# The Impacts of Family and Community Effects on the Inequality of Long Term Earnings Evidence from Siblings, Schoolmates and Neighbors' Correlations

Paul Bingley, SFI Copenhagen

Lorenzo Cappellari, Università Cattolica Milano

Konstantinos Tatsiramos, University of Nottingham and IZA

**Preliminary First Draft: June 2013**

**Abstract**

This paper develops a unified framework which enables disentangling the contribution of the family, the school, and the neighborhood in labor earnings over the life-cycle. This is achieved within a model of multi-person earnings dynamics distinguishing permanent from transitory earnings and allowing for heterogeneous earnings growth. The analysis is based on administrative registers from the Danish population, which allows connecting members of the same family and link each family member with other individuals at the community and school level. Our preliminary results, which are based on the comparison between family and school effects, suggest that family effects are by far the most relevant factor that shapes long-term incomes. This is true both for initial earnings and for earnings growth rates, where they represent 65 and 56 percent of total dispersion respectively. The other more relevant source of permanent inequality in earnings is the individual idiosyncratic component, accounting for approximately 20 percent of the dispersion in both initial earnings and earnings growth. The remaining shares are accounted for by school effects. Decomposing the sibling correlation into family and school effects, we find that the sibling correlation is driven by family effects, while school effects seem to play no role.

## 1. Introduction

Understanding what determines labor market outcomes over the life-cycle is important for explaining the driving forces of existing inequalities and for interventions that aim to reduce them. There are two main factors that shape the earnings potential of individuals: the family and the environment, or community, in which individuals grow up and live. The family affects outcomes through the transfer of abilities, preferences and resources. The community or neighborhood may also determine individual outcomes in many ways: through institutions such as the school and its quality, through peer influences, or through social norms and the influence of role models by others in the community.

This paper develops a unified framework which enables to disentangle the contribution of the family, the school and the neighborhood in labor earnings over the life-cycle. This is achieved within a model of multi-person earnings dynamics distinguishing permanent from transitory earnings and allowing for heterogeneous earnings growth. The analysis is based on administrative registers from the Danish population, which enables connecting member of the same family and link each family member with other individuals at the community and school level.

There exists some ambiguity in the existing literature as to the importance of the environment outside the family in explaining the variation of long-term earnings. There are two main approaches to gauge the importance of family and community factors. The first approach uses the decomposition of sibling correlation by Solon (1999) into the degree of intergenerational correlation of earnings (IGE) and other factors, suggesting an important role of community effects that are shared by siblings but are not related to parental income (e.g. Björklund and Jännti, 2009). The second approach to separate family from community factors is based on the estimation of correlations in long-run earnings for unrelated individuals who grew up in the same neighbourhood (e.g. Page and Solon, 2003a, 2003b) and their comparison with the sibling correlation. The findings suggest that family factors are the most important in explaining the similarity of siblings' earnings, while the effect of neighbourhood tends to be small.

Our approach extends the existing literature which models the correlation of siblings as an 'omnibus measure' for the importance of family and community effects in determining the inequality of earnings in the long-run. In particular, on top of the sibling correlation, we also consider the correlation of earnings between each sibling and his peers at school, and the correlation of earnings of each sibling with peer neighbors. This enables the decomposition of sibling correlations into family and school effects within a unified model.

## 2. Literature

There is a large literature using sibling correlations in long-run earnings as an omnibus measure of family and community influences in explaining permanent earnings inequality. The correlation of siblings' earnings captures all those factors which are shared between siblings transmitted through the family, the school and the neighborhood. Björklund and Jännti (2009) and Black and Devereux (2011) classify sibling studies as one of the methods for shedding light on the mechanisms behind intergenerational associations.

At the family level, the shared factors include income and values transmitted across generations. Following the seminar contribution of Becker and Tomes (1986) parents care about the lifetime earnings of their children and maximize utility by choosing between own consumption and investment in child earnings capacity. Offspring outcomes depend also on other productive endowments which are transmitted through the generations. As a result, lifetime earnings are transmitted across generations, through parental incomes and productive endowments.

Beyond what is transmitted through the family, any correlation of siblings' long-run earnings captures the exposure to similar influences outside the family, such as, neighbors, peers, or schools. Previous research has used two main ways to gauge the effect of those community factors in sibling correlations. The first approach uses the decomposition of sibling correlation by Solon (1999) into the degree of intergenerational correlation of earnings (IGE) and other factors. The larger the difference between sibling and intergenerational correlations, the more important are environmental factors that siblings share independently of the parents.

Björklund and Jännti (2009) report for Denmark an IGE of about 0.12 and a brother correlation of about 0.23. In both instances, Denmark ranks at the bottom of the correlation tables, i.e. it appears the country where the influence of family factors is the least important. Using the Solon (1999) decomposition to assess the importance of parental incomes in shaping sibling correlations it is shown that in Denmark the role of parental income is negligible, explaining 5 percent of the overall sibling correlation.

The second approach to separate family from community factors is based on the estimation of correlations in long-run earnings for unrelated individuals who grew up in the same neighbourhood (e.g. Page and Solon, 2003a, 2003b). The idea is that the correlation of long-run earnings of individuals who do not belong to the same family but live in the same neighbourhood will capture the common community influences that are not entirely driven by the family, although families living close will tend to be similar but not identical. The findings suggest that family factors are the most important in explaining the similarity of siblings' earnings, while the effect of neighbourhood tends to be small.

There are a number of methodological issues in the estimation of intergenerational correlation of earnings and sibling correlations, which are related to the measurement of parental and offspring income. In the presence of transitory shocks, a point-in-time measure of income is a mixture of long-term income and transitory income shocks. This leads to an underestimation of the intergenerational elasticity of incomes because the transitory shocks introduce classical measurement error (Solon, 1992 and Zimmerman, 1992). The common practice in the literature is to use averages over a limited number of time periods in order to mitigate the measurement error. When transitory shocks are not purely transitory but are characterized by serial correlation (even a moderate one), the measurement error on permanent incomes becomes more severe and harder to integrate out and it requires as many as 30 years of income (Mazumder, 2005).

Another source of bias in the estimation of IGEs and sibling correlations is the variation of long-term income over the life-cycle. This "life-cycle bias" stems from the fact that typically parents' and children's incomes are sampled at different phases of the life-cycle. For children, current income is measured too early in their life, while for parents measurement occurs too late, which leads to under- and over- estimate (respectively) long-term incomes (Jenkins, 1987 and Grawe, 2006). Haider and Solon (2006) show that if there is individual heterogeneity in life-cycle earnings growth, then the relationship between current and lifetime earnings varies over the life-cycle, and the bias incurred by using annual in place of lifetime measures is minimized in the 30-40 age range.

The findings in the literature addressing these measurement issues resulted in a substantial increase of the IGE from 0.2 to 0.4 (Solon,1992 and Zimmerman, 1992), which was increased to 0.6 when using data on sixteen-year income strings for fathers to capture serial correlation in the transitory shocks. (Mazumder, 2005). More recently, Anti-Nielsen et al. (2011) analyze intergenerational mobility in Norway using long strings of data on both fathers and sons. They show that taking averages of fathers' incomes over longer time windows than preceding studies increases the estimated elasticity. Increasing the starting age for fathers' income strings has the opposite effect. Transitory shocks and life-cycle biases can explain both findings. Finally, Nybom and Stuhler (2011) show how life-cycle bias gives rise to non-classical measurement error. They demonstrate that current strategies for estimating the IGE are still prone to substantial bias, and conclude calling for an explicit allowance for heterogeneous life-cycle growth across individuals in studies of intergenerational income associations.

Similar underestimation due to measurement error has been shown to exist for sibling correlations. Solon et al. (1991) and Altonji and Dunn (1991) report sibling correlations of about 0.35, which are much higher than those estimated in most of the preceding U.S. literature. Björklund et al. (2009) estimate a model of sibling earnings in Sweden. They draw data on siblings in the 30-40 age range, in order to minimize life-cycle bias, and model serially correlated transitory shocks. They show that the sibling correlation has been declining during the expansion of the Swedish welfare state in the 1960s.

One challenge faced by both approaches used to estimate the effect of community factors is that they capture many different aspects of the neighborhood including the effect of neighbors, the effect of school mates and the quality of schools. Identifying separately each of the community factors (school, neighbors, peers etc.) is challenging because one needs to be able to observe not only siblings but also their peers at school and their neighbors along with their long-run incomes.

## 3. Econometric model

Previous studies have used sibling correlations as 'omnibus measures' for the importance of family and community effects in determining the inequality of earnings in the long-run. Our purpose in this section is to develop an econometric model for assessing the relative importance of family and community within the sibling correlation. In particular, we will consider two dimensions of the community, namely schools and neighbors when the siblings were in grade 8 and aged 16, respectively, corresponding to the earlier points in time when we observe community affiliations. We define neighbors using parish of residence, which represents an intermediate level of geographical aggregation between the other alternatives available in the Danish registry, i.e. streets and zip codes.

Previous studies have highlighted two main challenges for the estimation of sibling earnings models. First, while the focus of theoretical models of intra-family earnings dependence is on permanent earnings representing long-term earning capacity (Becker and Tomes, 1986), available datasets provide information on current earnings, which are a mixture of permanent earnings and transitory earnings shocks, the latter generating measurement error bias (Solon, 1992; Zimmermann, 1992). It has also been noted that transitory earnings shocks are typically serially correlated, which implies that even the multi-period earnings averages commonly used as proxies of permanent earnings are measured with error (Mazumder, 2005). Second, when siblings are born far apart from one another, differences in their earnings may be an artifact of life-cycle earnings growth rather than a genuine differential in long-term earnings capacity, again resulting in downward biased estimates of the sibling correlation.

We tackle both estimating challenges with a model of multi-person earnings dynamics distinguishing permanent from transitory earnings and allowing for heterogeneous earnings growth. Our interest is in the effect of families, schools and neighbors on permanent earnings. We will base our estimation of school and neighbor effects on average earnings of schoolmates and neighbors assuming that they do not contain information on individual-level transitory earnings fluctuations. Therefore we do not allow for school and neighbor effects in the transitory component of earnings. Conversely, family effects exploit individual level variation of each of the two brothers (reflecting both from permanent and transitory earnings) so that we can allow for family effects in transitory earnings.

We specify individual earnings as:

$$w_{ifspa} = y_{ifspa} + v_{ifa} \; ; \; E(y_{ifspa}, v_{ifa}) = 0, \qquad (1)$$

where the indices *i, f, s, p* and *a* stand for individual, family, school, neighbor (parish) and age in deviation from the life cycle starting point (=age 21) respectively, *w* represent the log of age adjusted gross annual earnings, which are assumed to be the sum of a permanent (*y*) and transitory (*v*) components, orthogonal by definition. Separate identification of permanent and transitory earnings is granted by the availability of individual level panel data and ensures that we estimate sibling correlations in permanent earnings, avoiding measurement error biases due to transitory shocks.

*Specification of permanent earnings*

We model life-cycle dynamics of permanent earnings using a random growth specification, consisting of an individual-specific linear profile in age. While simple, the linear specification allows for heterogeneity both in starting earnings and earnings growth, the latter being crucial in avoiding life-cycle biases. Our permanent earnings model is:

$$y_{ifspa} = \lambda_t\left[\left(\alpha_i + \alpha_f + \alpha_s + \alpha_p\right) + \left(\beta_i + \beta_f + \beta_s + \beta_p\right)a\right]; \quad t = c(i) + 21 + a, \tag{2}$$

where $c(i)$ is the birth cohort of person $i$ and $\lambda_t$ is a calendar time shifter. We separate time and age effects exploiting earnings variances and covariances computed within 3-year birth cohorts, and using them jointly in estimation, where each cohort is conventionally imputed its central year of birth.

Both intercepts and slopes of the individual-specific linear profile are factored into four zero-mean components. The variance of each component captures idiosyncratic, family, school and neighbor heterogeneity in either initial earnings (the $\alpha$s) and life-cycle earnings growth (the $\beta$s). We assume earnings components to be correlated within each dimension of heterogeneity and uncorrelated across dimension. Correlation of earnings intercepts and slopes within each dimension of heterogeneity is allowed in order to capture the commonly observed fact that human capital investments at the beginning of the life-cycle lower the initial earnings of investors and raise their life-cycle growth rates, resulting in negative covariances of intercepts and slopes. This result in 'Mincerian cross-overs' of life cycle earnings profiles, and a u-shaped evolution of permanent earnings dispersion with age. The assumption of factor independence across dimensions of heterogeneity is done for simplicity and can be relaxed in robustness analyses. Our distributional assumptions are summarized as follows:

$$\left(\alpha_i\ \beta_i\right) \sim \left(0, 0;\ \sigma^2_{\alpha I}, \sigma^2_{\beta I}, \sigma_{\alpha\beta I}\right) \tag{3a}$$

$$\left(\alpha_f\ \beta_f\right) \sim \left(0, 0;\ \sigma^2_{\alpha\Phi}, \sigma^2_{\beta\Phi}, \sigma_{\alpha\beta\Phi}\right) \tag{3b}$$

$$\left(\alpha_s\ \beta_s\right) \sim \left(0, 0;\ \sigma^2_{\alpha\Sigma}, \sigma^2_{\beta\Sigma}, \sigma_{\alpha\beta\Sigma}\right) \tag{3c}$$

$$\left(\alpha_p\ \beta_p\right) \sim \left(0, 0;\ \sigma^2_{\alpha\Pi}, \sigma^2_{\beta\Pi}, \sigma_{\alpha\beta\Pi}\right) \tag{3d}$$

*Identification of permanent earnings*

The assumptions above fully specify the intertemporal and interpersonal distribution of permanent earnings. Identification of its parameters is achieved exploiting different types of moment restrictions generated by the model. For a given individual, moment restrictions for two non-necessarily different time periods *t* and *t'* are a function of all sources of earnings heterogeneity:

$$E\left(y_{ifspa}, y_{ifspa'}\right) = [(\sigma^2_{\alpha I} + \sigma^2_{\alpha\Phi} + \sigma^2_{\alpha\Sigma} + \sigma^2_{\alpha\Pi}) + \left(\sigma^2_{\beta I} + \sigma^2_{\beta\Phi} + \sigma^2_{\beta S} + \sigma^2_{\beta\Pi}\right)aa' + \\ (\sigma_{\alpha\beta I} + \sigma_{\alpha\beta\Phi} + \sigma_{\alpha\beta S} + \sigma_{\alpha\beta\Pi})(a+a')]\lambda_t\lambda_{t'} \tag{4}$$

Interpersonal moment restrictions do not depend on individual heterogeneity. Moment restrictions between siblings (different *i* but same *f*) depend on the family effect. Moreover, depending upon siblings sharing schools and neighbors, moment restrictions will also depend on school and neighbor effects.[1] Therefore, we distinguish among four types of between-sibling moments: sharing both school and neighbors, sharing school only, sharing neighbors only, not sharing school or neighbor. For $s \neq s'$ and $\neq p'$ , respective moment restrictions are as follows:

[1] This is one difference with PSID-based studies (e.g. Page and Solon, 2003) in which all sibling share the neighbor by sampling design.

$$E\left(y_{ifspa}, y_{i'fspa'}\right) = [(\sigma^2_{\alpha\Phi} + \sigma^2_{\alpha\Sigma} + \sigma^2_{\alpha\Pi}) + \left(\sigma^2_{\beta\Phi} + \sigma^2_{\beta\Sigma} + \sigma^2_{\beta\Pi}\right)aa' + (\sigma_{\alpha\beta\Phi} + \sigma_{\alpha\beta\Sigma} + \sigma_{\alpha\beta\Pi})(a+a')]\,\lambda_t\lambda_{t'} \tag{5.a}$$

$$E\left(y_{ifspa}, y_{i'fsp'a'}\right) = [(\sigma^2_{\alpha\Phi} + \sigma^2_{\alpha\Sigma}) + \left(\sigma^2_{\beta\Phi} + \sigma^2_{\beta\Sigma}\right)aa' + (\sigma_{\alpha\beta\Phi} + \sigma_{\alpha\beta\Sigma})(a+a')]\,\lambda_t\lambda_{t'} \tag{5.b}$$

$$E\left(y_{ifspa}, y_{i'fs'pa'}\right) = [(\sigma^2_{\alpha\Phi} + \sigma^2_{\alpha\Pi}) + \left(\sigma^2_{\beta\Phi} + \sigma^2_{\beta\Pi}\right)aa' + (\sigma_{\alpha\beta\Phi} + \sigma_{\alpha\beta\Pi})(a+a')]\,\lambda_t\lambda_{t'} \tag{5.c}$$

$$E\left(y_{ifspa}, y_{i'fs'p'a'}\right) = [(\sigma^2_{\alpha\Phi}) + \left(\sigma^2_{\beta\Phi}\right)aa' + (\sigma_{\alpha\beta\Phi})(a+a')]\,\lambda_t\lambda_{t'} \tag{5.d}$$

The above moment conditions are sufficient for identifying family, school, and neighbor effects. In particular, identification of school and neighbor effects is ensured by the presence of siblings that went to different schools or grew up in different neighbors. In order to avoid relying exclusively on these specific groups of siblings for the identification of school and neighbor effects, we exploit population data to recover inter-personal moment restrictions linking the two brothers to their peers in grade 8 or living in the same parish at 16. We define grade 8 peers as those born in the same year and attending grade 8 in the same school. Similarly, peer neighbors are born in the same year and lived in the same parish at age 16. There are three relevant sets of peers: those attending the same school and living in the same neighbor; those attending the same school but not living in the same neighbor; and those living in the same neighbor but attending different schools. Empirically, we exploit intertemporal earnings covariances between individuals and their average peers, where averages exclude own brothers. Respective moment restrictions are as follows, overbars denoting averages over the relevant sets:

$$E\left(y_{ifspa}, \bar{y}_{spa'}\right) = [(\sigma^2_{\alpha\Sigma} + \sigma^2_{\alpha\Pi}) + \left(\sigma^2_{\beta\Sigma} + \sigma^2_{\beta\Pi}\right)aa' + (\sigma_{\alpha\beta\Sigma} + \sigma_{\alpha\beta\Pi})(a+a')]\ \lambda_t\lambda_{t'} \qquad (6.a)$$

$$E\left(y_{ifspa}, \bar{y}_{sp'a'}\right) = [(\sigma^2_{\alpha\Sigma}) + \left(\sigma^2_{\beta\Sigma}\right)aa' + (\sigma_{\alpha\beta\Sigma})(a+a')]\ \lambda_t\lambda_{t'} \qquad (6.b)$$

$$E\left(y_{ifspa}, \bar{y}_{s'pa'}\right) = [(\sigma^2_{\alpha\Pi}) + \left(\sigma^2_{\beta\Pi}\right)aa' + (\sigma_{\alpha\beta\Pi})(a+a')]\ \lambda_t\lambda_{t'} \qquad (6.c)$$

*Specification of transitory earnings*

We model transitory earnings using an AR(1) process in order to capture serial correlation of transitory shocks. We allow the distribution of transitory earnings to be brother-specific, and we account for age effects in transitory shocks through an exponential spline. We allow for contemporaneous correlation of shocks across persons. Our model for transitory earnings is as follows:

$$\begin{gathered} v_{ifa} = \eta_t u_{ifa};\ u_{ia} = \ \rho_b u_{ifa-1} + \varepsilon_{ifa}; \\ \varepsilon_{ifa} \sim \left(0, \sigma^2_{\varepsilon b} \exp\left(g_b(a)\right)\right), u_{if21} \sim (0, \sigma^2_{21b}); cov\left(\varepsilon_{ifa}\varepsilon_{i'fa}\right) = \sigma_f, \end{gathered} \qquad (7)$$

where $b$=1,2 denotes brother specific parameters, $\eta_t$ is a time loading factor and $g_b(\ )$ is a brother specific linear spline with knots at 25, 30 and 35 years of age.

## 4. Data

We use data from registers of the Danish population. Register data enable connection of individuals with members of their families as well as their school peers and neighbours. Because the latter information is unavailable at the time of writing the present note, in what follows we will focus the data description and the analysis on school and family effects, leaving neighbour effects for future versions of this research.

We define school effects in terms of schools of grade 8 attendance. Register data on earnings span the 1980 – 2009 period. Similarly, information on schools is available starting in 1980, corresponding to individuals born 1966 onwards. We group individuals into 3-year birth cohorts imputing the central age to all cohort members, and conventionally fix the initial point of life cycle observation at age 21. Thence, the first year of valid observations on earnings is 1988, when the 1967 cohort (individuals born 1966 to 1968) turns 21.

We connect siblings exploiting information on their parents' personal identifiers, which we use to derive a sample of matched brothers. We consider only the first two brothers in the family and do not consider third and onwards brothers. Families with less than three brothers represent 97% of the population of families with male children. In keeping with the sibling correlations literature, we also include singletons in the sample. We consider full biological brothers sharing both parents according to the medical birth register. We consider brothers whose age difference is between one and ten years. We study men's earnings and do not consider mother/daughter, mother/son, father/daughter, brother/sister or sister/sister earnings associations. Brothers' ordering is determined irrespective of the presence of sisters: for example, we do not make any distinction for whether there is one sister born in-between the two brothers, before or after.

We use pre-tax annual earnings, i.e. income from labour. In order to model life-cycle dynamics we require observation of individual earnings strings over time. We focus on prime age men observed from age 21 onwards. We select birth cohorts so that each cohort is observed at least 5 times. In practice the shortest span of observation is the one of cohort 1982, corresponding to the seven years between 2003 and 2009. We require availability of at least 3 consecutive earnings observations at the individual level.

To identify school effects we use the earnings of school peers, defined as males who attended grade 8 in the same school and were born in the same year. For each individual, we generate the earnings profile of the average peer, i.e. the yearly average of earnings of his school peers, conditional on groups including at least 10 peers, and dropping individuals who could not be matched to peers.

**5. Descriptive statistics**

In this section we provide a description of the earnings variances and covariances in our data. There are three main perspectives from which one can look at raw earnings variances and covariances: individual, cross-siblings, cross-school peers. We start in Figure 1 with a description of individual earnings moment over age and plot the profile of the variance and of the 5 year-lag and 10 year-lag covariances. While the first provides a measure of both permanent and transitory sources of variation in individual earnings, "long" covariances are supposed to be approximation of dispersion in permanent earnings, the longer the lag the better the approximation. After an initial increase, the variance profile shows a marked decline between the ages of 25 and 30, and stabilizes thereafter. Such a decline is not evident in the covariance profile which displays a moderate increase over age. If we take the difference between the variance and the covariances as a rough estimate of the transitory component, the figure suggests that annual earnings are more unstable at the start of the life-cycle compared with intermediate ages, confirming findings from previous research on individual earnings such as the ones of Baker and Solon (2003). The graph also shows that the decline between the 5 and 10 year lags covariance is relatively negligible, suggesting that transitory shocks are not very persistent.

To better gauge at the age profile of the long term component, Figure 2 focuses only on the profiles of "long" covariances. The figure shows that the initial compression of earnings differentials is also evident in long covariances, while an increase with age is evident afterwards. Both features may be a symptom of Mincerian cross-overs of long-term earnings differentials and can be captured by the random growth specification discussed earlier in the paper.

In Figure 3 we move to cross-brother covariances. The covariances are computed when brothers have the same age, considering three groups alternatively: all brothers, brothers born more than five years apart from one another, and brothers born more than eight years apart. The three plots overlap pretty closely and show that brother covariances of permanent earnings are relatively high at young ages but then decline pretty rapidly before age 30 and stabilize thereafter. In principle, high earnings covariances between brothers at young ages may reflect both permanent and transitory earnings shocks; in the latter case the effect would result from the fact that earnings instability is larger at young ages and brothers tend to be young in the same years. However, we can note that while the declining shape is resembling the age profile of the variance of earnings of Figure 1, the level of the brother covariance is closer to the one of "long" covariances in Figure 1 and 2, suggesting that brother covariances are likely driven by long-term determinants of earnings (i.e. family and community effects) rather than by transitory shocks. The plots for brothers born five or eight years apart tend to support this interpretation, because for these brothers are less likely to share the economic environment at labour market entry.

Figure 4 provides additional insights into the characteristics of siblings' covariances by contrasting covariances of brothers that have the same age with those obtained fixing the age of the elder brother at 30. The latter plot underestimates the same-age covariance by a large extent during the first ten years of the earnings life-cycle, whereas the two plots roughly coincide after age 30. This illustrates the life-cycle bias.

In figure 5 we consider the earnings correlation between schoolmates, which is obtained by connecting individual earnings profiles with the average earnings profile of his grade 8 schoolmates (individuals that attended grade 8 in the same school). The level of covariances is lower compared with the level of brothers' correlations. The time pattern is instead remarkably similar.

## 6. Results

We report estimates of model parameter in Table 1a (Permanent component), Table 1b (transitory component) and Table 1c (time effects on both components). Parameters estimates of the permanent component indicate that family effects are by far the most relevant factor that shapes long-term incomes. This is true both for initial earnings and for earnings growth rates, where they represent 65 and 56 percent of total dispersion respectively. The other more relevant source of permanent inequality in earnings is the individual idiosyncratic component, accounting for approximately 20 percent of the dispersion in both initial earnings and earnings growth. Remaining shares are accounted for by school effects. Long-term earnings display the Mincerian cross-over property, i.e. the overall covariance between initial earnings and earnings growth rates is negative, corresponding to a long-term earnings distribution that first compresses and then fans out over the life cycle (with a turning point located at about age 29). Such compression/decompression occurs through the family and school components, whereas the covariance term of idiosyncratic component is positive but relatively negligible in size. Usually the Mincerian cross-over is taken as a sign of human capital investments occurring early in the career; our results point to a limited role for idiosyncratic factors in shaping these investments once family and school effects have been taken into account.

Parameter estimates of transitory earnings show a clear age pattern of transitory shocks, whose variance is increasing in the first five years observed into the labour market, and then steeply decreases between the ages of 25 and 35, while the decrease slows down after age 35. The sharp decline followed by a leveling-off is consistent with the patterns reported by Baker and Solon (2003), who find the variance of transitory shocks to be declining at decreasing rates between the ages of 25 and 45. The increase that we find between the ages of 21 and 25 may reflect new entries into the labour market and into our sample. These patterns look similar between brothers. Also, the autoregressive coefficient is very similar between brothers and of a moderate size, roughly 0.5. Finally, transitory shocks are contemporaneously correlated between brothers, although the size of the covariance looks negligible if compared with the variance of shocks, the implied correlation coefficient at age 22 being only 0.018.

We can use parameter estimates to predict the effects of families and schools in determining the overall variance of permanent earnings. This is done in Figure 5 which plot the age profile of permanent earnings variance and its components. The line labeled “Sibling” is the component of permanent variance shared by siblings and is the sum of family and school effects. The line labeled “Total” is the sum of the sibling component and of the idiosyncratic component. There is an u-shaped profile of all variance components, picking up the Mincerian cross-over effects discussed above. The turning point of total variance occurs few years earlier compared with family and school profiles, due to cross-over effects being absent in the idiosyncratic component. This result in the idiosyncratic component accounting for a larger share of permanent after age 30 compared with before 30.

Figure 6 performs the decomposition of siblings’ correlations into family and school effects. Benchmark estimates of the sibling correlation of permanent earnings for Denmark are provided by Bjorklund et al (Journal of Pop Econ 2002) and equal 0.23 using data in the 25-42 age range and a model without age and time effects in the permanent and transitory component. Our estimate matches theirs around the mid-point of their age window (0.228 at age 33) while our average estimate in the 25-42 interval is 0.33. In general, Figure 6 shows that there is considerable age variation in the sibling correlation, which almost at 0.8 at age 21, decreases steeply by the mid-30s and then increases again towards the early 40s, although at a moderate pace. Comparing these predictions with the raw sibling correlations of Figure 3, we can observe that the model matches the age evolution quite well (note that the initial increase of raw figures is picked up by the age spline in the transitory component of the model). There is a difference in the levels of the sibling correlation which is due to the fact that raw figure correlate current earnings (not permanent ones) and thus suffer from downward biases coming from transitory earnings fluctuations. As we saw in Table 1a, the permanent earnings determinants shared between siblings at labour market entry largely dominate idiosyncratic factors, which results in the high initial level of the sibling correlation. We also saw that cross-over within the permanent earnings distribution are driven by family and school effects, resulting in the decline of the sibling correlation towards the mid-30s, and in its increases afterwards.

The other interesting evidence reported in Figure 6 concerns the relative importance of family and school effects in shaping the sibling correlation. School effects are negligible and close to null in the years of the earnings compression. The sibling correlation is –essentially—entirely driven by family effects.

Our results indicate that there is little room for school effects once family effects have been accounted for. The natural question then is if school effects found in studies that ignore family effects are upwardly biased, i.e. are indeed a result of the fact that students in the same school tend to come from similar family environment. We can provide some evidence on this by estimating a model of schoolmates correlations only and contrasting its predictions with the school effects of Figure 6. Figure 7 plots the schoolmates correlation estimated from a schoolmates only model. The age pattern of the school correlation is similar to the one from the main model; there is some evidence of an upward bias, but it does not appear to be quantitatively relevant.

## 7. Summary

We develop a unified framework which enables to disentangle the contribution of the family, the school and the neighborhood in labor earnings over the life-cycle. This is achieved within a model of multi-person earnings dynamics distinguishing permanent from transitory earnings and allowing for heterogeneous earnings growth. The analysis is based on administrative registers from the Danish population, which enables connecting member of the same family and link each family member with other individuals at the community and school level. In particular, we consider two dimensions of the community, namely schools and neighbors (using the parish of residence) when the siblings were in grade 8 and aged 16, respectively, corresponding to the earlier points in time when we observe community affiliations.

Our preliminary results, which are based on the comparison between family and school effects, suggest that family effects are by far the most relevant factor that shapes long-term incomes. This is true both for initial earnings and for earnings growth rates, where they represent 65 and 56 percent of total dispersion respectively. The other more relevant source of permanent inequality in earnings is the individual idiosyncratic component, accounting for approximately 20 percent of the dispersion in both initial earnings and earnings growth. The remaining shares are accounted for by school effects. Decomposing the sibling correlation into family and school effects, we find that the sibling correlation is driven by family effects, while school effects seem to play no role. The next step in the analysis is to use data for neighbors and estimate the full model, which will allow decomposing the sibling correlation into family, school and neighborhood effects.

## References

Altonji, Joseph G. and Dunn Thomas A. “Family incomes and labor market outcomes of relatives.” In Ehrenberg, R., ed., Research in Labor Economics, Vol. 12, Greenwich CT: Jai Press, 1991, pp. 246-69. +

Anti Nilsen, Øivind, Vaage, Kjell, Aakvik, Arild and Jacobsen, Karl. “Intergenerational earnings mobility revisited: estimates based on lifetime earnings.” Scandinavian Journal of Economics, 114(1), 2012, pp. 1-23.

Becker, Gary S. and Tomes Nigel. “An equilibrium theory of the distribution of income and intergenerational mobility.” Journal of Political Economy 87 (), 1979, pp. 1153-89.

Black, Sandra E. and Devereux, Paul J.. “Recent developments in intergenerational Mobility.” in O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, Vol. 4A. Amsterdam: Elsevier Science, North Holland, 2011.

Baker, Michael and Solon, Gary. “Earnings dynamics and inequality among Canadian men, 1976-1992: evidence from longitudinal income tax records.” Journal of Labor Economics, 2003, 21(2), pp. 267-88.

Björklund, Anders and Jännti, Markus. “Intergenerational income mobility and the role of family background.” in W. Salverda, B. Nolan and T. Smeeding, eds., The Oxford Handbook of Economic Inequality, Oxford: Oxford University Press, 2009, pp. 491-521.

Björklund, Anders, Jäntti, Markus and Lindquist, Matthew J. “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, 2009, 93(5-6), pp. 671-80.

Grawe, Nathan D.. “Lifecycle bias in estimates of intergenerational earnings persistence.” Labour Economics, 2006, 13(5), pp. 551-70.

Haider, Steven J. and Solon, Gary. “Life-Cycle variation in the association between current and lifetime earnings.” American Economic Review, 2006, 96(4), pp. 1308-20.

Jenkins, Stephen J.. “Snapshots versus movies: ‘Lifecycle biases’ and the estimation of intergenerational earnings inheritance.” European Economic Review, 1987, 31(5), pp. 1149–58.

Mazumder, Bhashkar. “Fortunate sons: new estimates of intergenerational mobility in the United States using social security earnings data.” Review of Economics and Statistics, 2005, 87(2), pp. 235–55.

Nybom, Martin and Stuhler, Jan. “Heterogeneous income profiles and life-cycle bias in intergenerational mobility estimation.”, IZA DP No. 5697, 2011.

Page, Marianne and Solon, Gary. “Correlations between sisters and neighbouring girls in their subsequent income as adults.” Journal of Applied Econometrics, 2003, 18(5), pp. 545-62.

Page, Marianne and Solon, Gary. “Correlations between brothers and neighboring boys in their adult earnings: the importance of being urban.” Journal of Labor Economics, 2003, 21(4), pp. 831-56.

Solon, Gary, Corcoran, Mary, Gordon, Roger and Laren, Deborah. “A longitudinal analysis of sibling correlations in economic status.” Journal of Human Resources, 1991, 26(3), pp. 509-34.

Solon Gary. “Intergenerational income mobility in the United States”, American Economic Review, 1992, 82, pp. 393-408.

Solon, Gary. “Intergenerational mobility in the labor market.” in O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, Vol. 3A. Amsterdam: Elsevier Science, North Holland, 1999, pp. 1761-1800.

Zimmermann, David J. “Regression toward mediocrity in economic stature”, American Economic Review, 1992, 82, pp. 409-429.

**Table 1a: Parameter estimates – Permanent component**

| Variance of starting levels | | |
|---|---|---|
| | Coeff. | s.e. |
| $\sigma^2_{\alpha I}$ (Individual) | 0.0204 | 0.0024 |
| $\sigma^2_{\alpha \Phi}$ (Family) | 0.0621 | 0.0051 |
| $\sigma^2_{\alpha \Sigma}$ (School) | 0.0129 | 0.0023 |

| Variance of growth rates | | |
|---|---|---|
| | Coeff. | s.e. |
| $\sigma^2_{\beta I}$ (Individual) | 0.0002 | 0.00003 |
| $\sigma^2_{\beta \Phi}$ (Family) | 0.0004 | 0.00004 |
| $\sigma^2_{\beta \Sigma}$ (School) | 0.0001 | 0.00002 |

| Covariance | | |
|---|---|---|
| | Coeff. | s.e. |
| $\sigma^2_{\alpha\beta I}$ (Individual) | 0.0005 | 0.0002 |
| $\sigma^2_{\alpha\beta \Phi}$ (Family) | -0.0042 | 0.0004 |
| $\sigma^2_{\alpha\beta \Sigma}$ (School) | -0.0012 | 0.0002 |

**Table 1b: Parameter estimates – Transitory component**

| | Brother 1 | | Brother 2 | |
|---|---|---|---|---|
| | Coeff. | s.e. | Coeff. | s.e. |
| $\sigma_{0b}^2$ Variance of initial condition (age 21) | 0.4362 | 0.0108 | 0.4018 | 0.0100 |
| $\sigma_{\varepsilon b}^2$ Baseline variance of shocks (age 22) | 0.4556 | 0.0117 | 0.4334 | 0.0118 |
| Spline coefficients | | | | |
| Age 23-25 | 0.0307 | 0.0027 | 0.0567 | 0.0047 |
| Age 26-30 | -0.0819 | 0.0019 | -0.0993 | 0.0034 |
| Age 31-35 | -0.0585 | 0.0029 | -0.0711 | 0.0055 |
| Age 36+ | -0.0262 | 0.0048 | -0.0307 | 0.0061 |
| $\rho_b$ Autocorrelation coefficient | 0.5090 | 0.0020 | 0.4980 | 0.0025 |
| $\sigma_{12}$ Cross brother covariance | 0.0084 | 0.0021 | | |

**Table 1c: Parameter estimates – Time effects**

| | Permanent component | | Transitory component | |
|---|---|---|---|---|
| | Coeff. | s.e. | Coeff. | s.e. |
| 1988=1 | | | | |
| 1989 | 1.1949 | 0.0472 | 0.9541 | 0.0111 |
| 1990 | 1.4218 | 0.0649 | 0.9916 | 0.0137 |
| 1991 | 1.2135 | 0.0547 | 1.0357 | 0.0141 |
| 1992 | 1.4562 | 0.0717 | 0.9788 | 0.0153 |
| 1993 | 1.5850 | 0.0809 | 0.9997 | 0.0165 |
| 1994 | 1.5222 | 0.0733 | 1.0223 | 0.0164 |
| 1995 | 1.5062 | 0.0722 | 0.9161 | 0.0150 |
| 1996 | 1.5514 | 0.0754 | 0.9049 | 0.0152 |
| 1997 | 1.3896 | 0.0636 | 0.9453 | 0.0140 |
| 1998 | 1.5014 | 0.0701 | 0.8885 | 0.0140 |
| 1999 | 1.5430 | 0.0703 | 0.9152 | 0.0139 |
| 2000 | 1.4189 | 0.0638 | 0.9533 | 0.0139 |
| 2001 | 1.4983 | 0.0684 | 0.9337 | 0.0141 |
| 2002 | 1.4864 | 0.0689 | 0.9703 | 0.0146 |
| 2003 | 1.5123 | 0.0708 | 1.0191 | 0.0158 |
| 2004 | 1.4645 | 0.0696 | 0.9754 | 0.0152 |
| 2005 | 1.4182 | 0.0690 | 0.9643 | 0.0153 |
| 2006 | 1.3670 | 0.0667 | 0.9381 | 0.0149 |
| 2007 | 1.2689 | 0.0622 | 0.9273 | 0.0144 |
| 2008 | 1.1310 | 0.0552 | 0.9090 | 0.0139 |
| 2009 | 1.0615 | 0.0528 | 1.0665 | 0.0157 |

**Figure 1: Raw earnings variances and covariances of individual earnings by age**

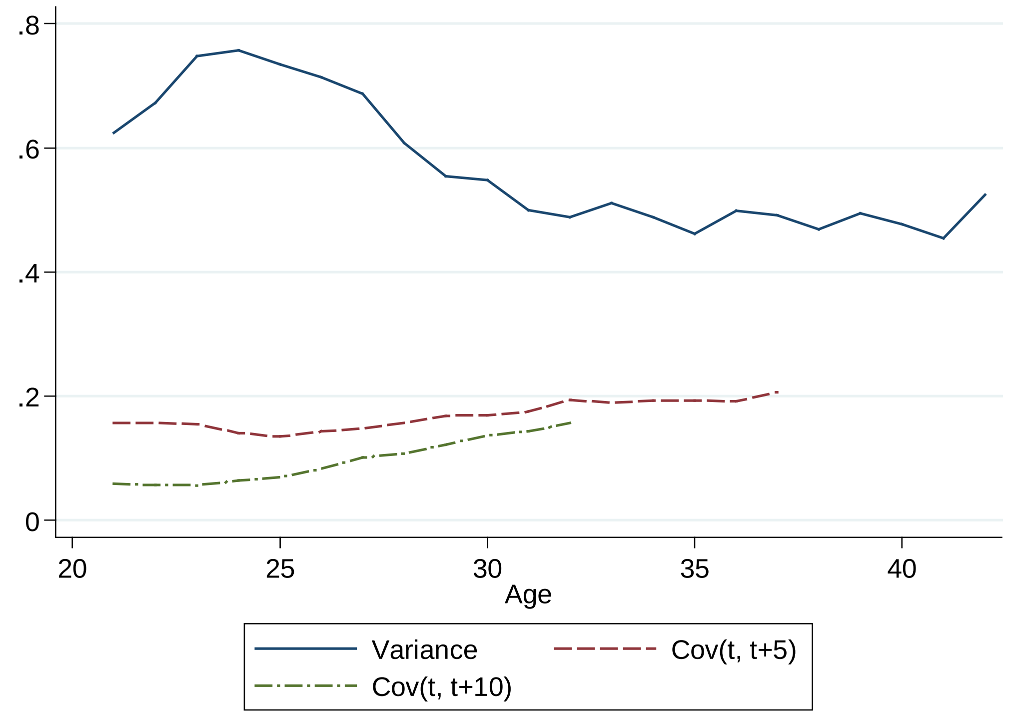

**Figure 2: Raw earnings covariances of individual earnings by age**

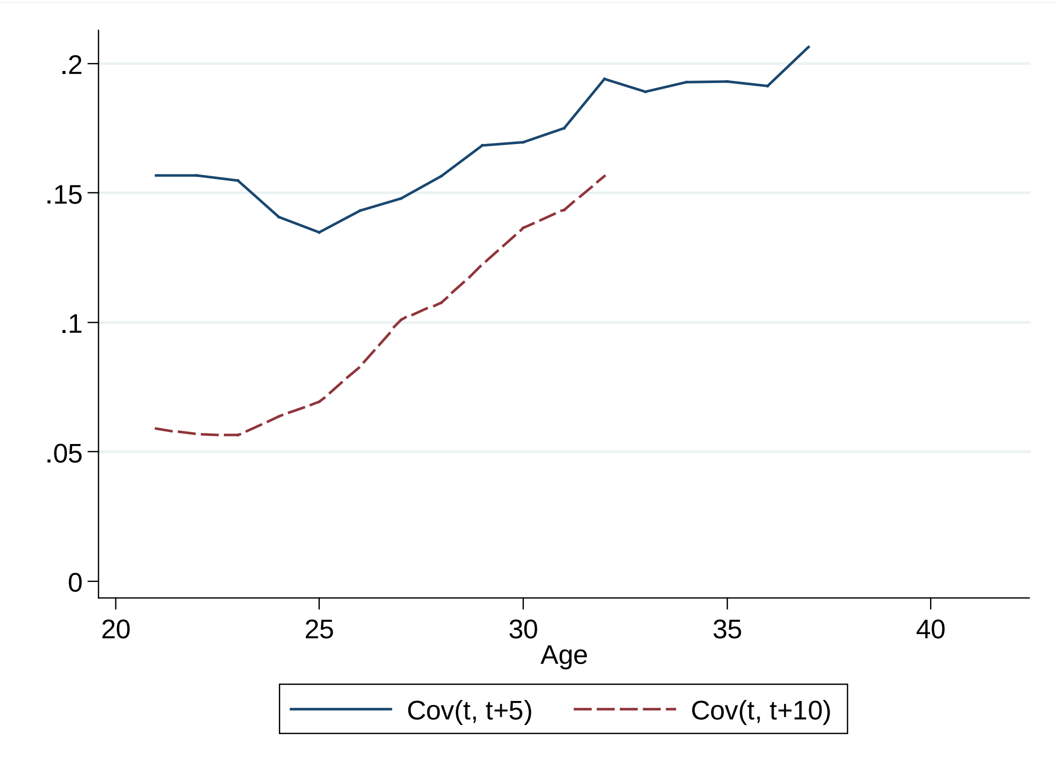

**Figure 3: Raw earnings covariances of brothers' earnings by age**

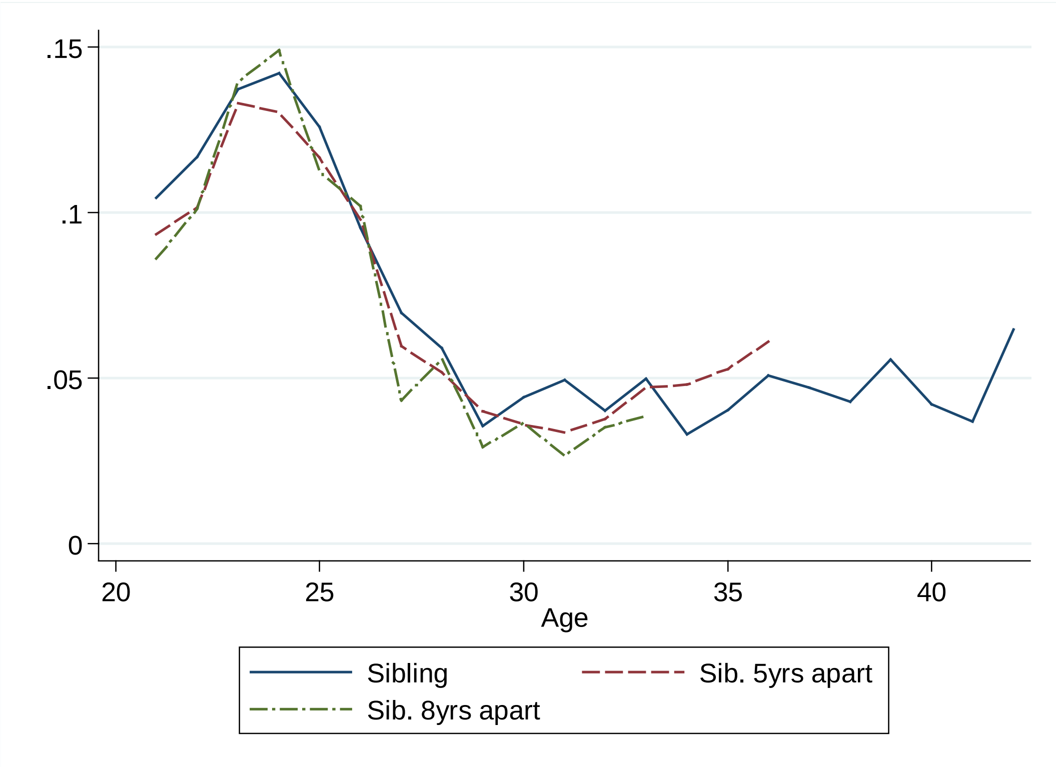

**Figure 4: Raw earnings covariances of brothers' earnings by age**

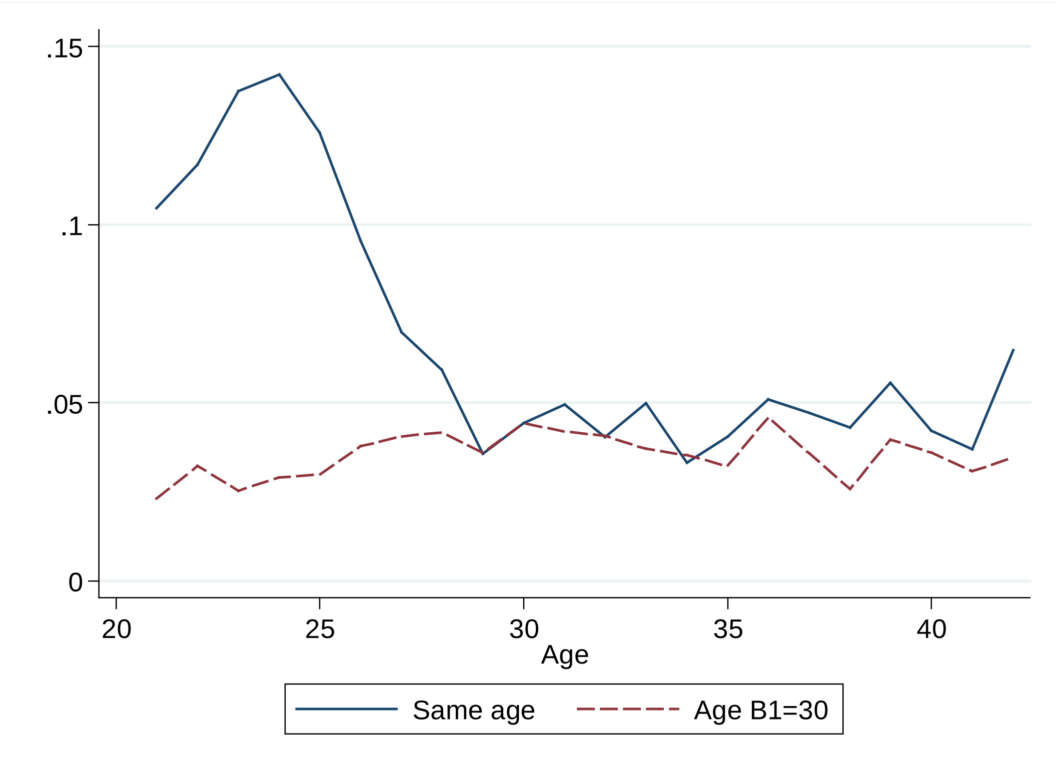

**Figure 4: Raw earnings covariances of grade 8 schoolmates**

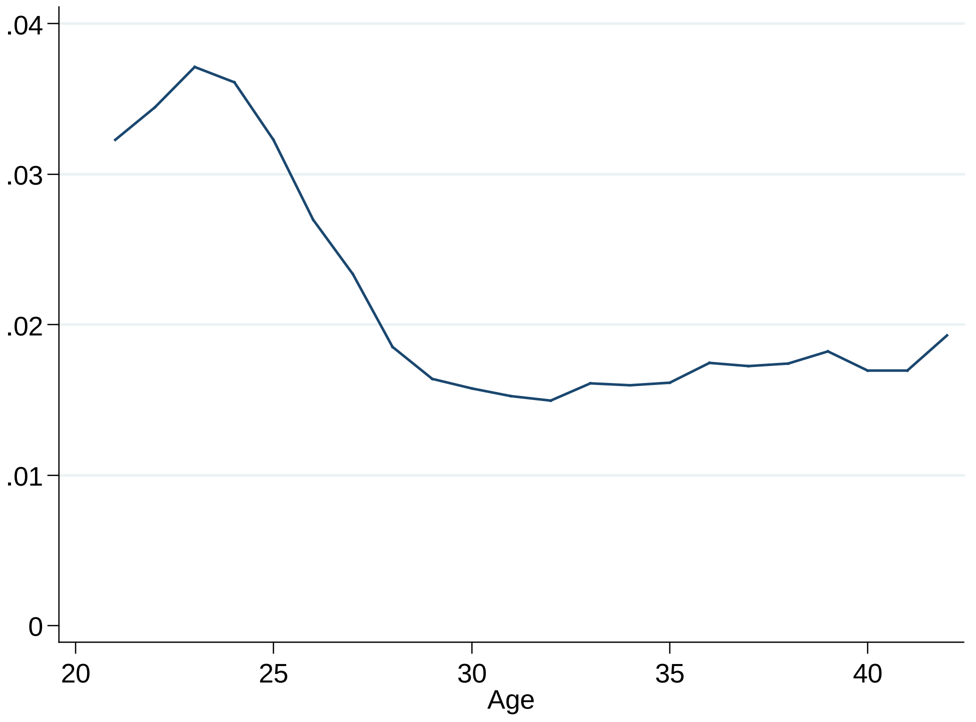

**Figure 5: Decomposition of predicted permanent variance into school and family effects**

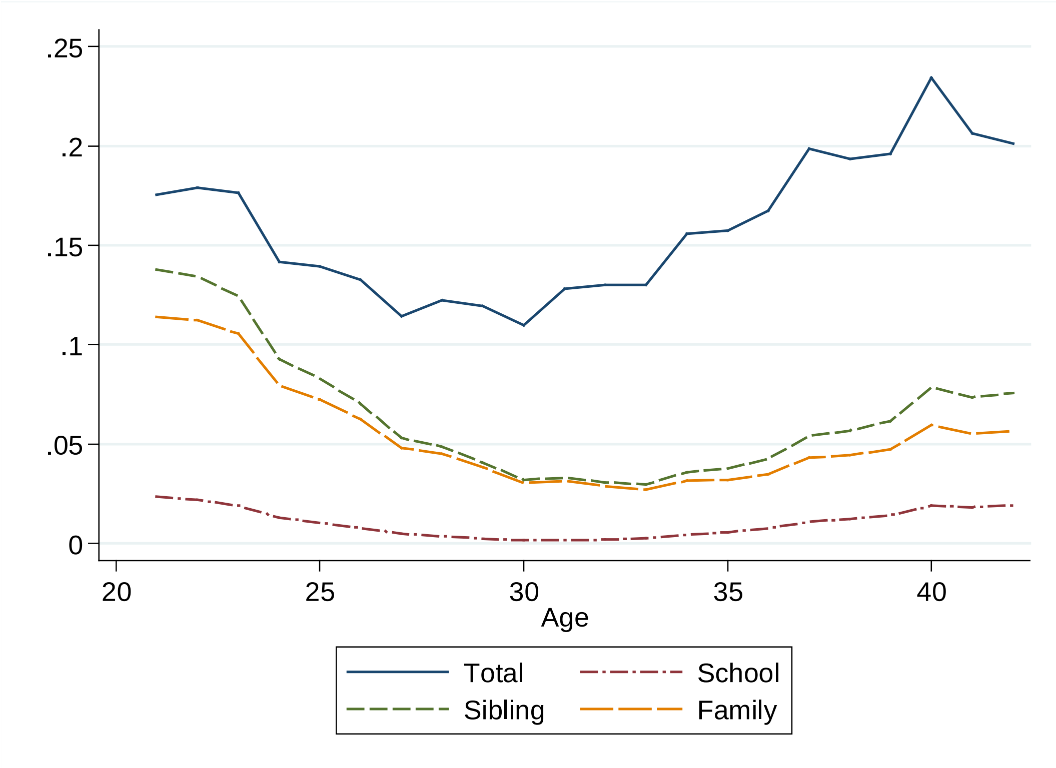

**Figure 6: Decomposition of predicted sibling correlation into school and family effects**

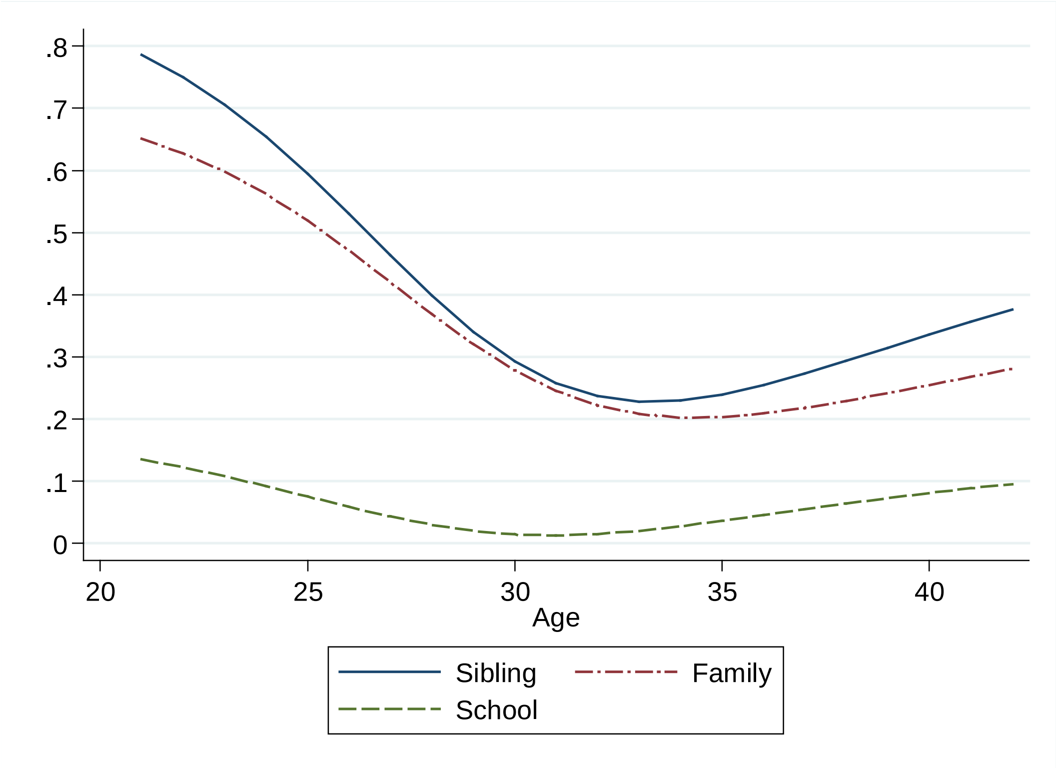

**Figure 7: Predicted schoolmate correlation from school-only model**