample for this group remained by

[19] Since re-entry is possible, the slope of these functions reflect both the hazard of leaving the sample in each year and the hazard of re-entry. Re-entry rates are, however small, ranging from .0025 to .0064.

1989.[20]

Chart 3 shows the patterns for sons and daughters while Chart 4 disaggregates by age. Daughters were somewhat less likely to attrite than sons but the difference is not large (57 versus 51 percent remaining by 1989). Differences across age groups are somewhat larger, especially during the 1970's with older children having lower response rates. This undoubtedly reflects the fact that children 13 to 18 in 1968 were in their twenties during the 1970's and were, therefore, more likely to be setting up their own households. As we will show, newly formed households were considerably more likely to attrite in the immediate years after they split off from the original PSID household, which is consistent with higher attrition for older children but a narrowing of the gap between the proportion of younger and older children responding as these younger children also aged through the period when they set up their own households.

Chart 5 shows the hazard of not responding broken down by the type of attrition: whether the family unit in which the member resided refused to participate or could not be found (FU non-response), whether the member died or whether the member moved out of the family unit and could not be followed[21]. The family unit refusing to participate is the largest category in each year and death is by far the smallest category. The substantially higher hazard in 1969

---

[20] Prior to 1985 Hispanics were coded as a separate race. We coded Hispanics as white in these years to maintain comparability. For 1985 and later, Hispanic ethnicity is a variable separate from race.

[21] "Move out" indicates that the family was interviewed, but that the person had moved out and either could not be followed or was followed and refused to be interviewed. Since the later could

than in all the following years primarily reflects a high rate of attrition due to non-response, though the hazard of moving out is also somewhat elevated. After 1969 the overall hazard and its components varies in a narrow range with no clear trend.

In summary, the loss in the sample was largest between 1968 and 1969 for all age, race, and sex groups. This largely reflected the high hazard of family unit non-response. As a result of the high overall hazard in the first year and the steady erosion over the remaining years, sample sizes in 1989 are roughly 40 to 60 percent of what they were in 1968 depending on the demographic group.

III. Representativeness of the Second Generation

In this section we explore the question of whether the second generation remains cross-sectionally representative in spite of the substantial attrition documented in the previous section. We address this question in two parts. First we show the extent to which attritors differed from non-attritors based on their 1968 characteristics and the characteristics of the households in which they resided in 1968.[22] Next we examine whether these differences lead to bias in the estimates of unconditional means by comparing the 1989 characteristics of non-attritors in the PSID to a corresponding sample from the CPS.

A. Matching Children with Parents in the PSID

We begin by presenting tabulations of the mean 1968

---

reflect a death not reported to the PSID, the proportion classified as “move out” is potentially overstated.

[22] In the case of unconditional means, tests of differences in means between attritors and non-attritors is equivalent to a test of differences between non-attritors and the full sample.

characteristics of children and their parents for the sub-sample of children who later attrited and those who did not. Because we are interested in the family background of attriting and non-attriting children it is necessary to first identify the parents of the children in the PSID. We, therefore, take a short detour to explore the issues raised by having to match children with their parents in this data set.

Matching children with their parents is straightforward in most data sets. For example, the National Longitudinal Survey of Youth (NLSY), includes information on parents' characteristics in the child's record. However, the structure of the PSID is more complicated because households are the unit of analysis and the only relationship coded for each person is his or her relationship to the household head. If the head of the household is the parent, it is straightforward to link parents and their children in years in which they are living in the same household. However, it is more difficult to link generations when another person, such as a grandparent, uncle or unrelated individual is the head of the household, or when the child has already moved out of the parental household.

Matching is straightforward for sample members 0-18 who are classified as a child or stepchild of the head in the 1968 interview. The PSID offers two sources of information to match the remaining children with their parents. First, the identifying numbers of the mother and father were appended in 1985 to the child's record.[23] However, these

[23] The PSID included a supplement in 1985 containing questions on the timing of demographic events of PSID family members, including childbirth, marriage, separation and divorce. See Hill (1992)

variables are missing for all children whose parents attrited before 1985. For children with missing data on these variables it is necessary to use either the "Relationship file", or the variable giving "relationship-to-head" in each year after 1968 to identify the child's parents[24]. If the latter is used it is necessary to identify those years in which the child is classified as "child or stepchild" of the head[25]. In those years it is possible to identify at least one parent and possibly both parents by examining the identification number of the head and wife (if married) in the household.

The question of how to treat stepchildren raises conceptual and measurement issues[26]. Any intergenerational study must decide whether to limit the analysis to the correlations between the outcomes of children and their biological parents or whether to include stepparents[27]. If hereditary links are the object of interest then it would be appropriate to include only children for whom it is possible to identify biological parents. However, since "parental" characteristics are often used to capture the home environment and since characteristics of stepparents would seem to be equally good measures of home environment, we

---

[24] The relationship file gives the blood, marital or cohabitation relationship between all pairs of individuals descending from the original 1968 sample families. See Hill(1992)

[25] Note that this misses the children of non-heads. In a small number of cases the relationship to head variables are inconsistent across years (e.g. the respondent is classified as child of a male head but the identity of the male head changes). In these cases we use the earliest match.

[26] Children not living with parents or stepparents in any year could not be matched.

[27] The PSID distinguishes between children and stepchildren only after 1982.

include stepparents in our analysis[28]. For children who move between the homes of biological parents, stepparents and custodial parents, no single match will capture the home environment in which these children were raised.

Table 1 presents the proportion of children 18 years old or younger in 1968 matched with both parents, the mother only, the father only or with neither parent. The top panel shows the distribution for sons and the bottom for daughters. Several patterns emerge. First, the proportion of children not matched with either parent is small, ranging from 1.6 percent for white sons to 11.3 for sons of other races, and the overall proportion of children not matched with either parent is only 3.2 percent. Second, whites are substantially more likely to be matched with both parents than either blacks or children of other races. The matching rates with both parents are roughly ninety percent for whites but only sixty percent for blacks and seventy percent for other races. This, undoubtedly reflects a greater number of single parents among non-whites. Third, when only one parent is identified it is much more likely to be the mother than the father. The fact that roughly thirty percent of black sons and daughters could only be matched with their mothers suggests that a large part of the missing matches reflect the actual family structure and not the inability of the PSID to identify a parent living in the household.

Whether the latter is relevant depends on the question asked. For example, estimates of the effects of outcomes of

[28] Solon (1992) uses the same rational for examining the correlation between the earnings of the male head of household and the earnings of any male child in that household. He, therefore, includes stepchildren as well as

household members, such as head's education, on children's later outcomes should be based on the observed characteristics of household members. Since non-custodial parents are not in the household it is irrelevant whether or not they can be matched with their children. Estimates of the effects of custodial and non-custodial parents on their children's outcomes would be biased if the custodial parents were not a random subsample of all parents. Existing studies are seldom clear on the relevant population.[29] For example, studies of intergenerational correlations in education could either refer to correlations between children and their custodial parents or to all parent children pairs.

B. Mean Characteristics of Attritors

Tables 2a and 2b present mean 1968 characteristics of children and their parents for mother-daughter pairs and father-son pairs, respectively, according to attrition status of the child, parent or both. The columns labeled Always In include persons who were in the PSID in all years between 1969 and 1989, the last year for our sample. Ever Out indicates attrition for reason other than death in at least one year[30]. The sample includes all mother-daughter (or father-son) pairs for which we have valid data in 1968. The first column of each table shows the 1968 characteristics for all parent-child pairs, whether or not either attrited. Column 2 shows the 1968 characteristics of the parent-child pairs in which the child remained in the panel for all years,

---

other relationships, such as heads who are uncles of the children in their households.

[29] An exception is Solon (1992, p. 398) who explicitly states that his focus is on the correlation in economic status of sons and the status of the households in which they grew up.

whether or not the parent attrited. Column 3 includes parent-child pairs in which the child attrited, either alone or with the parental family. The next two columns (4 and 5) show the corresponding information according to the parent’s attrition status.[31] While the attrition status of the parent is not directly relevant to most studies of the second generation they are relevant to studies that require information on the parents later in life, for example studies of the living arrangements of the second generation and their elderly parents. Columns 6 and 7 include parent-child pairs in which both members remained in the sample until 1989 and those in which either parent or child attrited.

Columns 2 and 3 in the top panel indicate that 87.1 percent of daughters who did not attrite were white while only 79.9 percent of daughters who did attrite were white. This under representation of whites among attritors also applies to mothers who attrited (86.7 versus 82.1). Daughters who attrited were more likely to come from disadvantaged households by almost all measures. For example, 84.2 percent of the mothers of daughters who attrited were married in 1968 versus 89.6 percent for non-attritors. Nearly half of the daughters who left the panel grew up in families where the mother had less than twelve years of education, compared to roughly one third for non-attritors. Daughters who later attrited lived in households in 1968 that were twice as likely to have received public assistance (4.0 versus 2.3) and in which the mother had

[30] See our companion paper for a discussion of mortality and attrition.
[31] The samples in these columns overlap with those in the prior two columns.

nearly one less year of education (10.1 versus 11.0).

This pattern of greater attrition among daughters of less advantaged mothers is corroborated in the family income and income needs ratio shown in the bottom panel. Mean family income was $27,703 for daughters who remained in the panel while the mean income of families in which the daughter attrited was $23,305. Similarly the mean income-needs ratio is substantially lower for families in which a daughter latter attrited.[32] These families had incomes that were 1.82 times the poverty line for their family size while the mean for families in which the daughter remained in the sample was 2.19. The standard deviation of income-needs ratios also shows that attritors had less dispersion around their lower mean. While total family income was lower, the mothers of attritors were as likely to have worked and their earnings were marginally higher.[33]

Table 2b shows patterns for the matched father-son pairs.[34] Our inability to match sons with non-custodial fathers is clearly reflected in the high proportion of sons coming from married households. However, even among sons drawn primarily from married households the attritors still differ from non-attritors in important dimensions. Attritors were substantially less likely to be white (84.2 versus 92.6) and were more likely to have fathers with low educational

[32] Needs is the family size adjusted poverty line for the household. For example an income needs ratio of 2 indicates the family’s income is twice as large as its poverty line.

[33] While this could reflect the higher probability of attrition among children from female headed households, the attrition probits presented later in the paper indicate that even after controlling for marital status, higher earnings of mothers are associated with a higher probability of attrition for their daughters.

attainment. Fully 46.8 percent of the sons who attrited came from families in which the father had less than a high school degree. This is substantially higher than for non-attritors (34.4). The result of these and other differences in characteristics led to mean earnings of fathers that were over $2,000 lower for sons who later attrited. After adding other sources of income, the difference is reduced to roughly $1,000.

Turning to the characteristics of the non-attriting parents of the second generation, shown in columns 4-7 of Tables 2a and 2b indicates that these parents were more advantaged than the parents of children who later attrited. Family incomes were uniformly higher for non-attritors than attritors. This holds, whether comparing parents who attrited with those who did not (columns 4 and 5) or parent-child pairs in which either parent or child attrited or neither did (columns 6 and 7). The higher incomes of non-attriting parents holds for mothers or fathers, whether or not we adjust for family size (by focusing on income-needs ratios). Likewise there is a systematic pattern in the dispersion of incomes. In all cases but one, the standard deviation of income is smaller for attritors than the non-attriting sample.

While these tables show the 1968 characteristics associated with later attrition, they do not provide information on the timing of attrition. Specifically, did children who attrited tend to leave before they splitoff from their parental families (to either form their own households or join a non-parental household)or were they lost after

---

[34] These patterns are similar to those reported for the sample used

leaving their parental households? The answer to this question is relevant not only to researchers studying household formation but also to PSID staff who may want to focus resources on maintaining contact with children when they leave the parental household, if that is when many of them are lost.

Table 3 presents the distribution of attriting children by whether they attrited before, during, or after they split off from their parental household[35]. Among the 4082 children who attrited from the PSID, roughly half (49.7 percent) attrited at the same time as their parents. Thus, parental characteristics are potential predictors of child attrition. The remaining half are roughly evenly divided between those who were lost to the survey in the year they left their parental household and those who attrited after leaving their parental home. These patterns are remarkably similar for daughters and sons. The fact that nearly a quarter of the attrition takes place in the year the child leaves the parental home suggests that moving is an important characteristic associated with attrition.

C. Attrition Probits

Thus far we have focused on individual characteristics associated with attrition. The multivariate counterpart to this tabular evidence is to estimate attrition equations corresponding to equation (2), including as regressors the 1968 characteristics of the parent and child (which necessarily excludes the adult outcome of the child). From

---

in Solon (1992, p. 398).

[35] Seven percent of children were not living with either of their parents in 1968. These are included as splitoffs when they move out of their 1968 household.

these equations we can determine whether some of the differences we found in the tabular analysis disappear when we control for other variables. Table 4 presents the probit coefficients and derivatives which can be used to test whether the 1968 characteristics have independent effects in predicting future attrition, holding other characteristics constant[36,37]. The dependent variable in these equations equals 1 if the child had attrited by 1989 and 0 if not. The relevance of these equations is clear given our stress on selection on observables. As we argued earlier, if the variables in these equations do not appear in equation 1, the primary equation of interest, and if these variable are correlated with the unobservables in the primary equation (i.e., ε_{ct}), then WLS can be used to obtain consistent estimates.

Separate equations are estimated for the impact of Father's 1968 characteristics (Column 1 and 2) and mother's characteristics (Column 3 and 4) on the child's attrition probability. Column 1 indicates that economic as well as demographic characteristics of fathers are important predictors of the child's later attrition. Being black or living in an SEO household increases the probability of attrition by .042 and .061 respectively. Father's age and education are associated with lower attrition probabilities.

---

36 The derivatives are calculated for each person and are averaged across persons. The expanded version of this paper presents attrition probits for mothers and fathers.

37 We also estimated hazard models which included a set of indicators for whether the child had splitoff from the parental household. As might be expected attrition probabilities are higher in the year immediately after the child splitoff from the parental household. But those children who survive through the

Likewise father's with no labor income had children who were more likely to attrite. Holding father's characteristics constant, attrition declines with the child's age after age 12.[38] The columns for the impact of mother's characteristics also show that mother's age and education are associated with lower attrition but mother's marital status is significant, while it was not for father's.

The results presented in this and the previous section clearly establish that attrition of the second generation is related to observable 1968 characteristics of the child and the child's family. To evaluate the relative importance of these observed factors we present R^2 described by Cameron et al (1997) for non-linear models[39]. These measures, which are all very small, indicate that there was substantial diversity of attrition experiences even among similar individuals. This is consistent with the findings in our companion paper which also shows that a large part of attrition is not explained by observables.

D. Comparison to the 1989 Current Population Survey

In this section we explore whether the surviving sample maintains its representativeness by comparing the mean 1989 characteristics for the non-attriting PSID sample with the characteristics of a corresponding sample drawn from the

---

splitoff period exhibit below average attrition probabilities later in the panel.

[38] This may either reflect state dependence (as children age they are less likely to attrite) or heterogeneity (the children still remaining after they reach 12 are the children with lower attrition probabilities at a given age).

[39] The R-squared equals one minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept. They show that is can be interpreted as the proportion of uncertainty explained by the regressors.

March Current Population Survey (CPS)[40]. While the CPS suffers from underreporting of income, undercount of minorities, and several other factors that may bias estimates of means of some variables, it is generally regarded as being sufficiently representative to serve as a benchmark for the large number of variables we examine[41].

Tables 5-7 show comparisons separately for male heads, wives, and female heads in 1989 who were 22-39 in that year (and hence 1-18 in 1968). The tables show mean characteristics from the CPS and three sets of means from the PSID: those from the SRC sample only, which are unweighted[42]; those from the combined SRC and SEO samples, using 1968 weights; and those form the combined SRC and SEO samples, using 1989 weights. The 1968 weights adjust the combined sample to account for the oversampling of the SEO sample.[43] The 1989 weights adjust the 1968 weights to account for differential attrition and mortality by a number of characteristics, hence represent a form of the weighting procedure we mentioned earlier in the context of our statistical model.[44] By necessity this weighting procedure can only adjust for pre-attrition observables; if selection

---

[40] This analysis can be motivated by the statistical model with no covariates in equation 1. We do not examine conditional means because of lack of space but the analysis in our companion paper indicates that, at least for that sample, similar results were found for conditional and unconditional means.

[41] For a more detailed discussion see Fitzgerald, Gottschalk and Moffitt (1997).

[42] The weights constructed by the PSID are based only on the combined SRC and SEO samples; no weights for the SRC alone have been constructed.

[43] SEO households were drawn with selection probabilities that depend on geographic location, age, race, and income (PSID User's Guide, p. E-2).

[44] The unweighted SRC sample can also be used to obtain unbiased estimates of population means in 1968 but not in future years if there is non-random attrition.

is only based on the observables used to construct the weights, and if the weight calculation is accurate, any remaining difference between the CPS and weighted PSID can be ascribed to selection on unobservables.[45]

Comparing columns 1 and 2 in Table 5 indicates that most of the demographic means for male heads are similar in the CPS and the 1989 weighted PSID. Mean age is identical and the educational, marital status and regional distributions differ by only a few percentage points. The largest difference comes in the proportion Hispanic (.04 in the PSID versus .08 in the CPS). This may partially reflect greater attrition of Hispanics but it also reflects the sample design of the PSID, which excludes recent immigrants.[46] Immigrants arriving after 1968, by definition, cannot be descendants of the 1968 families in the PSID[47].

Labor market outcomes show considerable similarity in these two data sets but some differences remain. For example, the PSID shows fewer male heads working zero weeks during the year (.02 versus .06) but this is offset by the lower number of weeks worked among those with positive weeks (46.8 versus 48.8), leading to the same unconditional mean weeks worked in the two data sets. Mean 1989 wage and salary income is 4.8 percent higher in the PSID than in the CPS

[45] However, the PSID weights do not systematically include many lagged values from the parental as well as child household.
[46] This is consistent with evidence in the PSID sample used in our companion paper, which also includes fewer Hispanics than in the CPS.
[47] The PSID added a Latino sample in 1990 that includes some immigrants but this sample is not included in our analysis which focuses on the children of the original families. Including the Latino sample would make the second generation more representative but this sample cannot be used for intergenerational analysis since the parents of these recently added respondents were not observed when the respondents were young.

($20,698 versus $19,751) and mean family income is 10.3 percent higher ($31,812 versus $28,836).[48] Since hours worked are very similar in the two data sets (the difference is less than one percent) the discrepancy in earnings reflects differences in wages not hours. While the higher reported wage and salary income in the PSID is consistent with attrition of lower wage sample members, this difference may also reflect less underreporting of income in the PSID than the CPS in each year. In fact, when we compare mean family income in the CPS and the PSID in 1968, which predates attrition, we find the PSID value 7.0 percent higher than the CPS value. If differences in under reporting did not change over time then much of the 1989 difference reflects under reporting in the CPS.

The other two columns in Table 5 indicate that using the 1968 weights changes the PSID values somewhat but the changes are seldom large, which indicates that the attrition component of the weights does not move the means very much. The SRC-only unweighted estimates are usually farther from the CPS (though not always, e.g., for the variance of wage and salary income) but usually not by a very large amount. The proportion Hispanics drops further and mean earnings in the PSID is 6.8 percent larger than the CPS value ($21,100 versus $19,751)when the unweighted SRC sample is used. Using the combined SEO and SRC sample with the 1968 weights (that adjust for the original sample but not later attrition) reduces the gap between the CPS and PSID to 6.0 percent, but it is still larger than the 4.8 percent when the 1989 weights are used. The relative gap between the CPS and PSID measures

[48] This is about twice as large as the difference in earnings we

of family income are likewise increased from 10.3 percent when 1989 weights are used to 11.1 percent when the 1968 weights are applied to the same sample and to 11.7 percent when the unweighted SRC sample is used.

Tables 6 and 7 show similar patterns for the 1968 children who became wives and female heads in 1989. Most demographic 1989 weighted means from the PSID are similar to the means from the CPS. The percent Hispanics, however, continues to be underrepresented in the second generation of the PSID. This is particularly pronounced for female heads. The CPS shows 8 percent Hispanic while the PSID ranges from a high of 5 percent for the 1989 weighted SEO and SRC sample to a low of 2 percent for the unweighted SRC sample.

Economic characteristics likewise show similar patterns to those for male heads. The PSID consistently shows higher wage and salary income and family income, though hours are very similar in the two data sets. Turning to welfare participation we find that the PSID shows somewhat lower participation rates for female heads (.18 versus .21 when 1989 weights are used). This may reflect higher attrition of welfare recipients since better reporting of income in the PSID would lead to higher, not lower participation rates.

For female heads the PSID also shows substantially different racial distributions. The CPS shows more whites than the PSID when the 1989 sample weights are applied to the combined SRC and SEO sample (72 percent versus 66 percent), but the same proportion when the 1968 weighted sample is used. The fact that the adjustment of weights for non-random attrition increases the difference is puzzling. This

---

find for male heads 25-59 in our companion paper (Table 18).

pattern, however, is not limited to the second generation since we also find it in the broader sample analyzed in the extended version of our companion paper.

We conclude from these tables that there is in general good correspondence between the PSID and CPS with notable exceptions for race and welfare participation of female heads, percent Hispanic and mean earnings and family income for both heads and wives. Some of these differences are explainable by other factors, but attrition would seem to play a role, especially in the low welfare participation rates. While we have focused on the differences this should not obscure the fact that the large majority of measures are quite similar in the PSID and CPS.

Tables 5-7 also shed light on the importance of using sample weights. The fact that the gap between the CPS means and the 1989 weighted PSID means are almost always smaller than the gap between the CPS means and the 1968 weighted means shows the value of updating the weights to reflect the non-random attrition on the basis of observables. Note that this closing of the gap is not a necessary consequence of reweighting since the weights are recalibrated only on a subset of the variables in Tables 5 to 7 and the weights are not recalibrated on the sample in this age range[49].

The fact that no weights are available to correct for selection on observables is one potential explanation for the large gaps between the CPS and this sample. While some researchers have chosen to use only the SRC arguing that it

---

[49] For example, race is one of the variables used to construct the weights. But these weights are based on a number of factors, and are not likely to match the percent black in our age range. The fact that the 1989 weighted percent of male heads who are black in

was initially a random sample, this rationale becomes increasingly questionable as the SRC loses its representativeness and no weights are used to account for selection on observables. This problem will become increasingly important as a majority of the remaining SEO sample members are dropped from the sample as part of a cost saving effort[50].

Three factors help reconcile our previous finding that attritors and non-attritors had substantially different 1968 characteristics but that most measures of adult characteristics of the second generation in 1989 are similar in the CPS and PSID. First, the difference between attritors and non-attritors in Tables 2a and 2b will always be larger than the difference between non-attritors and the full sample, which is what the PSID/CPS comparison capture. For example, if attritors make up half of the combined sample then the difference between attritors and non-attritors is twice as large as the difference between non-attritors and the full sample. The latter is the relevant difference, since we are interested in the impact of using the non-attritors to make inferences on the full sample. Second, for adult outcomes that differ between parents and their children, such as earnings, the size of the effect of selection on parent's 1968 outcomes on children's 1989 outcomes will depend on the size of the intergenerational

---

the PSID is .11 while it is only .09 in the CPS indicates that the weights have over-adjusted for attrition of blacks in our sample.

[50] The PSID sample size has grown as the number of sample members lost through death or attrition has been less than the number of new offspring of the original sample members. In order to reduce the cost of following an increasing number of sample members the PSID has undertaken a number of cost savings measures. Among these is the decision to reduce the sample size by dropping 70 to 80 percent of the SEO sample, starting in the Spring of 1997.

correlation in outcomes. As long as there is some intergenerational reversion to the mean (i.e. the correlation between children's parent's outcomes is less than one) then the impact of selection on parents' outcomes will have a reduced impact on children's outcomes. Finally, the low R^2's indicate that even though some characteristics do predict attrition, much of the attrition is not associated with the variables in our equations.

IV. Impact of Attrition on Intergenerational Analysis

Thus far we have focused on the impact of attrition on the 1968 and 1989 mean characteristics of the children in the original PSID sample. We now turn to the impact of attrition on estimates of the intergenerational relationship between adult outcomes of the second generation and attributes of their parents.

We test whether the key intergenerational coefficient, α, in equation (8) is significantly different for non-attriting children and all children.[51] But as we argued earlier, the comparison group cannot include the full initial sample of children in the PSID, since some of these children will have attrited before becoming adults and, hence, cannot be used to estimate 9 and 9'. We must, therefore, limit our analysis to a test of differences in coefficients between equations estimated on a sample of all children who reached adulthood (i.e. those who survived the pre-period)and a sample of children who remained in the PSID through 1989. As noted

[51] We cannot test the difference between the PSID estimates and estimates from other data sets because any other longitudinal data set also potentially suffers from attrition bias and those cross sectional data sets that do provide information on both parents and their children rely on recall, another potential source of bias.

earlier, this test requires considerably stronger identifying assumptions than those used in the previous section.

The focus on the relationship between the adult outcomes of the second generation and their parents raises several measurement issues which we now address. We then turn to estimates of the relationship between parents' and adult children's earnings, education and welfare participation.

A. Sample Definition

Several issues of sample definitions arise in studying the effect of parental characteristics on child outcomes. One concerns a tradeoff between sample size and the age range of the children in the sample. If interest centers on the relationship between the second generation and parental characteristics (such as welfare receipt or marital status) during the child's formative years, then the sample should only include children who were young at the beginning of the panel. On the other hand, many of the child outcomes of interest can only be observed after the child becomes an adult. This argues for using a high initial age to insure that a sufficient number of children are observed as adults by the end of the panel. For example, with 20 years of data one can only estimate the impact of parental characteristics when the child was 0 to 3 on the child's adult outcome at 20 using three birth cohorts. Larger samples can be obtained, but only by increasing the child's age when the parent's outcome is measured or by lowering the age at which the second generation's outcome is observed.

This tradeoff between age span and sample size becomes more severe when studying relatively rare events, such as welfare receipt. With a limited number of children, there may be insufficient variation in the outcomes of interest to gain precise estimates, even in a longitudinal data set that covers as many years as the PSID.

A second issue is whether to define the pre-period used to estimate the intergenerational coefficient (α_1 in equation 9) using child outcomes in a specified year or when the child reaches a specified age. If more than one cohort of children is used these two methods are not equivalent. The advantage of defining the pre-period as a fixed number of years (at the end of which the child outcome is measured) is that all children must survive exactly S years to be included in the sample and all children have an equal number of years (T-S years) to attrite in the post-period. The disadvantage of this method is that equation (9) is estimated on a sample of children of varying ages. Since the relationship between parental and child outcomes may change as the child ages, this approach requires controls for the child's initial age, interacted with parental outcomes[52].

An alternative is to estimate equation (8) on a sample of children at a fixed age.[53] However, this implies that the pre-period is of different lengths for different children (because their ages differ in 1968) and that, therefore, some children have had more opportunity to attrite than others.[54]

---

[52] For example, Reville (1995) shows that the coefficient on father's earnings increases with the age of the son.

[53] Studies differ in this dimension, for example, Solon (1992) measure children's outcomes in 1984 and parent's outcome in 1967. Corcoran et al (1992) measure the child's outcome at age 25 and parents' outcomes at fixed years.

[54] As an example, consider estimating equation (9) when the child reaches age 24. A child who is 18 in 1968 is 24 in 1974 and thus will only have to stay in the PSID for six years to be included in the sample. This person will then have 14 years to potentially attrite in the post-period. In contrast, a person 12 years old in 1968 will have to survive 12 years to be included in the sample of

To control for the possibility that shortening the pre-period affects estimates of the coefficient on parental characteristics, we include the length of exposure to attrition , which is equivalent to conditioning on age in 1968, and its interaction with parental characteristics.

A related question concerns the appropriate length (in terms of either age or time) of the pre-period. A short pre-period allows less time for attrition to take place but also leads to small samples. For example, including only children who reach 24 by 1980 yields a substantially smaller sample than limiting the sample to children who reach 24 by 1985. Findings of insignificant differences between non-attritors and the full-sample in the post-period might be as much a result of sampling variability as evidence against attrition bias. Lengthening the pre-period increases sample size but also increases the possibility that the coefficient estimates in the pre-period are already contaminated by attrition. Our approach to these various sample definition issues, for which no single approach is preferable, is to conduct sensitivity tests by trying alternative approaches.

B. Attrition Probits

We begin by estimating a set of probit models on the sample of children who had not attrited before they reached age 24 and who reached 24 no later than 1980.[55] This gives a minimum of six years in the pre-period (for children who were 18 in 1968) and a minimum of nine years in the post-period (for children who reached

24 year-olds. This person will then have only 8 years in which to attrite in the post-period.

[55] In the probits that focus on average earnings we include children who did not attrite before age 26 since we calculate the child's average earnings between the ages of 24 and 26.

24 in 1980).[56] The dependent variable is equal to one if the child attrited after age 24. Since we want to preserve the symmetry between the regressions we estimate later and the probit equations we estimate in this section, we estimate separate probit equations that focus on education, earnings and welfare.

Table 8a presents probit coefficients and derivatives (averaged across all individuals) from attrition equations that include the child's education at age 24 and a set of control variables, including the parent's education in 1968, as covariates. Separate equations are estimated using father's characteristics (columns 1 and 2) and mother's characteristics (columns 3 and 4).

As we argued earlier, if lower education is associated with higher attrition probabilities then attrition will alter the density of child's education, conditional on parent's education and, hence, the conditional expectation function estimated in the next section. The results in this section show that attrition probabilities decline both with the child's education and the parent's education, whether measured by father's or mother's education. The statistically significant coefficient on child's education indicates that each additional year of education decreases the probability of attrition during the post-period by .02.

Columns 1 and 2 of Table 8b present estimated probit coefficients and derivatives for daughters' attrition after age 24, where the key covariates are daughter's welfare participation at age 24 and mother's welfare participation in 1968. While mother's education has a statistically significant effect on daughters' attrition in the post-period, the coefficient on

[56] The extended version of our paper provides estimates for

daughter's welfare receipt is not significantly different from zero.

Columns 3 and 4 focus on labor market incomes. In these equations we follow Solon (1992) by using a three year average of earnings for both father and son. These probit equations indicate that son's earnings is not a significant predictor of future attrition.

Based on the evidence in this section we conclude that attrition continues to be associated with observable characteristics, even in the sample we use to examine the relationship between adult outcomes of children and their parents. While attrition is random with respect to daughter's welfare and son's earnings at age 24, children's education is a significant predictor of future attrition. However, the R^2's also continue to be small, indicating that much of the attrition is not associated with the variables in our equations.

C. Impact of Attrition in the Post-Period

We next focus directly on estimates of equations (9) and (9') for the same three sets of outcomes of the second generation: education and welfare participation at age 24 (for persons who reach 24 by 1980) and average earnings between the ages of 24 and 26 (for persons who reach 26 by 1980). In order to gauge the effect of limiting the amount of attrition during the pre-period we estimate the models using two alternative ending dates (1977 and 1985). Including children who reach the specified age by 1977 results in a short (nine year) pre-period but it limits the sample size because the only children who will have reached 24 by 1977

---

children who reached 24 by 1977 and by 1985. Results are similar.

are children who were 15 to 18 in 1968. In order to increase the sample size we also estimate models for children who reached the specified age by 1980 and 1985. Each model is estimated for the full sample of children who had not attrited by the indicated year and the subset of that sample who did not attrite during the post-period (i.e. were still in the sample in 1989).

1. Father-Son Earnings

Table 9 presents a specimen regression in which the dependent variable is the log of the son's average earnings when he was 24 to 26[57]. The sample includes all sons who had not attrited and who were at least 26 years old by 1980. The first three columns present estimates from the full sample of sons who had not attrited by 1980. The remaining three columns present the same models estimated on the sample that had not attrited by 1989. White standard errors (in italics) as well as the OLS standard errors are shown.[58]

Columns (1) and (4) include only the log of the father's average 1968 to 1970 earnings as a regressor. The estimated coefficient on father's earnings for the full sample is .307, while the estimated coefficient among non-attritors is .336. The resulting difference in coefficients of .029 shows that the intergenerational coefficient is larger for the non-attriting sample but the difference is not significantly different from zero.[59] Model 2 controls for the race and

[57] Earnings is measured as annual labor income.
[58] Since these are very similar and since the computational cost of estimating White standard errors is high for the large number of models estimated in this section, we provide OLS standard errors in the remaining tables.
[59] It can be shown that the variance of the difference in coefficients for the total sample (column 1) and the non-attriting sub-sample (column 4)is equal to the difference in the variances.

education of the son in 1980, additional 1969 characteristics of the father, and as noted earlier, for the effect of the pre-period. These additions reduce the coefficient on father's earnings to .189 for the full sample and .218 for the non-attriting sub-sample. The difference is of the same sign but is again not significantly different from zero. Finally, columns (3) and (6) add interaction of years in the pre-period with father's earnings. Since the partial effect of father's earnings on sons earnings now depends on this interaction, derivatives are evaluated at the mean pre-year at the bottom on the table. The partial effects are .179 for the full sample and .196 for the non-attriting sample. The difference is again of the same sign but is not statistically different from zero.

Similar regressions are estimated for sons 26 by 1977 and by 1985 using both the full SRC and SEO samples and the unweighted SRC sample alone. The key partial effects of father's earnings on son's earnings for each of these regressions are shown in Table 10.[60] The coefficients on father's earnings for the model with no covariates (Model 1) are shown for the full sample and the non-attriting sample in columns (1) and (2). The differences in coefficients and their standard errors are shown in column (3). The partial effects for the model with the full set of covariates(Model 3) are shown in columns (4) and (5), with the differences and

---

This result is a special case of Hausman's (1978) result that the variance of the difference between two consistent estimators, when one is efficient, is the difference in the variance of the estimators. The standard error of the difference is, therefore, .029 based on White standard errors and .024 when based on OLS standard errors.

[60] A similar summary table for sons earnings in 1977, 1980 and 1985 and the full set of regression coefficients for all equations are

their standard errors in column (6). The top panel shows estimates based on the weighted SRC and SEO sample; the bottom panel presents results of unweighted regressions estimated on the SRC sample.

These models show that the partial effects of father's earnings on son's earnings are somewhat higher for non-attritors than the combined sample but the differences are never statistically significant. This is consistent with the attrition probits in Table 8b which showed no significant effect of son's earnings on later attrition. Most differences are relatively small. A few differences are large but this seems to a result of high sampling variability. The point estimates are also considerably larger for the weighted combined SEO/SRC sample than for the unweighted SRC sample, a pattern we will find in all the outcomes examined in this section.

We conclude that while attrition during the post-period does seem to increase the intergenerational coefficients in the sons' earnings regressions these differences are not statistically significant. As we argued earlier, a finding of no additional attrition bias during the post-period must be interpreted with caution since it provides no evidence about attrition bias in the pre-period. All that we can conclude is that our analysis has not uncovered evidence of statistically significant attrition bias in estimates of the intergenerational relationship between fathers' and sons' earnings.

---

available in the extended version of this paper. Results are similar.

2. Parent-Child Education

Tables 11 and 12 present summary information on the relationship between child education at age 24 and either the father's education (Table 11) or mother's education (Table 12.) These partial effects are again from equations with no other covariates (columns 1 and 2) and equations that control for race, sex of child and characteristics of the 1968 household in which the child resided (SEO, number of children, and marital status of head) as well as the length of the pre-period (columns 4 and 5)[61].

The results in Table 11 indicate that the estimates of the partial effect of the father's education on child's education are somewhat larger in the sample of non-attritors than in the full sample, which is again consistent with the significant effect of child's education on later attrition in Table 9. While these differences are statistically different from zero in five of the twelve regressions we estimate, these differences are not large. For example, the coefficient in the 1980 equation with covariates increase from .233 to .242 when the sample is limited to persons who do not later attrite. For the combined SEO and SRC sample (top panel) the statistically significant difference between non-attritors and the full sample (attritors plus non-attritors) are never more than five percent. Excluding the SEO increases these differences, yet the differences are never larger than seven percent.

The estimates of the partial effect of mother's education on child's education in Table 12 are again

[61] Estimates for the relationship between fathers' and sons' education as well as between mother's and daughter's education are

statistically significant in five out of the twelve equations but the differences between coefficients estimated for non-attritors and for the full sample, are not large. For the model with no covariates estimated on the combined SEO and SRC sample in 1985, the difference in partial effects (between non-attritors and the full sample) is .015, or five percent. With controls this difference is .013.

This evidence of attrition bias in estimates of the intergenerational relationship between educational attainment of children and their parents is consistent with the significant coefficients on child's education in the attrition probits shown in Table 8. More to the point, the small R^2's in Table 8 are also consistent with the small effect of attrition on estimates of the intergenerational coefficients in Tables 11 and 12.

3. Mother-Daughter Welfare

Table 13 presents Probit estimates of the relationship between the mother's and daughter's welfare participation. Similar to the previous tables, the differences in estimated coefficients for the non-attriting sample and the full sample are generally positive. With two exceptions, these point estimates indicate that, if anything, attrition tends to raise estimates of the coefficient on mother's welfare participation. None of these differences are, however, significantly different from zero for the combined SEO and SRC samples.

The bottom panel, for the unweighted SRC sample, however shows differences that are significant at the 10 percent

---

available in the extended version of this paper, which also provides the full set of coefficients for all equations.

level for two models. For example in the model for 1985 without other covariates (column 3), the estimated partial effect of mother's participation on daughter's participation is raised from .134 to .151, or by 13 percent, when the sample is limited to non-attritors. In the corresponding model with covariates the partial effect is raised from .106 to .109. These differences are not large in a substantive sense[62].

Summarizing the results from all three outcomes, there is evidence of statistically significant attrition bias for some outcomes (education and possibly welfare) but not others (earnings). Coefficients are almost always higher for the non-attritors than the full sample, indicating that attrition during the post-period leads to estimates of intergenerational coefficients that are too high. But the bias is not usually large.

The other consistent pattern in these tables is that the point estimates of the differences between the full sample and non-attritors are consistently larger in the SRC than in the SRC/SEO, no matter whether we examine earnings, education, or welfare participation. It is possible that this reflects greater attrition on the part of SEO sample members. However, an alternative explanation is that the use of sample weights mitigates attrition bias. Since the PSID provides sample weights that take attrition into account, weighted estimates can partially correct for selection on observables. Since similar weights are not provided in the PSID for the

[62] However, it should be noted that the decision whether to use the combined SEO and SRC samples would create substantially larger differences (compare the estimate of the partial effect of .116 for the combined sample in column 4 with the .064 for the SRC sample).

SRC sample, the SRC estimates are unweighted. If it is the lack of weights that is responsible for the difference in results then users should be cautious in limiting themselves to the SRC sample and the PSID staff should be encouraged to provide universal SRC weights, especially in light of the decision to drop most of the SEO sample.[63] Alternatively, as we noted above, because model-specific weights are preferable to universal weights, users should construct the former to correct for selection on observables.

V. Conclusions

Our study of the impact of sample attrition on the second generation of respondents in the PSID has led us to several conclusions. The first set of conclusions focus on the characteristics of the non-attriting children:

- Attrition has been high among the children of the original PSID sample members. By 1989 roughly half of the 1968 children were no longer in the panel.

- Attrition of children was associated with observable characteristics of their 1968 families. Children living in less advantaged households in 1968 were significantly more likely to attrite than children living in more advantaged households in 1968.

- While some observable characteristics are significant predictors of attrition, the low R^2's in predictive equations indicate that observed covariates do not account for much of the observed

[63] As we noted in our statistical discussion previously, neither universal nor model-specific weights correct for selection on unobservables; in their presence, it is in principle possible for weights to increase bias (Horowitz and Manski, 1997).

variation in attrition.

- The 1989 characteristics of the second generation of the PSID are similar to the characteristics of a sample of persons of the same age drawn from the CPS, a data set that does not suffer from attrition bias and is generally considered to be representative[64]. There is a close correspondence in characteristics for most demographic variables, especially when sample weights are used.

- The close correspondence between the CPS and PSID indicates that the PSID continues to be a useful data set for studies that focus on characteristics of children of the original sample families.

The second set of conclusions focus on the limitations in identifying the impact of attrition on estimates of the relationship between parents' characteristics, such as welfare participation, and the adult outcomes of their children:

- It is not possible to estimate the relationship between adult outcomes of parents and their adult children before any attrition occurs since, by definition, the second generation in the PSID were children in 1968. Therefore, all analysis of the effect of attrition on intergenerational coefficients require further identifying assumptions.

---

[64] The PSID non-response rate for the initial interview is also considerably higher than the non-response rate in the CPS.

- We develop some strong identifying assumptions under which we could reject the hypothesis that there was no attrition bias. However, the converse is not possible since we cannot rule out bias from attrition that occurred while the children were becoming adults. At best we can say that we have not found evidence of attrition bias for some outcomes.

These caveats must be kept in mind when interpreting the final set of conclusions that focus on the effect of attrition on estimates of intergenerational relationships:

- The intergenerational relationship between the earnings, education and welfare participation of parents and their adult children is larger for the subsample of children who do not attrite by the end of the panel than for the sample that includes all children who did not attrite before their mid-20's (but may have attrited afterwards).

- The differences in intergenerational coefficients are small in magnitude and not statistically different from zero for welfare and earnings. The differences for education are, however, statistically different from zero.

- The statistically significant differences are primarily found in the unweighted SRC results.

We have explored a limited set of outcomes and have found evidence of attrition bias in estimates of some coefficients (e.g. intergenerational education equations) but not others. This should come as no surprise since attrition may be random with respect to some outcomes but not others. One of the aims of this paper has been to provide a method for detecting the effects of attrition in intergenerational analysis. We strongly urge that this or alternative procedures be followed in any intergenerational analysis.

REFERENCES

Antel, John (1992). "The Intergenerational Transfer of Welfare Dependency." Review of Economics and Statistics 3, 467-73.

Altonji, Joseph G. (1994). "The Use of the Panel Study of Income Dynamics for Research on Intergenerational Transfers." Mimeographed, Northwestern University.

Becketti, S.; W. Gould; L. Lillard; and F. Welch. (1988). "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation." The Journal of Labor Economics 6: 472-492.

Behrman, Jere, R.; Paul Taubman. "The Intergenerational Correlation Between Children’s Adult Earnings And Their Parents’ Income: Results From The Michigan Panel Survey Of Income Dynamics." Review of Income and Wealth, 36, no. 2, June 1990: 115-127.

Couch, Kenneth, A.; Thomas A. Dunn. "Intergenerational Correlations in Labor Market Status: A Comparison of the United States and Germany." The Journal of Human Resources, 32: 210-232.

Cameron, A.C.; F.A.G. Windmeijer. "An R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models." Journal of Econometrics, 77, April 1997: 329-342.

Duncan, G.; M. Hill; S. Hoffman, (1988). "Welfare Dependence Within and Across Generations", Science, Vol. 239.

Fitzgerald, John; Peter Gottschalk; Robert Moffitt, (1996). "An Analysis of Sample Attrition in Panel Data: The Panel Study of Income Dynamics." mimeo, Johns Hopkins University.

Gottschalk, Peter (1996). "Is the Correlation in Welfare Participation Across Generations Spurious?" Journal of Public Economics 63: 1-25.

Hausman, J.; David Wise, (1981). "Stratification on Endogenous Variables and Estimation: The Gary Experiment", in C. Manski and D. McFadden (eds). Structural Analysis of Discrete Data: with Econometric Applications, (Cambridge: MIT Press)

Hausman, J. "Specification Tests in Econometrics." Econometrica 46, 1978.

Hill, M. The Panel Study of Income Dynamics: A User's Guide. Newbury Park, Ca.: Sage Publications, 1992.

Horowitz, J. and C. Manski. “Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations.” Journal of Econometrics, forthcoming.

Institute for Social Research. A Panel Study of Income Dynamics: Procedures and Tape Codes, 1989 Interviewing Year, Vol. I, Procedures and Tape Codes. Ann Arbor, Michigan, 1992.

Menchik, Paul, L. "Inter-generational Transmission of Inequality: An Empirical Study of Wealth Mobility." Econometrica, November 1979, 46: 349-62.

Reville, Robert, T. "Intertemporal and Life Cycle Variation in Measured Intergenerational Earnings Mobility." Brown University, May 1995.

Solon, G. "Intergenerational Income Mobility in the United States." American Economic Review 82, 1992: 393-408.

Solon, G.; M. Corcoran; R. Gordon; D. Laren, (1987). "A Longitudinal Analysis of Sibling Correlations in Economic Status." The Journal of Human Resources, Summer 1991.

Solon, Gary; Mary Corcoran; Roger Gordon; Deborah Laren. "The Association Between Men's Economic Status and Their Family and Community Origins." The Journal of Human Resources, Fall 1992, 27: 575-601.

Table 1: Numbers of Children Matched to Parents

I. Sons Aged 0-18 in 1968

Matched to	White	Black	Other
Both Parents	2068 (90.7)	1349 (61.4)	36 (67.9)
Mother Only	168 (7.7)	718 (32.7)	11 (20.8)
Father Only	9 (0.4)	26 (1.2)	0 (0.0)
Neither	36 (1.6)	105 (4.8)	6 (11.3)
Total	2281	2198	53

II. Daughters Aged 0-18 in 1968

Matched to	White	Black	Other
Both Parents	1897 (88.8)	1381 (60.9)	29 (72.5)
Mother Only	161 (7.5)	715 (31.6)	10 (25.0)
Father Only	9 (0.4)	22 (1.0)	0
Neither	70 (3.3)	148 (6.5)	1 (2.5)
Total	2137	2266	40

Notes: Column percent in parenthesis.

Table 2A: Mother/Daughter Pairs: 1968 Characteristics by Attrition Status

Variable	All Pairs Combined (1)	Daughter: Always In (2)	Daughter: Ever Out (3)	Mother: Always In (4)	Mother: Ever Out (5)	Both Always In (6)	Either Ever Out (7)
I. Daughter's race							
White	84.3	87.1	79.9*	86.7	82.1*	87.7	81.0*
Black	15	12.5	19.0*	13	16.6*	11.9	18
Other	0.7	0.4	1.1*	0.3	1.4*	0.3	1.0*
II. Mother's characteristics							
1. Marital status							
Married	87.5	89.6	84.2*	89	86.6*	90.6	84.8*
Single	1.1	1	1.3	1.2	1	1.1	1.2
Widowed	3.2	3.7	2.6	3.5	1.0*	3.2	2.5
Divorced/ Separated	7.4	5.3	10.5*	5.9	9.8*	4.7	10.3*
Married, spouse absent	0.8	0.4	1.5*	0.4	1.6*	0.4	1.3*
2. Welfare participation	2.7	1.9	4.0*	2.3	3.3*	1.8	3.9*
3. Education							
Missing or 0	4	3.7	4.6	3.2	5.0*	3.1	4.8*
< 12 yrs.	40.7	35.9	48.2*	36.3	47.1*	34.7	47.3*
= 12 yrs.	41	42.3	38.6*	43.4	38.4*	43.6	38.8*
13-15 yrs.	8.1	10	5.2*	9.4	5.3*	9.8	5.5*
16+ yrs.	6.2	8.1	3.4*	7.7	4.2*	8.8	3.7*
Education (years)	10.6	11	10.1*	11.0	10.1*	11.2	10.1*
4. Age	36	36.2	35.5*	35.9	34.6*	35.8	35.7
5. Whether positive labor income	0.43	0.432	.430	0.439	0.401*	0.429	0.425
6. Labor inc. for those w/ labor Inc. > 0	6173	6075	6357	6100	6627*	6055	6389
7. Std. dev. of labor inc. for those w/labor inc. >0	17861	19138	16303	18552	16496	19163	16584

III. Family characteristics							
Family income							
Mean	25966	27703	23305*	27190	24208*	27894	23827*
Std. Dev.	62749	66263	57551	60923	68449	64513	60972
Percentile Points							
P20/P50	0.61	0.60	0.62	0.61	0.63	0.63	0.61
P40/P50	0.86	0.89	0.9	0.89	0.9	0.89	0.89
P60/P50	1.14	1.14	1.15	1.15	1.15	1.14	1.15
P80/P50	1.51	1.52	1.58	1.53	1.56	1.5	1.55
Income/ Need ratio							
Mean	2.04	2.19	1.82*	2.15	1.92*	2.21	1.87*
Standard dev.	4.83	5.43	3.89	5.03	4.48	5.29	4.22
Percentile points							
P20/P50	0.54	0.57	0.53	0.56	0.54	0.59	0.52
P40/P50	0.84	0.84	0.86	0.84	0.86	0.85	0.84
P60/P50	1.14	1.13	1.16	1.13	1.14	1.13	1.15
P80/P50	1.58	1.53	1.6	1.54	1.58	1.54	1.61

Notes: Ever out indicates missing in at least one year for reasons other than death. P indicates the percentile for parent child pairs in the indicated column. Asterik indicates a difference in means between Always In and Ever Out at the 10 percent significance level. Significance levels are not shown for standard deviations or percentile points.

Table 2B: 1968Characterisitcs of Father Son Pairs by Ever Attrite

Variable	All Pairs Combined (1)	Son: Always In (2)	Son: Ever Out (3)	Father: Always In (4)	Father: Ever Out (5)	Both Always In (6)	Either Ever Out (7)
I. Son's race							
White	88.8	92.6	84.2*	92.2	85.0*	93.8	84.2*
Black	10.5	7.1	14.5*	7.5	13.8*	6	14.7*
Other	0.7	0.3	1.2*	0.3	1.2*	0.2	1.1*
II. Father's characteristics							
1. Marital status							
Married	95	95.3	94.6	96.7	92.5*	96.7	93.1*
Single	0.8	1.0	0.6	0.6	1.3*	0.8	1.0
Widowed	0.6	0.8	0.5	0.8	0.4	0.8	0.5
Divorced/ Separated	2.8	2.2	3.5*	0.9	5.2*	0.8	4.5*
Married, spouse absent	0.8	0.7	0.8	0.9	0.5	0.8	0.8
2. Welfare participation	1.9	1.6	2.3	1.1	3.0*	0.9	2.7*
3. Education							
Missing or 0	9.1	8.7	9.3	6.9	13.1*	6.5	11.8*
< 12 yrs.	40.1	34.4	46.8*	31.6	46.9*	30.1	46.4*
= 12 yrs.	27.1	27.9	26.3	30.9	24.0*	31.2	24.7*
13-15 yrs.	10.5	11.8	9.3*	12.4	8.7*	12.7	9.0*
16+ yrs.	13.2	17.2	8.4*	18.1	7.2*	19.5	8.0*
Average Years	10.2	10.8	9.6*	11.2	9.0*	11.4	9.2*
4. Age	38.6	38.7	38.7	37.6	37.5	37.5	38.3*
5. Whether positive labor income	0.919	0.922	0.917	0.951	0.882*	0.958	0.895
6. Labor inc. for those w/ labor inc. > 0	20520	21519	19316*	21872	19084*	22200	19132*
7. Std. dev. of labor inc. for those w/ labor inc. >0	34658	37243	31342	36503	31319	38035	31440

III. Family characteristics							
Family income							
Mean	27132	27666	26557*	28705	25749*	28819	25980*
Standard. dev.	65831	62389	69240	63667	73662	67123	66577
Percentile Points							
P20	0.66	0.65	0.65	0.66	0.64	0.65	0.65
P40	0.89	0.9	0.88	0.9	0.88	0.87	0.88
P60	1.15	1.15	1.12	1.13	1.11	1.1	1.12
P80	1.52	1.51	1.5	1.49	1.46	1.44	1.52
Income/ Need Ratio							
Mean	2.13	2.2	2.06*	2.3	2.01*	2.33	2.02*
Standard. dev.	5.09	5.4	4.73	5.55	4.86	5.83	4.63
Percentile Points							
P20	0.6	0.61	0.57	0.63	0.57	0.64	0.56
P40	0.87	0.87	0.86	0.88	0.87	0.87	0.85
P60	1.14	1.15	1.14	1.12	1.18	1.14	1.16
P80	1.6	1.62	1.62	1.62	1.59	1.6	1.62

Notes: Ever out indicates missing in at least one year for reasons other than death. P indicates the percentile for parent child pairs in the indicated column. Asterik indicates a difference in means between Always In and Ever Out at the 10 percent significance level. Significance levels are not shown for standard deviations or percentile points.

Table 3: Distribution of Attriting Chidren by Whether Attrited After Splitoff

Attrited	Daughters and Sons		Daughters		Sons	
	Number (1)	Percent (2)	Number (3)	Percent (4)	Number (5)	Percent (6)
After splitoff	1083	26.4	507	26.8	575	26.3
During splitoff	970	23.8	448	23.7	522	23.9
Before splitoff	2029	49.7	939	49.6	1090	49.8
Total	4082	100.0	1894	100.0	2187	100.0

Notes: "During splitoff" indicates that the person moved out and was not successfully interviewed after the move.

Table 4: Attrition Probit for Children Ever Out--Children 0-18 in 1968

	Child (Parent=Father)		Child (Parent =Mother)	
	Coefficient	$\partial P/\partial X$	Coefficient	$\partial P/\partial X$
Intercept	.406*** (.136)	.158	.475*** (.126)	.183
Income/Needs	.028 (.029)	.011	-.048*** (.017)	-.018
SEO	.158*** (.042)	.061	.136*** (.038)	.052
Black	.111*** (.041)	.042	.152*** (.037)	.059
Parent's Education	-.021*** (.005)	-.008	-.024*** (.005)	-.009
Parent's Labor Income $x10^{-6}$	-.009*** (.003)	-.004	.017*** (.004)	.006
Parent No Labor Income	-.272*** (.078)	-.106	.215*** (.037)	.083
Number Children in Family	-.0001 (.009)	-.0001	-.017*** (.007)	-.007
Parent Never Married	-.309** (.164)	-.120	.046 (.084)	.018
Parent Widowed	.124 (.193)	.048	-.067 (.071)	-.026
Parent Divorced/Separated	-.003 (.100)	-.001	.228*** (.042)	.088
Parent's Age in 1968	-.007*** (.002)	-.003	-.011*** (.002)	-.004
Child's Age ≤12 (Spline)	.011** (.005)	.004	.015*** (.005)	.005
Child's Age >12 (Spline)	-.015*** (.006)	-.006	-.024*** (.006)	-.009
Child is Male	.138*** (.032)	.053	.169*** (.028)	.065
Sample Size	6303		8088	
Number Ever Out	2921		3926	
Log Likelihood	-4264.1		-5434.8	
R^2	.021		.030	

Notes: All characters measured in 1968. Standard errors in parenthesis. Asteriks denote significance level 1% (***), 5% (**) or 10% (*). Derivatives evaluated for each individual and average. R^2 equals one minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept.

Table 5

Characteristics of Male Heads 22-39: 1989

PSID and CPS

	CPS	PSID		
		Current Weights (SRC and SEO)	1968 Weights (SRC and SEO)	Unweighted (SRC only)
Age	31.5	31.5	31.3	31.3
Race				
White	.88	.88	.91	.93
Black	.09	.11	.08	.07
Hispanic	.08	.04	.03	.02
Education				
Less than 12	.12	.14	.13	.12
12	.38	.34	.34	.34
13-15	.22	.25	.24	.25
16+	.28	.28	.28	.29
Marital Status				
Never married	.19	.18	.19	.18
Married	.74	.73	.73	.73
Divorced/separated	.07	.09	.08	.08
Widowed	0	0	0	0
Region				
Northeast	.19	.20	.21	.19
North Central	.25	.27	.28	.29
South	.34	.32	.31	.33
West	.22	.19	.19	.18
Own Home	.56	.54	.55	.56
Labor Force				
Positive weeks worked	.94	.98	.98	.99
Conditional weeks worked	48.8	46.8	46.9	47.0
Conditional annual hours worked	2162	2182	2191	2220
Earnings				
Conditional real wage and salary	19751	20698	20940	21100
Conditional real labor income	--	20732	21323	21065

Table 5 continued

	CPS	PSID		
		Current Weights (SRC and SEO)	1968 Weights (SRC and SEO)	Unweighted (SRC only)
Family Income	28836	31812	32024	32220
Wage and Salary Distribution				
(Earners Only)				
Variance of log	.605	.752	.738	.687
Percentiles				
20th Percentile/Median	.565	.541	.554	.562
40th Percentile/Median	.870	.813	.843	.843
60th Percentile/Median	1.152	1.138	1.125	1.125
80th Percentile/Median	1.522	1.504	1.536	1.526
Welfare Participation	.02	.02	.02	.01

Table 6

Characteristics of Wives 22-39: 1989

PSID and CPS

	CPS	PSID		
		Current Weights (SRC and SEO)	1968 Weights (SRC and SEO)	Unw d (S only)
Age	31.4	31.4	31.2	3
Race				
White	.88	.89	.91	.
Black	.08	.08	.07	.
Hispanic	.09	.05	.04	.
Education				
Less than 12	.11	.12	.11	.
12	.44	.40	.40	.
13-15	.22	.26	.26	.
16+	.23	.22	.22	.
Region				
Northeast	.20	.23	.24	.
North Central	.25	.26	.26	.
South	.35	.32	.32	.
West	.21	.17	.17	.
Own Home	.68	.70	.71	.
Labor Force				
Positive weeks worked	.75	.80	.80	.
Conditional weeks worked	43.3	41.8	41.9	4
Conditional annual hours worked	1588	1571	1576	1
Earnings				
Conditional real wage and salary	11199	–	–	

Conditional real labor income	--	11641	11740	11

Table 6 continued

	CPS	PSID		
		Current Weights (SRC and SEO)	1968 Weights (SRC and SEO)	Unw d (S only)
Family Income	32949	38058	38402	38
Wage and Salary Distribution*				
(Earners Only)				
Variance of log	1.442	1.090	1.088	1.
Percentiles				
20th Percentile/Median	.375	.385	.385	.
40th Percentile/Median	.833	.769	.769	.
60th Percentile/Median	1.250	1.153	1.154	1.
80th Percentile/Median	1.833	1.769	1.780	1.
Welfare Participation	.02	.02	.02	.

* Labor Income is used for wives because wage and salary are not available.

Table 7

Characteristics of Female Heads 22-39: 1989

PSID and CPS

	CPS	PSID		
		Current Weights (SRC and SEO)	1968 Weights (SRC and SEO)	Unw d (S only)
Age	30.5	31.1	30.8	3
Race				
White	.72	.66	.72	.
Black	.26	.33	.27	.
Hispanic	.08	.05	.04	.
Education				
Less than 12	.15	.19	.17	.
12	.38	.34	.35	.
13-15	.23	.26	.26	.
16+	.24	.21	.22	.
Marital Status				
Never married	.53	.54	.54	.
Married	0	0	0	
Divorced/separated	.45	.43	.43	.
Widowed	.02	.03	.03	.
Region				
Northeast	.21	.21	.20	.
North Central	.24	.26	.27	.
South	.33	.33	.33	.
West	.22	.19	.19	.
Own Home	.23	.25	.27	.
Labor Force				
Positive weeks worked	.82	.86	.87	.
Conditional weeks worked	45.7	44.0	44.2	4

Conditional annual hours worked	1833	1862	1862	1

Table 7 continued

	CPS	PSID		
		Current Weights (SRC and SEO)	1968 Weights (SRC and SEO)	Unw d (S only)
Earnings				
Conditional real wage and salary	13393	14118	14521	14
Conditional real labor income	--	14250	14659	14
Family Income	14247	17063	17647	17
Wage and Salary Distribution (Earners Only)				
Variance of log	1.185	1.032	1.018	.
Percentiles				
20th Percentile/Median	.425	.471	.494	.
40th Percentile/Median	.844	.812	.823	.
60th Percentile/Median	1.194	1.118	1.117	1.
80th Percentile/Median	1.659	1.518	1.559	1.
Welfare Participation	.21	.18	.16	.

Table 8a: Attrition Probit: Child Attrites after Age 24. Focus on Education

Variable	(1) Coefficient	(2) ∂P/∂X	(3) Coefficient	(4) ∂P/∂X
Intercept	1.059*** (.383)	.281	.819*** (.309)	.232
Father's Education	-.021* (.012)	-.006		
Mother's Education			-.021* (.012)	-.006
Black	.192** (.096)	.051	.315*** (.081)	.089
Number of Children in FU in 1968	.013 (.019)	.003	.004 (.017)	-.001
Parent Married in 1968	-.157 (.170)	-.042	-.260*** (.081)	-.074
Child's Education at age 24	-.075*** (.023)	-.020	-.066*** (.019)	-.019
Years Pre	-.089*** (.020)	-.023	-.066*** (.017)	-.019
Child is Male	.142* (.080)	.038	.205*** (.069)	.058
Number of Observations	1334		1709	
Number who Attrited	266		390	
R^2	.048		.059	
Log Likelihood	-634.496		-863.806	

Notes: Sample of children who had not attrited by age 24 and who were 24 in 1980 or earlier. Child's characteristic measured at age 24 and parents characteristics measure in 1968. Standard errors in parenthesis. Asterisks denote statistically significant differences from zero at the 1% (***), 5%(**) 0 10% (*) level. Derivatives evaluated for each individual and averaged. R^2 equals one minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept.

Table 8b: Attrition Probit: Child Attrites after Age 24. Focus on Welfare Receipt of Daughters and Earnings of Sons

	Welfare Mothers/Daughters		Earnings Fathers/Sons	
	Coefficient	∂P/∂X	Coefficient	∂P/∂X
Intercept	.241 (.558)	.064	3.25 (3.88)	.872
Mother's Welfare in 68	.278 (.221)	.074		
Black	.124 (.121)	.033	.504** (.214)	.135
Number of Kids in FU in 68	.020 (.026)	.005	.001 (.043)	.0003
Parent Married in 68	-.165 (.118)	-.044	.089 (.421)	.024
Mother's Education	-.041** (.017)	-.011		.006
Mothers Worked in 68	.161 (.103)	.043		.083
Parent's Age in 68	-.004 (.008)	-.001	-.034 (.125)	-.009
Year Prior to 24	-.068 (.027)	-.018	-.143** (.062)	-.038
Daughter's Welfare Age 24	.110 (.144)	.029		
Father's Labor Income (avg. 68, 69, 70)			.003 (.157)	.0008
Father's Age^2 in 68			.0003 (.001)	.000
Son's Education at Age 26			-.117** (.049)	-.031
Son's Earnings (avg at age 24, 25, 26)			-.055 (.169)	-.015
Sample Size (Number Out)	847 (168)		315 (68)	
R^2	.038		.078	
Log Likelihood	-405.76		-151.442	

Notes: Sample of children who had not attrited by age 24 and who were 24 in 1980 or earlier. Child's characteristic measured at age 24 and parents characteristics measure in 1968. Standard errors in parenthesis. Asterisks denote statistically significant differences from zero at the 1% (***), 5%(**) 0 10% (*) level. Derivatives evaluated for each individual and averaged. R^2 equals one minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept.

Table 9: Son's Log Earnings at Age 26 on Father's Log Earnings in 1968--Son 26 by 1980

	All			Non-Attrite		
	Model 1 (1)	Model 2 (2)	Model 3 (3)	Model 1 (4)	Model 2 (5)	Model 3 (6)
Intercept	6.57*** (.433) *(.465)*	8.19*** (1.22) *(1.04)*	13.7*** (3.48) *(3.76)*	6.31*** (.494) *(.546)*	7.66*** (1.42) *(1.12)*	14.9*** (3.82) *(4.21)*
Father's Log Earnings (3 yr. Avg)	.307*** (.045) *(.048)*	.189*** (.053) *(.053)*	-.386 (.346) *(.365)*	.336*** (.051) *(.056)*	.218*** (.063) *(.062)*	-.548 (.384) *(.404)*
Black		-.177** (.084) *(.082)*	-.175** (.084) *(.082)*		-.243** (.109) *(.117)*	-.251** (.108) *(.117)*
Child's Education in 1980		.048*** (.016) *(.017)*	.049*** (.016) *(.017)*		.052*** (.018) *(.019)*	.053** (.017) *(.018)*
Number of Children in FU in 68		-.002 (.016) *(.015)*	-.005 (.016) *(.015)*		.006 (.019) *(.019)*	.002 (.019) *(.019)*
Head Married		-.171 (.156) *(.135)*	-.153 (.156) *(.137)*		-.300 (.188) *(.147)*	-.267 (.188) *(.152)*
Father's Age in 1968		-.034 (.043) *(.036)*	-.030 (.043) *(.036)*		-.029 (.051) *(.044)*	-.019 (.051) *(.051)*
Father's Age Squared in 1000s		.345 (.448) *(.381)*	.307 (.448) *(.373)*		.326 (.530) *(.445)*	.224 (.529) *(.426)*
SEO		-.045 (.078) *(.079)*	-.048 (.078) *(.079)*		-.021 (.095) *(.103)*	-.020 (.094) *(.103)*
Years Pre		-.006 (.021) *(.022)*	-.538* (.318) *(.351)*		.004 (.024) *(.026)*	-.700** (.349) *(.391)*
Years Pre * Father's Labor			.055 (.033) *(.036)*			.072** (.036) *(.040)*
$\partial E_s/\partial E_f$ at Mean Years Prior			.179*** (.054)			.196*** (.063)
R^2	0.13	0.19	0.2	0.15	0.22	0.23
Sample Size	315	315	315	247	247	247

Notes: Son at least age 26 in 1980 (Age 14-18 in 1968). Son's labor income is 3 years average at age 24-26. OLS standard errors in parenthesis. White standard errors in italics.