Family, Community and Long-Term Socioeconomic Inequality: Evidence from Siblings and Youth Peers §

Paul Bingley
VIVE Copenhagen

Lorenzo Cappellari
Università Cattolica Milano

Konstantinos Tatsiramos
University of Luxembourg and LISER

September 2019

Abstract

Using administrative data for the population of Danish men and women, we develop an empirical model which accounts for the joint earnings dynamics of siblings and youth community peers. We provide the first decomposition of the sibling correlation of permanent earnings into family and community effects allowing for life-cycle dynamics, and extend the analysis to consider other outcomes. We find that family is the most important factor influencing sibling correlations of earnings, education and unemployment. Community background matters for shaping the sibling correlation of earnings and unemployment early in the working life, but its importance quickly diminishes.

Keywords: Sibling correlation; Neighborhoods; Schools; Life-cycle earnings; Unemployment; Education
JEL codes: D31, J62

§ We thank for their comments an anonymous referee, the Editor Kjell Salvanes, as well as Anders Björklund, Stephen Gibbons, Markus Jäntti, Bash Mazumder, Andreas Peichl, Josef Zweimüller, seminar audiences at IZA in Bonn, SOFI in Stockholm, Lund University, Paris School of Economics, JRC in Ispra, University of Padua, University of Venice, ZEW in Mannheim, European University Institute in Florence, IAB in Nuremberg, and conferences audiences at Royal Economic Society in Manchester, European Society of Population Economics in Braga, 16th IZA/CEPR European Symposium in Labour Economics (ESSLE), Fourth SOLE/EALE World Conference in Montreal, DynIPer in Marseille, the Intergenerational Mobility workshop in Madrid and the Kraks Fond Institute for Urban Economic Research workshop at Copenhagen Business School. The usual disclaimer applies. The research for this paper was supported by a British Academy Research Grant (award no. SG113192) and by an Independent Research Fund Denmark – Social Sciences Grant (DFF-6109-00226). Addresses for correspondence: Bingley, VIVE - Danish Centre for Social Science Research, Herluf Trolles Gade 11, DK-1052, Copenhagen, pab@vive.dk; Cappellari, Department of Economics and Finance, Università Cattolica Milano, Largo Gemelli 1, I-20123 Milano, lorenzo.cappellari@unicatt.it; Tatsiramos, University of Luxembourg and LISER, 6, rue Richard Coudenhove-Kalergi, L-1359, Luxembourg, Konstantinos.Tatsiramos@uni.lu.

1. Introduction

Socioeconomic background is a key factor influencing individual outcomes. Parents, by transmitting abilities, preferences and resources, determine the accumulation of human capital and economic success (Becker and Tomes, 1979). Economic success, however, may not only be influenced by traits transmitted by parents within the family but also by interactions outside the family at the level of communities, neighborhoods and schools, taking the form of peer effects, role models, norms of behavior, or exposure to unemployment and crime (Benabou, 1996; Durlauf, 1996). Several studies from various countries measure the overall importance of socioeconomic background based on the sibling correlation, an *omnibus measure* of the contribution to inequality of all factors (observed and unobserved) that siblings share in the family and in the community.[1] However, still little is known about the relative importance of families *versus* communities in shaping the inequality of earnings and other outcomes. Understanding how much families and communities contribute to inequality is crucial for public policies aiming to provide a level playing field of equal opportunities, but measuring their relative importance is complicated because the two sources of influence tend to reinforce each other through segregation and sorting of families into communities.

In this paper, using administrative registers for the population of Danish men and women, we provide the first decomposition of the sibling correlation of life-cycle labor earnings into family and community effects. The decomposition is based on estimates from an empirical model which accounts for the joint earnings dynamics of siblings and youth community peers (neighbors and schoolmates). The model extends the framework developed in Bingley and Cappellari (2019) for life-cycle earnings dynamics within a three-person family to the case of dynamics for community groups of arbitrary size. The model also accounts for heterogeneous human capital investments and returns by including individual specific levels and growth rates of earnings (similar to Baker and Solon, 2003, and Magnac, Pistolesi and Roux, 2018), which depend on heterogeneity among families and communities, and allows for permanent and transitory shocks to individual earnings

[1] For reviews, see Solon (1999), Björklund and Jäntti (2009), Black and Devereux (2011).

trajectories. We also adapt this general framework to time-invariant or censored outcomes with applications to education and life-cycle unemployment, which are two important dimensions of long-term success that, in addition to life-cycle earnings, may be influenced by families and communities.

Our study contributes to the literature on social background and long-term inequalities in several ways. First, by modeling jointly the outcomes of siblings and peers we identify family effects *net of any* community influences. Second, by exploiting within-family community variation we show that it is possible to identify community influences separately from sorting of families into communities. Third, we define communities using both neighborhoods and schools, so we can investigate any differential influences between the two factors. Fourth, by observing earnings trajectories up to age 45 we trace the relative influence of these attributes on earnings for a relevant part of the life cycle. With long earnings histories, we can assess the importance of families and communities in shaping permanent earnings inequality while avoiding measurement error and life-cycle biases. Finally, we develop a Tobit version of the dynamic variance components model to analyze unemployment, which is novel in the literature.

The focus on long-term outcomes, such as lifetime earnings, is in line with recent research which has shifted attention away from current outcomes as the appropriate measure to capture the economic effects of human capital investments (e.g. Bhuller, Mogstad and Salvanes, 2017; Nybom, 2017). In the context of youth communities, taking the long-term perspective seems particularly appropriate since some early influences may be longer lasting than others, and their relative impacts may depend on the specific part of the life cycle considered. We analyze the inequality of individual life-cycle earnings, instead of total household income inequality, for the following three reasons. First, persistent inequality of earnings is the main component of the intergenerational correlation of welfare (Piketty, 2000). Second, labor earnings provide a broad measure for the market value of individual human capital, whose investments and returns may be affected by families and communities. Third, household incomes between the mid-20s and early 30s (when our life-cycle analysis starts) may change for other reasons, such as household formation, and we want to abstract from such mechanisms in our assessment of family and community effects in life-cycle inequality.

Analyzing total household inequality requires modeling changes in family composition, assortative mating and the tax and welfare-benefit system, which are beyond the scope of this paper.[2]

We show that for both genders the sibling correlation of earnings is *U-shaped* in age, which is consistent with human capital theory suggesting that heterogeneous investments in human capital induce an inverse relationship between initial earnings and earnings growth rates (e.g. Bhuller, Mogstad and Salvanes, 2017; Magnac, Pistolesi and Roux, 2018). We also find a similar pattern for the sibling correlation of the unemployment process for women, which is consistent with the idea that human capital investments, although reducing labor market experience of young investors, eventually pays off later in life through better employment opportunities. For men, we find a positive association between initial unemployment and life-cycle unemployment across sibling pairs, which points towards the prevalence of state dependence in family-specific unemployment shocks. These sibling correlation patterns in earnings and unemployment highlight the relevance of heterogeneity *between sibling pairs*.

Our decomposition result for earnings shows that family is the most important factor explaining sibling correlations throughout the life cycle for both men and women. On average, between ages 24 to 45, community background accounts for less than a tenth of the sibling correlation of labor earnings. However, we find that community matters most early in the working life, accounting for as much as a fifth of the sibling correlation at age 25, but its importance quickly diminishes and becomes negligible after age 30. These findings hold both when we define community solely on the basis of the neighborhood or the school, although neighborhood effects are slightly larger than school effects; and these findings are robust to the age we measure youth communities and to various sample selection choices. The diminishing community influence over the life cycle highlights the importance of observing long earnings histories beyond the first years of the working life; measuring earnings only at relatively young ages overstates the long-term relevance of community effects in explaining earnings variation. Finally, extending our analysis to education and unemployment, we also find that family accounts for most of the sibling correlation

[2] For an analysis of household income inequality and its relationship with inequality in labor earnings see Blundell et al. (2018).

of these outcomes for both men and women. For women, we find a larger community influence on unemployment than for men, while for both genders community effects decline over the life cycle.

Our paper relates to the approach in the literature which aims to separate the influence of family from the influence of community by comparing the correlation of sibling outcomes with the correlation of outcomes among unrelated neighbors (e.g., Solon, Page and Duncan, 2000 on educational attainment for the US; Page and Solon, 2003a,b, on earnings for the US; Raaum, Sörensen and Salvanes, 2006, on education and earnings for Norway). The main idea of this approach is that while siblings share both the family and the neighborhood, unrelated neighbors share only the neighborhood but not the family. Findings from these studies suggest that the neighborhood of residence during childhood is an important factor in explaining the resemblance in adult outcomes among siblings, accounting for a substantial portion of the sibling correlation, which for earnings can be as much as half in the US case.

However, these estimated neighborhood effects are recognized to be upper bounds because of non-random sorting of families into communities, which leads to positive correlation between the two factors. Exploiting quasi-random assignment of families to public housing projects in Toronto, which eliminates sorting, Oreopoulos (2003) finds instead a zero influence of neighborhood quality in the total variance of income and wages. Within our proposed framework we can estimate family and community effects net of sorting. Our paper also complements the analysis of Bingley and Cappellari (2019) who use brothers and their fathers to investigate the impact of father-son transmission on the brother correlations of earnings and education, showing that sibling similarities are largely a reflection of intergenerational effects. Our findings on the limited importance of communities relative to families in shaping sibling correlations are consistent with the findings in Bingley and Cappellari (2019) regarding the prevalence of intergenerational transmission.

Outside the sibling correlation literature, our paper relates to the large and growing literatures on the impact of schools and youth neighborhoods on adult outcomes. Some recent examples include Chetty and Hendren (2018) and Chetty, Hendren and Katz (2016) on the impact of age of moving to a higher income neighborhood during childhood, in observational and experimental frameworks respectively; Fredriksson, Öckert and Oosterbeek (2012) and Chetty, et al. (2011) on the impact of class size and school peers, also in observational and experimental frameworks. Evidence emerging from these studies supports the view that environmental circumstances during youth may have impacts when adult. Our communities encompass schools and neighborhoods, enabling us to speak to both literatures.

The paper is structured as follows. In Section 2 we describe the data and contrast our community definition with that used in comparable studies. In Section 3 we present descriptive statistics on earnings of siblings and youth community peers over the life cycle. In Section 4 we develop the econometric model based on the joint analysis of life-cycle earnings for siblings and youth peers. In Section 5 we present the main results for the earnings process, discuss a series of robustness checks and compare our findings to the previous sibling correlation literature. In Section 6 we extend the analysis to education and unemployment. We conclude in the last section.

2. Data

2.1 Sample Selection

We use data from administrative registers of the Danish population. The civil registration system was established in 1968 and everyone resident in Denmark then and since has been registered with a unique personal identification number, which has subsequently been used in all national registers enabling accurate linkage. In our analysis, we consider men and women separately and we construct our datasets as follows. First, for each gender we create a sample of siblings by sampling fathers and finding their first and second children born in the years 1965-1985 who share both legal parents from registration at birth and are not adopted. The year of birth selection starts in 1965 to allow use of address information from age 11 in 1976, and stops in 1985 to allow observation of individual earnings up to the late 20s for younger cohorts (earnings data described below are available until 2014).[3] We exclude from the sample children who are also observed as parents of

[3] Subsequent sons (daughters) beyond the first two are only 4 percent (4 percent) and are not considered in the analysis. Including third born siblings in the analysis produced results entirely consistent with those presented. The son (daughter) birth order is determined irrespective of daughters (sons) present in the family.

children born 1965-85, siblings born less than 12 months apart and siblings born more than 12 years apart.

For sampled individuals we obtain pre-tax annual labor earnings between 1990 and 2014 from the Statistics Denmark Income Register, measured in 2012 prices; see Baadsgaard and Quitzau (2011) for a detailed description of the earnings data.[4] We select all valid observations on earnings and exclude zero-earnings observations. Assuming that earnings are missing at random, the exclusion of zero-earnings observations is common with most of the earnings dynamics literature and is also applied in the sibling correlation literature (for instance, see Björklund, Jäntti and Lindquist, 2009). We 'trim' a quarter of a percentile from each tail of the annual earnings distribution and require at least three consecutive earnings observations for an individual to be included in the sample. This selection rule is intermediate between the one used by Baker and Solon (2003), that is, continuous earnings strings for each individual within a cohort, and the approach of Haider (2001), who allows individuals to move in and out of the sample only requiring two positive but not necessarily consecutive valid observations on earnings. Overall, we obtain 97,596 (88,050) brother (sister) pairs. We keep singletons – children without a younger sibling – in the sample; there are 315,325 singleton sons (308,178 singleton daughters) giving a total of 510,517 men (484,278 women) who meet the selection criteria above and are either a first or second brother (sister), or a singleton.

Next, for each gender we match individuals from the selected birth cohorts depending on whether they are born in the same year and share their youth communities by either being neighbors, schoolmates, or both. To match every person in the sample to members of their community of the same gender we choose either a specific age or an age interval within which the persons share their community. In contrast to our treatment of siblings, peers are included in the analysis irrespective of birth order and age spacing from their own siblings. This adds 68,956 men (61,881 women), giving a total sample of 579,473 men (546,159 women). Note that the vast majority of peers appears also in the sibling sample, either as singletons, sibling 1 or sibling 2. In

[4] In Section 6, we broaden the analysis and also take into account the impacts of families and communities on other relevant outcomes, namely education and unemployment; we provide there a description of the specific data sources.

substance, the two samples largely overlap in terms of *individuals*, but the *pairwise matches* that are used to estimate correlations are unique to each case.

Using information on individual addresses from the central person register we define neighborhoods as the parish of residence.[5] Given data constraints, we can measure neighbors at any age level or age interval starting at age 11. For the descriptive analysis presented in Section 3, and for the baseline estimates reported in Section 5 for earnings and in Section 6 for education and unemployment, we link persons to neighbors on 31 October of the calendar year they turn 15. There are 2,123 parishes in our sample containing on average 14 males (2,124 parishes containing on average 12 females) turning 15 in the same year.

To characterize the degree of heterogeneity in socio-economic background between neighbors, Figure 1 presents a map of Danish parishes shaded according to mean level of fathers' gross annual labor earnings (in 2012 prices) when the child is age 15. Highest earnings are concentrated in the biggest towns, but otherwise there is substantial paternal earnings dispersion around the country.

Using information from the educational register we link pupils to schoolmates on 31 October of the calendar year they turn 15, which is in the academic year they would normally attend 9th grade; the last year of compulsory education.[6] School enrolment rules were such that pupils should start in first grade in the August of the calendar year they turn 7. The national pupil database was established to monitor compliance with the 1972 school reform, which made 8th and 9th grade schooling compulsory in 1972/3 and 1973/4, respectively. Beginning in August 1973, the database links pupils to the schools they are enrolled from 8th grade and above.[7] School identifiers are consistent over time and schools are classified according to whether they are publicly run (77% of schools and 89% of pupils in our estimation sample) or privately run, and whether they are

[5] Complete information on municipality of residence is available from 1970 and full addresses are complete from 1976 (see Pedersen, et al. 2006 for details). We use an intermediate aggregation of locality as our neighborhood indicator – parish of residence. Individuals are required to report changes of address to the municipality within five days.
[6] In 1990, 95 percent of pupils began 9th grade during the year they turned 15. In recent years delays have been more common – in 2007, 13 percent of pupils delayed their school start by a year and 4 percent repeated the same grade the following year. See Jensen and Rasmussen (2011) for a description on the Statistics Denmark education register.
[7] The national pupil database was first extended to cover grades K-7 from August 2007. Hence, we are unable to match pupils to schoolmates in earlier grades to look at long run outcomes.

exclusively for pupils with special educational needs (10% of schools and 1% of pupils that we exclude from the estimation sample). Our sample contains 1,821 schools with males (1,789 with females) aged 15; within each school there are on average 19 males (18 females) born in the same calendar year.

During our sample period, pupils were assigned to public schools on a catchment area basis according to place of residence. Primary and lower secondary education usually takes place in the same school and most pupils attend the same school for all grades. From 2007 we can see that 92% of pupils in grades 1-8 were enrolled in the same school the following year. Due to the organization of primary and lower secondary schools largely as a single unit, there is likely to be less pupil mobility between schools than in other countries. This institutional setting makes Denmark a good place to look for school effects, because of the coherence of the schoolmate group.

An important Danish institutional feature is that parishes and public-school catchment areas do not completely overlap. As a consequence, neighbors may attend different schools, and schoolmates may come from different neighborhoods. Amongst school-birth-cohort clusters, 94 percent have men (93 percent have women) from more than one parish, and amongst parish-birth-cohort clusters 87 percent have men (85 percent have women) from more than one school.

Because communities are defined on the basis of individual year of birth, not all siblings will share the community at a given age, mainly because of family mobility. That is, neighbors of sibling 1 at a given age will be drawn from a different neighborhood than neighbors of sibling 2 at that same age if there is family mobility. Similarly, schoolmates of sibling 1 will be drawn from a different school than schoolmates of sibling 2 if they attend different schools. For example, among our brother (sister) pairs, 72 (72) percent share both school and neighborhood at age 15, 4 (5) percent share only the school, 16 (15) percent share only the neighborhood, while the remaining 8 (8) percent do not share either of the two community affiliations.

Table 1 presents the cohorts we include in the sample, the years for which we observe earnings and sample sizes in various dimensions. We group data in three-year birth cohorts, as shown in Column 1, and we compute age by imputing each cohort with its central year of birth.[8] The selection of birth cohorts and time window ensures that we observe each cohort starting at age 24 (23-25) for at least 7 years (last cohort 1983-85) and for as long as 25 years (first cohort 1965-67) up to age 48 (47-49). Columns 5 and 6 show the number of earnings observations and number of men (women) used in estimation. The number of communities into which we group men (women) is shown in Columns 7 and 8, totaling 1,821 (1,789) schools and 2,123 (2,124) parishes.[9] The number of parishes and schools is quite stable, which reflects an absence of administrative unit reform during the exposure period.[10] The falling number of earnings observations by birth year is due to later cohorts having less time to accumulate earnings histories.

2.2 Sample comparison to other studies

It is informative to contrast our community definition with that used in comparable studies such as Page and Solon (2003a, b), Raaum, Sörensen and Salvanes (2006) and Oreopoulos (2003).[11] We focus the comparison on neighborhoods because this is the community definition used in these studies. Table 2 characterizes ours alongside these four other studies according to characteristics of the different types of neighborhoods and exposures considered, and outcomes observed. Neighborhood geography and exposure group – an area and an age range – together define the cluster of individuals within which later outcomes are correlated. Neighbors in study (4) have the closest proximity because of the medium-to-high density of housing projects, followed by studies (1+2) because of the clustered PSID sampling frame. Interestingly, studies (1+2) find neighborhood effects only for urban areas, where neighbors are in closest proximity. Our Danish parishes cover a wider area than the neighborhoods used in studies (1+2) and (4) but are only about half the size of Norwegian census tracts used in study (3). For Denmark in the year 2000 we can

[8] We could form only few sibling matches within the first birth cohort (1965-68), therefore we exclude families in which both siblings belong to this cohort. As a consequence, the first birth cohort contains only elder brothers or singletons.

[9] There are 30,634 school-cohorts for men (29,628 for women) and 41,748 parish-cohorts for men (41,430 for women).

[10] A 2007 reform changed the number of municipalities from 273 to 95. Responsibility for primary and lower secondary schools changed accordingly. This reform comes after the last year of youth community affiliation in our data – 2002.

[11] In what follows we refer to Page and Solon (2003a, b) as studies (1+2), Raaum, Sörensen and Salvanes (2006) as study (3) and Oreopoulos (2003) as study (4).

calculate the distribution of distances between the different residences of neighbors within parish: 25 percent of distances are within 0.5 km, 50 percent within 1.1km, and 75 percent within 1.9km.

The other four studies pool neighbors together of different ages – with up to 9- and 11-year age differences – to form neighborhood clusters. In the main part of the analysis we consider neighbors at 15 years of age as belonging to the same cluster. Neighborhood affiliation at age 15 is at the upper end of the 5-16 age range together considered in the other studies. All else equal, if neighbors of the same age are more likely to interact than neighbors of different ages, then we would expect to find stronger neighbor correlations with our definition. In sensitivity checks we show robustness of results to neighborhood definitions based on affiliation down to age 11 or using age ranges between 11 and 15 rather than a single age.

The number of persons in each of our neighborhood clusters is in the middle of the range for the other studies. The estimation sample in studies 1-4 comprises a similar percentage of the total population of individuals in each cluster with 4.6, 3.1, and 4.8 percent respectively. Due to our narrower age range for clustering neighbors for Denmark the estimation sample covers only 0.5 percent of the cluster population. Although our neighbors are more homogeneous in terms of age, they represent only between one eighth and one fifth of the within-cluster sampling density of the other studies. However, this sparser sampling should reduce precision rather than introduce any bias.

3. Descriptive statistics on earnings of siblings and community peers

To motivate the model, which we present in the next Section, we first provide a description of the interpersonal covariance structure of earnings in our sample. There are two types of cross-person relationships that are of interest for the analysis: i) between siblings and ii) between community peers.

Figure 2 (solid line) – for men in Panel A and for women in Panel B – shows that the sibling correlation of earnings – when siblings are at the same point in their life cycle – is high at the beginning of the working life at age 24, declines until age 30 and remains relatively stable thereafter with a slight increase up to about age 40. This *U-shaped pattern* of earnings correlation suggests that the sources of initial earnings heterogeneity that siblings share are negatively correlated with heterogeneity in earnings growth. As we discuss in the next Section, human capital theory predicts that investments in education or training induce such a negative correlation.

In Figure 2 (dashed line) we also report the sibling correlation of earnings by age of the younger sibling (brother in Panel A and sister in Panel B) after we fix the age of the elder sibling at 35. We find that the earnings correlation starts close to zero when the younger sibling is age 24 and it increases sharply, so that by the early-30s it matches the "same age" correlation. This pattern illustrates that the earnings correlation between siblings at different ages is an underestimate of the actual correlation when they are at the same point in their life cycle. This pattern is a form of life-cycle bias similar to that highlighted by Haider and Solon (2006), which we can detect in the data and control for in estimation.

Besides human capital investments, the large contemporaneous associations at the early stage of the life cycle depicted in Figure 2 may also reflect the correlation of transitory shocks. It is well known that earnings instability is large in the beginning of the working life (for instance, see Baker and Solon, 2003). It is also plausible that siblings may be subject to common shocks, for example, due to similar local economic conditions at labor market entry. To assess if the relatively large sibling correlation at labor market entry is driven by differences in permanent earnings or transitory fluctuations, we compute the earnings correlation for not-closely-spaced siblings. The larger the age-spacing between siblings, the less likely that they enter the labor market at the same time and are influenced by common transitory fluctuations. Figure 3 shows that the declining pattern of the sibling correlation between the mid-20s and the early-30s persists for brothers (Panel A) and sisters (Panel B) born at least five, eight or ten years apart. This pattern suggests that the source of convex evolution of sibling correlations in Figures 2 and 3 is in the long-term component of earnings.

Figure 4 shows the correlation of earnings for community peers who are neighbors residing in the same parish, or schoolmates who attend the same school. To obtain the between-peers correlation of earnings (at each relevant age), we first compute the within-community correlation and then we average between communities using the weighting scheme of Page and Solon (2003a, pp. 841), which gives greater importance to more populous communities and makes inference person-representative. These empirical correlations pick-up all sources of peer similarity, which include correlated family effects within communities and also influences independent of the family. The magnitude of the earnings correlation of community peers is roughly one tenth of the correlation of sibling earnings, though higher at the beginning of the life cycle (up to age 30) and negligible thereafter. Figure 4 also shows that the correlation of earnings for "unrelated" individuals – who are neither siblings nor community peers – is zero for all ages.[12] This contrast suggests that the evolution of sibling and community peer correlations over the life cycle is not simply an artifact of aging, but is picking up factors attributable to families and communities.

4. Econometric model

We develop a model to assess the relative influence of family and community background on earnings inequality over the life cycle. Motivated by the empirical facts presented in Section 3, we exploit the linked earnings records of siblings and community peers within a model of multi-person earnings dynamics where we distinguish permanent from transitory earnings and allow for heterogeneous earnings growth. The model extends the joint earnings dynamics model of Bingley and Cappellari (2019) for three persons (a father and two sons) to multi-person groups. We use the parameter estimates from this extended model to decompose the sibling correlation of life-cycle earnings into family and community influences.

4.1 The model of earnings dynamics

To separate life-cycle effects from calendar-time trends, we consider the distribution of earnings within three-year birth cohorts. In particular, we assume that residual log earnings (w) – after regressing log real gross annual labor earnings on year dummies and a quadratic age trend by birth cohort group – are the sum of two components: (i) a permanent component (y) and (ii) a transitory component (v), which are orthogonal by definition and we write as

$$w_{ifca} = y_{ifca} + v_{ifca}\ ;\ E\left(y_{ifca}v_{ifca}\right) = 0, \tag{1}$$

[12] We compute this correlation by randomly matching each individual in the sample with 1,000 individuals born in the same year who do not share family and/or community.

where the indices *i, f, c* and *a* stand for individual, family, community and age, respectively.[13]

Since our model allows for a transitory component in earnings and life-cycle effects, we can tackle the two well-known biases in the estimation of correlations in permanent earnings between persons due to measurement error. The first source of bias is related to transitory income shocks, which make current earnings a poor measure of permanent earnings (Solon, 1992; Mazumder, 2005). Separate identification of permanent and transitory earnings is granted by the availability of individual-level longitudinal data. The second source of bias is related to life-cycle bias due to age differences between family members and the heterogeneous earnings variation over individual life cycles (Jenkins, 1997; Haider and Solon, 2006; Bohlmark and Lindquist, 2006; Nybom and Stuhler, 2016).

4.2 Specification of permanent earnings

We allow the permanent component of earnings (y) in equation (1) to depend on both *shared* and *idiosyncratic* factors. Shared factors capture the determinants of permanent earnings that are common between siblings and community peers. The idiosyncratic factors represent individual-specific sources of variation in permanent earnings.

We model life-cycle dynamics of shared factors using a specification based on *heterogeneous income profiles* (HIP); also known as a *random growth* model. The HIP specification for the shared factors is consistent with human capital theory suggesting that differential investments generate heterogeneity of *initial earnings* and *earnings growth* (Mincer, 1958; Ben-Porath, 1967; Baker and Solon, 2003; Magnac, Pistolesi and Roux, 2018). In the model of Becker and Tomes (1979), parents influence the human capital of their offspring by transmitting abilities, preferences and resources, and thereby affecting offspring earnings. Community background can also influence individual outcomes through institutions such as the school and its quality (Hanushek, 2006), or through the quality of neighborhood, or peer influences, social norms and role models in the neighborhood (Wilson, 1987, Benabou, 1996; Durlauf, 1996).

[13] We measure age in deviation from 24, which we set as the life cycle starting point.

Differences between families in the availability of these traits, resources and exposure to the community environment would lead to differences in human capital accumulation. Human capital theory predicts that these heterogeneous investments induce an inverse relationship between initial earnings and earnings growth rates, because investors trade off initial earnings against earnings growth throughout the life cycle. For example, in the Ben-Porath (1967) model, initial earnings of investors are lower because they are devoting much of their time to education rather than to work, while non-investors are fully participating in the labor market.

Such heterogeneity in earnings due to differences in work experience is one of the mechanisms we want to capture with our model. Another mechanism arises if one considers only full-time workers at a given early age. Among them, those who stayed longer in school had less time to accumulate experience, which lowers their earnings compared to early school leavers who, at that age, will have already accumulated some years of labor market experience (Bhuller, Mogstad and Salvanes, 2017). In both cases, due to their greater human capital, investors will experience faster earnings growth.

The resulting negative covariance of initial earnings and growth rates generates a U-shaped evolution of earnings dispersion by age due to the 'Mincerian cross-over' of earnings profiles. These observations motivate the specification for the determinants of shared earnings, which reflects the idea that cross-person resemblance of earnings stems from similarities in social background and human capital investments. The life-cycle patterns of earnings correlations between siblings and community peers shown in Section 3 are consistent with these mechanisms.

Besides the earnings profile shared by siblings and peers, we allow for idiosyncratic permanent shocks (ω_{ia}) to capture persistent individual deviations from the shared profile. To model these permanent shocks, we add a *random walk* process starting at age 24 (a *restricted income profile* – RIP –model).

Overall, our permanent earnings model is specified as follows:

$$y_{ifca} = \pi_t\left[\left(\mu_f + \mu_c\right) + \left(\gamma_f + \gamma_c\right)a + \omega_{ia}\right]; \quad \omega_{ia} = \omega_{i(a-1)} + \xi_{ia}; \quad t = b + 24 + a, \quad (2)$$

where b is the birth cohort of person i and π_t is a calendar time shifter that allows for aggregate changes of the permanent earnings process over time.[14] We factor the intercept and the slope of the individual-specific linear profile of earnings into two zero-mean components, where their variances capture family (*f*) and community (*c*) heterogeneity in initial earnings (denoted by μ_f, μ_c) and life-cycle earnings growth (denoted by γ_f, γ_c).

To summarize, the assumptions on the variance-covariance structure of permanent earnings are the following:

$$(\omega_{i24}, \xi_{ia}) \sim \left(0{,}0; \sigma^2_{\omega_{24}s}, \sigma^2_{\xi s}\right),\ s = 1{,}2; \tag{3.a}$$

$$\left(\mu_f, \gamma_f\right) \sim \left(0{,}0; \sigma^2_{\mu\mathrm{F}}, \sigma^2_{\gamma\mathrm{F}}, \sigma_{\mu\gamma\mathrm{F}}\right); \tag{3.b}$$

$$(\mu_c, \gamma_c) \sim \left(0{,}0; \sigma^2_{\mu\mathrm{C}}, \sigma^2_{\gamma\mathrm{C}}, \sigma_{\mu\gamma\mathrm{C}}\right), \tag{3.c}$$

where we denote the specific dimensions of heterogeneity of the variance-covariance parameters by F (for family) and C (for community). Assumption (3a) allows the idiosyncratic parameters to vary by sibling birth order, denoted by s, to increase model flexibility, and we include singletons among the first born. Assumptions (3.b) and (3.c) specify the distribution of shared factors and allow for an unrestricted covariance of initial earnings and earnings growth heterogeneity within each factor, which we denote by $\sigma_{\mu\gamma\mathrm{F}}$ and $\sigma_{\mu\gamma\mathrm{C}}$, respectively. Mincerian crossovers of earnings profiles would result in negative estimates for these covariances. We also model the sorting of families across communities by allowing for the covariance between family and community effects through the intercept of the individual-specific profiles:

$$cov\left(\mu_f, \mu_c\right) = \sigma_{\mathrm{FC}}\ . \tag{3.d}$$

The covariance (σ_{FC}) is non-zero if families sort themselves across communities.[15]

---

[14] We integrate out cohort effects in mean earnings via flexible first stage regressions (described in Section 4.1) and we do not include cohort effects in the earnings dynamics model which already includes age and time effects in both permanent and transitory earnings. Robustness checks that included cohort shifters on the earnings components of Equation (1) produced results entirely in line with the ones obtained from the model without cohort shifters. Similarly, estimating the baseline model only on older cohorts in our data, namely 1975 or earlier who reach at least age 38 in the sample period, did not change the substance of our findings.

[15] In preliminary analyses we experimented with a specification that allowed for correlation between the family and community components of the HIP *slopes*, obtaining a positive and non-significant estimate of their covariance and virtually unaffected estimates of remaining model parameters. Also, estimating a more general model with heterogeneous *quadratic* profiles yields results very similar to the ones of the linear HIP. We focus our analysis on the linear HIP for simplicity and ease of parameter interpretation.

4.3 Specification of transitory earnings

To capture any serial correlation of transitory shocks we model transitory earnings (v) in equation (1) using an AR(1) process. We allow siblings to draw shocks from sibling-specific distributions and we account for age effects in the variance of these shocks through an exponential spline in age. The model for transitory earnings can be summarized as follows:

$$v_{ifca} = \eta_t u_{ifca}; \qquad u_{ifca} = \rho_s u_{ifc(a-1)} + \varepsilon_{ifca}; \tag{4}$$
$$\varepsilon_{ifca} \sim \big(0, \sigma_{\varepsilon s}^2 \exp\big(g_s(a)\big)\big), \qquad u_{ifc24} \sim \big(0, \sigma_{u_{24}s}^2\big),$$

where η_t is a time loading factor and u_{ifca} is the sibling-specific AR(1) process. The autoregressive process begins at age 24 and we specify the variance of the initial condition ($\sigma_{u_{24}s}^2$). The process evolves through the arrival of white noise shocks (ε_{ifca}) whose variance is age-and-sibling-specific ($\sigma_{\varepsilon s}^2 \exp\big(g_s(a)\big)$, where $g_s(a)$ denotes a linear spline in age for sibling s with knots at 28, 33, 38 and 43.

We also allow for cross-person correlation of transitory shocks because they may be correlated not only across time but also between individuals. For siblings, the sibling-specific distribution of shocks enables identification of the contemporaneous correlation of AR(1) innovations. For two individuals i and i', the sibling covariance of AR(1) innovations is specified as follows:

$$E\big(\varepsilon_{ifca}\varepsilon_{i'fc'a'}\big) = \sigma_f, \quad \forall\, c, c', a = a' \pm |b - b'|, \tag{5}$$

where b' is the birth cohort of i'. That is, for siblings observed in the same time period (at different ages) the innovations of transitory earnings are allowed to co-vary with parameter σ_f. This covariance of shocks between siblings is transmitted over time through the autoregressive structure of the model.

For community peers, we follow a different approach to that used for pairs of siblings due to the high dimensionality that would result from parameterizing the covariance of transitory shocks between numerous peers belonging to different families (f and f'). Specifically, we allow for a catch-all "mass-point" covariance (λ) collapsing all the parameters of the underlying stochastic processes and allow the covariance to fade away over time. For any two (not necessarily different) ages a and a', the covariance of transitory shocks across community peers is specified as follows:

$$E\left(u_{ifca}\, u_{i'f'ca'}\right) = \lambda^{1+|t-t'|}, \quad |\lambda| < 1. \tag{6}$$

4.4 Identification of permanent earnings decomposition

Identification of the parameters determining permanent earnings relies on three sets of moment restrictions: for a given individual over time, and cross-person moment restrictions for siblings and for community peers. Compared to previous studies in the sibling literature which consider separately sibling and peer correlations, the modeling approach we propose has two advantages. First, we can identify family effects *net of any* community influences because we exploit *jointly* the moment restrictions for siblings and community peers. Second, we show that it is possible to separately identify sorting from community and family effects by exploiting between-sibling community variation.[16]

Earnings covariances for an individual are a function of all sources of earnings heterogeneity which include: (i) family influences, (ii) community influences, (iii) the sorting of families into communities and (iv) the idiosyncratic component. Individual moment restrictions for two (not necessarily different) ages, a and a', can be written as follows:

$$\begin{gathered} E\left(y_{ifca}, y_{ifca'}\right) = \\ \left\{\sigma^2_{\mu \mathrm{F}} + \sigma^2_{\mu \mathrm{C}} + \left(\sigma^2_{\gamma \mathrm{F}} + \sigma^2_{\gamma \mathrm{C}}\right)aa' + \left(\sigma_{\mu\gamma \mathrm{F}} + \sigma_{\mu\gamma \mathrm{C}}\right)(a + a') + 2\sigma_{\mathrm{FC}} \right. \\ \left. + \sigma^2_{\omega_{24}s} + \sigma^2_{\xi s} min(a, a')\right\}\pi_t\pi_{t'}. \end{gathered} \tag{7}$$

Cross-person moments, such as those between siblings and between community peers, do not depend on idiosyncratic heterogeneity. Moment restrictions for siblings (different i but same f) depend on family, community and sorting effects and can be written as follows:

$$E\left(y_{ifca}, y_{i'fc'a'}\right) = \left\{\sigma^2_{\mu \mathrm{F}} + \sigma^2_{\gamma \mathrm{F}} aa' + \sigma_{\mu\gamma \mathrm{F}}(a + a') + \right. \tag{8}$$

[16] In Page and Solon (2003a, b), due to the sampling design in the PSID, siblings always share the community. Raaum, Sörensen and Salvanes (2006) use a linear projection of earnings on neighborhood characteristics and neighborhood fixed effects to derive an approximation for the contextual term. Oreopoulos (2003) accounts for sorting of families into neighborhoods by using quasi-random assignment of neighbors.

$$I(c = c')\left[\sigma_{\mu\mathrm{C}}^2 + \sigma_{\gamma\mathrm{C}}^2 aa' + \sigma_{\mu\gamma\mathrm{C}}(a + a')\right] + 2\sigma_{\mathrm{FC}}\}\pi_t\pi_{t'} ,$$

where $I(.)$ is the indicator function. For community peers the covariance function depends only on community and sorting effects, but not on family effects, and can be written as follows:

$$E\left(y_{ifca}, y_{i'f'ca'}\right) = \left[\sigma_{\mu\mathrm{C}}^2 + \sigma_{\gamma\mathrm{C}}^2 aa' + \sigma_{\mu\gamma\mathrm{C}}(a + a') + 2\sigma_{\mathrm{FC}}\right]. \quad (9)$$

Combining the moment restrictions defined in equations (7)-(9), we can identify community effects including sorting from equation (9), family effects from equation (8) and idiosyncratic effects from equation (7). Because we exploit sibling and peer moments jointly in the model, we can identify family effects *net of any* community and sorting effects.

However, because families sort into communities, the estimated community influences will be upward biased according to the extent of sorting. Although this bias does not affect the estimated family effects, to identify separately the sorting parameter (σ_{FC}) from community effects we need an additional moment restriction. One possible source for the additional moment restriction is to consider community variation between siblings. Equation (8) nests moment restrictions for two types of siblings: (i) those who share the community, that is, $I(c = c') = 1$ and (ii) those who do not share the community, that is, $I(c = c') = 0$. Unlike the previous discussion, where we considered siblings who only share the community, with between-sibling community variation the additional moment restriction permits the identification of community effects separately from sorting.

To illustrate this identification argument, consider the covariance function for siblings sharing the community, which depends on family, sorting, and community effects. Whereas, for siblings who do not share the community, the covariance function depends only on family and sorting effects. The difference in the covariance functions between these two types of siblings identifies the community influences. Because community effects are identified by the moment restrictions for siblings, moment restrictions in (9) effectively identify sorting of families across communities. Finally, we can identify family effects by combining the moment restrictions of equation (8) and equation (9).

Siblings can be exposed to different communities at any given age because of family mobility. For example, when we measure community at age 15, family mobility after sibling 1 turns 15 (but before sibling 2 turns 15) generates between-sibling community variation because, at that age, the siblings reside in different neighborhoods or attend different schools. Instead, for immobile families, the two siblings share the community at that age. This between-sibling community variation identifies the sorting parameter separately from community effects.

The between-sibling community variation based on family mobility rests on the assumption that families who move are comparable to immobile families. However, these two types of families may be different due to underlying unobserved characteristics that may also affect earnings. To address possible biases coming from the comparison of moving families with stayer families, we also estimate the model using only moving families and exploit the timing of family mobility as a source of between-sibling community variation. These are all families who move, so it is only the difference in the timing of mobility that exposes siblings to different environments at that specific age. We argue that this source of between-sibling community variation is less prone to selection than contrasting movers and stayers. The main findings we present in Section 5 are unchanged when we restrict the analysis to movers and exploit the timing of family mobility.

4.5 Estimation and decomposition of the sibling correlation

We estimate parameters by Minimum Distance where we match moment restrictions implied by the model to the empirical moments derived from the data, using empirical moments based on at least 100 observations.[17] There are three types of empirical moments entering into the estimation. First, there are individual moments, which include the variances and intertemporal covariances of individual earnings. Second, there are sibling moments, which are defined in our sample only for the first two siblings in a family. This second set of moments implies that each family contributes at most once in the estimation of sibling empirical moments, while families with singletons do not contribute. To match the two different moment restrictions nested in equation (8) we estimate separate empirical moments for siblings depending on whether they share the community. Finally,

[17] Moment restrictions for transitory earnings are given in the Appendix. The orthogonality assumption between permanent and transitory earnings in equation (1) implies that moment restrictions of the full model are the sum of moment restrictions for permanent and transitory earnings. We use Equally Weighted Minimum Distance (see, for example, Haider, 2001).

there are empirical moments for community peers who by definition share the community. In contrast to families, the number of peers varies over communities. We account for the varying importance of communities using the weighting scheme we described in Section 3.

Using parameter estimates from the model we predict the contributions of family and community to the sibling correlation of permanent earnings over the life cycle. Specifically, we use the moment conditions of equations (8) and (9) and attribute the sorting parameter in equal parts to family and to community as follows:

$$\begin{aligned} r^F(a) &= \frac{E\left(y_{ifca}, y_{i'fc'a}\right) - \pi_t \pi_{t'}(\sigma_{\mathrm{FC}})}{E\left(y_{ifca}, y_{ifca}\right)}\ ; \\ r^C(a) &= \frac{E\left(y_{ifca}, y_{i'f'ca}\right) - \pi_t \pi_{t'}(\sigma_{\mathrm{FC}})}{E\left(y_{ifca}, y_{ifca}\right)}, \end{aligned} \tag{10}$$

where r denotes correlation coefficients of permanent earnings. It is worth noting that correlations vary with age because they are estimated from a model of life-cycle earnings. Given the model assumptions, the sibling correlation of permanent earnings for siblings sharing the community is the sum of the two components:

$$r^S(a) = r^F(a) + r^C(a). \tag{11}$$

5. The impact of family and communities on life-cycle earnings

In this part we present results for the impact of families and communities on life-cycle earnings. In Section 5.1 we discuss the 'core' parameter estimates of the permanent and transitory components; we report estimates of calendar time shifters of the two components in Appendix Table A1a for men and Table A1b for women. In Section 5.2, we decompose the sibling correlation of earnings into family and community effects. In Section 5.3 we present checks of robustness to various alternative community definitions, to different sources of community variation between siblings, to different measures of community affiliation, to the use of earnings ranks rather than log earnings, and in Section 5.4 we compare the main findings with the existing evidence in the literature.

5.1 Parameter estimates

Permanent earnings in equation (2) depend on shared factors and on idiosyncratic determinants. Panel A of Table 3a for men, and Panel B for women, show that family is the most important shared factor determining inequality of long-term earnings, both for initial earnings (intercept) and for earnings growth rates (slope). We find families to be the most important for the baseline model (Column 1), where community is defined as sharing the neighborhood, the school or both dimensions, as well as when community is defined based only on sharing the neighborhood regardless of school sharing (Column 2), or sharing the school regardless of neighborhood sharing (Column 3). For men, we still find positive and significant estimates of the variance of the community factor for both initial earnings and earnings growth, which are lower for the school-based definition of community. For women, instead, the estimated dispersion of the intercept for the community factor is very small and not significantly different from zero.[18]

For both genders and for all community definitions, Table 3a also shows negative covariances between intercepts and slopes of earnings profiles ($\sigma_{\mu\gamma\mathrm{F}}, \sigma_{\mu\gamma\mathrm{C}}$), which suggests that families and communities associated with low initial earnings (at age 24) are also associated with faster growth in life-cycle earnings. That is, shared determinants of long-term earnings display the "Mincerian crossover" property, implying that the variance of permanent earnings across these factors is U-shaped in age because it falls in the years of earnings catch-up and increases after the point of crossover. The point of crossover is the year the earnings variance is minimized, which is located – for the estimates in Column 1 – at age 33 (age 35) for the between-family earnings distribution for men (women) and at age 35 (age 34) for the between-community earnings distribution for men (women).[19]

[18] For women, because the effect of the community factor is very small, we obtain a negative but insignificant estimate for the intercept variance when we define community based only on sharing the neighborhood or only on sharing the school. Negative variances may occur in our framework because we perform unconstrained estimation, but a constrained estimator imposing positive variances just delivers estimates of constrained parameters on the boundaries of the parameter space. When we estimate these models restricting the sorting parameter to zero, we obtain positive and significant variances of intercepts, which reflect the positive sorting of families into communities.

[19] Taking into account that sorting contributes to the variance of intercepts, the *correlation* between intercept and slope for the baseline estimates equals -0.841 for men and -0.821 for women. Estimating a model imposing no sorting we can compute this correlation separately for the family and community factors, which for men equal -0.861 and -0.840, respectively, while the corresponding figures for women are -0.781 and -0.808.

Following the discussion in Section 4.4, to estimate the covariance between the two shared determinants of earnings (σ_{FC}), which captures sorting of families into communities, we exploit between-sibling community variation comparing mobile to immobile families. We find that sorting is positive and statistically significant for both men and women, which implies that a high draw from the distribution of family effects in permanent earnings is associated with a similarly high draw in the distribution of community effects.

Table 3b reports the estimates of the idiosyncratic source of permanent inequality in earnings. For both genders we find differences by birth order, such that dispersion of the initial condition of the random walk is larger for elder siblings, while the dispersion of life-cycle shocks is larger for younger siblings.

Table 4a for men and Table 4b for women show the parameter estimates of transitory earnings where there is a clear age pattern of transitory shocks whose variance decreases between the mid-20s and the mid-30s, while the decrease slows down around age 35. This sharp decline followed by a leveling-off is consistent with the patterns reported by Baker and Solon (2003) who find that the variance of transitory shocks declines at decreasing rate between the ages of 25 and 45. The age pattern of transitory shocks, as well as the autoregressive coefficients, are very similar between siblings. Finally, the correlation of transitory shocks between siblings is positive and significant. For community peers, the correlation of transitory shocks is generally negative. However, it is not always significant when the community definition includes the neighborhood and is positive and significant when community is based only on schools.

5.2 Decomposition of sibling correlation of earnings

To assess the relative importance of family and community effects in explaining life-cycle earnings inequality, we start by using the baseline parameter estimates of the model to decompose the sibling correlation of permanent earnings over the life cycle. We generate predictions of the sibling correlation and its decomposition based on the formulae provided in section 4.5 (equations 10 and 11). In particular, we consider the scenario represented by equation (11) of two siblings who share the community at age 15. The resulting sibling correlation is the sum of family and community effects, where we impute the estimated sorting parameter to the two factors in equal parts.

Figure 5 shows that the life-cycle pattern of the sibling correlation is U-shaped in age for both men, in Panel A, and women in Panel B. As discussed earlier, the U-shape pattern is the result of the "Mincerian crossover" of earnings profiles, which implies that the sibling correlation first shrinks and then fans out over the life cycle.[20] The decomposition of the sibling correlation in Figure 5 shows that family is the most important factor influencing the dispersion of permanent earnings. The average sibling correlation over the life cycle equals 0.313 for men (column 1 of Table 5a, Panel A) and 0.280 for women (column 1 of Table 5b, Panel A), which are in line with previous estimates for Denmark.[21] The average community correlation of earnings for men equals 0.023, while the estimate for women is smaller, 0.012, and not significant.

Figure 5 also shows that community effects are relatively more important at the beginning of the working life, but their influence in explaining earnings dispersion quickly diminishes over time. These results indicate that community effects play little role in shaping the sibling correlation in the long run. The family is the main factor that generates a substantial correlation in permanent earnings between siblings throughout the life cycle for both men and women. Furthermore, measuring earnings at relatively young ages exaggerates the relevance of community effects. For example, if we measured earnings only up to the beginning of the working life (age 25), the correlation of earnings between community peers is equal to 0.14 for men (Table 5a, column 1, panel B) and 0.079 for women (Table 5b, column 1, panel B). Instead, if we measured earnings only up to age 30, the community peers' correlation of earnings reduces to 0.079 for men and to 0.035 for women, while by age 35 it reduces further to 0.041 for men and to 0.015 for women. This evidence highlights the importance of analyzing earnings beyond the early part of the working life.

[20] The same U-shaped pattern is also a feature of the raw cross-person correlations shown in Figures 2 to 4, and especially in Figure 3, which depicts the earnings correlation for widely spaced siblings.

[21] Björklund et al. (2002, p. 765) report for men aged 25-42 a sibling correlation of earnings equal to 0.29 for a model without community effects. We obtain also an average estimate of 0.29 if we limit our sample to cover the same age range. In their study of sibling correlations and intergenerational transmission, Bingley and Cappellari (2019) report a sibling correlation of earnings equal to 0.22 using a sample that includes also earlier birth cohorts compared to the ones in our study. Applying our model to the same cohorts of Bingley and Cappellari (2019), we also obtain a sibling correlation of 0.22.

5.3 Robustness checks

5.3.1 Alternative community definitions

We check the robustness of the baseline results to alternative community definitions. For both men and women, we find a U-shaped pattern of the sibling correlation of earnings for the neighborhood-only (Figure 6) and school-only (Figure 7) definitions of community. In both cases family remains the most important factor explaining earnings inequality throughout the life cycle. For men, the estimates of Column 2 in Table 5a (Panel A) show that neighborhoods account for a small share of the sibling correlation of earnings with an estimate equal to 0.017 relative to a sibling correlation of 0.308, while the share accounted for by schools in Column 3 is lower with an estimate equal to 0.008 relative to a sibling correlation of 0.297. Similar to the baseline estimates, Columns 2 and 3 in Table 5a (Panel B) show for both community definitions that the average community correlation diminishes as we measure earnings for a greater part of the life cycle. In both cases, the share of sibling correlation of earnings accounted for by community effects declines when earnings are measured up to age 35, which is similar to the pattern for the baseline estimates. For women, in Table 5b, we obtain very similar patterns as for men, although the correlations are much lower and, in most cases, are not significantly different from zero.

5.3.2 Alternative between-sibling community variation

Next, we test the robustness of the baseline estimates to the source of between-sibling community variation which we exploit to separate community from sorting effects. Instead of comparing mobile to immobile families we re-estimate the model restricting the sample of siblings only to those coming from mobile families. Because we can perform this robustness check only for the neighborhood-based community definition, the proper comparison with the baseline results is with the decompositions in Figure 6, and the correlations in Column 2 of Tables 5a and 5b. We find for both men and women that the sibling correlation of earnings is still U-shaped (Figure 8) and that family accounts for most of the sibling correlation, with a diminishing importance of community effects as earnings are measured closer to the middle of the life cycle (Tables 5a and 5b – column

4 – for men and women, respectively). Although the share of sibling correlation accounted for by community – on average over the life cycle – increases somewhat, it still remains relatively small.

Although our main findings remain robust, in the absence of a random assignment of families to communities we cannot entirely rule out any bias due to endogenous family mobility. However, as we noted in Section 4.4, family mobility is necessary only if we aim to identify community effects separately from sorting. If not, we can still estimate within our model family effects net of any community influences using only siblings who always share the community (immobile families). In this case we expect community effects to be upward biased because they measure both the pure community effect and the sorting of families into communities. This estimate would give us an upper bound for community influences. Figure 9 and Column 5 in Tables 5a and 5b (for men and women, respectively) report the predicted sibling correlation for earnings and its decomposition over the life cycle from the immobile family model; they confirm our general finding that family accounts for most of the sibling correlation and that community effects do not persist in the long run. Compared to the baseline estimates, the share of the sibling correlation accounted for by community in the immobile family model increases, which as expected is upward biased, but nevertheless explains a very small portion of the resemblance between siblings.[22]

5.3.3 Alternative measures of community affiliation

It is possible that the community effects we report above may appear low because of measurement error, which tends to reduce their importance relative to family effects. By measuring communities at a single age (e.g. age 15) we might miss part of the community effects due to potentially limited exposure (for a similar discussion see also Chetty and Hendren, 2018). To check the sensitivity of the estimated community correlation to different ages of community affiliation we estimate a community-only model in which we ignore family ties and allow for community as the only factor determining cross-person correlations in permanent earnings. For this sensitivity exercise we vary the age we measure community from 11 to 15, and we also consider different age intervals.

[22] Parameter estimates from the estimation of the model restricted to mobile and immobile families are available in Appendix Tables A2a and A2b for men and Tables A2c and A2d for women.

Appendix Tables A3a, for men, and A3b for women show that the estimated average community correlations over the life cycle for all different ages and age intervals are very similar and are roughly equal to 0.06. These community estimates, although upward biased because in this case they capture both community and family influences, do not appear sensitive to the specific age we measure community affiliation.

5.3.4 Rank correlations

We further assess the robustness of our findings, particularly the greater importance of family relative to community in shaping sibling similarities, by considering earnings percentile ranks rather than log earnings. Rank correlations are becoming increasingly popular in analyses of intergenerational income mobility (see e.g. Chetty and Hendren, 2018), but there are no applications to siblings. We cannot apply the life-cycle model or the permanent/transitory decomposition to percentile ranks because the variance of percentiles is a fixed number, while our variance components model exploits changes in the variance of earnings with time and age. Therefore, we focus on the average rank correlation of *total* (not *permanent*) earnings and on its decomposition into family and community effects, leaving aside life-cycle aspects. We proceed by first estimating the (Spearman) intertemporal correlation structure of percentile ranks for individuals, sibling and peers. Next, we regress these correlations on dummies for whether they refer to individual earnings over time, or to pairs of observations sharing family, community or both, controlling for time, cohort and lags.

Results for rank correlations are reported in Appendix Table A4. In Panel A we report estimated coefficients; regressions do not include a constant such that estimates associated with the dummy variables described above are readily interpretable as levels of the rank correlations, not as deviations from the overall average of the correlation. Among shared factors, family is much more important than community, which is in line with results from life-cycle models. Another similarity with baseline results is that community is more important for men than for women. Panel B reports the implied percentile rank correlation for the case of two siblings sharing the community, that is, the sum of the “Family” and “Community” coefficients from Panel A. These estimates are roughly in the range of the correlations of raw earnings in Figure 2 and are lower than the correlations of permanent earnings because they refer to total earnings that are more volatile. Overall, the analysis of rank correlations confirms the findings that family is the most important factor for shaping the sibling correlation of earnings.

5.4 Discussion

Our estimated community effects are smaller than those found in observational studies and are closer to estimates accounting for sorting of families into neighborhoods using quasi-random assignment of neighbors, such as Oreopoulos (2003) who finds a zero correlation of earnings between neighbors. Page and Solon (2003a,b), without taking sorting into account, find a neighbor earnings correlation of 0.16 for men and 0.12 for women, which accounts for half of the estimated sibling correlation for men and one third for women. Raaum, Sörensen and Salvanes (2006) report for Norway an earnings correlation among men (women) youth neighbors of 0.028 (0.020) and a sibling correlation of 0.185 (0.164), which imply that neighborhood effects in Norway account for a smaller share compared to the US. For Denmark, a country in many respects similar to Norway, we find the influence of communities to be even weaker. Moreover, when we use the neighborhood-only definition of communities, which is prevalent in the literature, the share of sibling correlation accounted for by communities reduces further.

Within our model, because we consider both factors jointly, we estimate community net of family effects. Estimates of community effects in previous observational studies instead pick up also the influence of families because they compare the sibling correlation of earnings with the correlation of earnings among neighbors. To demonstrate this confounding, we return to the estimates from the community-only model, which is the closest equivalent to the ones reported in the sibling correlation literature, where community is defined as neighborhood-only and includes both community and family influences. Compared to the baseline, the community effects – reported in Appendix Tables A3a and A3b – when we do not account for family effects are consistently overestimated throughout the life cycle by a factor of three and are similar in size to the findings for Norway.

Our estimated life-cycle pattern of community effects concentrated before age 30 is consistent with the results of Bingley and Cappellari (2019). Using a different research design based on father-first-son/second-son triads, they show that father-son transmission is the main explanatory factor for brothers' resemblance in life-cycle earnings, while residual shared factors independent of paternal earnings – that include community effects—are only relevant at early stages of the life cycle.

Finally, our long follow-up period – beyond age 35 – also helps reconcile our estimated community effects with those in the literature. Our decomposition estimates show that observing earnings only during the early part of the working life inflates the importance of community. Taken together, these findings provide strong evidence that through the proposed model we can isolate a pure community influence which is lower than most previous estimates.

6. The impacts of families and communities on education and unemployment

In this section, we broaden the scope of the analysis by investigating the impacts of families and communities on other socioeconomic outcomes, namely educational attainment and unemployment. Sibling studies document sizeable correlations in education (see for example Björklund and Jäntti, 2012), but there is still limited evidence on how these compare with educational correlations among youth peers. Solon, Page and Duncan (2000), for the US, find a limited role for neighborhood factors in accounting for inequality in educational attainment. Raaum, Sörensen and Salvanes (2006), for Norway, find that both families and neighborhoods are important, but neighborhoods are less important than families. For unemployment, we provide the first comparison of correlations between siblings and youth peers.

6.1 Educational attainment – model and results

We observe qualifications in the education register, and use the norms produced by the Ministry of Education to calculate the time that would be taken to obtain each qualification by the shortest route. Educational attainment is imputed from the qualification with the highest normed duration on October 31 in the year of turning 29.

Because our measure of educational attainment is time-invariant, to estimate family and community effects we use a restricted version of the empirical model presented in Section 4 without time- or age-related variation of the outcome and, consequently, without a transitory component. To account for the fact that data refer to individuals attending school over a 20-year window, we allow for heterogeneity in educational attainment across cohorts through a set of cohort shifters. We specify years of education, y_{ifc}, for person i belonging to family f and community c, to be the sum of time-invariant factors that can be individual-specific (ω_i), family-specific (μ_f) or community-specific (μ_c) and loaded by cohort-specific factor δ_b:

$$y_{ifc} = \delta_b(\omega_i + \mu_f + \mu_c); \quad \omega_i \sim (0; \sigma^2_{\omega s}), \; s = 1,2; \quad \mu_f \sim \left(0; \sigma^2_{\mu \mathrm{F}}\right); \mu_c \sim \left(0; \sigma^2_{\mu \mathrm{C}}\right). \tag{12}$$

We admit cross-factor dependence between family and community to allow for the possibility of sorting of families into communities with parameter σ_{FC}. As with the earnings model, parameters are identified from moment restrictions. Namely, the covariance of years of education between siblings who share the community depends on family, community and sorting effects; for siblings who do not share the community, their covariance only depends on family and sorting; while for community peers who belong to different families their covariance of education depends on community and sorting effects. Finally, education variances will depend on all four sources of heterogeneity in the model (idiosyncratic, family, community and sorting). Using cohort-specific empirical moments identifies cohort shifters.

Considering the particular nature of the outcome under investigation, in Table 6 we present results where community is defined using the school affiliation.[23] Parameter estimates of the variance components, reported in Panel A, clearly indicate that, for both men and women, heterogeneity between families is a more important determinant in shaping educational inequalities than differences between communities. The model also features a sorting parameter whose estimate is positive but small in size and statistically insignificant. It should be emphasized that the model is estimated on relatively few empirical moments (68 versus more than 10,000 in the earnings case) such that the identification of the sorting parameter, although theoretically feasible, may turn out

[23] Estimates from models with community defined as neighborhood produced results virtually identical to those of Table 6, with the exception of the sorting parameter that in that case is estimated to be negative and insignificant.

difficult in practice. The estimates also show a declining trend of educational inequalities for men moving from older to younger cohorts, while for women the pattern is less pronounced with the exception of the youngest cohort.

Panel B of Table 6 reports estimates of the sibling correlation in education of 0.33 for men and 0.35 for women, which are in line with previous evidence for Denmark (Bredtmann and Smith, 2018). The novelty of our results is that we can decompose the sibling correlation into family and community factors. We find a community correlation in educational attainment equal to 0.065 for men and 0.057 for women. Compared to our findings for earnings these community effects are larger. Nevertheless, we still find that family is the most important factor of the sibling correlation (accounting for more than four fifths) and that the prevalence of the family effect is stronger among women than it is for men.

6.2 Unemployment – model and results

For unemployment, we use social security records of time spent on unemployment benefits during the year for those who are insured against unemployment. This measure takes the value of 0 for those who are not registered unemployed throughout the year and reaches a maximum of 1,000 for those on unemployment benefits for the whole year. Because the unemployment variable is available to us only up to 2009, to ensure that life-cycle unemployment trajectories can be observed at least up to age 30 – as in the case of earnings – we exclude from the analysis the two youngest cohorts, 1981 and 1984.[24]

The variance components analysis for the unemployment measure is challenging because the vast majority of cases in the data (81% for men and 73% for women) score 0 on the measure (equivalent to no registered unemployment), which makes the linear model used for earnings dynamics inappropriate for the unemployment case.[25] We overcome this issue by taking a limited dependent variables approach assuming the data are generated by an underlying latent linear unemployment propensity that maps into observed realizations through a Tobit specification. To

[24] Recall from footnote 14 that excluding younger cohorts does not alter the results for life-cycle earnings.
[25] Right-side censoring at the value of 1,000 (full-year unemployment) is negligible, 0.02% for men and 0.03% for women.

operationalize this approach, we develop the Tobit version of the dynamic variance components model, which is novel in the literature.

Let ϖ_{ifca} be the observed unemployment variable, where the indices *i, f, c* and a stand for individual, family, community and age, respectively. The unemployment indicator depends on an underlying latent unemployment propensity w^*_{ifca} according to the following relationship

$$\varpi_{ifca} = w^*_{ifca} \times I(w^*_{ifca} > 0). \tag{12}$$

The latent propensity depends on a quadratic polynomial in age, which is period and cohort specific, and on a composite error term w_{ifca}, which is defined in equation (1) of Section 4, and can be written in the following way:

$$w^*_{ifca} = \beta^{(0)}_{bt} + \beta^{(1)}_{bt} a + \beta^{(2)}_{bt} a^2 + w_{ifca}. \tag{13}$$

Assuming that all components of the composite error term, w_{ifca}, are zero mean normally distributed random variables with second moments specified in equations (1) to (6) of Section 4, equations (12) and (13) result in a Tobit model for the censored unemployment indicator, and the empirical covariance structure of w_{ifca} can be estimated from the second moments of Tobit errors. In particular, the empirical covariances of w_{ifca} across different time periods, or across different individuals (siblings or peers), can be estimated as the error covariance of two-equation Tobit systems.[26] In this way, we obtain estimates of the empirical counterparts of $E(w_{ifca}, w_{ifca'})$ (intertemporal individual covariance), $E(w_{ifca}, w_{i'fca'})$ (sibling covariance for siblings sharing community), $E(w_{ifca}, w_{i'fc'a'})$ (sibling covariance for siblings not sharing community), and $E(w_{ifca}, w_{i'f'ca'})$ (peers covariance).

Due to the orthogonality assumption of equation (1) these covariances factor into the sum of covariances of the permanent and the transitory component, whose analytical expressions are given in Section 4 (permanent component) and in the Appendix (transitory component). The model features (sibling-specific) unit root processes, consistent with the idea that individual

[26] When estimating covariances across peers, we weight the maximum likelihood estimator using the weighting scheme of Page and Solon (2003a).

unemployment is highly persistent. In addition, it includes heterogeneous profile specifications of the shared components. In principle, one would expect a positive covariance between intercept and slope because of state dependence in unemployment, such that those hit by an adverse shock at the start of their working life – originating from the family or community – will remain scarred over the life cycle. Still, the human capital interpretation of shared HIP profiles suggested for earnings is applicable also to the unemployment case; initial investments in human capital may delay the school-to-work transition resulting in higher levels of unemployment in the beginning but catching up later in life through more stable employment patterns. This mechanism may be historically more relevant for women to the extent that non-investors select into less career-oriented jobs, which are associated with more unstable employment patterns over the life cycle.

We estimate the unemployment model by EWMD, matching the empirical moments from the covariance structure of Tobit errors to the theoretical moments implied by the model of Section 4.[27] Results are reported in Table 7 (permanent components), Table 8 (transitory components) and Table 9 (siblings and peers correlations).[28] For women, parameters of the permanent component are generally precisely estimated (the variance of the slopes for the community component is significant at the 20 percent confidence level) and display patterns that are largely consistent with those observed in the earnings case; in particular, a negative covariance of HIP intercepts and slopes supporting the human capital interpretation of family- or community- specific heterogeneity in life-cycle unemployment profiles. For men, we find an impact of communities on initial unemployment but not on life-cycle dynamics since the variance of community-specific HIP slopes and the intercept-slopes covariance are imprecisely estimated. On the other hand, families do affect the life-cycle unemployment profile of men, and the positive estimate of the intercept-slope

[27] We estimate the cross-sectional variances as the variance of the errors from single-equation Tobit models. Haider and Solon (2006) used the variances and covariances of Tobit errors to retrieve the empirical covariance structure of earnings from censored data and simulate earnings profiles based on that. In contrast to Haider and Solon, we use the empirical covariance structure of the latent process to estimate the parameters of the underlying dynamic process via EWMD. Because we do not observe the realizations of the linear unemployment propensity, we cannot estimate individual-level deviations from average population moments that we used to obtain the full fourth moment matrix in the earnings analysis. In turn, this implies that we can retrieve the variance of each estimated moment, but not the cross-moments covariance, yielding a *diagonal* fourth moments matrix that we use in the EWMD estimator for unemployment.

[28] Estimates of the time shifters are reported in Appendix Table A5a for men and Table A5b for women.

covariance points to the prevalence of state dependence in family-specific shocks to the unemployment trajectories of siblings. For both genders, family is by far the most relevant shared factor of unemployment dynamics. Also, in both cases, there is a positive and sizeable covariance across components reflecting the fact that families with a high unemployment propensity sort into high unemployment communities.

Estimates of transitory unemployment shocks in Table 8 show that the non-stationary AR(1) process is generally precisely estimated also in the unemployment case, but, in contrast to earnings, it is less clear that the dispersion of shocks declines over the life cycle, especially for women. Transitory unemployment shocks are equally persistent for both genders, but they are more positively correlated between brothers than between sisters, which is further evidence that family-specific unemployment persistence is more pronounced for men. Finally, we do not find a statistically significant covariance of shocks among youth peers.

We use parameter estimates to retrieve sibling correlations and their decompositions over the life cycle. On average, sibling correlations in unemployment are similar in size to those for earnings and education, 0.29 and 0.32 for men and women, respectively. For both genders, family is by far the most relevant factor in unemployment correlations. Community influences are more prevalent for women compared to men, reflecting the greater impact of communities on unemployment trajectories shown in Table 7. The more pronounced influence of community among women is distinctive for the unemployment variable relative to education and earnings. We also find evidence of a life-cycle decline of the sibling correlation, for both men and women, suggesting that over time individual idiosyncratic factors become the relevant drivers of unemployment dynamics.

To summarize the estimates of the unemployment model we plot in Figure 10 the life-cycle profile of sibling correlations in unemployment and their components. The patterns are very similar to those obtained for earnings, but there are differences. In particular, community effects seem to be more substantive, particularly at the beginning of the life cycle – a feature that is more evident for women.

7. Conclusion

Using population-based administrative data for Denmark, by virtue of which we can link records on earnings, education and unemployment of siblings, schoolmates and parish neighbors, we analyze the relative influence of family and youth community on life-cycle inequality of earnings and unemployment as well as inequality of educational attainment. We develop and estimate a model which accounts for the joint earnings dynamics of multiple groups of individuals, which allows us to decompose the sibling correlation of earnings into family and community effects over the life cycle. Modelling jointly the earnings of siblings and peers we identify family effects net of any community effects, and we show that it is possible to identify community influences separately from sorting of families into communities. We adapt this earnings model to the joint unemployment dynamics of siblings and peers and to educational attainment.

We find that the sibling correlation of earnings is U-shaped for both men and women, consistent with the prediction of human capital theory that heterogeneous investments in human capital induce an inverse relationship between initial earnings and earnings growth rates. Our decomposition of the sibling correlation of earnings shows that community background matters more in the early part of the working life, but its importance quickly diminishes and becomes negligible after age 30. On average, community accounts for less than a tenth of the sibling correlation of earnings. Within the youth environment, family is the most important factor influencing sibling correlations of earnings in the long run. These findings are robust to the measurement of youth communities and to various sample selection choices. The diminishing community influence over the life cycle highlights the importance of observing long earnings histories beyond the first years of the working life. We find similar life-cycle patterns for unemployment, and results for education also highlight the relevance of the family. We conclude that family is the most important factor influencing sibling correlations not only of earnings but also of education and unemployment for both men and women.

Our findings are based on data from Denmark with a welfare system promoting equality of opportunity. However, as highlighted recently by Landersø and Heckman (2016), there is less educational mobility than income mobility in Denmark, with low private financial returns to schooling discouraging educational investments among the children of less educated parents. These family influences are consistent with our finding that family is the most important determinant of long run earnings similarities across siblings. Communities seem to affect earnings early in the working life, for example through peers influencing youth behaviors, but these influences are short lived deviations from an earnings profile that – apart from idiosyncratic factors – mainly reflects characteristics and choices of the family. Our findings for education and unemployment confirm that family is the most important shared factor for inequalities in the long run. As administrative datasets and cohort studies mature in other countries, our approach to modeling group-wise dynamics could be applied to measure family and community effects on long run outcomes in other contexts.

REFERENCES

Baadsgaard, Mikkel, and Jarl Quitzau. 2011. "Danish Registers on Personal Income and Transfer Payments." *Scandinavian Journal of Public Health* 39 (7): 103-105.

Baker, Michael, and Gary Solon. 2003. "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records." *Journal of Labor Economics*, 21 (2): 267-88.

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy*, 87 (6): 1153-89.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75 (4): 352-65.

Bénabou, Roland. 1996. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review*, 86 (3): 584-609.

Bertdtmann, Julia and Nina Smith. 2018. "Inequalities in Educational Outcomes: How Important Is the Family?" *Oxford Bulletin of Economics and Statistics*, 80 (6): 1117-1144.

Bhuller, Manudeep, Magne Mogstad and Kjell Salvanes. 2017. "Life-Cycle Earnings, Education Premiums, and Internal Rates of Return." *Journal of Labor Economics*, 35 (4): 993-1030.

Bingley, Paul, and Lorenzo Cappellari. 2019. "Correlations of Brothers' Earnings and Intergenerational Transmission." *Review of Economics and Statistics*, 101 (2): 370-383.

Björklund, Anders, and Markus Jäntti. 2009. “Intergenerational Income Mobility and the Role of Family Background.” in *Oxford Handbook of Economic Inequality*, edited by Wiemer Salverda, Brian Nolan, and Timothy Smeeding, 491-521. Oxford: Oxford University Press.

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): 671-80.

Björklund, Anders, Tor Eriksson, Markus Jäntti, Oddbjrön Raaum and Eva Österbacka. 2002. “Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States.” *Journal of Population Economics*, 15:757–72.

Black, Sandra E., Paul J. Devereux. 2011. “Recent Developments in Intergenerational Mobility.” in *Handbook of Labor Economics, Vol. 4A*, edited by Orley Ashenfelter and David Card, 1487-541. Amsterdam: Elsevier Science, North Holland.

Blundell, Richard, Robert Joyce, Agnes Norris Keiller and James P. Ziliak. 2018. “Income Inequality and the Labour Market in Britain and the US.” *Journal of Public Economics*, 126: 48-62.

Bohlmark, Anders, and Matthew J. Lindquist. 2006. “Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden.” *Journal of Labor Economics*, 24 (4): 879-900.

Chetty, Raj, and Nathaniel Hendren. 2018. “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects.” Quarterly Journal of Economics 113 (3).

Chetty, Raj, Nathalien Hendren, and Lawrence F. Katz. 2016. “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.” *American Economic Review*, 106 (4): 855-902.

Chetty, Raj, John N, Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *Quarterly Journal of Economics*, 126 (4): 1593-660.

Durlauf, Steven N. 1996. “A Theory of Persistent Income Inequality.” *Journal of Economic Growth*, 1 (1): 75-93.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. “Long-Term Effects of Class Size.” *Quarterly Journal of Economics*, 128 (1): 249-85.

Haider, Steven J. 2001. “Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991.” *Journal of Labor Economics*, 19 (4): 799-836.

Haider, Steven J., and Gary Solon. 2006. “Life-Cycle Variation in the Association between Current and Lifetime Earnings.” *American Economic Review*, 96 (4): 1308-20.

Hanushek, Erik A. 2006. “School Resources.” In *Handbook of the Economics of Education, Volume* 2, ed. Eric A. Hanushek, and Finis Welch, 865-908. Amsterdam: Elsevier Science, North Holland.

Jenkins, Stephen J. 1987. “Snapshots vs. Movies: ‘Lifecycle biases’ and the Estimation of Intergenerational Earnings Inheritance.” *European Economic Review*, 31 (5): 1149–58.

Jensen, Vibeke Myrup and Astrid Wurtz Rasmussen. 2011. “The Danish Education Registers.” *Scandinavian Journal of Public Health* 39 (7): 91-94.

Landersø, Rasmus and James J. Heckman. 2016. “The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S.” Forthcoming, *Scandinavian Journal of Economics*.

Magnac, Thierry, Nicolas Pistolesi, and Sébastien Roux. 2018. "Post-Schooling Human Capital Investments and the Life Cycle of Earnings." *Journal of Political Economy*, 126 (3): 1219-1249.

Mazumder, Bhashkar. 2005. “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data.” *Review of Economics and Statistics* 87 (2): 235–55.

Mincer, Jacob. 1958. “Investment in Human Capital and Personal Income Distribution” *Journal of Political Economy*, 66 (4): 281-302.

Nybom, Martin, and Jan Stuhler. 2016. “Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation.” *Journal of Human Resources*, 51 (1): 239-68.

Nybom, Martin. 2017. “The Distribution of Lifetime Earnings Returns to College.” *Journal of Labor Economics*, 35 (4): 903-52

Oreopoulos, Philip. 2003. “The Long-Run Consequences of Living in a Poor Neighborhood.” *Quarterly Journal of Economics*, 118 (4): 1533-75.

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): 831-56.

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.

Pedersen, Carsten Bøcker, Heine Gøtzsche, Jørgen Østrup Møller and Preben Bo Mortensen 2006. "The Danish Civil Registration System. A cohort of eight million persons." *Danish Medical Bulletin*, 53 (4): 441-9.

Piketty Thomas. 2000. "Theories of Persistent Inequality and Intergenerational Mobility." In *Handbook of Income Distribution*, Volume 1, ed. A.B. Atkinson and F. Bourguignon, 429-476. Amsterdam: Elsevier Science, North Holland.

Raaum, Oddbjörn, Erik Sörensen, and Kjell G. Salvanes 2006. "The Neighbourhood Is Not What It Used to Be." *Economic Journal*, 116 (508): 200-22.

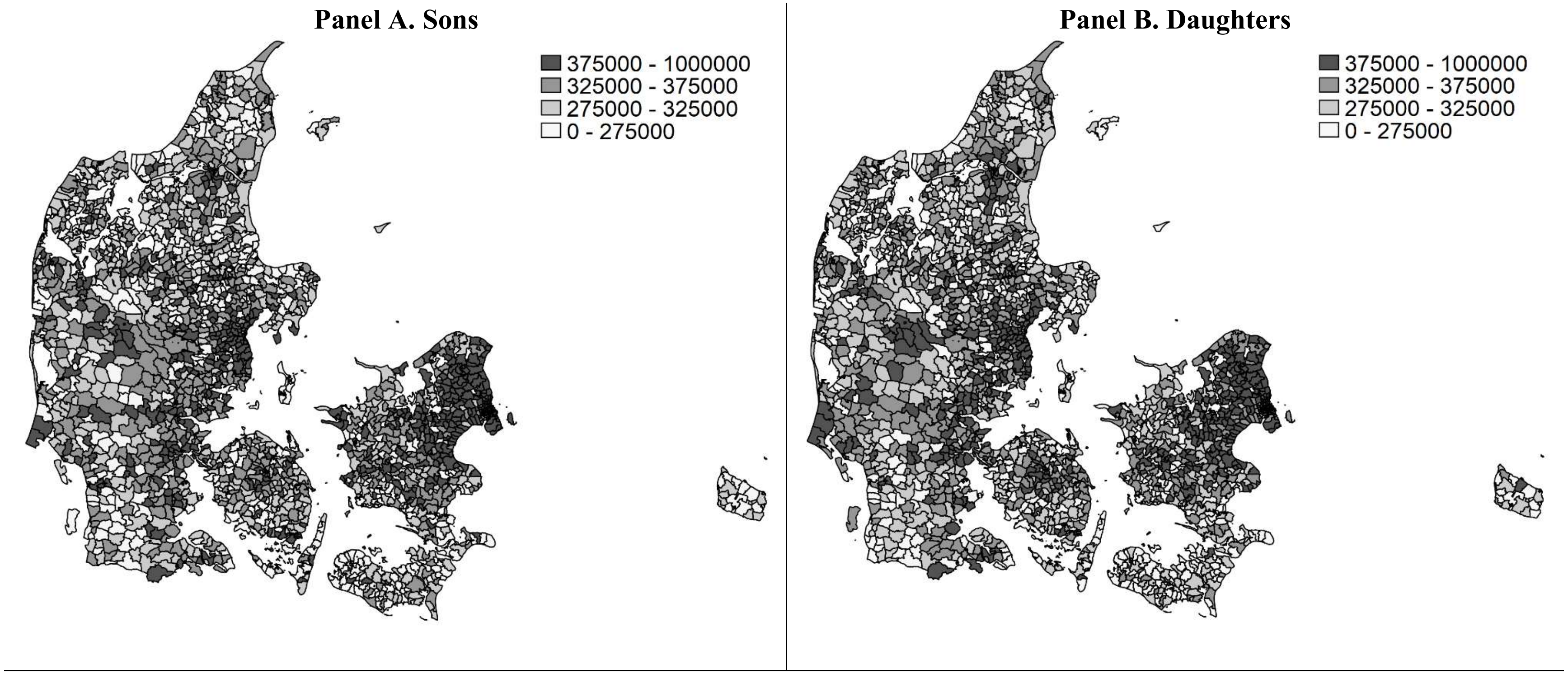
Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82 (3): 393–408.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." In *Handbook of Labor Economics, Vol. 3A*, ed. Orley Ashenfelter, and David Card, 1761-1800. Amsterdam: Elsevier Science, North Holland.

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

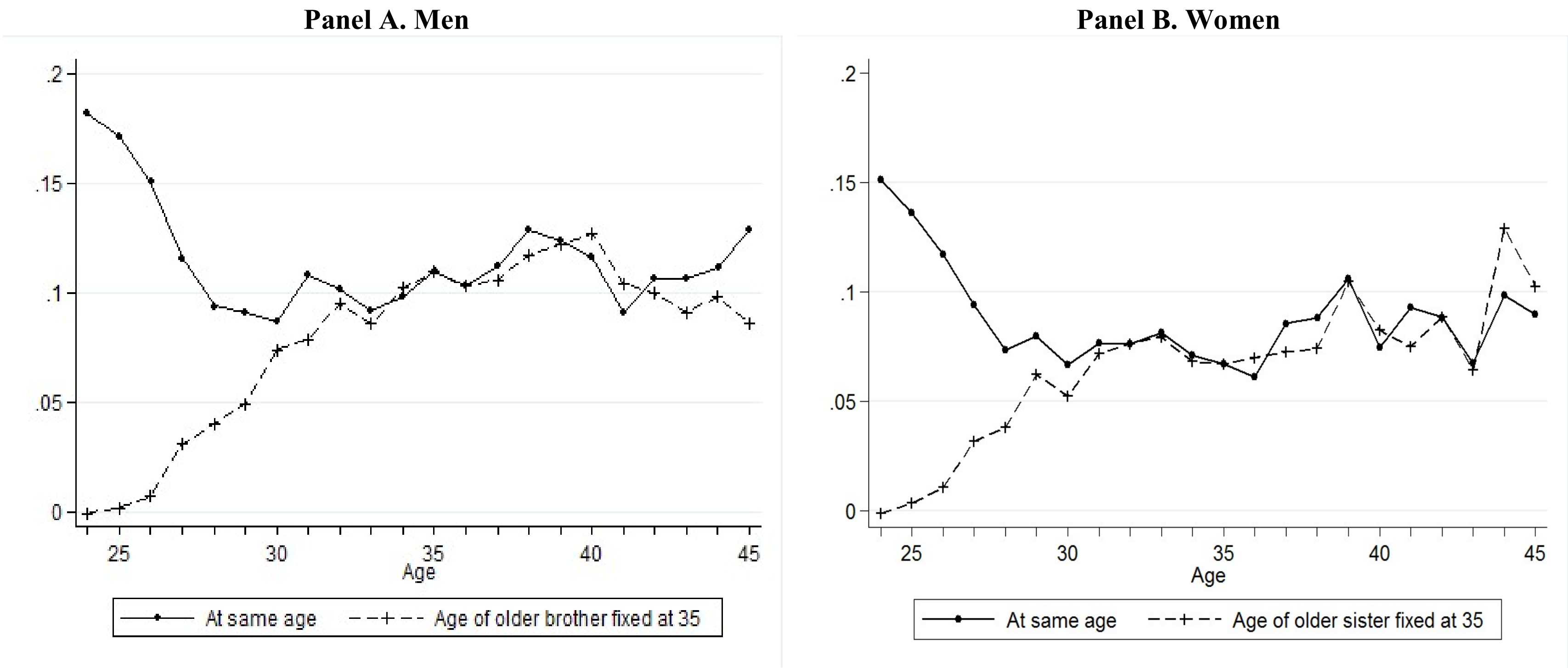
Wilson, William J. 1987. "The Truly Disadvantaged." Chicago: Chicago University Press.

Figure 1. Mean fathers’ earnings when child is age 15



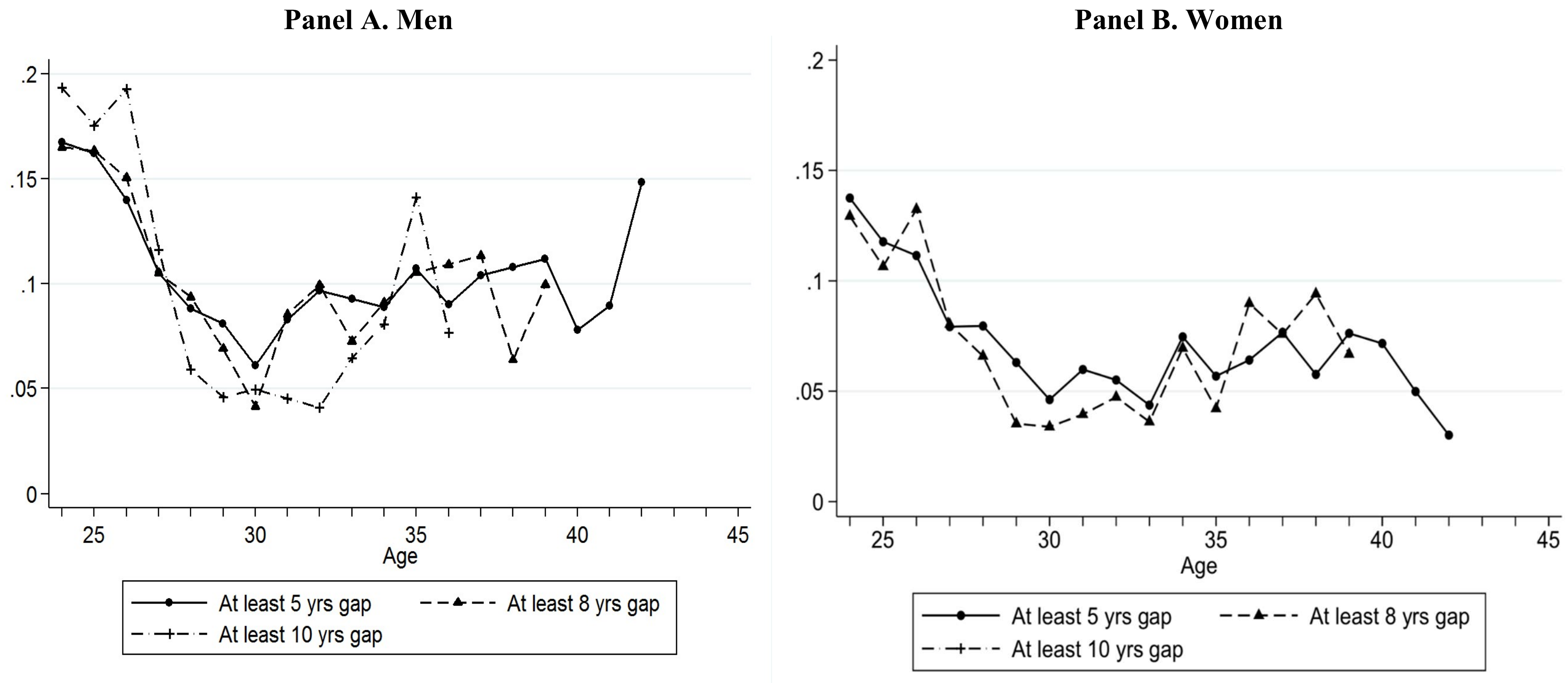
Note: Paternal annual labor earnings (DKK 2012 prices) across Danish parishes when son (Panel A) or daughter (Panel B) is age 15. Shading indicates different mean earnings with groupings approximately corresponding to earnings quartiles. The scale of the map is 300km from east to west and 200km from south to north.

Figure 2. Sibling correlation of annual earnings.



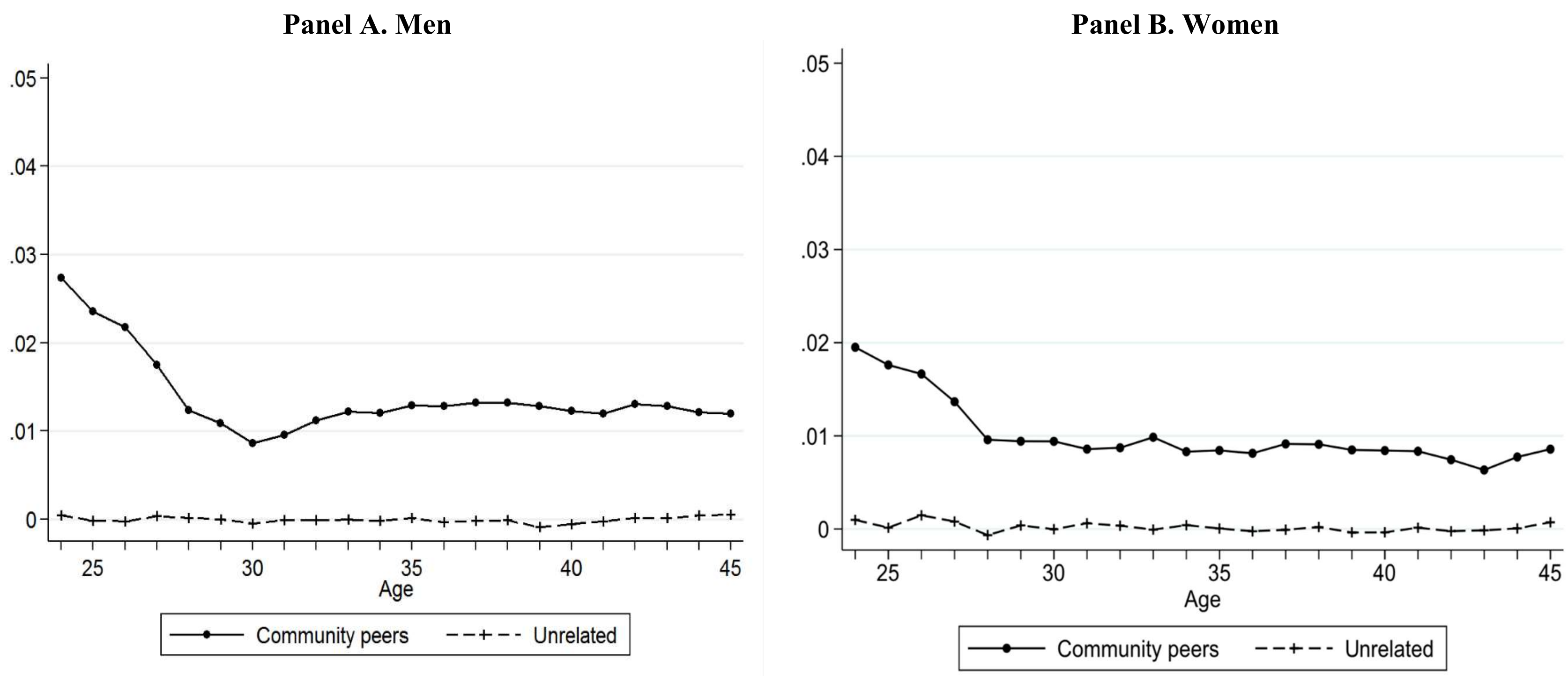
Note: The figure shows raw sibling correlations of earnings over the life cycle based on 89,738 brother pairs in Panel A and 79,785 sister pairs in Panel B born between 1965 and 1985. The solid line shows the sibling correlation when the siblings are at the same point in their life cycle, while the dashed line shows the sibling correlation when we fix the age of the elder sibling at 35.

Figure 3. Sibling correlation of annual earnings by siblings' age gap.



Note: The figure shows raw sibling correlations of earnings over the life cycle when siblings are at the same point in their life cycle, and refer to sibling pairs with an age gap of 5, 8 and 10 years born between 1965 and 1985.

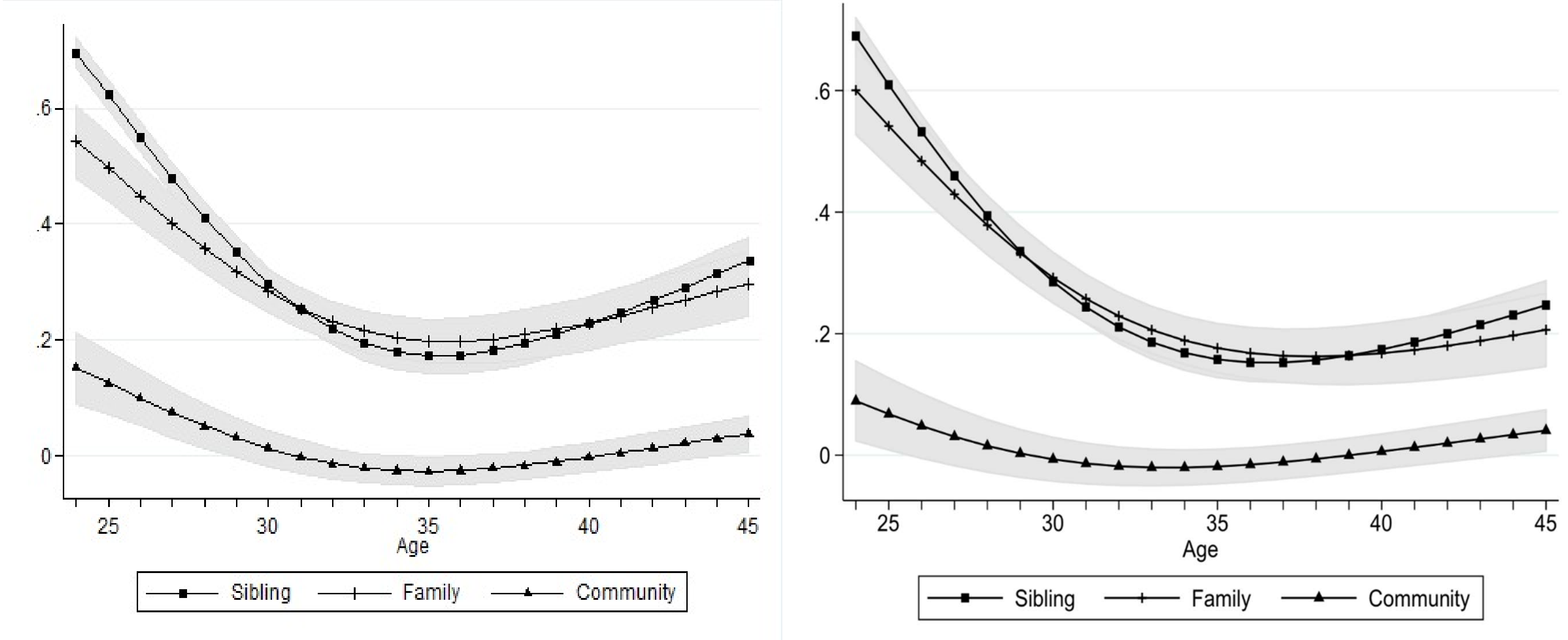
Figure 4. Correlation of annual earnings among community peers.



Note: The figure shows raw correlations of earnings for community peers (boys in Panel A and girls in Panel B) born between 1965 and 1985 over the life cycle. Community peers are defined as neighbors who reside in the same parish at age 15, or schoolmates who attend the same school at age 15. Unrelated individuals share neither the family nor the community. We compute this correlation by randomly matching each individual in the sample with 1,000 same gender individuals born in the same year, regardless of family and community.

Figure 5. Predicted sibling correlation of permanent earnings and factor decomposition: baseline model.

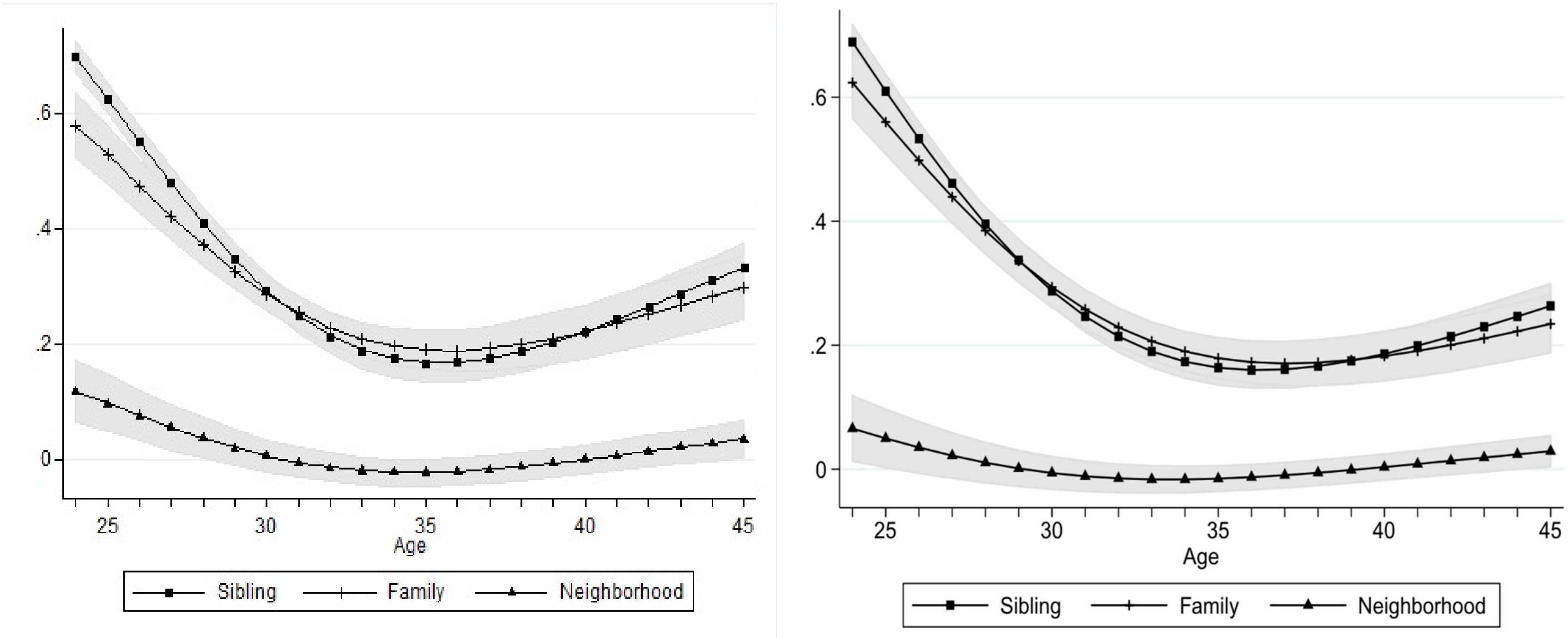
Panel A. Men **Panel B. Women**



Note: The figure shows the predicted sibling correlation of earnings over the life cycle and its decomposition into family and community effects in Panel A for men and in Panel B for women, using the baseline estimates of Tables 3a and 4a for men and Tables 3b and 4b for women. Community is defined as sharing the neighborhood, the school or both community dimensions at age 15. The sorting parameter is estimated using between-sibling variation of community comparing mobile to immobile families. Predictions are generated using the formulae provided in Section 4.5 for the case of siblings sharing the community**.** The resulting sibling correlation is the sum of family and community effects, where we impute the estimated sorting parameter to the two factors in equal parts.

Figure 6. Predicted sibling correlation of permanent earnings and factor decomposition: community defined based only on neighborhoods.

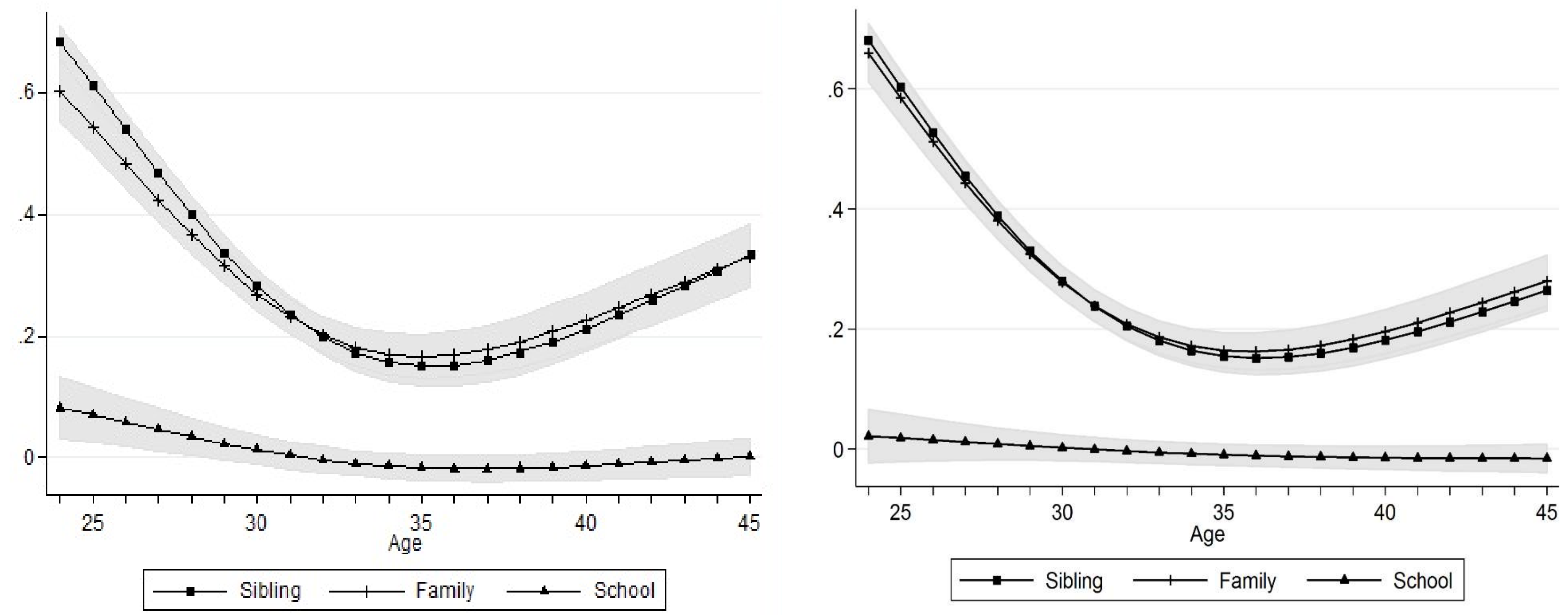
Panel A. Men **Panel B. Women**



Note: The figure shows the predicted sibling correlation of earnings over the life cycle and its decomposition into family and community effects when community is defined as sharing only the neighborhood regardless of school sharing at age 15 based on the estimates of Column 2 in Tables 3a and 4a. The sorting parameter is estimated using between-sibling variation of community comparing mobile to immobile families. Predictions are generated using the formulae provided in Section 4.5**.** The resulting sibling correlation is the sum of family and neighborhood effects, where we impute the estimated sorting parameter to the two factors in equal parts.

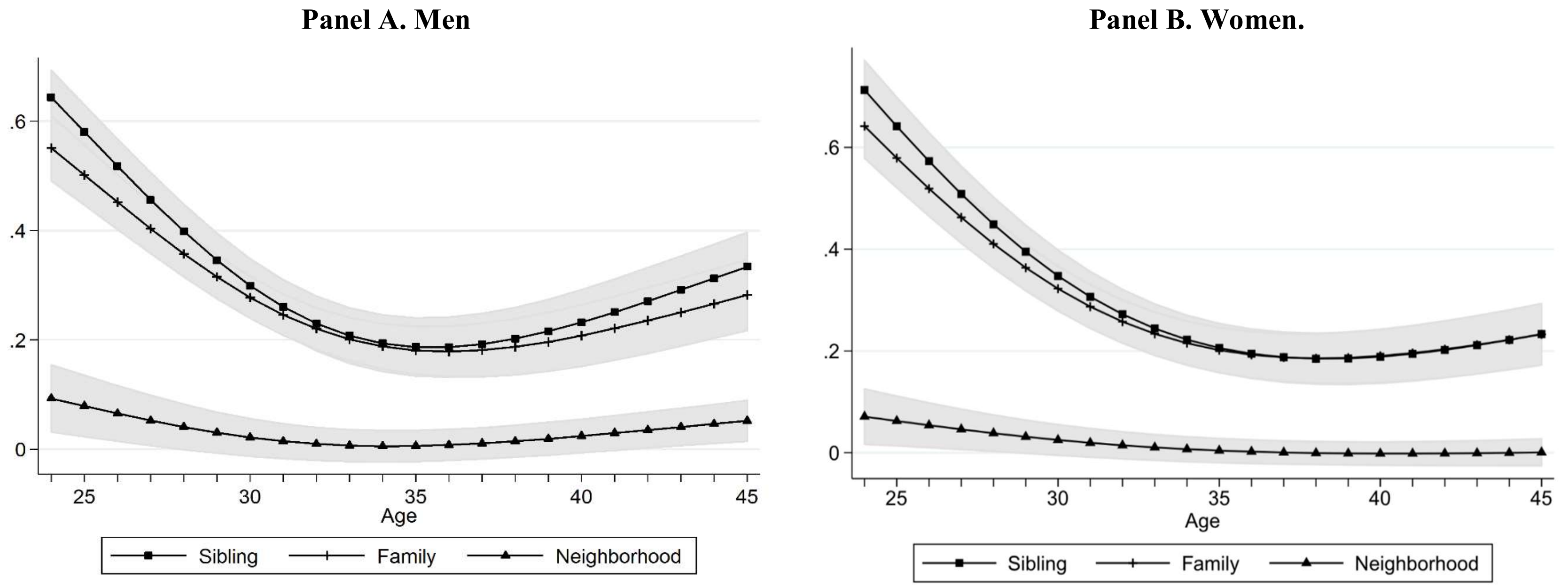
Figure 7. Predicted sibling correlation of permanent earnings and factor decomposition: community defined based only on schools.

Panel A. Men **Panel B. Women**



Note: The figure shows the predicted sibling correlation of earnings over the life cycle and its decomposition into family and community effects when community is defined as sharing only the school regardless of neighborhood sharing at age 15 based on the estimates of Column 3 in Tables 3b and 4b. The sorting parameter is estimated using between-sibling variation of community comparing mobile to immobile families. Predictions are generated using the formulae provided in Section 4.5**.** The resulting sibling correlation is the sum of family and school effects, where we impute the estimated sorting parameter to the two factors in equal parts.

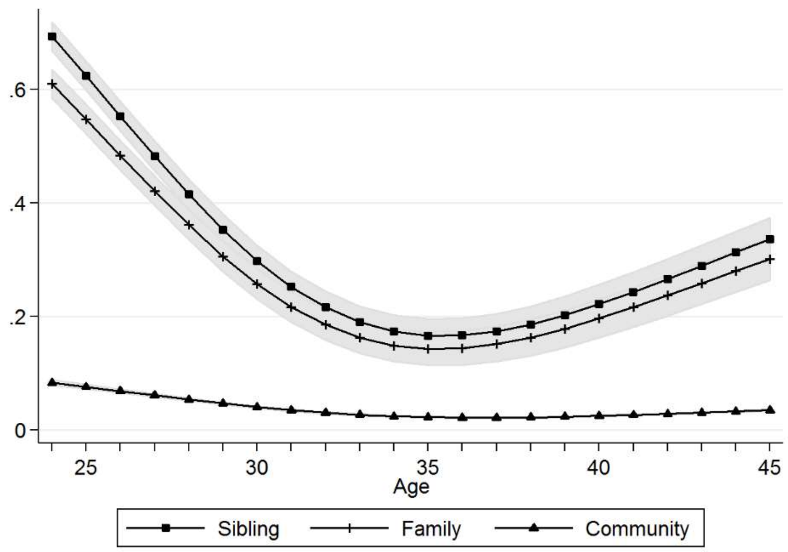
Figure 8. Predicted sibling correlation of permanent earnings and factor decomposition using only mobile families.



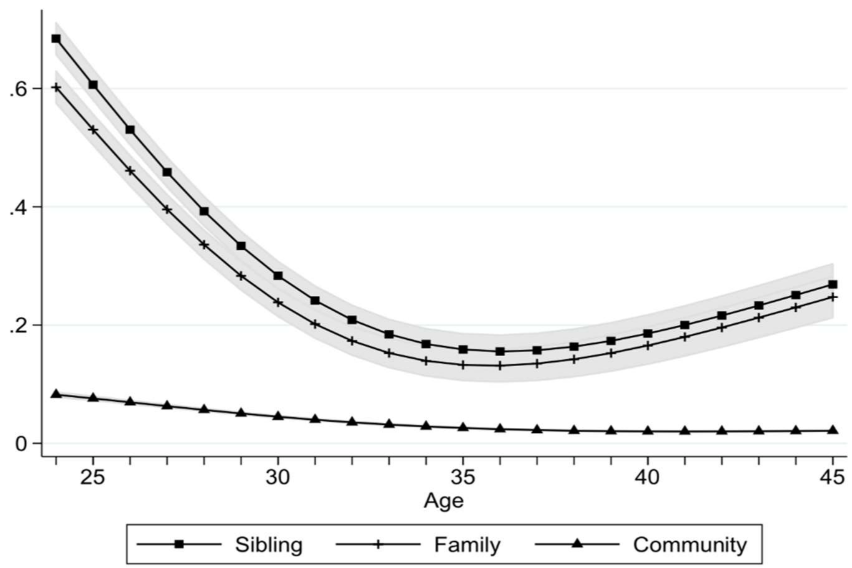
Note: The figure shows the predicted sibling correlation of earnings over the life cycle and its decomposition into family and community effects exploiting between-sibling community variation based only on mobile families. Community is defined as sharing only the neighborhood regardless of school sharing at age 15. Differently from Figure 6, where community is also defined based on neighborhood-only affiliation, we exclude families for whom the parish of the first sibling is the same between ages 11 and 15. So we compare only mobile families; those who move before sibling 1 turns 15 to those who move after sibling 1 turns 15. As a result the sorting parameter is estimated using between-sibling variation of community based only on mobile families. Predictions are generated using the formulae provided in Section 4.5. The resulting sibling correlation is the sum of family and community effects where we impute the estimated sorting parameter to the two factors in equal parts.

Figure 9. Predicted sibling correlation of permanent earnings and factor decomposition using only immobile families.

Panel A. Men

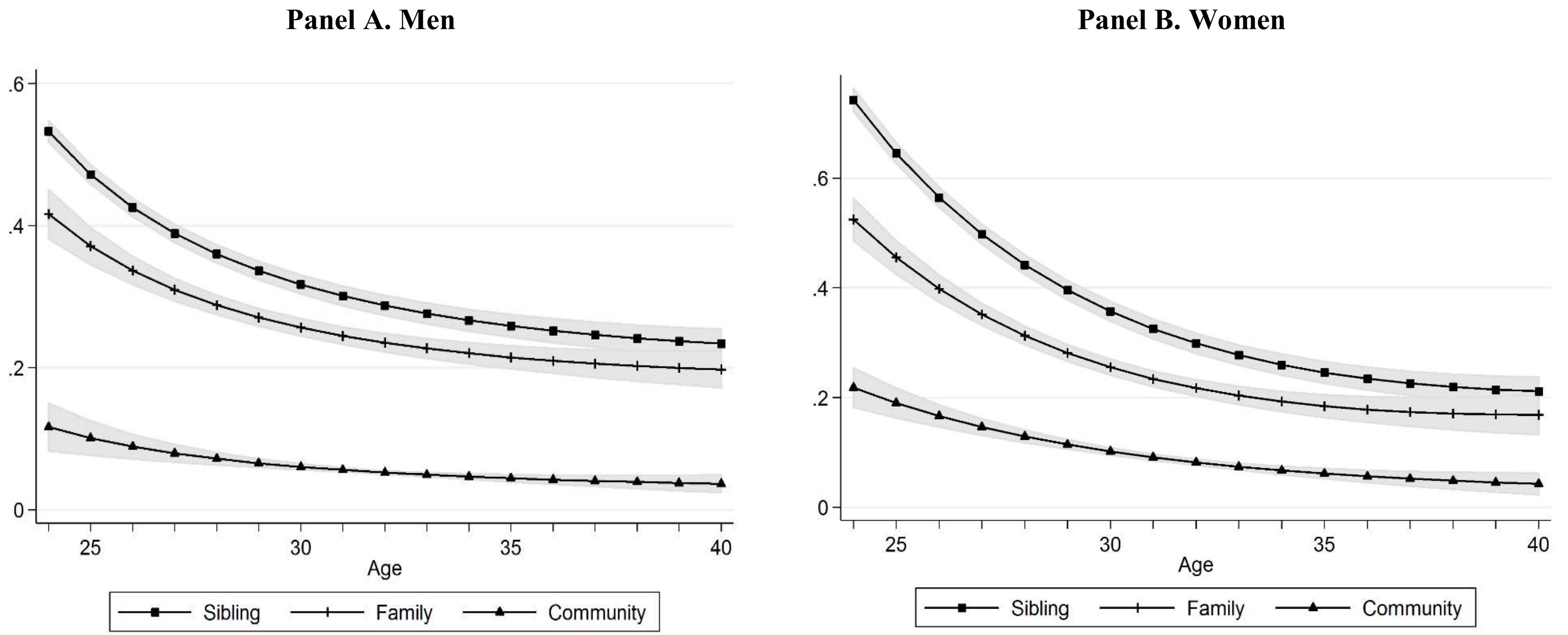


Panel B. Women



Note: The figure shows the predicted sibling correlation of earnings over the life cycle and its decomposition into family and community effects for immobile families for whom the parish of the first sibling remained the same between ages 11 and 15. The definition of community is based on sharing the neighborhood, the school or both community dimensions at age 15. Because for immobile families there is no between-sibling community variation, the sorting parameter cannot be estimated separately from community effects. Predictions are generated using the formulae provided in Section 4.5. The resulting sibling correlation is the sum of family and community effects.

Figure 10. Predicted sibling correlation of unemployment and factor decomposition



Note: The figure shows the predicted sibling correlation of unemployment over the life cycle and its decomposition into family and community effects using the estimates in Table 8 and Table 9, where community is defined as sharing the neighborhood, the school or both community dimensions at age 15. The sorting parameter is estimated using between-sibling variation of community comparing mobile to immobile families. Predictions are generated using the formulae provided in Section 4.5**.** The resulting sibling correlation is the sum of family and community effects, where we impute the estimated sorting parameter to the two factors in equal parts.

Table 1. Cohorts included in the sample.

(1) Birth cohorts	(2) First year observed	(3) # years observed	(4) Last age observed	(5) Earnings Observations		(6) Persons		(7) Schools	(8) Parishes
				Men	Women	Men	Women		
1965-67	1990	25	48	2,234,572	2,130,928	103,774	98,200	1,484	2,112
1968-70	1993	22	45	1,748,750	1,639,951	91,297	85,460	1,497	2,108
1971-73	1996	19	42	1,566,608	1,474,656	93,529	87,887	1,539	2,105
1974-76	1999	16	39	1,245,362	1,181,062	87,498	82,887	1,540	2,103
1977-79	2002	13	36	888,691	840,269	76,228	71,968	1,514	2,100
1980-82	2005	10	33	606,580	570,021	66,752	62,897	1,530	2,091
1983-85	2008	7	30	393,008	369,135	60,395	56,860	1,550	2,092
1965-85	1990-2008	7-25	30-48	8,683,571	8,206,022	579,473	546,159	1,821	2,123

Note: The table reports sample characteristics by birth cohort for men and women born 1965-1985. Schools are defined using school of enrollment at age 15; neighborhoods are defined using the parish of residence at age 15. Columns 7 and 8 report the number of schools and number of parishes within each group of male birth cohorts. Counts for females are very similar, totaling 1,789 schools and 2,124 parishes.

Table 2. Neighborhood and long run earnings – key study characteristics.

	(1+2) Page and Solon (2003a, b)	(3) Raaum, et.al. (2006)	(4) Oreopoulos (2003)	(5) Our study
Location	United States	Norway	Toronto, Canada	Denmark
Neighborhood	PSID cluster	Census tract	Housing project	Parish
Proximity	20-30 dwellings	44 km^2	20 buildings	20 km^2
#Clusters	120 (137)	7,996	81	41,748 (41,430)
#Men (women)	443 (516)	228,700 (195,889)	4,060	579,473 (546,149)
Persons/cluster	4 (4)	28 (24)	50	14 (13)
Others/cluster	86	430	1,036	2,671 (2,672)
Exposures				
Birth cohorts	1952-62	1946-65	1963-70	1965-1985
Years	1968	1960 and 1970	1978-86	1976-2002
Ages	6-16	5-15	8-16	11-15
Duration	snapshot	snapshot	1-9 years	1-6 years
Outcomes				
Measure	Earnings	Residual earn., education	Income	Residual earn., education, unemployment
Duration (years)	5	6	3	3-25 (mean 15)
Transformation	total mean	total mean	total mean	untransformed
Years observed	1987-91	1990-95	1997-99	Earn. 1990-2014 UIB 1990-2009
Ages observed	25-39	25-50	27-36	23-49

Table 3a. Parameter estimates of permanent earnings – shared components
(heterogeneous income profile – random growth)

	(1) Baseline		Baseline with alternative community definitions (2) Neighborhood-only		(3) School-only	
Panel A. Men	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
Variance of intercepts						
Family ($\sigma^2_{\mu\text{F}}$)	0.0758	0.0119	0.0756	0.0101	0.0786	0.0132
Community ($\sigma^2_{\mu\text{C}}$)	0.0163	0.0065	0.0112	0.0052	0.0053	0.0044
Variance of slopes						
Family ($\sigma^2_{\gamma\text{F}}$)	0.00054	0.00009	0.00053	0.00008	0.00062	0.00011
Community ($\sigma^2_{\gamma\text{C}}$)	0.00021	0.00005	0.00016	0.00004	0.00007	0.00003
Covariance intercepts-slopes						
Family ($\sigma^2_{\mu\gamma\text{F}}$)	-0.0051	0.0008	-0.0052	0.0007	-0.0061	0.0010
Community ($\sigma^2_{\mu\gamma\text{C}}$)	-0.0024	0.0005	-0.0018	0.0004	-0.0010	0.0004
Covariance between components						
Family-Community (σ_{FC})	0.0069	0.0022	0.0053	0.0018	0.0062	0.0017
Panel B. Women	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
Variance of intercepts						
Family ($\sigma^2_{\mu\text{F}}$)	0.0941	0.0149	0.0978	0.0127	0.0920	0.0112
Community ($\sigma^2_{\mu\text{C}}$)	0.0047	0.0081	-0.0010	0.0058	-0.0030	0.0043
Variance of slopes						
Family ($\sigma^2_{\gamma\text{F}}$)	0.00051	0.00009	0.00059	0.00008	0.00063	0.00008
Community ($\sigma^2_{\gamma\text{C}}$)	0.00020	0.00005	0.00011	0.00003	0.00000	0.00002
Covariance intercepts-slopes						
Family ($\sigma^2_{\mu\gamma\text{F}}$)	-0.0056	0.0009	-0.0062	0.0008	-0.0065	0.0008
Community ($\sigma^2_{\mu\gamma\text{C}}$)	-0.0020	0.0005	-0.0011	0.0004	-0.0002	0.0003
Covariance between components						
Family-Community (σ_{FC})	0.0109	0.0037	0.0087	0.0024	0.0062	0.0018

Note: The table reports Equally-Weighted Minimum Distance estimates for the parameters of the permanent component of the earnings process shared by siblings for men (Panel A) and for women (Panel B). In the baseline estimates of column (1) the community is defined as siblings sharing the neighborhood, the school or both community dimensions; in column (2) as sharing only the neighborhood regardless of school sharing, and in column (3) as sharing only the school regardless of neighborhood sharing. For all estimates there are two types of siblings: those who share the community (stayers) and those who do not share the community (movers). Estimates in the three columns in Panel A for men are derived using 14,012, 15,059 and 15,702 empirical variances and covariances (from left to right). Estimates in the three columns in Panel B for women are derived using 13,958, 14,540 and 15,696 empirical variances and covariances (from left to right).

Table 3b. Parameter estimates of permanent earnings - idiosyncratic components (restricted income profile-random walk)

			Baseline with alternative community definitions			
	(1) Baseline		(2) Neighborhood-only		(3) School-only	
Panel A. Men						
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
Initial condition (age 24)						
Brother 1 ($\sigma^2_{\omega 24,1}$)	0.0596	0.0089	0.0544	0.0072	0.0569	0.0095
Brother 2 ($\sigma^2_{\omega 24,2}$)	0.0337	0.0060	0.0304	0.0049	0.0324	0.0062
Variance of innovations						
Brother 1 ($\sigma^2_{\xi 1}$)	0.0081	0.0014	0.0075	0.0011	0.0067	0.0012
Brother 2 ($\sigma^2_{\xi 2}$)	0.0097	0.0016	0.0090	0.0013	0.0083	0.0015
Panel B. Women						
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
Initial condition (age 24)						
Sister 1 ($\sigma^2_{\omega 24,1}$)	0.0710	0.0094	0.0668	0.0080	0.0617	0.0071
Sister 2 ($\sigma^2_{\omega 24,2}$)	0.0376	0.0066	0.0352	0.0057	0.0332	0.0050
Variance of innovations						
Sister 1 ($\sigma^2_{\xi 1}$)	0.0117	0.0019	0.0104	0.0015	0.0092	0.0012
Sister 2 ($\sigma^2_{\xi 2}$)	0.0157	0.0024	0.0142	0.0019	0.0126	0.0016

Note: The table reports Equally-Weighted Minimum Distance estimates for the parameters of the permanent component of the earnings process which are sibling-specific for men (Panel A) and for women (Panel B). Community definitions and the number of empirical covariances and variances for each column are similar to the notes of Table 3a.

Table 4a. Parameter estimates of transitory earnings - men

	(1) Baseline		Baseline with alternative community definitions			
			(2) Neighborhood-only		(3) School-only	
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
	Initial condition (age 24)					
Brother 1 ($\sigma^2_{24,1}$)	0.7810	0.0299	0.7972	0.0246	0.7977	0.0312
Brother 2 ($\sigma^2_{24,2}$)	0.7976	0.0325	0.8147	0.0273	0.8166	0.0338
	Variance of innovations at age 25					
Brother 1 ($\sigma^2_{\varepsilon 1}$)	0.5710	0.0225	0.5863	0.0191	0.5950	0.0237
Brother 2 ($\sigma^2_{\varepsilon 2}$)	0.5534	0.0237	0.5666	0.0202	0.5754	0.0247
	Age splines in variance of innovations					
Brother 1						
26-28	-0.1380	0.0034	-0.1389	0.0034	-0.1368	0.0034
29-33	-0.1040	0.0027	-0.1039	0.0027	-0.1028	0.0027
34-38	-0.0337	0.0034	-0.0338	0.0035	-0.0333	0.0034
39-43	-0.0406	0.0064	-0.0386	0.0063	-0.0342	0.0062
44+	-0.0248	0.0099	-0.0234	0.0095	-0.0210	0.0090
Brother 2						
26-28	-0.1614	0.0063	-0.1625	0.0063	-0.1611	0.0064
29-33	-0.1172	0.0054	-0.1174	0.0054	-0.1165	0.0054
34-38	-0.0358	0.0076	-0.0351	0.0076	-0.0326	0.0076
39-43	-0.0528	0.0145	-0.0515	0.0144	-0.0477	0.0139
44+	0.0271	0.0490	0.0288	0.0482	0.0316	0.0453
	Autoregressive coefficient					
Brother 1 (ρ_1)	0.5000	0.0036	0.4979	0.0036	0.4895	0.0040
Brother 2 (ρ_2)	0.5239	0.0042	0.5238	0.0042	0.5175	0.0043
	Cross-person associations in transitory earnings					
Sibling covariance of innovations (σ_f)	0.0098	0.0018	0.0108	0.0020	0.0127	0.0021
Peers covariance of transitory earnings (mass point, λ)	-0.0020	0.0010	-0.0011	0.0009	0.0024	0.0010

Note: The table reports for men Equally-Weighted Minimum Distance estimates for the parameters of the transitory component of the earnings process. Community definitions for each column are similar to the notes of Table 3a. Estimates in the three columns are derived using 14,012, 15,059 and 15,702 empirical variances and covariances (from left to right).

Table 4b. Parameter estimates of transitory earnings – women

	(1) Baseline		Baseline with alternative community definitions (2) Neighborhood-only		(3) School-only	
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
	Initial condition (age 24)					
Sister 1 ($\sigma^2_{24,1}$)	0.7974	0.0309	0.8092	0.0262	0.8291	0.0234
Sister 2 ($\sigma^2_{24,2}$)	0.8013	0.0329	0.8141	0.0282	0.8347	0.0256
	Variance of innovations at age 25					
Sister 1 ($\sigma^2_{\varepsilon 1}$)	0.6599	0.0262	0.6727	0.0225	0.6914	0.0202
Sister 2 ($\sigma^2_{\varepsilon 2}$)	0.6572	0.0276	0.6710	0.0240	0.6894	0.0219
	Age splines in variance of innovations					
Sister 1						
26-28	-0.0883	0.0028	-0.0878	0.0028	-0.0876	0.0028
29-33	-0.1091	0.0021	-0.1083	0.0020	-0.1082	0.0020
34-38	-0.0513	0.0028	-0.0518	0.0028	-0.0526	0.0028
39-43	-0.0473	0.0049	-0.0476	0.0050	-0.0468	0.0049
44+	-0.0539	0.0092	-0.0504	0.0091	-0.0489	0.0089
Sister 2						
26-28	-0.1115	0.0054	-0.1112	0.0054	-0.1112	0.0054
29-33	-0.1210	0.0044	-0.1210	0.0044	-0.1214	0.0044
34-38	-0.0493	0.0066	-0.0494	0.0066	-0.0496	0.0065
39-43	-0.0548	0.0131	-0.0547	0.0131	-0.0530	0.0130
44+	-0.1955	0.0603	-0.1792	0.0585	-0.1776	0.0573
	Autoregressive coefficient					
Sister 1 (ρ_1)	0.4533	0.0022	0.4520	0.0021	0.4491	0.0021
Sister 2 (ρ_2)	0.4639	0.0031	0.4623	0.0031	0.4604	0.0031
	Cross-person associations in transitory earnings					
Sibling covariance of innovations (σ_f)	0.0085	0.0018	0.0067	0.0016	0.0094	0.0015
Peers covariance of transitory earnings (mass point, λ)	-0.0011	0.0007	0.0002	0.0007	0.0028	0.0006

Note: The table reports for women Equally-Weighted Minimum Distance estimates for the parameters of the transitory component of the earnings process. Community definitions for each column are similar to the notes of Table 3a. Estimates in the three columns are derived using 13,958, 14,540 and 15,696 empirical variances and covariances (from left to right).

Table 5a. Decomposition of sibling correlation of earnings into family and community effects – average over the life cycle and by age - Men

	(1) Baseline		Baseline with alternative community definitions: (2) Neighborhood-only		Baseline with alternative community definitions: (3) School-only		Alternative between-sibling community variation: (4) Mobile		Alternative between-sibling community variation: (5) Immobile	
Panel A. Average decomposition with earnings measured over the life cycle (age 24-45).										
	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.	Cor.	s.e.
Siblings (S)	0.313	0.014	0.308	0.015	0.297	0.015	0.309	0.024	0.310	0.013
Family (F)	0.289	0.017	0.291	0.015	0.289	0.015	0.277	0.020	0.271	0.013
Community (C)	0.023	0.012	0.017	0.011	0.008	0.010	0.032	0.014	0.038	0.001
Panel B. Average decomposition with earnings measured over parts of the life cycle (age 24 up to age 25, 27, 30, 35).										
	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.	Cor.	s.e.
Siblings										
Age 25	0.661	0.013	0.663	0.013	0.651	0.013	0.614	0.025	0.661	0.013
Age 27	0.544	0.024	0.544	0.012	0.533	0.012	0.513	0.023	0.544	0.012
Age 30	0.487	0.013	0.485	0.014	0.475	0.014	0.462	0.024	0.488	0.013
Age 35	0.355	0.012	0.352	0.013	0.341	0.013	0.344	0.022	0.354	0.012
Family										
Age 25	0.521	0.031	0.555	0.028	0.575	0.025	0.527	0.029	0.581	0.013
Age 27	0.437	0.026	0.462	0.023	0.473	0.020	0.445	0.024	0.477	0.012
Age 30	0.407	0.024	0.427	0.021	0.429	0.018	0.408	0.023	0.427	0.013
Age 35	0.313	0.018	0.323	0.015	0.315	0.014	0.309	0.018	0.308	0.012
Community										
Age 25	0.140	0.029	0.108	0.026	0.076	0.024	0.086	0.029	0.080	0.002
Age 27	0.107	0.012	0.082	0.022	0.060	0.020	0.068	0.025	0.067	0.002
Age 30	0.079	0.022	0.059	0.020	0.047	0.018	0.054	0.023	0.062	0.001
Age 35	0.041	0.016	0.029	0.014	0.025	0.013	0.034	0.017	0.046	0.001

Note: The table reports the predicted sibling correlation of labor earnings and its decomposition into family and community effects for men. Panel A shows the average decomposition over the life cycle (24-45). Panel B shows the predicted sibling correlation and its decomposition averaging up to the reported age. The decomposition in columns (1)-(3) refer to the parameters reported in Table 3a. The decomposition in column (4) is based on parameter estimates from mobile families with community defined as sharing only the neighborhood. The decomposition in column (5) is based on parameter estimates from immobile families; in this case the sorting parameter is not identified separately from community effects. Predictions are generated using the formulae provided in Section 4.5.

Table 5b. Decomposition of sibling correlation of earnings into family and community effects – average over the life cycle and by age - Women

			Baseline with alternative community definitions				Alternative between-sibling community variation			
	(1) Baseline		(2) Neighborhood-only		(3) School-only		(4) Mobile		(5) Immobile	
Panel A. Average decomposition with earnings measured over the life cycle (age 24-45).										
	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.	Cor.	s.e.
Siblings (S)	0.280	0.014	0.292	0.013	0.280	0.013	0.313	0.024	0.284	0.013
Family (F)	0.268	0.020	0.287	0.016	0.284	0.014	0.296	0.022	0.247	0.012
Community (C)	0.012	0.015	0.005	0.011	-0.004	0.009	0.017	0.012	0.037	0.001
Panel B. Average decomposition with earnings measured over parts of the life cycle (age 24 up to age 25, 27, 30, 35).										
	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.	Cor.	s.e.
Siblings										
Age 25	0.652	0.015	0.655	0.014	0.645	0.014	0.680	0.029	0.648	0.013
Age 27	0.538	0.013	0.541	0.012	0.531	0.012	0.576	0.027	0.534	0.012
Age 30	0.472	0.014	0.508	0.013	0.466	0.013	0.518	0.027	0.470	0.012
Age 35	0.343	0.012	0.352	0.013	0.332	0.013	0.389	0.024	0.341	0.011
Family										
Age 25	0.573	0.036	0.614	0.029	0.625	0.023	0.613	0.031	0.569	0.013
Age 27	0.482	0.030	0.513	0.024	0.515	0.019	0.521	0.027	0.466	0.012
Age 30	0.437	0.027	0.488	0.023	0.455	0.017	0.471	0.026	0.407	0.012
Age 35	0.328	0.021	0.344	0.017	0.332	0.013	0.358	0.021	0.293	0.011
Community										
Age 25	0.079	0.032	0.041	0.026	0.020	0.021	0.067	0.026	0.079	0.002
Age 27	0.057	0.027	0.028	0.022	0.016	0.018	0.055	0.022	0.067	0.002
Age 30	0.035	0.024	0.020	0.020	0.012	0.016	0.047	0.020	0.063	0.001
Age 35	0.015	0.019	0.005	0.014	0.005	0.011	0.032	0.016	0.048	0.001

Note: The table reports the predicted sibling correlation of labor earnings and its decomposition for women. Panel A shows the average decomposition over the life cycle (24-45). Panel B shows the predicted sibling correlation and its decomposition averaging up to the reported age. The decomposition in columns (1)-(3) refer to the parameters reported in Table 3b. The decomposition in column (4) is based on parameter estimates from mobile families with community defined as sharing only the neighborhood. The decomposition in column (5) is based on parameter estimates from immobile families; in this case the sorting parameter is not identified separately from community effects. Predictions are generated using the formulae provided in Section 4.5.

Table 6. Parameter estimates for educational attainment and decomposition of sibling correlation

Panel A – Parameter estimates

	(1) Men		(2) Women	
	Coef.	s.e.	Coef.	s.e.
Sibling 1 ($\sigma^2_{\omega 1}$)	3.3131	0.4071	2.8417	0.4538
Sibling 2 ($\sigma^2_{\omega 2}$)	3.1077	0.4446	2.6597	0.4846
Family ($\sigma^2_{\mu \mathrm{F}}$)	1.2759	0.1713	1.2433	0.1910
Community ($\sigma^2_{\mu \mathrm{C}}$)	0.3087	0.1151	0.2251	0.1283
Family-Community (σ_{FC})	0.0067	0.0948	0.0198	0.1057
Cohort 1969	1.1442	0.1680	1.1141	0.2133
Cohort 1972	1.1079	0.1641	1.1564	0.2191
Cohort 1975	1.0599	0.1590	1.1695	0.2209
Cohort 1978	0.9949	0.1523	1.1220	0.2144
Cohort 1981	0.9458	0.1473	1.1009	0.2115
Cohort 1984	0.5562	0.1126	0.7149	0.1628

Panel B
Decomposition of sibling correlation of educational attainment into family and community effects

	(1) Men		(2) Women	
	Cor.	s.e.	Cor.	s.e.
Siblings (S)	0.3323	0.0326	0.3541	0.0427
Family (F)	0.2667	0.0298	0.2966	0.0397
Community (C)	0.0656	0.0205	0.0575	0.0255

Note: The table reports Equally-Weighted Minimum Distance estimates for the parameters of the model for education attainment measured as years of education accumulated up to age 29 in Panel A and the predicted sibling correlation of education and its decomposition into family and community effects in Panel B. The community is defined as siblings sharing only the school regardless of neighborhood sharing. Estimates in each column are derived using 68 empirical variances and covariances.

Table 7. Parameter estimates of permanent component of the unemployment process

	(1) Men		(2) Women	
Panel A. Shared components (heterogeneous profile –random growth)				
	Coef.	s.e.	Coef.	s.e.
	Variance of intercepts			
Family ($\sigma^2_{\mu \mathrm{F}}$)	7.6557	0.4950	7.6919	0.6154
Community ($\sigma^2_{\mu \mathrm{C}}$)	1.5035	0.3612	2.6502	0.3464
	Variance of slopes			
Family ($\sigma^2_{\gamma \mathrm{F}}$)	0.0108	0.0066	0.0286	0.0067
Community ($\sigma^2_{\gamma \mathrm{C}}$)	0.0021	0.0044	0.0054	0.0041
	Covariance intercepts-slopes			
Family ($\sigma^2_{\mu\gamma \mathrm{F}}$)	0.0878	0.0456	-0.2432	0.0446
Community ($\sigma^2_{\mu\gamma \mathrm{C}}$))	-0.0098	0.0346	-0.0909	0.0310
	Covariance between components			
Family-Community (σ_{FC})	0.8866	0.0282	0.9340	0.0298
Panel B. Idiosyncratic components (restricted income profile-random walk)				
	Coef.	s.e.	Coef.	s.e.
	Initial condition (age 24)			
Brother 1 ($\sigma^2_{\omega 24,1}$)	9.3769	0.4851	3.3913	0.3704
Brother 2 ($\sigma^2_{\omega 24,2}$)	9.8203	0.5739	5.0941	0.4658
	Variance of innovations			
Brother 1 ($\sigma^2_{\xi 1}$)	2.8020	0.1619	2.0320	0.1664
Brother 2 ($\sigma^2_{\xi 2}$)	2.8574	0.1682	2.2104	0.1761

Note: The table reports Equally-Weighted Minimum Distance estimates for the parameters of the permanent component of the unemployment process for men in Column 1 and for women in Column 2. Panel A reports parameter estimates for the unemployment components shared by siblings, whereas Panel B reports parameter estimates for sibling-specific components. The community is defined as siblings sharing the neighborhood, the school or both community dimensions. Estimates in the two columns are derived using 6,933 and 6,830 empirical variances and covariances.

Table 8. Parameter estimates of transitory component of the unemployment process

	(1) Men		(2) Women	
	Coef.	s.e.	Coef.	s.e.
	Initial condition (age 24)			
Sibling 1 ($\sigma^2_{24,1}$)	21.6385	1.1246	21.1981	1.2230
Sibling 2 ($\sigma^2_{24,2}$)	22.4442	1.2752	21.6616	1.3189
	Variance of innovations at age 25			
Sibling 1 ($\sigma^2_{\varepsilon 1}$)	11.8458	0.6631	9.5050	0.5657
Sibling 2 ($\sigma^2_{\varepsilon 2}$)	12.8508	0.7611	10.6911	0.6588
	Age splines in variance of innovations			
Sibling 1				
26-28	-0.0030	0.0056	-0.0836	0.0048
29-33	0.0308	0.0030	0.0125	0.0024
34-38	-0.0167	0.0042	0.0556	0.0032
39+	-0.0214	0.0086	-0.0146	0.0083
Sibling 2				
26-28	-0.0368	0.0104	-0.1089	0.0083
29-33	0.0426	0.0066	0.0245	0.0053
34-38	-0.0264	0.0117	0.0194	0.0092
39+	0.0350	0.0533	0.0477	0.0658
	Autoregressive coefficient			
Sibling 1 (ρ_1)	0.5738	0.0045	0.5872	0.0041
Sibling 2 (ρ_2)	0.5552	0.0064	0.5409	0.0048
	Cross-person associations in transitory component			
Sibling covariance of Innovations (σ_f)	0.5727	0.0734	0.2931	0.0351
Peers covariance of transitory component (mass point, λ)	0.0550	0.0391	0.0286	0.0285

Note: The table reports Equally-Weighted Minimum Distance estimates for the parameters of the transitory component of the unemployment process for men in Column 1 and for women in Column 2. Community definitions and the number of empirical covariances and variances for each column are similar to the notes of Table 7.

Table 9. Decomposition of sibling correlation of unemployment into family and community effects – average over the life cycle and by age

	(1) Men		(2) Women	
Panel A. Average decomposition with unemployment measured over the life cycle (age 24-45).				
	Cor.	s.e.	Cor.	s.e.
Siblings (S)	0.2985	0.0079	0.3276	0.0099
Family (F)	0.2440	0.0087	0.2425	0.0113
Community (C)	0.0545	0.0044	0.0850	0.0061
Panel B. Average decomposition with unemployment measured over parts of the life cycle (age 24 up to age 25, 27, 30, 35).				
	Cor.	s.e.	Cor.	s.e.
Siblings				
Age 25	0.5021	0.0073	0.6937	0.0101
Age 27	0.4546	0.0067	0.6125	0.0094
Age 30	0.4046	0.0064	0.5207	0.0090
Age 35	0.3303	0.0059	0.4005	0.0080
Family				
Age 25	0.3935	0.0154	0.4898	0.0176
Age 27	0.3583	0.0120	0.4324	0.0142
Age 30	0.3212	0.0090	0.3684	0.0109
Age 35	0.2647	0.0066	0.2855	0.0078
Community				
Age 25	0.1086	0.0146	0.2039	0.0160
Age 27	0.0964	0.0111	0.1801	0.0124
Age 30	0.0833	0.0076	0.1522	0.0087
Age 35	0.0656	0.0046	0.1150	0.0053

Note: The table reports the predicted sibling correlation of unemployment and its decomposition into family and community effects for men in Column (1) and for women in Column (2). Panel A shows the average decomposition over the life cycle (24-40) and Panel B shows the predicted sibling correlation and its decomposition averaging up to the reported age. Predictions are generated using the formulae provided in Section 4.5. For comparability with earnings results, life-cycle averages are computed over the 24-45 age span using parameters estimated on the 24-40 span to predict correlations over the 41-45 interval.

Appendix: Moment restrictions for transitory earnings

Considering two (not necessarily different) age levels a and a', the intertemporal covariance structure of the transitory component of *individual* earnings from the birth order specific AR(1) process is as follows:

$$\begin{gathered} E\left(v_{ifca}v_{ifca'}\right) = [I(a = a' = 24)\sigma_{24s}^2 + \\ I(a = a' > 24)\left(\exp\left(g_s(a)\right) + var\left(u_{ifc(a-1)}\right)\rho_s^2\right) + \\ I(a \neq a')\left(E\left(u_{ifc(a-1)}u_{ifca'}\right)\rho_c\right)]\eta_t\eta_{t'} \ . \end{gathered} \tag{A.1}$$

Allowing for correlation of AR(1) innovations across siblings, the model yields restrictions on transitory earnings also for cross-siblings moments:

$$\begin{gathered} E\left(v_{ifca}v_{i'fc'a'}\right) = \\ \sigma_f\left(\frac{\left(1-\left(\rho_1\rho_2^{|t-t'|}\right)^P\right)}{1-\rho_1\rho_2^{|t-t'|}}\right)^{I(t\leq t')}\left(\frac{\left(1-\left(\rho_2\rho_1^{|t-t'|}\right)^P\right)}{1-\rho_2\rho_1^{|t-t'|}}\right)^{I(t>t')}\eta_t\eta_{t'}, \end{gathered} \tag{A.2}$$

where *P* is the number of overlapping years the two siblings are observed in the data.

We also model the correlation of transitory earnings across *non-sibling peers*. Differently from the case of siblings, we do not model the correlation of AR(1) innovations among peers because it would require distinguishing idiosyncratic components of transitory earnings for each member of school or neighborhood clusters, generating dimensionality issues. We, therefore, collapse all the cross-peers covariance structure of the transitory component into catch-all "mass point" factors absorbing all the parameters of the underlying stochastic process. For any two (not necessarily different) age levels a and a', covariances of transitory earnings across non-sibling peers are as follows:

$$E\left(v_{ifca}, v_{i'f'ca'}\right) = \lambda^{1+|t-t'|}\eta_t\eta_{t'} \tag{A.3}$$

The moment restrictions above characterize the inter-temporal distribution of transitory earnings for each individual and between siblings and peers. The orthogonality assumption between permanent and transitory earnings in equation (1) implies that moment restrictions of the full model are the sum of moment restrictions for permanent and transitory earnings, the former being discussed in Section 4.2 of the paper. In general, these restrictions are a non-linear function of a parameter vector θ. We estimate θ by Minimum Distance (see Chamberlain, 1984; Haider, 2001). We use Equally-Weighted Minimum Distance (EWMD) and a robust variance estimator $Var(\theta) = (G'G)^{-1}G'VG(G'G)^{-1}$, where V is the fourth moments matrix and G is the gradient matrix evaluated at the solution of the minimization problem.

Appendix Tables

Table A1a. Parameter estimates of time effects for earnings – Baseline model – Men

	Permanent Component (π_t)		Transitory Component (η_t)	
	Coef.	s.e.	Coef.	s.e.
t=				
1991	1.0314	0.0721	0.9825	0.0186
1992	1.0472	0.0856	1.0087	0.0230
1993	1.1210	0.0980	1.0535	0.0247
1994	1.0723	0.0923	1.0303	0.0245
1995	1.0629	0.0900	0.9653	0.0243
1996	1.1228	0.0908	0.9601	0.0221
1997	1.0606	0.0895	0.9590	0.0230
1998	1.0888	0.0901	0.9526	0.0235
1999	1.1075	0.0879	0.9797	0.0217
2000	1.1633	0.0930	0.9668	0.0226
2001	1.1258	0.0889	1.0172	0.0233
2002	1.2036	0.0930	1.0278	0.0223
2003	1.2073	0.0945	1.0869	0.0239
2004	1.1363	0.0892	1.0774	0.0237
2005	1.1469	0.0899	1.0493	0.0226
2006	1.0450	0.0832	1.0162	0.0224
2007	1.0315	0.0833	0.9965	0.0229
2008	0.9951	0.0807	1.0001	0.0224
2009	0.9982	0.0821	1.1160	0.0250
2010	1.0135	0.0835	1.1829	0.0267
2011	0.9756	0.0810	1.1938	0.0276
2012	0.9195	0.0769	1.2267	0.0284
2013	0.8811	0.0735	1.2487	0.0290
2014	0.8695	0.0718	1.2588	0.0285

Note: The table reports for men Equally-Weighted Minimum Distance estimates for the time shifters from the baseline model. Estimates are derived using 14,012 empirical variances and covariances.

Table A1b. Parameter estimates of time effects for earnings – Baseline model – Women

	Permanent Component (π_t)		Transitory Component (η_t)	
	Coef.	s.e.	Coef.	s.e.
t=				
1991	0.9162	0.0750	1.0204	0.0208
1992	0.8921	0.0748	1.0399	0.0224
1993	0.8733	0.0712	1.0604	0.0229
1994	0.9360	0.0765	1.1555	0.0259
1995	0.9603	0.0761	1.0269	0.0236
1996	0.8509	0.0684	1.1010	0.0238
1997	0.8919	0.0708	1.0186	0.0227
1998	0.8803	0.0681	1.0201	0.0226
1999	0.8859	0.0679	0.9999	0.0218
2000	0.9630	0.0727	1.0264	0.0224
2001	0.9323	0.0697	1.0421	0.0225
2002	0.8845	0.0659	1.0624	0.0223
2003	0.9266	0.0696	1.1953	0.0253
2004	0.8852	0.0657	1.0768	0.0227
2005	0.8647	0.0650	1.0711	0.0225
2006	0.8565	0.0641	1.0224	0.0217
2007	0.7933	0.0599	1.0051	0.0216
2008	0.7546	0.0568	0.9912	0.0209
2009	0.7276	0.0545	1.0373	0.0218
2010	0.7196	0.0536	1.0967	0.0230
2011	0.7183	0.0538	1.1109	0.0235
2012	0.6958	0.0519	1.1201	0.0238
2013	0.7061	0.0528	1.1327	0.0243
2014	0.7162	0.0539	1.1306	0.0245

Note: The table reports for women Equally-Weighted Minimum Distance estimates for the time shifters from the baseline model. Estimates are derived using 13,958 empirical variances and covariances.

Table A2a. Parameter estimates of permanent earnings for specifications based on alternative between-sibling-community-variation – Men

	(1) Mobile		(2) Immobile	
Panel A. Shared components (heterogeneous income profile –random growth)				
	Coef.	s.e.	Coef.	s.e.
	Variance of intercepts			
Family ($\sigma^2_{\mu\mathrm{F}}$)	0.0726	0.0117	0.0979	0.0139
Community ($\sigma^2_{\mu\mathrm{C}}$)	0.011	0.0057	0.0134	0.0018
	Variance of slopes			
Family ($\sigma^2_{\gamma\mathrm{F}}$)	0.0005	0.0001	0.0007	0.0001
Community ($\sigma^2_{\gamma\mathrm{C}}$)	0.0001	0.00003	0.0001	0.00001
	Covariance intercepts-slopes			
Family ($\sigma^2_{\mu\gamma\mathrm{F}}$)	-0.0046	0.0008	-0.0072	0.001
Community ($\sigma^2_{\mu\gamma\mathrm{C}}$)	-0.0012	0.0004	-0.0009	0.0001
	Covariance between components			
Family-Community (σ_{FC})	0.0015	0.002	Not Identified	
Panel B. Idiosyncratic components (restricted income profile-random walk)				
	Initial condition (age 24)			
Brother 1 ($\sigma^2_{\omega 24,1}$)	0.0507	0.0088	0.0626	0.0088
Brother 2 ($\sigma^2_{\omega 24,2}$)	0.0453	0.0091	0.036	0.006
	Variance of innovations			
Brother 1 ($\sigma^2_{\xi 1}$)	0.0065	0.0012	0.0079	0.0012
Brother 2 ($\sigma^2_{\xi 2}$)	0.0068	0.0013	0.0094	0.0014

Note: The table reports for men Equally-Weighted Minimum Distance estimates for the parameters of the permanent component of the earnings process for mobile and immobile families corresponding to the decompositions presented in Table 5a, columns (4) and (5), respectively.

Table A2b. Parameter estimates of transitory earnings for specifications based on alternative between-sibling-community-variation – Men

	(1) Mobile		(2) Immobile	
	Coef.	s.e.	Coef.	s.e.
	Initial condition (age 24)			
Sibling 1 ($\sigma^2_{24,1}$)	0.6072	0.0223	0.5701	0.0215
Sibling 2 ($\sigma^2_{24,2}$)	0.5967	0.0269	0.5538	0.0226
	Variance of innovations at age 25			
Sibling 1 ($\sigma^2_{\varepsilon 1}$)	0.4933	0.0060	0.4938	0.0034
Sibling 2 ($\sigma^2_{\varepsilon 2}$)	0.5069	0.0080	0.5179	0.0041
	Age splines in variance of innovations			
Sibling 1				
26-28	-0.1324	0.0041	-0.1345	0.0033
29-33	-0.0981	0.0030	-0.1027	0.0026
34-38	-0.0349	0.0039	-0.0339	0.0033
39-43	-0.0442	0.0075	-0.0422	0.0057
44+	-0.0270	0.0118	-0.0318	0.0093
Sibling 2				
26-28	-0.1349	0.0125	-0.1579	0.0062
29-33	-0.0865	0.0097	-0.1162	0.0054
34-38	-0.0519	0.0132	-0.0358	0.0075
39-43	-0.0408	0.0238	-0.0562	0.0143
44+	-0.0058	0.0793	0.0198	0.0491
	Autoregressive coefficient			
Sibling 1 (ρ_1)	0.8191	0.0274	0.7719	0.0291
Sibling 2 (ρ_2)	0.7938	0.0325	0.7884	0.0314
	Cross-person associations in transitory earnings			
Sibling covariance of Innovations (σ_f)	0.0154	0.0031	0.0105	0.0017
Peers covariance of transitory earnings (mass point, λ)	-0.0002	0.0011	0.0019	0.0001

Note: The table reports for men Equally-Weighted Minimum Distance estimates for the parameters of the transitory component of the earnings process for mobile and immobile families corresponding to the decompositions presented in Table 5a, columns (4) and (5), respectively.

Table A2c. Parameter estimates of permanent earnings for specifications based on alternative between-sibling-community-variation – Women

	(1) Mobile		(2) Immobile	
Panel A. Shared components (heterogeneous income profile –random growth)				
	Coef.	s.e.	Coef.	s.e.
	Variance of intercepts			
Family ($\sigma^2_{\mu\mathrm{F}}$)	0.0978	0.0127	0.0945	0.0078
Community ($\sigma^2_{\mu\mathrm{C}}$)	-0.0010	0.0058	0.0129	0.0010
	Variance of slopes			
Family ($\sigma^2_{\gamma\mathrm{F}}$)	0.0006	0.0001	0.0006	0.0001
Community ($\sigma^2_{\gamma\mathrm{C}}$)	0.0001	0.0000	0.0000	0.0000
	Covariance intercepts-slopes			
Family ($\sigma^2_{\mu\gamma\mathrm{F}}$)	-0.0062	0.0008	-0.0065	0.0005
Community ($\sigma^2_{\mu\gamma\mathrm{C}}$)	-0.0011	0.0004	-0.0006	0.0000
	Covariance between components			
Family-Community (σ_{FC})	0.0087	0.0024	Not Identified	
Panel B. Idiosyncratic components (restricted income profile-random walk)				
	Initial condition (age 24)			
Brother 1 ($\sigma^2_{\omega 24,1}$)	0.0668	0.0080	0.0645	0.0054
Brother 2 ($\sigma^2_{\omega 24,2}$)	0.0352	0.0057	0.0346	0.0043
	Variance of innovations			
Brother 1 ($\sigma^2_{\xi 1}$)	0.0104	0.0015	0.0096	0.0010
Brother 2 ($\sigma^2_{\xi 2}$)	0.0142	0.0019	0.0132	0.0013

Note: The table reports for women Equally-Weighted Minimum Distance estimates for the parameters of the permanent component of the earnings process for mobile and immobile families corresponding to the decompositions presented in Table 5b, columns (4) and (5), respectively.

Table A2d. Parameter estimates of transitory earnings for specifications based on alternative between-sibling-community-variation – Women

	(1) Mobile		(2) Immobile	
	Coef.	s.e.	Coef.	s.e.
	Initial condition (age 24)			
Sibling 1 ($\sigma^2_{24,1}$)	0.6727	0.0225	0.6823	0.0160
Sibling 2 ($\sigma^2_{24,2}$)	0.6710	0.0240	0.6805	0.0176
	Variance of innovations at age 25			
Sibling 1 ($\sigma^2_{\varepsilon 1}$)	0.4520	0.0021	0.4498	0.0020
Sibling 2 ($\sigma^2_{\varepsilon 2}$)	0.4623	0.0031	0.4605	0.0031
	Age splines in variance of innovations			
Sibling 1				
26-28	-0.0878	0.0028	-0.0868	0.0028
29-33	-0.1083	0.0020	-0.1081	0.0020
34-38	-0.0518	0.0028	-0.0522	0.0027
39-43	-0.0476	0.0050	-0.0477	0.0047
44+	-0.0504	0.0091	-0.0520	0.0088
Sibling 2				
26-28	-0.1112	0.0054	-0.1102	0.0054
29-33	-0.1210	0.0044	-0.1211	0.0044
34-38	-0.0494	0.0066	-0.0496	0.0065
39-43	-0.0547	0.0131	-0.0549	0.0130
44+	-0.1792	0.0585	-0.1848	0.0584
	Autoregressive coefficient			
Sibling 1 (ρ_1)	0.8092	0.0262	0.8194	0.0180
Sibling 2 (ρ_2)	0.8141	0.0282	0.8253	0.0204
	Cross-person associations in transitory earnings			
Sibling covariance of Innovations (σ_f)	0.0067	0.0016	0.0091	0.0015
Peers covariance of transitory earnings (mass point, λ)	0.0002	0.0007	0.0018	0.0001

Note: The table reports for women Equally-Weighted Minimum Distance estimates for the parameters of the transitory component of the earnings process for mobile and immobile families corresponding to the decompositions presented in Table 5b, columns (4) and (5), respectively.

Table A3a. Community correlation of earnings – sensitivity to measurement of community affiliation by level and age intervals – Men

Panel A. Average community correlation of earnings for community measured at various age levels.

	Age 11		**Age 12**		**Age 13**		**Age 14**		**Age 15**	
	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.	Cor.	s.e.
Earnings measured over ages 24-45	0.057	0.001	0.057	0.001	0.057	0.001	0.058	0.001	0.058	0.001
Earnings measured up to										
Age 25	0.209	0.003	0.211	0.003	0.210	0.003	0.212	0.003	0.213	0.003
Age 27	0.177	0.002	0.180	0.002	0.178	0.002	0.180	0.002	0.181	0.002
Age 30	0.132	0.002	0.134	0.002	0.133	0.002	0.134	0.002	0.135	0.002
Age 35	0.075	0.001	0.076	0.001	0.075	0.001	0.076	0.001	0.077	0.001

Panel B. Average community correlation of earnings for community measured over various age intervals.

	Age 11-15		**Age 12-15**		**Age 13-15**		**Age 14-15**	
	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.
Earnings measured over ages 24-45	0.057	0.001	0.057	0.001	0.057	0.001	0.058	0.001
Earnings measured up to								
Age 25	0.214	0.003	0.215	0.003	0.215	0.003	0.214	0.003
Age 27	0.182	0.003	0.183	0.002	0.182	0.002	0.182	0.002
Age 30	0.135	0.002	0.136	0.002	0.136	0.002	0.135	0.002
Age 35	0.076	0.001	0.077	0.001	0.077	0.001	0.077	0.001

Note: The table reports for men the estimated average community correlation of earnings over the life cycle and for different segments of the life cycle from the community-only model in which we allow for community as the only factor determining permanent earnings. Community peers are defined as parish male neighbors born in the same year and residing in the parish at the given age level or age interval.

Table A3b. Community correlation of earnings – sensitivity to measurement of community affiliation by level and age intervals – Women

Panel A. Average community correlation of earnings for community defined at various ages.

	Age 11		**Age 12**		**Age 13**		**Age 14**		**Age 15**	
Earnings measured	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.	Cor.	s.e.
over ages 24-45	0.0660	0.0010	0.0659	0.0010	0.0666	0.0010	0.0664	0.0010	0.0666	0.0010
Earnings measured up to										
Age 25	0.2392	0.0024	0.2395	0.0024	0.2399	0.0024	0.2393	0.0024	0.2399	0.0024
Age 27	0.2072	0.0021	0.2075	0.0021	0.2080	0.0021	0.2074	0.0021	0.2080	0.0021
Age 30	0.1597	0.0017	0.1599	0.0017	0.1604	0.0017	0.1600	0.0017	0.1605	0.0017
Age 35	0.0945	0.0011	0.0946	0.0011	0.0952	0.0011	0.0950	0.0011	0.0953	0.0011

Panel B. Average community correlation of earnings for community defined over various age ranges.

	Age 11-15		**Age 12-15**		**Age 13-15**		**Age 14-15**	
Earnings measured	Cor.	s.e.	Cor.	s.e	Cor.	s.e.	Cor.	s.e.
over ages 24-45	0.0666	0.0010	0.0664	0.0010	0.0666	0.0010	0.0667	0.0010
Earnings measured up to								
Age 25	0.2416	0.0025	0.2404	0.0025	0.2408	0.0025	0.2402	0.0024
Age 27	0.2094	0.0022	0.2084	0.0022	0.2087	0.0022	0.2082	0.0022
Age 30	0.1616	0.0018	0.1607	0.0018	0.1611	0.0017	0.1606	0.0017
Age 35	0.0959	0.0012	0.0954	0.0012	0.0956	0.0012	0.0954	0.0011

Note: The table reports for women the estimated average community correlation of earnings over the life cycle and for different segments of the life cycle from the community-only model in which we allow for community as the only factor determining permanent earnings. Community peers are defined as female parish neighbors born in the same year and residing in the parish at the given age level or age interval.

Table A4. Decomposition of percentile rank correlation

	Men		Women	
	Panel A			
	Cor.	s.e.	Cor.	s.e.
Idiosyncratic	0.3757	0.0033	0.3270	0.0031
Family	0.0963	0.0031	0.0781	0.0029
Community	0.0142	0.0079	0.0058	0.0075
	Panel B			
Sibling	0.1105	0.0038	0.0839	0.0037
Number of moments	13,813		13,759	

Notes: Panel A reports estimated coefficients from regression of earnings percentile correlations on dummies for whether the correlation refers to individual earnings over time, or to pairs of observations sharing family, community or both. The regression does not include a constant; it includes dummies for calendar time and birth cohort, and controls for the time lag over which the correlation is computed. Adjusted S.E via EWMD are reported. Panel B reports the implied rank correlation for two siblings sharing both the family and the community.

Table A5a. Parameter estimates of time effects for unemployment - Men

	Permanent Component (π_t)		Transitory Component (η_t)	
	Coef.	s.e.	Coef.	s.e.
t=				
1991	0.8826	0.0249	1.1550	0.0352
1992	0.8284	0.0227	1.2091	0.0359
1993	0.7902	0.0194	1.1110	0.0296
1994	0.7817	0.0197	1.1969	0.0330
1995	0.7787	0.0198	1.1937	0.0341
1996	0.7957	0.0194	1.1655	0.0317
1997	0.7893	0.0200	1.2641	0.0355
1998	0.7604	0.0196	1.2519	0.0362
1999	0.8169	0.0208	1.2904	0.0362
2000	0.8113	0.0216	1.3745	0.0399
2001	0.8094	0.0218	1.3786	0.0412
2002	0.8062	0.0215	1.4148	0.0402
2003	0.8010	0.0215	1.4116	0.0410
2004	0.8081	0.0219	1.4096	0.0420
2005	0.7964	0.0218	1.4134	0.0426
2006	0.7765	0.0213	1.4265	0.0434
2007	0.7649	0.0209	1.4416	0.0444
2008	0.7778	0.0211	1.5714	0.0491
2009	0.6671	0.0179	1.6191	0.0481

Note: The table reports for men Equally-Weighted Minimum Distance estimates for the time shifters from the baseline model. Estimates in the three columns are derived using 6,933 empirical variances and covariances.

Table A5b. Parameter estimates of time effects for unemployment - Women

	Permanent Component (π_t)		Transitory Component (η_t)	
	Coef.	s.e.	Coef.	s.e.
t=				
1991	0.9152	0.0407	1.1517	0.0420
1992	0.8271	0.0361	1.2408	0.0431
1993	0.8489	0.0322	1.0841	0.0337
1994	0.7691	0.0297	1.2357	0.0383
1995	0.7171	0.0279	1.2887	0.0402
1996	0.6967	0.0263	1.2452	0.0374
1997	0.6496	0.0249	1.3485	0.0407
1998	0.6278	0.0245	1.3808	0.0421
1999	0.6388	0.0244	1.3991	0.0418
2000	0.6314	0.0245	1.5013	0.0452
2001	0.6361	0.0249	1.5142	0.0460
2002	0.6531	0.0255	1.5278	0.0459
2003	0.6568	0.0256	1.5613	0.0470
2004	0.6365	0.0251	1.5568	0.0471
2005	0.6097	0.0243	1.5539	0.0471
2006	0.5230	0.0209	1.7099	0.0510
2007	0.5050	0.0204		
2008	0.9323	0.0352		
2009	0.9488	0.0356		

Note: The table reports for women Equally-Weighted Minimum Distance estimates for the time shifters from the baseline model. Estimates in the three columns are derived using 6,830 empirical variances and covariances.