The B.E. Journal of Economic Analysis & Policy

Contributions

Volume 10, *Issue* 1 2010 *Article* 102

What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents

Anders Björklund* Lena Lindahl†

Matthew J. Lindquist‡

*Stockholm University, anders.bjorklund@sofi.su.se

†Stockholm University, lena.lindahl@sofi.su.se

‡Stockholm University, matthew.lindquist@sofi.su.se

Recommended Citation

Anders Björklund, Lena Lindahl, and Matthew J. Lindquist (2010) "What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents," *The B.E. Journal of Economic Analysis & Policy*: Vol. 10: Iss. 1 (Contributions), Article 102.

Available at: http://www.bepress.com/bejeap/vol10/iss1/art102

What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents*

Anders Björklund, Lena Lindahl, and Matthew J. Lindquist

Abstract

Sibling correlations are broader measures of the impact of family and community influences on individual outcomes than intergenerational correlations. Estimates of such correlations in income show that more than half of the family and community influences that siblings share are uncorrelated with parental income. We employ a data set with rich family information to explore what factors in addition to traditional measures of parents' socio-economic status can explain sibling similarity in long-run income. Measures of family structure and social problems account for very little of sibling similarities beyond that already accounted for by income, education and occupation. However, when we add indicators of parental involvement in schoolwork, parenting practices and maternal attitudes, the explanatory power of our variables increases from about one-quarter (using only traditional measures of parents' socio-economic status) to nearly two-thirds.

KEYWORDS: family background, intergenerational mobility, long-run income, parents, sibling correlations, siblings, socio-economic status

---

*We would like to thank Paul Devereux, Bhashkar Mazumder, John van Reenen and seminar participants at the London School of Economics, SOFI, the University of Edinburgh and the University of Zurich for their valuable comments and suggestions. Björklund and Lindahl acknowledge financial support from the Swedish Council for Working Life and Social Research (FAS).

1. Introduction

During the past 10 to 20 years, there has been an upsurge in empirical research by economists concerning the relationship between family background and income during adulthood. Most of this research has focused on the intergenerational relationship between parents' and offspring's long-run income and, most notably, the relationship between fathers' and sons' income.[1] Many researchers are motivated to study this in order to gauge the degree to which a society promotes equality of opportunity. Using Roemer's (1998) terminology, the argument is that family background represents "circumstances" that members of the offspring generation have not chosen themselves, in contrast to their own "effort."[2] Thus, a strong dependence of outcomes, such as income during adulthood, on family background implies low equality of opportunity.

As important as this is, it is somewhat surprising that relatively little research has been devoted to exploring sibling correlations with regard to income. The literature on intergenerational mobility has recognized for quite some time that a correlation between siblings is, in fact, a broader measure of the importance of family background and community effects than the parent-offspring association.[3] This is mainly due to the simple fact that siblings share much more than their parents' income. Siblings represent a broad set of "circumstances" in life that people have not chosen themselves.

A sibling correlation in an outcome such as income has two properties that make it particularly informative and useful for a discussion about the importance of family background and community influences. First, from a simple decomposition of permanent income into a family and an individual component, it follows that a sibling correlation tells us what fraction of total inequality is attributable to the family and community component shared by siblings. Second, the relationship between the sibling correlation in income and the corresponding intergenerational correlation (IGC) is as follows:

Sibling correlation = $(\text{IGC})^2$ + other shared factors that are uncorrelated with parental income.

[1] See Björklund and Jäntti (2009) for a recent survey.
[2] See Bourguignon et al. (2007) for an illuminating empirical application of Roemer's (1998) approach.
[3] This insight goes back at least to Corcoran et al. (1976). See also Erikson (1987) and Sieben and De Graaf (2003) for sociological approaches using occupational and educational variables. Solon (1999) offers a formal exposition of the interpretation of the sibling correlation and its relationship to intergenerational associations discussed here.

The few studies of sibling correlations in long-run income have estimated them to be around 0.45 for the United States and around 0.25 for the Nordic countries.[4] Estimates of the IGC have centered around 0.4 for the United States and around 0.2 for the Nordic countries. Plugging these numbers into the above equation shows us that more than half of the family and community background influences that siblings share are not even correlated with parental income. Thus, the strong focus by economists in the current literature on intergenerational relationships in income is like focusing only on the tip of the iceberg. There is much more below the surface that needs to be explored in order to understand the circumstances that are important for generating sibling similarities in long-run outcomes such as income. The goal of this study is to fill some of this gap in the literature.

One reasonable hypothesis would be that it is neighborhood characteristics shared by siblings that dominate among the "other" shared factors that explain income. However, a few recent studies, covering Norway, Sweden and the United States, have all found that such factors are (relatively) unimportant.[5] Therefore, it must be something within the family in addition to parental income that accounts for the relatively high sibling correlations in income. Investigating the contents of this "something" is what we intend to do in this paper. More specifically, we attempt to answer the following question: What factors beyond traditional measures of socio-economic status (income, education and occupation) are responsible for generating positive and significant sibling correlations in income?

To do this, we make use of data from the *Stockholm Birth Cohort Study* (SBC). This data source contains a rich set of variables concerning individual, family, social and neighborhood characteristics. The data set includes all children who were born in 1953 and living in the greater Stockholm metropolitan area on November 1, 1963. Cohort members' siblings have been matched onto the data set along with income during adulthood. One feature of the SBC study that is particularly interesting for the problem at hand is that it includes survey data from interviews with a selected sample of mothers to cohort members. The Family Study was conducted in 1968 and provides information concerning parental involvement in schoolwork, child rearing strategies and maternal attitudes.

[4] See Solon et al. (1991) and Mazumder (2008) for U.S. estimates and Björklund et al. (2002) for a comparative study of the U.S. and the Nordic countries. Björklund et al. (2009) report recent Swedish estimates.

[5] See Raaum et al. (2006) for Norway, Solon et al. (2000) and Page and Solon (2003a, 2003b) for the United States, and Lindahl (forthcoming) for Sweden. The underlying idea in these studies (proposed by Solon et al. 2000) is that a correlation in adult outcomes among children who have grown up in the same neighborhood is an upper bound on the importance of factors that neighbors share. This upper bound, in turn, is found to be low compared to the sibling correlation that captures family as well as neighbor factors. Neighbor factors are thus small relative to family factors.

The remainder of the paper is structured as follows: In Section 2, we offer a more detailed explanation of what a sibling correlation is (and isn't) and explain how it can be estimated. We also describe our econometric approach to disentangling the determinants of this correlation and we discuss how this approach differs from the more traditional regression approach used in previous studies. Section 3 contains a discussion of previous related studies, which guide us in our search for explanatory variables. Section 4 describes the data source in more detail. Our empirical results are reported in Section 5, followed by a series of sensitivity analyses in Section 6. We conclude in Section 7 with a summary and brief discussion.

2. Exploring Sibling Correlations: Models and Methods

To clarify the useful interpretation of the sibling correlation, suppose that we have an outcome measure such as long-run income at our disposal. This variable, y_{ij}, for sibling j in family i can be modeled as

$$y_{ij} = \mu + a_i + b_{ij}, \tag{1}$$

where μ is the population mean, a_i is a permanent component common to all siblings in family i, and b_{ij} is a permanent component unique to individual j in family i, which captures individual deviations from the family component. If we assume that these two components are independent, then the variance of y_{ij} can be written as the sum of the variances of the family and individual components:

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2. \tag{2}$$

The share of the variance in the outcome variable, y_{ij}, which can be attributed to family background effects, is

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{3}$$

This share coincides with the correlation in the outcome variable of randomly drawn pairs of siblings, which is why ρ is called a sibling correlation.[6]

[6] Although this common formulation of the sibling correlation is a well-defined statistical entity, it is important to keep in mind that it is still a somewhat arbitrary model of the process which

A sibling correlation can thus be thought of as an omnibus measure of the importance of family background and community effects. It includes anything shared by siblings: parental income and parental influences such as aspirations and cultural inheritance, as well as things not directly experienced in the home, such as school, church and neighborhood effects. Genetic traits not shared by siblings, differential treatment of siblings, time-dependent changes in neighborhoods, schools, etc. are captured by the individual component b_{ij} . If such non-shared factors are relatively more important than shared factors for incomes, the variance of the family effects will be small relative to the variance of the individual effects, and the sibling correlation will be low. The more important the effects that siblings share are, the larger is the sibling correlation.

In order to calculate the sibling correlation in long-run income, ρ, we need estimates of the within-family variation, σ_b^2, and the between-family variation, σ_a^2. These can be obtained using the following mixed-effects model

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + a_i + b_{ij} . \tag{4}$$

This formulation allows for the inclusion of multiple control variables $\mathbf{x}_{ij}$. In the present exercise, for example, all of our estimates of the sibling correlation include gender and birth-year dummies in the vector $\mathbf{x}_{ij}$. Such a model is the prototypical one used in the previously discussed studies on sibling correlations. A minor exception is that we directly use a measure of long-run income, whereas most previous studies employ annual income and therefore sometimes add a transitory error component to model (4).

Our contribution in this paper is to include potentially important family-wide variables in the $\mathbf{x}_{ij}$ vector, either one at a time or simultaneously. For example, consider the inclusion of parental income in $\mathbf{x}_{ij}$. This additional control variable should reduce the residual variation in the outcome variable and produce a lower estimate of the between-family variation, σ_a^{2*}, than the estimate produced without the added control for parental income. Abstracting from measurement error (for the moment) enables us to interpret the difference between these two estimates, $\sigma_a^2 - \sigma_a^{2*}$, as an upper bound on the amount of the variance in the family component that can be explained by parental income. It is only viewed as an upper bound on the importance of parental income, since it includes other factors that are correlated with parental income. In the presence of measurement error, this difference, $\sigma_a^2 - \sigma_a^{2*}$, is more correctly viewed as a downwardly biased estimate of the upper bound on the amount of the variance in the family component that can be explained by parental income.

---

generates sibling similarities. It rules out, for example, the fact that sibling correlations could, in theory, be negative.

This experiment also produces a new sibling correlation ρ*. From what we know about the relationship between intergenerational and sibling correlations, we expect this new sibling correlation to be lower, but still substantial in magnitude.

The central question that we address in this paper is: What family characteristics other than traditional measures of socio-economic status (income, education and occupation) help us to explain sibling similarities in long-run income? To answer this question, we continue adding variables to the $\mathbf{x}_{ij}$ vector in order to produce new estimates of the between-family variation, σ_a^{2}**. We interpret the difference (σ_a^{2}* - σ_a^{2}**) as the added importance of the new variable(s) above and beyond those already accounted for by previously included variables.[7]

There is an interesting difference between our approach and the more standard regression approach used to study the importance of parent and family characteristics for children's outcomes (e.g., Datcher 1982, Yeung et al. 2000, Fryer and Levitt 2004, Mason 2007).[8] Our focus is on the subset of parental and family characteristics that make siblings more similar to each other (i.e., variables that influence outcomes through a_i). After all, we are trying to "explain" the sibling correlation in long-run income. As we shall see, not all family-wide characteristics that help predict (average) children's outcomes also lower the sibling correlation. In this sense, the purpose of our paper is somewhat more focused than previous studies which use standard regression techniques. Our goal is to identify a set of parental characteristics that matter for long-run outcomes *and* make the long-run outcomes of siblings more similar. We are interested in identifying background characteristics which perpetuate inequality across families.

We now turn to a discussion of previous literature that can guide us in the search for such variables.

[7] Mazumder (2008) has inspired us to pursue this approach. He adds a two-year average of parental income as one additional variable and finds a 36 percent reduction in the sibling correlation. He then also adds non-monetary characteristics of various types, but these are all variables pertaining to the offspring (the siblings). Our interpretation is that the addition of offspring variables in the $\mathbf{x}_{ij}$ vector addresses a different question than the one we are interested in. Mazumder's approach sheds light on the question via which variables parental income has an impact, whereas our approach explores which parental characteristics are important for sibling similarities in long-run income.

[8] Although Mayer (1997) is mainly concerned with estimating the causal impact of parental income on children's income, in Chapter 4 of her book, she discusses at length the traditional regression approach for studying the impact of parental characteristics on children's outcomes.

3. Previous Literature and Our Choice of Family-Wide Variables

The challenge is to find family background characteristics that can account for sibling similarities beyond those created by parental income, education and occupation. Because most favorable parental characteristics that have an impact on children's income are also likely to affect parents' own income, educational attainment and choice of occupation, it is hard to say *a priori* what characteristics are particularly important. This fact also underscores the exploratory nature of our investigation.

One group of potentially important variables can be labeled *family structure*. Variables such as parental separation, number of siblings and mother's age at first birth are strong correlates of children's achievements during adulthood (Mayer 1997).[9] Interestingly, Fryer and Levitt (2004), in their search for variables that can explain black-white differences in early test scores in the United States, find that such variables have explanatory power, even conditional upon the socio-economic status of the parents. However, there appears to be a growing consensus in the literature that many variables concerning family structure (e.g., parental separation and teenage motherhood) are correlated with child outcomes not primarily due to causal effects, but by serving as an indicator of underlying characteristics that predict weak outcomes.[10] In our data set, presented in the next section, we explore and include a number of variables concerning family structure.[11]

One could also argue that a standard measure of parental income, or even a broader measure of parents' socio-economic status, does not fully capture the intergenerational impact of an array of *social problems* that some families suffer from. Variables such as social assistance recipiency, drug and alcohol abuse, mental illness and father's criminality are generally found to be intergenerationally connected.[12] Such variables might serve as indicators of underlying characteristics that have negative effects on offspring's labor market

[9] Hill et al. (2001) discuss how the timing of changes in family structure over the course of childhood is relevant, and that different types of changes may affect girls and boys differently.

[10] See, e.g., Geronimus and Korenman (1992) and Ginther and Pollak (2004) for U.S. evidence and Holmlund (2005) and Björklund and Sundström (2006) for Swedish evidence. See Mayer (1997) for a more general discussion.

[11] As discussed in Section 2 (and in Hill et al. 2001), some family-wide variables may affect children differently. Parental separation, for example, may affect an older sibling (who may have already left the home) less than a younger sibling who experiences the separation more directly. In fact, it is this type of age-related, differential treatment effect that motivates us (in our empirical analysis) to focus on siblings that are born no more than nine years apart from each other.

[12] See, for example, Case and Katz (1991) and Duncan et al. (2005) for U.S. evidence. Hjalmarsson and Lindquist (2009, 2010) find strong intergenerational patterns in drunk driving and other types of criminal behavior using the same Swedish data set we use.

achievement. These intergenerational associations may also reflect causal effects of the specific characteristic; for example, parents' social assistance recipiency or criminal behavior might affect their children directly through role modeling (Mayer 1997, Duncan et al. 2005, Hjalmarsson and Lindquist 2009). Such behavior, in turn, may have deleterious effects on children's incomes during adulthood. We include indicators of social problems in our analysis.

A third group of variables that are reasonable to explore for our purposes refer to *parenting style*. Some styles of parenting are most likely more conducive to children's labor market success than others. Duncan et al. (2005) offer an interesting discussion of the intergenerational implications of parenting styles, a discussion that is based on insights from developmental psychology. Their U.S. data set allows them to consider five parenting practices denoted as parental involvement, parental monitoring, child autonomy, emotional warmth and child stimulation. But much to their surprise, such indicators are generally insignificant, or at least not very important, in explaining a number of different child outcomes. Fryer and Levitt (2004) also experiment with some parenting indicators such as the use of spanking. They do not find any strong intergenerational impact of such parenting indicators. Nevertheless, we find it useful to consider such variables in our exploration of Swedish data. To this end, we make use of a set of questions taken from the Family Study that we label *parenting firmness* and another set of questions that we call *parental involvement in schoolwork.*

Parents can also help and influence their children by offering a home environment that is conducive to school performance and further learning. A concrete example is to keep useful books available in the home or, more generally, by encouraging reading. Mayer (1997), for example, finds a strong correlation between her TV-Read index and children's performance on the Peabody Picture Vocabulary Test Revised.[13] Many surveys include a question about how many books are available in the home. Both Fryer and Levitt (2004) and Mason (2007) find significant coefficients for such a variable in explaining child outcomes. Although the causal interpretation of these results is unclear, we include a variable called *number of books in the home* in our analysis.

Finally, children are likely to inherit family values (i.e., attitudes and preferences) of different types that are more or less conducive to labor market success. In an interesting study, Dohmen et al. (2006) employ German intergenerational data and demonstrate quite strong parent-child associations in willingness to take risks and willingness to trust other people. Their measures of risk and trust are validated in several ways. For example, they plug one of their risk measures into a standard Mincer earnings equation and find that wages are 20 percent higher for those who are fully prepared to take risks than those who are

[13] Mayer's (1997) TV-Read index is based on mothers' reports of how often they read to their children and how many hours per day the television is on.

completely unwilling to do so (on their 11-point scale). However, since they do not set up a statistical horse-race between income or socio-economic status and their family value variables, one cannot rule out that the intergenerational risk and trust associations that they find mainly capture the same mechanisms as an intergenerational income association would capture. Using data from the United States, Mason (2007) finds that indicators of family values are significantly related to the economic outcomes of young adults even after controlling for parents' socio-economic status. Yeung et al. (2000) test whether fathers' attitudes and behaviors have any predictive power over and beyond a wide set of family (and mother) controls. They found significant and positive impacts of fathers' precautionary behavior (reports of using seat belts, having car insurance and having precautionary savings) on sons' schooling and wages.

Many social scientists would naturally argue that attitudes and preferences can be transmitted from one generation to the next through a process of learning (be it conscious or unconscious). But there are also several recent papers that study the genetic inheritance of attitudes and preferences such as giving and risk-taking (Cesarini et al. 2009a), overconfidence (Cesarini et al. 2009b) and financial decision making (Cesarini et al. 2010). In line with the existing literature that examines intergenerational correlations in attitudes and preferences, we have identified a set of questions in the Family Study concerning parents' willingness to plan ahead and to postpone financial gains into the future. We call this set of indicators *parental patience*.[14]

[14] Becker and Tomes' (1979) discussion of the importance of "family culture" includes (among other things) family values and/or parental attitudes, parenting styles, helping with schoolwork, family ethnicity and family connectedness. As such, many of the variables that we have mentioned in this section and that we include in our empirical analysis would fall under their umbrella of "family culture." Mayer (1997) discusses standard, economic theories of parental investment in the human capital of their children as well as theories of "good parenting," which are more common outside of the economics literature. Such theories include the effects of parental stress, role modeling and parenting practices. All of the variables that we list in this section are included in her discussions of the parental characteristics researchers think should matter for children's outcomes. She also discusses several variables that we do not have data on, such as parents' cognitive and non-cognitive abilities. Using data from the PSID, Datcher (1982) studied the impact of family background and community influences on the black-white achievement gap. Her explanatory variables included parental education, family income, number of siblings, and five variables concerning family attitudes and expectations. She also included information concerning neighborhood location and quality.

4. Data

Our data come from the *Stockholm Birth Cohort* (SBC), which was created in 2004, by means of a probability matching of two previously existing longitudinal datasets.[15] The first is the *Stockholm Metropolitan Study 1953-1985* (SMS), which consists of all children born in 1953 who were living in the Stockholm metropolitan area on November 1, 1963. The SMS contains a rich set of variables concerning individual, family, social and neighborhood characteristics. The second is the *Swedish Work and Mortality Database* (WMD), which consists of administrative register information on income, work, unemployment, in-patient and mortality data for all individuals who were born before 1985 and living in Sweden in 1980 or 1990.

Data from the WMD for the years 1990 – 2001 were matched to data from the SMS.[16] WMD includes information on income, which is our object of interest. The outcome variable that we want to "explain" is the sibling correlation in long-run income. Our measure of long-run income is the log of average annual labor market income for the years 1990 – 2001. Annual labor market income comes from registers based on employers' compulsory reports to the tax authorities. It includes sickness benefits, parental-leave benefits and income from self employment (including farming). It excludes capital income, pensions, unemployment benefits and social assistance. Average labor market income is calculated using only those positive income years that exceed 10,000 SEK in 2001 prices (≈ 1,400 USD). In Section 6, we examine how sensitive our results are to this treatment of low, zero and missing incomes.

The *Stockholm Birth Cohort* dataset also includes income data from the WMD for most siblings.[17] This is what allows us to calculate sibling correlations in long-run income. Siblings of the original SMS cohort members were identified using Statistic Sweden's *Multi-Generational Register*. Cohort members and siblings are identified through their mother, which means that the data include biological full siblings as well as half-siblings on the mother's side. The data also include children that are adopted by the mother. Unfortunately, in this particular dataset, we cannot distinguish half-siblings and adopted children from full biological siblings.

[15] See Stenberg and Vågerö (2006) for a full description of the dataset and the matching procedure. Codebooks describing all of the data are available online at http://www.stockholmbirthcohort.su.se/.

[16] 722 of the original 15,117 SMS cohort members were lost in this matching process.

[17] The original SMS cohort included 15,117 individuals. The new SBC cohort is comprised of 14,395 individuals. Of these, 1122 were not successfully matched with their siblings, leaving us with a sample of 13,273 SBC cohort members. In total, we have 35,220 individuals (cohort members and their siblings) in our data set.

When calculating sibling correlations in income, we only use data for closely spaced siblings. Since all of the data were collected with the SMS cohort member in mind, using closely spaced siblings raises the probability that these same data are also valid for the cohort member's siblings.[18] For this same reason, we center the age of older and younger siblings on the SMS cohort members' age (who were all born in 1953). The youngest siblings are born in 1957 and the oldest are born in 1949. Thus, the maximum possible age difference between any pair of siblings is nine years. In Section 6, we examine how sensitive our findings are to changes in these age limits.[19]

Once we have our sibling correlation in long-run income in hand, the goal of this study is to see how much of the family component (i.e., what siblings share) can be explained, or accounted for, by adding in a series of control variables to our $\mathbf{x}_{ij}$ vector. We examine the importance of seven different categories of control variables: (1) traditional measures of socio-economic status, (2) family structure, (3) social problems, (4) parenting firmness, (5) parental involvement in schoolwork, (6) number of books in the home and (7) parental patience.[20] All these variables are taken from the original *Stockholm Metropolitan Study*. Descriptive statistics for all variables included in categories (1) – (3) are presented in Table 1. Variables in categories (4) – (7) are presented in Table 2.

Our traditional measures of parental socio-economic status include income, education and occupational category. We use the log of total market income in 1963 for the mother and father (separately). These were taken from the official tax register. We have information on the education of both parents taken

[18] Some family characteristics, such as family structure, may change over time and will most likely change by even more if we look at siblings spaced very far apart. Differences in family structure experienced by different siblings within the same family make it somewhat harder to interpret our results. Ideally, for all siblings, we want to compare siblings holding, for example, family structure constant. This is most likely approximated by studying closely spaced siblings.

[19] Since we have income data for the years 1990 – 2001, our age limits imply that we observe income for ages 33 – 52. According to Böhlmark and Lindquist (2006), this means that our measure of long-run income is appropriate for the women in the sample, but that it is probably too high for the men in our sample given that we want to mimic long-run income. For men, one would prefer to have it centered around (or, at least, closer to) age 34. However, since we are dealing with closely spaced brothers, this bias should be roughly equal for both. This potential life-cycle bias may affect the level of our baseline correlation (slightly), but it should not affect our efforts towards explaining sibling correlations using a set of common family background variables.

[20] In an earlier discussion paper (Björklund et al. 2008), we also included controls for neighborhood fixed-effects. If one controls for neighborhood fixed-effects, then the sibling correlation becomes a "tighter upper bound" on the importance of common family background variables (see, e.g., Page and Solon 2003a or Raaum et al. 2006 for more on this). However, if sorting into neighborhoods by income and education is important, then one may, actually, be controlling for some of what one would (instead) like to explain. In this paper, we have chosen not to control for neighborhood fixed-effects, since neighborhood effects account for almost none of what siblings share and, at the same time, demand a lot from our data.

from the 1960 census.[21] Education is given by four categories: (1) missing, (2) grade school, (3) high school and (4) college. We also include a variable for the father's occupational category in 1953 and 1963. This measure is collapsed into five strata: (1) upper and upper middle class, (2) lower middle class officials and non-agricultural employees, (3) lower middle class, entrepreneurs, (4) working class, skilled workers and (5) working class, unskilled workers.[22] Missing values for this variable tend to be strong predictors of negative outcomes, so we include missing as a separate stratum.[23]

Our second category of "exploratory" variables, *family structure*, includes the mother's age at the birth of her first child and the number of siblings a cohort member has. Number of siblings is treated as a continuous variable, while mother's age at first birth is entered as a categorical variable: 15-19, 20-24, 25-29, 30-34, 35-39 and 40-44. We also include two variables that are meant to reflect the type of family that our siblings grew up in. The first of these variables is concerned with the marital status of the head of the household and is taken from the 1960 census. This variable includes six categories: (1) missing, (2) married, not cohabitating, (3) married and cohabitating, (4) single, (5) widow/widower, and (6) divorced. The second variable is taken from the 1964 register of population and income and refers to the family type in 1963: (1) father and mother living together, (2) mother living alone, (3) father living alone, (4) mother living together with other than child's biological father, (5) father living together with other than child's biological mother, and (6) other.

Our variables indicating *social problems* include an indicator whether the family received any social assistance (welfare payments) during the period 1953 – 1972.[24] We have information on "incidents of drunkenness" and alcoholism for

[21] Note that 960 of the original 15,117 SMS cohort members (6.4 percent) were not included in, or did not respond to, the 1960 census. Thus, any variables taken from this census (e.g., parental education) are missing for these 960 individuals.

[22] The original occupation data no longer exist. They were coded into these five categories by the group of researchers in charge of collecting the original SMS data. The five-group occupational classification used in the SMS data is a refinement of the three-group system that was used by Statistics Sweden in its official reports concerning general elections from 1911 to 1956.

[23] Missing is because the father was unemployed, was in jail, or for some other reason could not be categorized. When the father was missing altogether from the family, information on the mother's occupational status was used instead.

[24] Unlike welfare in the United States, means-tested social support in Sweden is not primarily aimed at single-mothers. In Sweden, all single-parents receive support through a system of family support that is (for the most part) not means-tested. Furthermore, we could have included means-tested social support as an indicator of socio-economic status. But we believe that it is a better signal of social problems than of long-run socio-economic status. Our reasoning follows that of Stenberg (2000), who has studied intergenerational inheritance of welfare recipiency using SBC data. He argues: "Because the main part of Swedish social benefits is universal, families who fall through this economic safety net and must rely on means-tested assistance as their last resort are

both parents. Furthermore, we have knowledge about serious mental health problems of either parent and also if either parent died before 1972 (i.e., before the cohort member turned 19). All of these variables were taken directly from the files of the original SMS cohort members kept by the local Child Welfare Committees. We code them as dummies (yes = 1 and no = 0).

Our variables indicating *social problems* also include two measures of fathers' criminality. We have official police register data concerning the number and type of the fathers' criminal convictions (if any). We use these data to construct an extensive margin dummy variable equal to 1 if the father has one or more convictions, and 0 if he has no conviction. We also construct a continuous, intensive margin variable which is equal to the number of convictions that a father has.

Besides the data mentioned above, that are sourced from official census and/or register data, the original *Stockholm Metropolitan Study* consisted of separate surveys. In this paper, we make use of information taken from the 1968 Family Study, that, among others, includes information on parental involvement in schoolwork, parenting styles and firmness, and maternal attitudes such as patience.

The Family Study was conducted in the following manner: In 1968, a sample of the cohort members' mothers was interviewed. Of the original 15,117 SMS cohort members, 4,021 were included in the sample that was interviewed for the Family Study.[25] This sample, however, was not drawn randomly. First, the SMS cohort members that were still living in the greater Stockholm metropolitan area as of November 1, 1967 were listed. Some 525 original cohort members had been lost since November 1, 1963. The IQ test scores (from the 1966 School Study) of the remaining 14,592 were placed into five groups: (1) low, (2) medium, (3) high, (4) non-response, and (5) incomplete. All cohort members in the "high" group were kept in the sample. High was defined as the top 5 percent of scores. All cohort members in the "low" group were also kept. Low was defined as the lowest 5 percent of the scores. The cut-off points for high and low were set separately for boys and girls. In each of the other three groups, one in five cohort members was chosen at random. Of these 4021 cohort members, 370 chose not to participate, so 3651 persons are included in the Family Study sample.

---

likely to be a more negatively selected group with respect to different types of social problems. Therefore, we could expect to find a greater representation of non-economic problems here than among welfare families in the United States (p. 233)." Hjalmarsson and Lindquist (2009) report that the five strongest predictors of receiving social support in the original SMS data are psychological problems of parents, alcoholism of parents, single household (predominantly single mothers), incidents of drunkenness by the mother, and father's criminality.

[25] 1,972 females and 2,049 males.

We use this sample of individuals, together with their siblings, in our estimations, since we feel that the types of questions asked in this particular survey fit our research purpose. We take the original sampling design into consideration and re-weight the Family Study sample when running our regressions. Families in groups (1) and (3) receive an inverse probability weight equal to 1. Families in groups (2), (4) and (5) receive an inverse probability weight equal to 5, which reflects the fact that only 1 in 5 families in these groups were chosen to participate in the Family Study.

Table 1 gives us some feel for just how selective the Family Study sample actually is. Descriptive statistics for the original members of the full Stockholm Birth Cohort are shown in column 1. These can be compared with the (unweighted) descriptive statistics of each of the three separate groups in the Family Study sample.

The descriptive statistics of the large, mid-range test score group in column 3 (stratum 2, 4 and 5) closely resemble those of the full sample. The descriptive statistics of the smaller, low- and high-range groups (strata 1 and 3, respectively), however, differ from those in the full sample in a manner that one might expect. For example, the average log income of the high-IQ group is higher than the average log income of the full sample, which, in turn, is higher than the average log income of the low IQ-group. More importantly, family-wide variables, such as parents' incomes and educations, differ across these groups as well. Because of these differences, and because the high-IQ and low-IQ groups were over-sampled, we run all of our empirical exercises on the weighted Family Study sample.

The rest of our family-wide variables are taken from the 1968 Family Study. First, we use a set of questions concerned with *parental involvement in schoolwork*. The interviewers asked a series of questions of each mother (or substitute mother) including, for example: *Do you and your daughter/son ever talk about what she/he has read/done in school?* Mothers could choose from the following answers: (1) very often, (2) rather often, (3) now and then, (4) rather seldom, (5) almost never, or (6) do not know. Mothers were then asked to answer the same set of questions in terms of their husband's involvement in their child's schoolwork.

The second set of questions that we make use of deals with *parenting firmness*, which is intended to reflect a particular parenting style. Mothers were asked whether or not they agreed with a set of statements, for example: *Children must have firm rules.* Mothers were allowed to choose from the following answers: (1) quite right, (2) generally, right, (3) neither right nor wrong, (4) generally, wrong, (5) quite wrong, or (6) do not know.

Table 1. Descriptive Statistics.

	Full Sample (1)		Family Study Sample, Stratum 1 (2)		Family Study Sample, Strata 2, 4, 5 (3)		Family Study Sample, Stratum 3 (4)	
	Mean (s.d.)	#Obs.	Mean (s.d.)	#Obs.	Mean (s.d.)	#Obs.	Mean (s.d.)	#Obs.
			Outcome Variables					
Long-run income males	12.39 (.575)	6475	12.10 (.485)	230	12.42 (.572)	1053	12.69 (.558)	307
Missing		1244						
Long-run income females	12.07 (.470)	6247	11.85 (0.403)	250	12.08 (.441)	1005	12.28 (.514)	279
Missing		1151						
			Family-Wide, Explanatory Variables					
			Traditional Measures of Socio-Economic Status					
Log father income 1963	10.13 (.545)	12618	9.95 (.467)	434	10.12 (.540)	1844	10.44 (.518)	530
Missing		2499		46		214		56
Log mother income 1963	8.73 (.909)	7730	8.56 (.869)	272	8.73 (.925)	1127	8.75 (1.141)	283
Missing		7387		208		931		303
Father's education								
Missing	0.06	947	0.04	21	0.04	85	0.03	16
Grade school	0.70	10567	0.88	423	0.72	1476	0.45	264
High school	0.16	2351	0.06	27	0.16	336	0.25	148
College	0.08	1252	0.02	9	0.08	161	0.27	158
Mother's education								
Missing	0.06	947	0.04	21	0.04	85	0.03	16
Grade school	0.87	13215	0.94	453	0.90	1853	0.75	440
High school	0.05	695	0.01	6	0.04	91	0.15	90
College	0.02	260	0.00	0	0.01	29	0.07	40
Father's occupation 1953								
Missing	0.04	579	0.04	19	0.03	60	0.02	13
Upper & upper middle class	0.14	2016	0.05	24	0.12	245	0.34	199
Lower middle class - employees	0.32	4648	0.18	88	0.32	653	0.36	210
Lower middle class - entrepreneurs	0.06	927	0.05	22	0.06	128	0.06	38
Working class - skilled	0.28	4131	0.35	167	0.29	592	0.15	87
Working class -unskilled	0.19	2816	0.33	160	0.19	380	0.07	39
Father's occupation 1963								
Missing	0.03	413	0.03	14	0.02	36	00.01	5
Upper & upper middle class	0.18	2587	0.05	24	0.16	331	0.41	238
Lower middle class - employees	0.36	5240	0.24	116	0.35	722	0.37	216
Lower middle class-entrepreneurs	0.08	1155	0.06	28	0.08	168	0.07	43
Working class - skilled	0.23	3306	0.30	146	0.24	489	0.09	53
Working class -unskilled	0.16	2416	0.32	152	0.15	312	0.05	31

Table 1 (continued)

				Family Structure				
Mother's age at first birth								
Missing	0.11	1719						
15-19	0.04	615	0.16	77	0.10	220	0.03	15
20-24	0.22	3247	0.49	234	0.39	805	0.25	146
25-29	0.28	4254	0.23	111	0.33	673	0.45	264
30-34	0.22	3341	0.08	37	0.13	269	0.22	131
35-39	0.10	1488	0.03	15	0.04	72	0.04	22
40-47	0.03	453	0.01	6	0.01	19	0.01	8
Family Type in 1960								
Missing	0.06	960	0.04	21	0.04	85	0.03	16
Married, but not cohabitating	0.02	297	0.03	12	0.02	36	0.01	8
Married, cohabitating	0.85	12790	0.85	407	0.89	1826	0.92	536
Single	0.01	215	0.01	7	0.01	29	0.01	3
Widow(er)	0.02	238	0.01	6	0.01	24	0.01	8
Divorced	0.04	617	0.06	27	0.03	58	0.03	15
Family Type in 1963								
Missing	0.00	4						
Father & mother living together	0.89	13499	0.88	423	0.92	1883	0.94	551
Single mother	0.09	1420	0.10	50	0.08	161	0.06	34
Single father	0.01	129	0.01	3	0.00	7	0.00	1
Mother & step father	0.00	38	0.00	2	0.00	6	0	0
Father & step mother	0.00	8	0.00	1	0.00	1	0	0
Other	0.00	19	0.00	1	0	0	0	0
Number of Siblings	2.66 (1.21)	13398	3.11 (1.51)	480	2.64 (1.12)	2058	2.48 (1.00)	586
Missing		1719						
				Indicators of Social Problems				
Social support	0.21 (0.41)	15117	0.36 (0.48)	480	0.19 (0.39)	2058	0.05 (0.22)	586
Alcoholic	0.04 (0.21)	15117	0.08 (0.27)	480	0.04 (0.19)	2058	0.01 (0.11)	586
Drunken behavior	0.03 (0.17)	15117	0.07 (0.25)	480	0.03 (0.16)	2058	0.01 (0.12)	586
Mother record of mental illness	0.04 (0.20)	15117	0.08 (0.27)	480	0.03 (0.16)	2058	0.01 (0.09)	586
Father record of mental illness	0.03 (0.17)	15117	0.05 (0.22)	480	0.02 (0.15)	2058	0 (0.01)	586
Father's crime extensive margin	0.12 (0.33)	15117	0.21 (0.41)	480	0.12 (0.33)	2058	0.06 (0.23)	586
Father's crime intensive margin	0.27 (1.09)	15117	0.41 (1.14)	480	0.28 (1.13)	2058	0.09 (0.45)	586
Mother died	0.01 (0.01)	15117	0	480	0	2058	0	586
Father died	.02 (.12)	15117	.02 (.14)	480	.01 (.11)	2058	0	586

The five strata were constructed using children's 6th grade IQ test scores. The five groups are defined as follows: (1) lowest 5%, (2) middle 90%, (3) highest 5%, (4) non-response, and (5) incomplete. The missing observations on *Mother's age at first birth* in the SBC sample are due to the fact that this variable can only be constructed for those individuals who were successfully matched to their siblings.

Table 2. Questions and Answers Taken from the Family Study.

Parental Involvement in Schoolwork						
	(1) very often	(2) rather often	(3) now and then	(4) rather seldom	(5) almost never	(6) do not know
Do you and your daughter/son ever talk about what she/he has read/done in school?	1796	1618	1249	196	118	3
Have you read in your daughter's/son's schoolbooks to see what she/he is learning in school and to keep up a little yourself?	733	1026	1925	734	553	7
Do you usually help her/him with her/his homework by questioning, etc.?	368	601	1286	869	1847	4
Do your husband and your daughter/son ever talk about what she/he has read/done in school?	900	1033	1461	553	458	17
Does your husband read in your daughter's/son's schoolbooks to see what she/he is learning in school and to keep up a little himself?	400	714	1402	819	1046	42
Does your husband usually help her/him with her/his homework by questioning, etc.?	264	450	1189	684	1824	7
	(0) no	(1) yes, once	(2) yes, several times	(3) do not know		
Have you been to a Parent Teacher Association meeting this school year and if so, have you been more than once?	2303	1884	789	5		
Has your husband been to a Parent Teacher Association meeting this school year and if so, has he been more than once?	2962	1271	461	6		
Parenting Firmness						
	(1) quite right	(2) generally, right	(3) neither right nor wrong	(4) generally, wrong	(5) quite wrong	(6) do not know
Children must learn to obey.	2926	1574	310	119	53	4
Children must have firm rules.	3247	1439	235	54	5	6
Children must respect their parents.	1552	1584	949	482	390	29
Children should be taught to control themselves.	1209	2133	1053	443	141	7
You have to be consistent when raising children.	3363	1403	153	32	11	24
Parental Patience						
	(1) yes, definitely	(2) yes, perhaps	(3) do not know	(4) no, perhaps not	(5) no, definitely not	
If you could choose between 1,000 SEK now and 10,000 SEK in five years, would you choose 1,000 SEK now?	811	472	298	413	2991	
Do you think one gets more out of life if one thinks matters over carefully first?	1216	1453	353	1098	864	
Do you like to make long-term plans?	1461	1589	130	855	951	

Table 2 (continued)

Do you think it is worth planning for the future?									
	2086	1582	408	498	412				
Do you often think about the future?									
	1437	1413	89	1152	892				
Do you think your future mainly depends on chance?									
	815	1399	601	925	1241				
Do you like doing things you have not planned ahead of time?									
	2009	1647	187	651	492				
Do you like saving up money for something big?									
	2672	1368	210	375	357				
Do you think that you yourself can influence your future through your present actions?									
	1835	1844	585	404	312				

	Number of Books in the Household									
	none	≈ 1	≈ 3	≈ 6-10	≈ 30	≈ 100	≈ 300	≈ 1000	>= 3000	do not know
How many books do you think there are in this apartment/house?										
	10	9	57	441	1580	1770	922	186	11	1

We also make use of questions about the future. We label these questions as a measure of *parental patience*. Mothers were asked a series of questions, for example: *If you could choose between 1,000 SEK now and 10,000 SEK in five years, would you choose 1,000 SEK now?* The answers that they could choose from were: (1) yes, definitely, (2) yes, perhaps, (3) do not know, (4) no, perhaps not, or (5) no, definitely not. The last question that we use is the (now) classic book question: *How many books do you think there are in this apartment/house?* The possible answers and the responses to these four sets of questions are shown in Table 2.

5. Results

We begin by considering our estimated sibling correlations with controls only for siblings' gender and birth year. These are reported in the upper panel of Table 3. We report these estimates for men and women both separately and pooled, where pooling allows us to also include mixed gender siblings in the identification of the two variance components.

Our baseline sibling correlations are 0.250 (0.042) for men, 0.227 (0.046) for women and 0.203 (0.023) for the pooled sample.[26] Estimates of the same

[26] Family and individual variance components are estimated in a mixed-effects model using the *gllamm* package in Stata. As discussed in Section 4, we use inverse probability weights in order to reweight the Family Study sample in accordance with the original sampling design. All standard errors of fixed and random (variance) components are robust to this weighting procedure. Robust standard errors are calculated using the Huber/White/sandwich estimator of the covariance matrix of the estimated model parameters. *gllamm* estimates the model using an Iterative Generalized

sibling correlations using all available SBC cohort members and their siblings (not just the Family Study sample) are 0.255 (0.018) for men, 0.148 (0.019) for women and 0.175 (0.010) for the pooled sample. The two brother correlations are nearly identical and very similar in magnitude to previous estimates cited in our introduction. The two sister correlations are different from each other, but not significantly so.[27]

In the lower panel of Table 3, we add parental income, parental education and father's occupation to our $\mathbf{x}_{ij}$ vector.[28] We first add each variable separately in order to explore their individual importance and then we add all of the variables simultaneously. Father's income, education and occupation matter most for brothers. Only father's occupation appears to matter for sisters. Pooled effects match those for brothers quite closely. Mother's education has only a small, gender-neutral effect on the size of the sibling correlations.[29]

Least Squares algorithm (see Rabe-Hesketh et al. 2002). In our model, these estimates are equivalent to maximum likelihood estimates, which are biased in small-samples. However, in large samples such as ours, they are not. We know this since we have also estimated all of our models using the Restricted Maximum Likelihood (REML) method found in Stata's *xtmixed* command. This method does not suffer from small sample bias. All parameter estimates were virtually identical to those reported in this paper. Although REML may be the preferred (unbiased) method, *xtmixed* does not allow for probability weights. Running *xtmixed* on replicated data that mimics probability weights produces the correct point estimates, but does not produce robust standard errors, which are necessary for correct inference in our application.

Due to the large number of included variables, in combination with the need to specify probability weights, we were not able to estimate several of the regressions presented in Table 6 in a single step using *gllamm*. We have, therefore, adopted a two-step estimation procedure which (for the sake of consistency) is used in *all* regressions reported in Tables 3 - 6. In the first step, we regress our measure of adult income on all variables included in the $\mathbf{x}_{ij}$ vector. This is done using weighted OLS with weights equal to the inverse of the probability of being selected into the sample. The residuals from this regression are saved and then passed on to *gllamm* which estimates the necessary variance components and produces robust standard errors. In 75 out of 87 of the regressions reported in Tables 3-6, we were able to implement the estimations in a single step. The results from these estimations were virtually identical to their two-step counterparts.

The sibling correlation is calculated using Stata's *nlcom* command, which uses the delta method to calculate approximate standard errors (reported in parentheses).

[27] In Section 6, we examine the potential effects that this rather large difference may have on our experiment.

[28] Regression coefficients are presented in Appendix Table A.

[29] Mother's education is one of several variables that does matter for children's outcomes (see Appendix Table A), but at the same time it does not appear to play a large, direct role in explaining sibling similarities. Recall our discussion in Section 2.

Table 3. The Importance of Parental Income, Parental Education and Father's Occupation for Sibling Correlations in Income.

	Men			Women			Pooled		
				Baseline Estimates					
Sibling Correlation	0.250			0.227			0.203		
(s.e.)	(0.042)			(0.046)			(0.023)		
Family Component	0.078			0.045			0.052		
(s.e.)	(0.014)			(0.010)			(0.006)		
		↓	%↓		↓	%↓		↓	%↓
				Father's Income					
Sibling Correlation	0.219	0.032	13	0.215	0.012	5	0.183	0.021	10
(s.e.)	(0.044)			(0.046)			(0.024)		
Family Component	0.065	0.013	16	0.042	0.003	6	0.045	0.007	13
(s.e.)	(0.014)			(0.009)			(0.006)		
				Mother's Income					
Sibling Correlation	0.244	0.006	2	0.224	0.003	1	0.202	0.002	1
(s.e.)	(0.043)			(0.046)			(0.023)		
Family Component	0.077	0.002	3	0.044	0.001	1	0.051	0.001	1
(s.e.)	(0.014)			(0.009)			(0.006)		
				Father's Education					
Sibling Correlation	0.207	0.043	17	0.214	0.013	6	0.178	0.025	12
(s.e.)	(0.044)			(0.045)			(0.024)		
Family Component	0.061	0.017	22	0.041	0.003	7	0.044	0.008	15
(s.e.)	(0.013)			(0.009)			(0.006)		
				Mother's Education					
Sibling Correlation	0.235	0.015	6	0.210	0.016	7	0.187	0.016	8
(s.e.)	(0.042)			(0.046)			(0.023)		
Family Component	0.072	0.007	8	0.041	0.004	9	0.047	0.005	10
(s.e.)	(0.013)			(0.009)			(0.006)		
				Father's Occupation					
Sibling Correlation	0.198	0.052	21	0.206	0.021	9	0.167	0.036	18
(s.e.)	(0.042)			(0.045)			(0.023)		
Family Component	0.058	0.020	25	0.039	0.005	12	0.041	0.011	21
(s.e.)	(0.013)			(0.009)			(0.006)		
				Income, Education & Occupation					
Sibling Correlation	0.180	0.070	28	0.198	0.029	13	0.158	0.045	22
(s.e.)	(0.044)			(0.045)			(0.025)		
Family Component	0.052	0.026	34	0.037	0.007	16	0.038	0.014	26
(s.e.)	(0.013)			(0.009)			(0.006)		

The family variance component and the individual variance component (not reported) are estimated in a mixed-effects model using *gllamm* in Stata. Standard errors (in parentheses) are robust. All control variables except log income are entered as dummy variables. Gender and birth-year dummies are included in all regressions. The sibling correlation is calculated using the *nlcom* command which uses the delta method to calculate approximate standard errors. ↓ gives the absolute decrease in the sibling correlation or family component. %↓ gives the percentage decrease. Percentage decreases ≥ 10% are shaded.

So how much of the family variance component (and sibling correlation) can we account for when we control for all of these variables at once? When we do this, we can account for 34 percent (28 percent) of the family variance component (sibling correlation) for brothers and 16 percent (13 percent) for sisters. These magnitudes of the importance of parents' income, education and occupation are in line with what would be expected from previous estimates of sibling correlations and intergenerational correlations in permanent income, according to the formal relationship between these parameters stressed in the introduction.[30]

We now turn to the basic question addressed by our study: What more than parental income, education and occupation do siblings get from their parents? Table 4 contains the results we obtain when adding our indicators for family structure. Mother's age at first birth is the most important variable.[31] As stressed in the existing literature (Geronimus and Korenman 1992, Holmlund 2005) teenage motherhood has a negative association with children's outcomes, while the children of older, first-time mothers tend to have above average long-run incomes (see the regression coefficients in Appendix Table C). Despite this apparent importance for outcomes, it accounts for only 7 to 10 percent of the family variance component.

The number of siblings in the family does not seem to matter much, which is in line with previous findings presented in Lindahl (2008). We find it somewhat surprising, however, that family type (married, divorced, etc.) does not explain much of sibling similarities despite the fact that it does matter for individual outcomes (see the regression coefficients in Appendix Table C). This implies that family structure may have heterogeneous effects on children.[32]

Taken together, our family structure variables account for just over half of what the indicators of parents' socio-economic status (in Table 3) accounted for. However, if we examine the bottom rows in Table 4 (labeled *All Controls Used in Tables 3 and 4*), we see that the total amount of variation explained in addition to the variation already accounted for by parental income, education and occupation is quite small: between 3 and 7 percent. These results suggest that, to a large extent, both groups of variables – family structure and socio-economic status – capture the same underlying mechanisms.

[30] Mazumder (2008) found a 36 percent reduction in the family variance component after adding a two-year average of parental income.

[31] Mother's age at first birth is controlled for using a set of age-category dummies: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44.

[32] The potentially heterogeneous effects of changes in family structure on sibling outcomes is stressed by Hill et al. (2001) and Conley (2004).

Table 4. The Importance of Family Structure for Sibling Correlations in Income.

	Men			Women			Pooled		
		↓	%↓		↓	%↓		↓	%↓
				Mother's Age at First Birth					
Sibling Correlation	0.231	0.019	8	0.214	0.013	6	0.189	0.014	7
(s.e.)	(0.043)			(0.046)			(0.023)		
Family Component	0.071	0.008	10	0.042	0.003	7	0.048	0.004	8
(s.e.)	(0.013)			(0.009)			(0.006)		
				Family Type 1960					
Sibling Correlation	0.238	0.012	5	0.222	0.005	2	0.197	0.006	3
(s.e.)	(0.044)			(0.045)			(0.023)		
Family Component	0.073	0.0045	6	0.043	0.001	3	0.050	0.002	4
(s.e.)	(0.014)			(0.009)			(0.006)		
				Family Type 1963					
Sibling Correlation	0.236	0.014	6	0.225	0.002	1	0.196	0.007	4
(s.e.)	(0.045)			(0.046)			(0.023)		
Family Component	0.073	0.005	7	0.044	0.000	1	0.050	0.002	4
(s.e.)	(0.014)			(0.009)			(0.006)		
				Number of Siblings					
Sibling Correlation	0.246	0.004	1	0.216	0.011	5	0.194	0.009	4
(s.e.)	(0.042)			(0.046)			(0.023)		
Family Component	0.077	0.002	2	0.042	0.002	6	0.049	0.003	5
(s.e.)	(0.014)			(0.009)			(0.006)		
				All Family Structure Controls					
Sibling Correlation	0.208	0.042	17	0.202	0.025	11	0.175	0.028	14
(s.e.)	(0.046)			(0.046)			(0.024)		
Family Component	0.063	0.016	20	0.039	0.006	13	0.043	0.008	16
(s.e.)	(0.014)			(0.009)			(0.006)		
				All Controls Used in Tables 3 and 4					
Sibling Correlation	0.170	0.080	32	0.184	0.042	19	0.148	0.055	27
(s.e.)	(0.046)			(0.044)			(0.025)		
Family Component	0.048	0.030	38	0.034	0.010	23	0.035	0.017	32
(s.e.)	(0.013)			(0.009)			(0.006)		

The family variance component and the individual variance component (not reported) are estimated in a mixed-effects model using *gllamm* in Stata.. Standard errors (in parentheses) are robust. Mother's age at first birth is controlled for using a set of age dummies: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44. Number of siblings is treated as a continuous variable. Family type in 1960 and 1963 are categorical variables. Gender and birth-year dummies are included in all regressions. The sibling correlation is calculated using the *nlcom* command which uses the delta method to calculate approximate standard errors. ↓ gives the absolute decrease in the sibling correlation or family component. %↓ gives the percentage decrease. Percentage decreases ≥ 10% are shaded.

Table 5. The Importance of Social Problems for Sibling Correlations in Income.

	Men			Women			Pooled		
		↓	%↓		↓	%↓		↓	%↓
				Social support					
Sibling Correlation	0.233	0.017	7	0.215	0.011	5	0.188	0.015	8
(s.e.)	(0.043)			(0.045)			(0.023)		
Family Component	0.071	0.007	9	0.042	0.003	6	0.047	0.005	9
(s.e.)	(0.013)			(0.009)			(0.006)		
				Alcohol					
Sibling Correlation	0.237	0.014	5	0.223	0.003	1	0.196	0.008	4
(s.e.)	(0.044)			(0.046)			(0.024)		
Family Component	0.073	0.005	7	0.044	0.001	2	0.050	0.002	5
(s.e.)	(0.014)			(0.010)			(0.006)		
				Mental Illness					
Sibling Correlation	0.244	0.006	2	0.221	0.005	2	0.197	0.006	3
(s.e.)	(0.042)			(0.047)			(0.023)		
Family Component	0.076	0.002	3	0.043	0.001	3	0.050	0.002	3
(s.e.)	(0.014)			(0.010)			(0.006)		
				Father's Criminality (extensive margin)					
Sibling Correlation	0.243	0.007	3	0.219	0.008	3	0.197	0.006	3
(s.e.)	(0.043)			(0.045)			(0.023)		
Family Component	0.075	0.003	4	0.043	0.002	4	0.050	0.002	4
(s.e.)	(0.014)			(0.009)			(0.006)		
				Father's Criminality (intensive margin)					
Sibling Correlation	0.241	0.009	4	0.218	0.009	4	0.196	0.008	4
(s.e.)	(0.043)			(0.046)			(0.023)		
Family Component	0.074	0.004	5	0.043	0.002	5	0.050	0.002	5
(s.e.)	(0.014)			(0.010)			(0.006)		
				Parental Deaths					
Sibling Correlation	0.249	0.001	0	0.226	0.000	0	0.203	0.000	0
(s.e.)	(0.042)			(0.046)			(0.023)		
Family Component	0.078	0.000	0	0.044	0.000	0	0.052	0.000	0
(s.e.)	(0.014)			(0.010)			(0.006)		
				All Controls for Social Problems					
Sibling Correlation	0.222	0.028	11	0.208	0.019	8	0.183	0.020	10
(s.e.)	(0.044)			(0.045)			(0.024)		
Family Component	0.067	0.011	14	0.040	0.005	10	0.046	0.006	12
(s.e.)	(0.014)			(0.009)			(0.006)		
				All Controls Used in Tables 3 and 5					
Sibling Correlation	0.168	0.082	33	0.188	0.039	17	0.151	0.052	26
(s.e.)	(0.045)			(0.045)			(0.025)		
Family Component	0.048	0.031	39	0.035	0.010	21	0.036	0.016	30
(s.e.)	(0.013)			(0.009)			(0.006)		
				All Controls Used in Tables 3, 4 and 5					
Sibling Correlation	0.159	0.091	36	0.178	0.049	22	0.143	0.060	30
(s.e.)	(0.046)			(0.044)			(0.025)		
Family Component	0.045	0.033	43	0.033	0.012	26	0.034	0.018	35
(s.e.)	(0.013)			(0.008)			(0.006)		

The family variance component and the individual variance component (not reported) are estimated in a mixed-effects model using *gllamm* in Stata. Standard errors (in parentheses) are robust. All variables concerning social problems are simple yes (=1) or no (=0) dummy variables except for father's criminality at the intensive margin, which is treated as a continuous variable. Gender and birth-year dummies are included in all regressions. The sibling correlation is calculated using the *nlcom* command which uses the delta method to calculate approximate standard errors. ↓ gives the absolute decrease in the sibling correlation or family component. %↓ gives the percentage decrease. Percentage decreases ≥ 10% are shaded.

Table 6. The Importance of Parental Involvement and Attitudes for Sibling Correlations in Income.

	Men			Women			Pooled		
		↓	%↓		↓	%↓		↓	%↓
Involvement in Schoolwork									
Sibling Correlation	0.202	0.048	19	0.194	0.032	14	0.170	0.033	16
(s.e.)	(0.046)			(0.045)			(0.024)		
Family Component	0.060	0.018	23	0.036	0.008	19	0.042	0.010	20
(s.e.), *t*-statistic	(0.014)			(0.009)			(0.006)		
Parenting-Firmness									
Sibling Correlation	0.234	0.016	7	0.198	0.029	13	0.186	0.017	8
(s.e.)	(0.041)			(0.049)			(0.023)		
Family Component	0.071	0.008	10	0.038	0.007	15	0.046	0.005	11
(s.e.), *t*-statistic	(0.013)			(0.010)			(0.006)		
Patience									
Sibling Correlation	0.204	0.046	18	0.187	0.040	18	0.169	0.034	17
(s.e.)	(0.046)			(0.046)			(0.023)		
Family Component	0.060	0.018	23	0.035	0.009	21	0.041	0.010	20
(s.e.), *t*-statistic	(0.014)			(0.009)			(0.006)		
Number of Books in Home									
Sibling Correlation	0.228	0.022	9	0.205	0.022	10	0.182	0.021	10
(s.e.)	(0.042)			(0.046)			(0.024)		
Family Component	0.069	0.009	12	0.039	0.005	12	0.045	0.006	12
(s.e.), *t*-statistic	(0.013)			(0.009)			(0.006)		
All Controls for Books, Involvement and Attitudes									
Sibling Correlation	0.144	0.106	42	0.120	0.107	47	0.125	0.078	38
(s.e.)	(0.048)			(0.049)			(0.024)		
Family Component	0.040	0.038	49	0.021	0.024	53	0.029	0.023	44
(s.e.), *t*-statistic	(0.013)			(0.009)			(0.006)		
All Controls Used in Tables 3 and 6									
Sibling Correlation	0.094	0.156	62	0.110	0.117	52	0.103	0.100	49
(s.e.)	(0.049)			(0.049)			(0.024)		
Family Component	0.025	0.053	68	0.019	0.026	58	0.023	0.029	55
(s.e.), *t*-statistic	(0.013)			(0.009)			(0.006)		
All Controls Used in Tables 3, 4 and 6									
Sibling Correlation	0.084	0.166	67	0.099	0.128	56	0.094	0.109	54
(s.e.)	(0.050)			(0.048)			(0.025)		
Family Component	0.022	0.056	72	0.017	0.028	62	0.021	0.031	59
(s.e.), *t*-statistic	(0.013)			(0.008)			(0.006)		
All Controls Used in Tables 3, 4, 5, and 6									
Sibling Correlation	0.073	0.177	71	0.095	0.132	58	0.092	0.111	55
(s.e.)	(0.051)			(0.048)			(0.025)		
Family Component	0.019	0.059	76	0.016	0.028	64	0.021	0.031	60
(s.e.), *t*-statistic	(0.013)			(0.008)			(0.006)		

The family-variance component and the individual-variance component (not reported) are estimated in a mixed-effects model using *gllamm* in Stata. Standard errors (in parentheses) are robust. Parental attitudes, involvement and the number of books at home are all categorical variables. Gender and birth-year dummies are included in all regressions. The sibling correlation is calculated using the *nlcom* command which uses the delta method to calculate approximate standard errors. ↓ gives the absolute decrease in the sibling correlation or family component. %↓ gives the percentage decrease. Percentage decreases ≥ 10% are shaded.

Next, in Table 5, we explore the impact of our indicators of social problems.[33] Social support is the most important indicator. However, it never accounts for more than 9 percent of the family variance component (nor is this decrease statistically significant). The total reduction of the sibling correlation from all these variables is somewhat lower than the corresponding reduction of the family structure variables. Furthermore, the additional explanatory power, above and beyond those variables that we have already included, is negligible.

Finally, we add variables for parental involvement in schoolwork, attitudes and the number of books in the home. Table 6 shows these results. It is interesting to note that these variables (when added simultaneously) can account for reductions in both the family variance component and the sibling correlation that are larger than the reductions produced by our indicators of socio-economic status shown in Table 3. These added variables seem to be particularly important for explaining sister similarities in adult income. Few other variables seemed to have much effect on the sister correlation. Parental involvement in schoolwork and parental patience are the two most important new variables. Parental firmness and the number of books in the home play only secondary roles.

In Tables 7 and 8, we report the regression coefficients associated with two variables concerning parental involvement in schoolwork and two variables concerning maternal patience. Examining regression coefficients directly may help us understand how these variables relate to individual outcomes and how they operate to make siblings more similar. The four questions reported in Tables 7 and 8 were chosen because each was particularly important for lowering either the brother correlation or the sister correlation. They were not chosen by looking at the regression coefficients themselves.

The first question concerning parental involvement, *Do your husband and your daughter/son ever talk about what she/he has read/done in school?,* was the single most important parental-involvement variable for lowering the brother correlation. It lowered the family variance component (on its own) by 8 percent. As shown in Table 7 (columns 1 and 2 for brothers and columns 4 and 5 for sisters), the answers (1) "very often" and (2) "rather often" are strongly positively correlated with children's long-run income, even after controlling for parents socio-economic status.

[33] Regression coefficients are reported in Appendix Table B.

Table 7. Selected Regression Coefficients from Models in Table 6 Concerning Parental Involvement in Children's Schoolwork.

	(1) Brothers	(2) Brothers	(3) Brothers	(4) Sisters	(5) Sisters	(6) Sisters
	Do your husband and your daughter/son ever talk about what she/he has read/done in school?					
(1) very often	0.188***	0.138***	0.120***	0.116***	0.105***	0.089***
	(0.040)	(0.040)	(0.041)	(0.030)	(0.030)	(0.031)
(2) rather often	0.086**	0.066**	0.040	0.049*	0.050*	0.046*
	(0.035)	(0.034)	(0.034)	(0.027)	(0.027)	(0.028)
(3) now and then	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
(4) rather seldom	-.018	-.002	0.014	0.000	0.007	0.013
	(0.039)	(0.039)	(0.039)	(0.033)	(0.032)	(0.033)
(5) almost never	-.025	-.006	0.036	-.041	-.030	-.015
	(0.044)	(0.042)	(0.044)	(0.039)	(0.039)	(0.039)
(6) do not know	-.468*	-.680*	-0.699*	-.127	-.162	-.231
	(0.286)	(0.399)	(0.378)	(0.169)	(0.163)	(0.166)
P-values from Wald tests	0.000	0.003	0.030	0.001	0.005	0.034
	Does your husband usually help her/him with her/his homework by questioning, etc.?					
(1) very often	-.144**	-.160**	-0.110	-.103**	-.107**	-0.087*
	(0.069)	(0.068)	(0.071)	(0.046)	(0.046)	(0.049)
(2) rather often	-.011	-.044	-.008	0.002	-.015	-.005
	(0.051)	(0.050)	(0.050)	(0.037)	(0.036)	(0.036)
(3) now and then	0.033	0.009	0.059*	0.063**	0.053**	0.078***
	(0.034)	(0.033)	(0.033)	(0.026)	(0.026)	(0.026)
(4) rather seldom	0.034	0.001	0.018	0.021	0.006	0.010
	(0.036)	(0.035)	(0.035)	(0.029)	(0.029)	(0.029)
(5) almost never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
(6) do not know	0.348	0.657*	0.551*	-.559**	-.433*	-0.358
	(0.268)	(0.370)	(0.325)	(0.256)	(0.228)	(0.286)
P-values from Wald tests	0.097	0.075	0.060	0.001	0.002	0.002
Controls:						
SES	NO	YES	YES	NO	YES	YES
IQ	NO	NO	YES	NO	NO	YES

Coefficients are weighted least squares regression coefficients. Controls for SES include parental income, parental education and father's occupation. Controls for IQ include the number of points scored (0-40) on three different standardized tests: verbal, numerical and spatial. These are entered as continuous variables. These IQ tests were taken in the 6th grade when the children were 13 years old. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. The p-value from a Wald test of the null hypothesis that coefficients are jointly equal to zero is reported below each set of coefficients.

The second question concerning parental involvement, *Does your husband usually help her/him with her/his homework by questioning, etc.?*, was the single most important parental involvement variable in terms of its impact on the sister correlation. It lowered the family-variance component (on its own) by 9 percent. As shown in Table 7 (columns 1 and 2 for brothers and columns 4 and 5 for sisters), the answer (1) "very often" is negatively correlated with children's long-run income. This negative correlation is statistically significant even after controlling for parents' socio-economic status. The answer that is most strongly positively correlated with children's long-run income is (3) "now and then."

Thus far, the message from these two variables seems clear. Parents should show interest in their children's progress in school, but should not help them too often with their schoolwork. However, this interpretation raises the obvious question of why helping your children with their homework should hurt their long-run outcomes? Although one could certainly invent stories consistent with this thought, the most likely explanation is simply that those children who receive help with their homework "very often" are those who *need* the most help. In turn, low ability in school could be due to inherited traits, environmental factors or just plain bad luck. In columns 3 and 6 of Table 7, we see that controlling for IQ test scores significantly weakens this negative relationship.

Does our other variable for parental involvement in schoolwork also suffer from a similar type of selection bias? Do parents talk more with their children about school and show more interest in their children's schoolwork if those children are doing well in school? Do good students bring up the topic of school with their parents more often than poor students do? To take a quick look at this issue, we looked at the mean of the IQ test score for all children within each of the categorical answers (1) – (5) and found that the mean test score within each answer falls (monotonically) as parents report having shown less interest in their children's schooling.

This result means that it may not be parental involvement in schoolwork *per se* that causes siblings to be similar. It implies that siblings may already be doing well in school for other reasons. But this does not preclude the possibility that this "something else" is parental involvement at an earlier stage in the child's life. In fact, controlling for children's IQ test scores does not change the strong, positive relationship between parental interest and their children's adult incomes (see columns 3 and 6 in Table 7).

We now turn to our two questions concerning maternal patience (see Table 8). The first question, *Do you like to make long-term plans?,* was the single most important variable concerning parental patience for lowering the brother correlation. It lowered the family variance component (on its own) by 10 percent. The regression coefficients in the brother equations are always strongly, negatively correlated with their long-run income whenever the mother answers

this question in the negative, answers (3), (4) or (5). The results for sisters, however, are less significant, but become more negative as the mother's answer becomes more negative.

The second question concerning parental (maternal) patience, *Do you think it is worth planning for the future?,* was the single most important variable concerning parental patience for lowering the sister correlation. It lowered the family-variance component (on its own) by 6 percent. The regression coefficients in these equations are negatively correlated with children's long-run income whenever the mother answers this question in the negative, answers (3), (4) or (5).

So what is maternal patience? Our preferred interpretation is one in which patience is an attitude that parents pass on to their children, which in turn raises their propensity to save for the future and invest in education. In terms of the Becker and Tomes (1979) model of intergenerational mobility, patience would be a component of "family culture." This interpretation is in line with the literature that says that children "inherit" certain preferences and attitudes from their parents and with models of learned behavior (see, e.g., Mayer 1997, Duncan et al. 2005, Dohmen et al. 2006, Cesarini et al. 2009a).

A second interpretation is that these questions are proxies for mothers' non-cognitive abilities which they can then pass on to their children. Such abilities have been shown to be highly correlated across generations. They are passed down from both mothers and fathers and are important for labor market outcomes (Grönqvist et al. 2010).[34]

A third interpretation is one related to measurement error. One could argue that our measures of parental income in 1963, parental education and father's occupation in 1953 and 1963 lead to downward biased estimates of the impact of socio-economic status on children's long-run income since they are measured with some degree of error. The fact that our measures of parental attitudes help explain sibling similarities could be due to the fact that these new variables are simply adding more information about parents' socio-economic status.

It is likely that this is part of the story, but unlikely that it is the whole story. In Table 8, we see that the regression coefficients for parental patience do not move by much when we control for parental income, education and occupation. This is especially true for the brother coefficients. This implies that it would require an implausibly large amount of measurement error in our variables for parents' socio-economic status to make a story based *solely* on measurement error float.

Since this is an exploratory study, and since we do not feel that our observational data will ever lend themselves to pinning down the correct

[34] Using data from the NLSY, Groves (2005) documents significant father-son correlations in personality (as measured by the Rotter locus of control scale). She then goes on to show that these correlations can account for a significant share of the father-son correlation in earnings.

interpretation with any large degree of certainty, we leave this as an open question for future research. Instead, we wrap up this section with a final look at the bottom rows of Table 6.

Table 8. Selected Regression Coefficients from Models in Table 6 Concerning Parental (Maternal) Patience.

	(1) Brothers	(2) Brothers	(3) Sisters	(4) Sisters	(5) Brothers	(6) Brothers	(7) Sisters	(8) Sisters
	Do you like to make long-term plans?				*Do you think it is worth planning for the future?*			
(1) yes, definitely	-0.023 (0.029)	-.001 (0.028)	-0.020 (0.024)	-.004 (0.024)	Ref.	Ref.	Ref.	Ref.
(2) yes, perhaps	Ref.	Ref.	Ref.	Ref.	-.031 (0.029)	-.020 (0.028)	-.020 (0.023)	-.007 (0.022)
(3) do not know	-.178** (0.077)	-.141* (0.077)	0.049 (0.056)	0.062 (0.055)	-.093** (0.044)	-.066 (0.044)	-.143*** (0.039)	-.117*** (0.038)
(4) no, perhaps not	-.107*** (0.037)	-.111*** (0.036)	0.006 (0.027)	0.009 (0.026)	-.084** (0.040)	-.049 (0.039)	-.034 (0.032)	-.016 (0.032)
(5) no, definitely not	-.172*** (0.037)	-.150*** (0.036)	-.061** (0.031)	-.061** (0.030)	-.068 (0.047)	-.019 (0.0247)	-.089** (0.037)	-.047 (0.037)
P-values from Wald tests	0.000	0.000	0.154	0.120	0.093	0.514	0.002	0.041
Controls for SES	NO	YES	NO	YES	NO	YES	NO	YES

Coefficients are weighted least squares regression coefficients. Controls for SES include parental income, parental education and father's occupation. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. The p-value from a Wald test of the null hypothesis that coefficients are jointly equal to zero is reported below each set of coefficients.

The total reduction in the family-variation component that we can account for in this study is between 60 and 76 percent. These reductions account for 55 to 71 percent of the sibling correlation in the Family Study sample. This implies that more than half of the iceberg (that we alluded to in our introduction) can now be viewed from above the surface. In addition to parental income, education and occupation, we have also discovered that parental involvement in schoolwork and parental willingness to postpone financial gains and plan for the future (i.e., their patience) are factors that matter for children's adult outcomes and also work to make siblings more similar in terms of long-run income.

6. Sensitivity Analysis

Our baseline specification and sample definition involve a number of choices. In this section, we investigate the robustness of our main findings to several of these choices. For example, all sibling correlations in income are estimated using data for siblings born between 1949 and 1957, i.e., the maximum possible age

difference between any pair of siblings is nine years. If we remove these age limits altogether, then our new sibling correlations are 0.200 for brothers, 0.200 for sisters and 0.164 using the pooled data. Compare these with our baseline correlations of 0.250, 0.227 and 0.203, respectively. In this experiment, the family variance component is lowered by up to 64 percent for brothers, 52 percent for sisters and 58 percent for all siblings pooled. In our baseline experiment, these decreases were equal to 76, 64 and 60 percent respectively. Thus, as we expected, widening the allowable age gap between siblings lowers the shared component of our new, added variables. In contrast to this, the share of the family variance component explained by income, education and occupation is actually the same in both experiments.

Next, we perform a sensitivity analysis to check if our treatment of low, zero and missing incomes is important for our results. In our baseline specification, we calculated income averages using only those years in which income exceeds 10,000 SEK in 2001 prices. Here, we include all income years in these averages. Missing incomes are treated as zeros and included, so that all individuals now have 12 non-missing income years. Our new (old) sibling correlations are 0.187 (0.250) for brothers, 0.138 (0.227) for sisters and 0.146 (0.203) for the pooled data. Thus, treating low, zero and missing incomes in this fashion produces smaller sibling correlations. At the same time, we can explain a larger share of these smaller correlations. In this experiment, the family variance component is lowered by up to 84 percent for brothers, 76 percent for sisters and 58 percent for all siblings pooled. In our baseline experiment, these decreases were equal to 76, 64 and 60 percent, respectively. For brothers, the relative share of the family variance component explained by variables other than parental income, education and occupation, remains roughly constant. For sisters, on the other hand, the relative share of the family variance component that is explained by variables other than parental income, education and occupation, increases quite substantially.

As mentioned in Section 4, we ran all of the experiments using the unweighted Family Study data. Whenever possible, we also ran our experiments using all available SBC observations (not just those individuals who were selected into the Family Study). Although some of the quantitative results do change, none of the variables that we stress in this paper were unimportant in these alternative exercises. It is always the same set of variables in all exercises that matter most. For more details concerning these exercises see our discussion paper (Björklund et al. 2008). In particular, compare Tables 3A and 3B with Table 3 in this paper, Tables 4A and 4B with Table 4 in this paper, Tables 5A and 5B with Table 5 in this paper and Table 6 in that paper with Table 6 in this paper.

At the beginning of Section 5, we pointed out that there was a significant difference between our baseline sister correlation (0.227) and the sister correlation

calculating using all available SBC data (0.148). Tables 3B, 4B and 5B in our discussion paper show that the full sample does not produce different results from the ones reported in this paper concerning which variables and to what extent we can account for the sister correlation.[35]

A final concern that we have is the fact that the models in Table 6 are based on estimates of a very large number of variables (and dummies). Many more variables are included in Table 6 than in Table 3. For example, the models represented by the last rows of Table 6 include up to 173 estimated coefficients, while the last rows in Table 3 include only 24 estimated coefficients.

Imagine an extreme case in which all of our "new" variables and dummies included in Tables 4, 5 and 6 are simply random noise. How much of the variation in income could they explain and how far could we push the sibling correlation down using a large set of noisy, meaningless variables? To examine the relevance of this concern, we created 149 random variables (173 – 24 = 149). The same value for each of these random variables is shared by all siblings (just as all answers in the Family Study are shared by all siblings). We then "mimic" the exercise carried out in the bottom rows of Table 6 by adding these 149 random variables to the exercise in Table 3.

After controlling for parents' income, education and occupation, the family variance component in long-run income is 0.052 for brothers (see Table 3). After adding 149 extra random variables, we were only able to push this family variance component down to 0.043. This implies that only a small percentage of the increased explanatory power in Table 6 can be attributed solely to the large increase in the number of variables and dummies included in these models.

7. Conclusion

We have explored what lies behind sibling correlations in long-run income. As in previous Swedish studies, we estimated such correlations to be around 0.23. From the interpretation of a sibling correlation it then follows that 23 percent of the variation in long-run income can be attributed to factors that siblings share. Similar to previous studies of intergenerational correlations in long-run income, we also found that parents' socio-economic status can only account for 13 percent (sisters) and 28 percent (brothers) of this 23 percent.

Our contribution has been to explore whether family characteristics other than parents' income, education and occupation can account for more of these family background effects. We first added quite rich sets of indicators for family structure and social problems, but (overall) these added variables accounted for

[35] Please note that the sister correlation reported in our discussion paper using the full sample is equal to 0.168 (and not 0.148), since in that earlier version of this work we had a tighter (7 year) age window on siblings. We have now loosened that window to 9 years.

very little of sibling similarities in adult income above and beyond that already accounted for by parents' socio-economic status. Mother's age at first birth and the family's receipt of social support do appear to matter somewhat, but their importance appears confounded with that of parents' socio-economic status.

We also find that several variables, such as family type, which do have significant impacts on individual outcomes do not necessarily make siblings more similar. This implies that some important family-wide variables have heterogeneous effects on children, which should be taken into consideration when discussing sources of overall inequality. It appears that some inequality may be generated at home.[36]

When we added a set of indicators for parental involvement in schoolwork, parenting firmness, maternal patience and the number of books in the home, the explanatory power of our set of family-wide variables increased from 13 to 58 percent for sisters and from 28 to 71 percent for brothers. Indicators of parents' involvement in schoolwork and maternal patience, i.e., willingness to postpone benefits into the future and propensity to plan ahead, proved to be particularly important. This implies that parental involvement and parental attitudes may play a role in explaining sibling similarities. It appears that the lack of such positive attributes may help to perpetuate inequality across families.

Although these results do give some guidance for researchers to investigate the role of parental involvement and attitudes in more detail, the direct policy implications are not immediate. Affecting parental attitudes and parenting practices is not an easy task for politicians.

Future research would benefit from surveys that measure parental attitudes (such as patience) with greater precision. Also useful would be data that measure parental attitudes at different occasions and separately for each child (sibling). We would also like to know more about the nature of parental involvement in schoolwork. Do younger children receive as much "face time" as older children? Do child spacing and parental time constraints matter? Do gifted children get more or less help with their schoolwork than less gifted children? Finally, finding sources of exogenous variation in our explanatory variables is necessary for causal inference about what more than parental income, education and occupation is important for child outcomes and for making the outcomes of siblings so similar.

[36] Conley (2004) develops this thesis in his book *The Pecking Order*. Such heterogeneity also underscores the idea that the sibling correlation may only be a lower bound of the importance of family and community factors.

Appendix

Appendix Table A. Regression Coefficients from Models in Table 3 (Men Only).

Log father income 1963	0.219***				0.080*
	(0.027)				(0.035)
Father's income missing	-0.106**				-0.057
	(0.034)				(0.038)
Log mother income 1963		0.018			0.005
		(0.018)			(0.018)
Mother's income missing		0.096***			0.063**
		(0.024)			(0.024)
Father's education					
missing		-0.099			0.000
		(0.068)			(0.000)
Grade school		Ref.			Ref.
High school		0.174***			0.043
		(0.033)			(0.040)
College		0.397***			0.177**
		(0.044)			(0.075)
Mother's education					
missing			-0.141**		-0.096
			(0.068)		(0.066)
Grade school			Ref.		Ref.
High school			0.331***		0.163***
			(0.060)		(0.064)
College			0.335***		0.122
			(0.097)		(0.101)
Father's occupational category 1953					
Missing				-0.154*	-0.128*
				(0.076)	(0.074)
Upper & upper middle class				0.089	0.024
				(0.058)	(0.063)
Lower middle class - employees				Ref.	Ref.
Lower middle class-entrepreneurs				0.032	0.056
				(0.055)	(0.054)
Working class - skilled				-0.025	-0.006
				(0.037)	(0.037)
Working class - unskilled				-0.066	-0.038
				(0.043)	(0.043)
Father's occupational category 1963					
Missing				-0.216***	-0.159**
				(0.072)	(0.081)
Upper & upper middle class				0.118***	-0.003
				(0.055)	(0.061)

Appendix Table A (continued)

Lower middle class - employees	Ref.	Ref.
Lower middle class- entrepreneurs	-0.075	-0.065
	(0.050)	(0.050)
Working class - skilled	-0.091**	-0.067*
	(0.036)	(0.037)
Working class - unskilled	-0.172***	-0.140***
	(0.044)	(0.044)

Coefficients are weighted least squares regression coefficients produced in the first stage of our variance decomposition exercise. All control variables except log income are entered as dummy variables. Gender and birth-year dummies are included in all regressions (but not reported). Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table B. Regression Coefficients from Models in Table 5 (Men Only).

Social support	-.20***						-.14***	-.06*
	(0.027)						(0.032)	(0.033)
Alcoholic		-.34***					-.21***	-.19***
		(0.067)					(0.068)	(0.067)
Drunken behavior		-0.04					0.05	0.09
		(0.081)					(0.084)	(0.081)
Mother record of mental illness			-.18***				-.03	-.03
			(0.061)				(0.065)	(0.064)
Father record of mental illness			-.18**				-.06	-.08
			(0.088)				(0.090)	(0.090)
Father's crime extensive margin				-.15***			0.04	.08*
				(0.035)			(0.042)	(0.042)
Father's crime intensive margin					-.06***		-.04***	-.04***
					(0.011)		(0.013)	(0.013)
Mother died						-.09	-.01	-.14
						(0.496)	(0.606)	(0.720)
Father died						-.13	-.07	-.02
						(0.097)	(0.094)	(0.099)
Controls for SES	NO	NO	NO	NO	NO	NO	NO	YES

Coefficients are weighted least squares regression coefficients produced in the first stage of our variance decomposition exercise. All variables concerning social problems are simple yes (=1) or no (=0) dummy variables except for father's criminality at the intensive margin, which is treated as a continuous variable. Gender and birth-year dummies are included in all regressions (but not reported). *Controls for SES* include parental income, parental education and father's occupation. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table C. Regression Coefficients from Models in Table 4 (Men Only).

Mother's age at first birth						
15-19	-0.10***				-0.082**	-0.043
	(0.036)				(0.036)	(0.035)
20-24	Ref.				Ref.	Ref.
25-29	0.139***				0.124***	0.054*
	(0.029)				(0.029)	(0.028)
30-34	0.117***				0.096**	0.022
	(0.038)				(0.039)	(0.039)
35-39	0.142**				0.103	0.070
	(0.066)				(0.066)	(0.068)
40-44	-0.019				-0.066	-0.050
	(0.148)				(0.151)	(0.147)
Family type in 1960						
Missing		-0.174**			-0.152**	0.000
		(0.068)			(0.068)	(0.000)
Married, but not cohabitating		-0.095			0.043	0.036
		(0.090)			(0.086)	(0.085)
Married, cohabitating		Ref.			Ref.	Ref.
Single		-0.006			0.111	0.136
		(0.076)			(0.093)	(0.092)
Widow(er)		0.095			0.226*	0.219**
		(0.094)			(0.105)	(0.103)
Divorced		-0.253***			-0.048	-0.063
		(0.084)			(0.088)	(0.086)
Family type in 1963						
Father & mother living together			Ref.		Ref.	Ref.
Single mother			-0.186***		-0.193***	-0.162**
			(0.046)		(0.053)	(0.076)
Single father			-0.007		0.039	0.028
			(0.167)		(0.156)	(0.155)
Mother & step father			-0.610**		-0.556	-0.436
			(0.309)		(0.355)	(0.311)
Father & step mother			0.103**		0.082	0.151
			(0.098)		(0.117)	(0.115)
Number of siblings				-0.039***	-0.020**	-0.023**
				(0.009)	(0.010)	(0.010)
Controls for SES	NO	NO	NO	NO	NO	YES

Coefficients are weighted least squares regression coefficients produced in the first stage of our variance decomposition exercise. All variables are treated as categorical variables except for *Number of siblings* which is treated as a continuous variable. Gender and birth-year dummies are included in all regressions (but not reported). *Controls for SES* include parental income, parental education and father's occupation. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

References

Becker, Gary and Nigel Tomes (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy* 87(6), 1153-1189.

Björklund, Anders, Tor Eriksson, Markus Jäntti, Oddbjörn Raaum and Eva Österbacka (2002), "Brother Correlations in Denmark, Finland, Norway and Sweden Compared to the United States," *Journal of Population Economics* 15(4), 757-772.

Björklund, Anders and Markus Jäntti (2009), "Intergenerational Income Mobility and the Role of Family Background," in W. Salverda, B. Nolan and T. Smeeding (eds.), *Oxford Handbook of Economic Inequality*, Oxford University Press, New York.

Björklund, Anders, Markus Jäntti and Matthew J. Lindquist (2009), "Family Background and Income during the Rise of the Welfare State: Brother Correlations in Income for Swedish Men Born 1932 – 1968," *Journal of Public Economics* 93(5-6), 671-680.

Björklund, Anders, Lena Lindahl and Matthew J. Lindquist (2008), "What More Than Parental Income? An Exploration of What Swedish Siblings Get from Their Parents," IZA Discussion Paper No. 3735.

Björklund Anders and Marianne Sundström (2006), "Parental Separation and Children's Educational Attainment: A Siblings Analysis on Swedish Register Data," *Economica* 73(292), 605-624.

Böhlmark, Anders and Matthew J. Lindquist (2006), "Life-Cycle Variation in the Association between Current and Lifetime Income: Replication and Extension for Sweden," *Journal of Labor Economics* 24(4), 879-96.

Bourguignon Francois, Francesco H. G. Ferreira and Marta Menéndez (2007), "Inequality of Opportunity in Brazil," *Review of Income and Wealth* 53(4), 585-618.

Case, Anne and Lawrence Katz (1991), "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth," NBER Working Paper 3705.

Cesarini, David, Christopher T. Dawes , Magnus Johannesson, Paul Lichtenstein and Björn Wallace (2009a), "Genetic Variation in Preferences for Giving and Risk-Taking," *Quarterly Journal of Economics* 124(2), 809-842.

Cesarini, David, Magnus Johannesson, Paul Lichtenstein and Björn Wallace (2009b), "Heritability of Overconfidence," *Journal of the European Economic Association* 7(2-3), 617-627.

Cesarini, David, Magnus Johannesson, Paul Lichtenstein and Örjan Sandewall, (2010), "Genetic Variation in Financial Decision Making," *Journal of Finance* 65(5), 1725-1754.

Conley, Dalton (2004), *The Pecking Order*, Vintage Books, New York.

Corcoran, Mary, Christopher Jencks and Michael Olneck (1976), "The Effects of Family Background on Earnings," *American Economic Review* 66(2), 430-35.

Datcher, Linda (1982), "Effects of Community and Family Background on Achievement," *Review of Economics and Statistics* 64(1), 32-41.

Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde (2006), "The Intergenerational Transmission of Risk and Trust Attitudes," IZA Discussion Paper 2380.

Duncan, Greg, Ariel Kalil, Susan E. Mayer, Robin Tepper and Monique R. Payne (2005), "The Apple Does Not Fall Far from the Tree," in S. Bowles, H. Gintis and M. Osborne Groves (eds.), *Unequal Chances*, Princeton University Press, Princeton.

Erikson, Robert (1987), "The Long Arm of the Origin: The Effects of Family Background on Occupational and Educational Achievement," in U. Bergryd and C-G. Janson (eds.), *Sociological Miscellany*, University of Stockholm, Stockholm.

Fryer Jr., Roland G. and Steven D. Levitt (2004), "Understanding the Black-White Test Score Gap in the First Two Years of School," *The Review of Economics and Statistics* 86(2), 447-464.

Geronimus, Arline T. and Sanders Korenman (1992), "The Socioeconomic Consequences of Teen Childbearing Reconsidered," *Quarterly Journal of Economics* 107(4), 1187-1214.

Ginther, Donna and Robert Pollak (2004), "Family Structure and Children's Educational Outcomes: Blended Families, Stylized Facts, and Descriptive Regressions," *Demography* 41(4), 671-96.

Grönqvist, Erik, Björn Öckert and Jonas Vlachos (2010), "The Intergenerational Transmission of Cognitive and Non-Cognitive Abilities," CEPR Discussion Paper No. 7908.

Hill, Martha S., Wei-Jun J. Yeung and Greg J. Duncan (2001), "Childhood Family Structure and Young Adult Behaviors," *Journal of Population Economics* 14(2), 271-299.

Hjalmarsson, Randi and Matthew J. Lindquist (2010), "Driving Under the Influence of Our Fathers," *The B.E. Journal of Economic Analysis & Policy* 10(1) Topics, Article 100.

Hjalmarsson, Randi and Matthew J. Lindquist (2009), "Like Godfather, Like Son: Explaining the Intergenerational Nature of Crime," Department of Economics, Stockholm University, Research Papers in Economics No 2009:18.

Holmlund, Helena (2005), "Estimating Long-Term Consequences of Teenage Childbearing: An Examination of the Siblings Approach," *Journal of Human Resources* 40(3), 716-743.

Lindahl, Lena (2008), "Do Birth Order and Family Size Matter for Intergenerational Income Mobility? Evidence from Sweden," *Applied Economics* 40(17), 2239-2257.

Lindahl, Lena (forthcoming), "A Comparison of Family and Neighborhood Effects on Grades, Test Scores, Educational Attainment and Income – New Evidence for Sweden," *Journal of Economic Inequality*.

Mason, Patrick L. (2007), "Intergenerational Mobility and Interracial Inequality: The Return to Family Values," *Industrial Relations* 46(4), 51-80.

Mayer, Susan E. (1997), *What Money Can't Buy: Family Income and Children's Life Chances*, Harvard University Press, Cambridge, Massachusetts.

Mazumder, Bhashkar (2008), "Sibling Similarities and Economic Inequality in the U.S.," *Journal of Population Economics* 21(3), 685-701.

Osborne Groves, Melissa (2005), "Personality and the Intergenerational Transmission of Economic Status," in S. Bowles, H. Gintis and M. Osborne Groves (eds.), *Unequal Chances*, Princeton University Press, Princeton.

Page, Marianne and Gary Solon (2003a), "Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban," *Journal of Labor Economics* 21(4), 831-855.

Page, Marianne and Gary Solon (2003b), "Correlations between Sisters and Neighboring Girls in Their Subsequent Income as Adults," *Journal of Applied Econometrics* 18(5), 545-562.

Raaum, Oddbjörn, Kjell Salvanes and Erik Sørensen (2006), "The Neighborhood Is Not What It Used to Be," *Economic Journal* 116(1), 200-222.

Rabe-Hesketh, Sophia, Anders Skrondal and Andrew Pickles (2002), "Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature," *Stata Journal* 2(1), 1-21.

Roemer, John E. (1998), *Equality of Opportunity*, Harvard University Press, Cambridge, MA.

Sieben, Inge and Paul M. De Graaf (2003), "The Total Impact of the Family on Educational Attainment: A Comparative Sibling Analysis," *European Societies* 5(1), 33-68.

Solon, Gary (1999), "Intergenerational Mobility in the Labor Market," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* 3A, Elsevier, Amsterdam, North Holland.

Solon, Gary, Marianne Page, and Greg Duncan (2000), "Correlations between Neighboring Children in Their Subsequent Educational Attainment," *Review of Economics and Statistics* 82(3), 383-392.

Solon, Gary, Mary Corcoran, Roger Gordon and Deborah Laren (1991), "A Longitudinal Study of Sibling Correlations in Economic Status," *Journal of Human Resources* 26 (3), 509-534.

Stenberg, Sten-Åke (2000), "An Intergenerational Study of Social Assistance Recipiency in Postwar Sweden," *Journal of Marriage and the Family* 62(1), 228-239.

Stenberg, Sten-Åke and Denny Vågerö (2006), "Cohort Profile: The Stockholm Birth Cohort of 1953," *International Journal of Epidemiology* 35(3), 546-548.

Yeung, W. Jean, Greg J. Duncan and Martha S. Hill (2000), "Putting Fathers Back in the Picture: Parental Activities and Children's Adult Outcomes," *Marriage and Family Review* 29(2), 97-113.