Opportunities, Preferences and Incomes

by William Nilsson[1]
Department of Economics, Umeå University

The distinction between circumstances that constrain an individual's opportunities and the individual choices also affecting a particular outcome is the main idea of theories of equality of opportunity. In this study, equality of opportunity is analyzed for Swedish data using a large set of variables indicating different circumstances likely to affect an individual's opportunities. A semiparametric model is estimated to allow for a possible nonlinear relation between parental income and the income of the adult child. The reason is a hypothesis that a constrained investment behavior would make the relationship nonlinear. The results indicate significant inequality of opportunities. However, they do not indicate a nonlinear relationship between parental income and the income of the adult child. Thus, the hypothesis that low income families will have a constrained investment behavior in human capital formation is brought into question as the explanation of intergenerational income correlation in Sweden.

JEL: J62, D63, C14
Keywords: equality of opportunity, intergenerational income mobility, preferences, tastes, semiparametric estimation

[1]I wish to thank seminar participants at the Department of Economics, Umeå University, Sweden, and CEMFI, Madrid, Spain, for valuable comments. In particular I would like to thank Karl-Gustaf Löfgren, Xavier de Luna, Kurt Brännäs, Anders Stenberg, Magnus Wikström, Erik Norlin, Samuel Bentolila and Manuel Arellano. I also want to thank Chris Hudson for improving the language. Financial support from Wallander-Hedelius-Browaldh Foundation, Kempe Foundation and Swedish Council for Working Life and Social Research are gratefully acknowledged. To contact the author, use william.nilsson@econ.umu.se.

1. Introduction

Sweden has a history of ambitions to build institutions that grant access to education for all regardless of income. In this spirit, the higher education system is free from tuition fees and there is a system of student grants and government loans. The main idea has been to reduce the importance of the individual's social background for his/her position later in life. The concept of equality of opportunity makes an important distinction between factors that the individual is unable to influence and consequences arising as a result of choices made by the individual. For example, children cannot choose their parents' income or the average income in the neighborhood where they live. However, such exogenous circumstances could have a constraining effect on their opportunities for earning a high income when they are adults. At the same time, the choices that individuals make can also influence their income later in life. Theories of equality of opportunity are different from theories of equality in that they include the preferences of the individual. These preferences could, for example, influence choices concerning education, migration and labor supply.

It is not possible to observe all choices or circumstances that might affect the individual's income as an adult. One approach to deal with this problem is to identify what are likely to be the most important factors influencing future income, and then let all the income differences that are unexplained by these, to be labelled as effort. Effort is, by assumption, a result of choices based on preferences. A dilemma with this way of analyzing is that, if too many background factors are included, it would mean that very few individuals would have experienced the same circumstances. On the other hand, if too few circumstances are included, this would leave unexplained income differences that would be labelled as effort.

The purpose of this study is to evaluate whether a set of circumstances that the individual is unable to influence are important for his/her income as an adult.[2] The intention is to keep the analysis as close to the concept of equality of opportunity as possible.

The theory of equality of opportunity, developed by Roemer (1998, 2002), provides an analytical framework within which it is possible to compare how different circumstances constrain the opportunities open to the individual when the degree of effort is kept constant. Most of the empirical literature has analyzed one particular factor, the income of the father, and has not focused on keeping the amount of effort exerted by the individual constant. Studies of the correlation between the father's income and that of the adult son are common.[3] The purpose of these studies is to examine whether children from poor families continue to have a low income as adult. The usual result from these studies is that the parents' income is positively correlated with the income of the adult child. The magnitude of the elasticity varies between different studies and countries.[4] Österberg (2000) finds, using a Swedish data set, an elasticity of between 0.125-0.185, which is relatively low compared with studies for other countries. Another finding is that the intergenerational elasticity may vary in accordance with different parental income levels. Becker & Tomes (1986) suggest that, if low income families are constrained to borrow for investments in human capital for their children, the correlation should be higher for these families than for families possessing the financial resources to invest optimally. The reason is that increased investments in

[2]From a practical point of view, these circumstances have to be measurable and available in order to be possible to analyze. Note that those characteristics influencing income, which the individual can to some extent choose, such as education, are not important in the analyses.
[3]See Solon, G. (1999) for an excellent survey.
[4]See Solon, G. (1999) and Björklund & Jäntti (2000) for an overview of different studies and see Björklund & Jäntti (1997) for a comparison of intergenerational income correlation in Sweden and in the United States.

human capital would lower the marginal rates of return, which would eventually reach the market rate on assets. Further increases in income would then be invested in assets and given as bequests to the children, rather than invested in human capital.

If education is important for income, the Swedish aim of reducing inequality of opportunity through access to higher education may reduce, or eliminate, the constraint to invest optimally for low income families. If no nonlinearities are found, the potential effects of income would not seem to work through a constrained investment decision. The income level can, however, still be important as the correlation may not necessarily be caused by constrained investment behaviour.

This article contributes to the literature on equality of opportunity in the following ways. Firstly, a semiparametric model is estimated which does not specify the functional form of the parents' income, and the elasticity is allowed to vary over their income. Secondly, several factors, not just the fathers' income, are included in the analysis. These relate the empirical analysis to the theory of equality of opportunity in a more consistent manner than has been typically done in the literature. The model also handles the dilemma that arises when too few individuals face the same conditions when several explanatory variables are included. Individuals are compared as if they faced the same circumstances except with regard to the dimension that is analyzed in an equality of opportunity-framework. The results indicate connections between the circumstances analyzed and the income of the adult child, and equality of opportunity clearly does not exist. It should, however, be noted that the included variables explain a very small part of the variation in income. The results also indicate that no specific nonlinearities, that would support a theory of constrained investment behavior for low income families in Sweden, are present.

The remainder of the paper is organized as follows. Section two describes the theoretical framework for equality of opportunity. The data set is described in section three. Section four illustrates the econometric considerations and introduces the semiparametric model. The results are presented in section five and section six consists of some concluding remarks.

2. Theoretical framework

This study is based on the theory of equality of opportunity described in Roemer's book Equality of Opportunity (1998).[5] The theory is built on a few key concepts; circumstances, type and effort, each of which will be presented below.

Circumstances are characteristics that the individual cannot influence, and therefore, cannot be held responsible for. They could refer to characteristics of the family and the neighborhood in which the individual grew up, as well as personal traits present from birth. The latter include, for example, gender and other possible genetic traits. Family characteristics that would support human capital formation might, for example, be parental income and the time parents spent with the child, in particular helping with homework. Parents' incomes could, for example, facilitate the formation of human capital through the purchase of books, computers or influence the quality of schooling the child receives. These family characteristics may also work through preferences. For example, time spent by parents in being supportive of the child's school work is likely to encourage his/her preference for education.

[5]The theory is also described in several different papers such as O'Neill *et al.* (2000) and Roemer (2002). The theory described by Roemer is not the only, nor the first theory of equality of opportunity. The theory is, however, suitable to base an empirical analysis on.

Theoretically, parental income could affect preferences where the standard of living in childhood is a reference point for adulthood. A luxurious life in childhood could stimulate a demand for a luxurious life later, which would require a high income. In a similar fashion, growing up in a low income family could induce a different attitude towards the importance of earning a lot of money. Occupational choice could, for example, be based on other values. Neighborhood characteristics that might influence preferences could work through peer pressure, role-models and social norms. Different neighborhoods could also have schools of differing quality that affect human capital formation.

The circumstances in which the child grows up can, potentially, affect the ability to earn a high income in several ways. This means it is, in general, difficult to distinguish between, for example, the direct effects of parental income on the individual's opportunities and the possible indirect effects of that income on preferences. One way to approach this problem is to view these kinds of attitudes as a result of inequality of opportunity. It is, however, not obvious that income differences due to different preferences, following from different circumstances, should be viewed as a state of inequality of opportunity. Dworkin (1981), for example, argues that differences based on preferences, with which the individual identifies and is satisfied, should be ethically acceptable. Taking this view, it becomes important to separate the effects of those circumstances that work through preferences from the more direct effects, such as investment in human capital formation.

Individuals can be classified into different categories, i.e. types, with respect to their circumstances, each with their own income distribution. If inequality of opportunity exists, then incomes and, ultimately, income distributions will be affected by different opportunities. Those types with more favourable circumstances will have distributions covering higher

incomes. The individual's placement in the income distribution for his/her type will be the result of a large number of unobserved life events. These events could be based on preferences, luck, etc. and, in the equality of opportunity-framework they are summarized into a "catch-all" term called effort. The income distribution can, accordingly, be seen as a mapping of a space of opportunities, hereafter called an indirect opportunity set, in which the individual can choose his/her effort and be rewarded with a particular income. The opportunity set is indirect because individuals do not choose their income, instead they decide the amount of effort they will put in and the type of education they will pursue. It is important to remember that effort, defined in this way, is also a result of environmental characteristics that have not been included in the list of circumstances. Controlling for only a few circumstances will naturally be too conservative as other relevant factors may possibly be left out. If this is the case, the results are likely to indicate that the differences in income are within the individual's scope of choice, even though this is not entirely true. On the other hand, including every event in life that is likely to affect income in adulthood would, however, change the focus of the analysis from equality of opportunity to just equality. The reason is that every individual would then constitute a unique type, and consequently all rankings would collapse. More specifically, all the differences in income would, ultimately, be the result of circumstances.

As explained previously, where an individual ends up within the income distribution of a type is the result of differences in effort. Different individuals are assumed to have used the same degree of effort if they end up in the same percentile of their respective type's distribution. This is an important assumption as it enables us, at least in theory, to keep the degree of effort constant while changing the circumstances. It allows us to investigate the extent to which unfortunate circumstances constrain the opportunities open to the individual.

2.1 Model

The theory utilized in this paper is based on Roemer (2002) and O'Neill *et al*. (2000). Within the equality of opportunity-framework, the focus in this study is on income.[6] The income of individual *i* is assumed to be determined by the following function:

$$y_i = y(e_i, C_{t(i)}) = y_{t(i)}(e_i) \tag{1}$$

where y_i is the income of individual *i, i = 1,…,n* and e_i is the value of relative effort for individual *i*. $C_{t(i)}$ is a vector of circumstances corresponding to the type to which *i* belongs. The income distribution of *i*'s type is labelled $G_{t(i)}$. $F_{t(i)}(e_{t(i)})$ is the distribution of effort in the type to which individual *i* belongs, where $e_{t(i)}$ is absolute effort. The difference between relative and absolute effort is explained below. Note that where *i* is within parentheses, this indicates that *i* belongs to that type, although *i* is, of course, not the only individual in that type. As every type is assumed to consist of a large number of individuals, a single individual cannot influence the distribution of effort within its type. The distribution is, therefore, a characteristic of *i*'s type in the same way as are the circumstances.

Further, $y_i = y(e_i, C_{t(i)})$ can be written as $y_{t(i)}(e_i)$, and y is strictly increasing in the relative effort, e_i. The absolute effort, $e_{t(i)} = e(v_1, v_2, \ldots, v_m)$ is a function of a finite number of measurable but, for the

[6]In the study, both labor income and disposable income are used as measure of income. Using labor income captures the opportunities to earn a high labor income, while disposable income is closer to capturing the opportunities to obtain a certain living standard.

researcher, unobservable events, $v_1, v_2, \ldots, v_m$. The efforts are assumed to be summarized in a scalar. Note that the effort distributions can be different for different types. For some types it is possible that the circumstances *together with* the unobservable events, $v_1, v_2, ..., v_m$, make the effort distribution to cover fairly low levels of absolute effort. In the same way it is possible that different circumstances influence the absolute effort positively. Roemer makes, in his theory, a subjective valuation that to compare the absolute effort would not be appropriate as certain circumstances have made it easier to exert a high level of absolute effort.[7] For this reason it is necessary to find a relative measure of effort. To proceed from here, it is, accordingly, necessary to find some way to identify the individual's relative effort, e_i, without knowing the unobservable events, $v_1, v_2, \ldots, v_m$. The Roemer Identification Axiom serves this purpose and works in the following manner.

Individuals in different types are assumed, as a definition, to have the same relative effort if they are at the same quantile in their types' effort distribution. For an individual with an absolute effort of, for example, $e^*_{t(i)}$, it is possible to find the corresponding relative effort by looking in the effort distribution, $F_{t(i)}(e_{t(i)})$, of its type. Accordingly, $F_{t(i)}(e_{t(i)}) = e_i$, where $e_i \in [0, 1]$ and is the quantile within the effort distribution. Thus, it is possible that the relative effort, $F_{t(i)}(e_{t(i)}) = e_i$, equals a relative effort in another type, for example, $F_{t(j)}(e_{t(j)}) = e_j$, even though the absolute efforts, $e_{t(i)}$ and $e_{t(j)}$ are different.

---

[7]Dworkin would probably prefer to write $y_i = y(e_{t(i)}, C_{t(i)})$, because then the absolute effort, i.e. all the preferences, would be the part that individuals would be responsible for. However, the absolute effort is influenced by the circumstances and Roemer does not want to keep individuals responsible for preferences that are based on circumstances that the individual cannot influence. Hence, the relevant effort is e_i and $C_{t(i)}$ can influence the income both directly and indirectly by influencing preferences.

This means that, for individual i and individual j who are in different types,

$$\pi_{t(i)} = G_{t(i)}(e_i) = F_{t(i)}(e_{t(i)}) = F_{t(j)}(e_{t(j)}) = G_{t(j)}(e_j) = \pi_{t(j)} \quad (2)$$

$$\text{iff } e_i = e_j$$

Ending up in the π-th quantile in the income distribution is equivalent to ending up in the π-th quantile in the distribution of effort. The reason is that only effort is assumed to influence the income when the circumstances are the same. Thus, individuals in the same quantile in the income distribution for their type, have the same level of relative effort, even though they may have different incomes. As assumed, once the circumstances are controlled for, the rank within the income distribution of a type can be used to identify the rank within the effort distribution. This means that an individual's effort is identified through the income differences that are left when all circumstances have been controlled for.

With this foundation, the next step is to define the indirect opportunity set. This is an illustration of the reachable outcomes for individuals with equal opportunities, depending on the level of effort they make. Let the rank within the income distribution for the type that i belongs to be,

$$\pi_{t(i)} = G_{t(i)}(y)$$

Monotonicity enables us to write,

$$G_{t(i)}^{-1}[\pi_{t(i)}] = y_{t(i)} = y[e_i, C_{t(i)}], \tag{3}$$

which is the indirect opportunity set. By varying the amount of relative effort, (e_i or π), it is possible to influence the income. The opportunity set is indirect as it is not possible for the individual to directly choose π. Income is affected by a large number of life events, such as the choice of education and occupation, whether to marry and settle down, to move etc. These consequently determine the amount of effort observed in the measure, e_i. The intuition behind this indirect opportunity set is that an individual makes many different choices that affect his/her income. This income is then used to find a measure of effort, which is the rank in the income distribution within the type.

The last step before proceeding to the empirical part is to define a situation in which there is equality of opportunity.

$$G_{t(i)}^{-1}[\pi_{t(i)}] = y = G^{-1}[\pi] \tag{4}$$

Equality of opportunity holds if the income distribution is independent of type. The same value of effort would, accordingly, mean the same income, even though the types, and thus the circumstances, are different. Equality of opportunity is simply characterized by a situation where the circumstances do not affect the income. Investigating this issue is the main task for the empirical part of this paper.

3. Data

The data set for the analysis consists of the complete cohort of individuals in Sweden born 1965. Biological and social parents are identified for each individual. Everyone who has been registered as a Swedish resident is included.

The data set is based on information from several different registers, which are linked and matched by Statistics Sweden. The register used to obtain the biological connection between children and parents is the Several Generations Register.[8] The longitudinal database, Louise, is used for the outcome years 1994-99. Background information for the years 1971, 1974, 1977, 1980, 1983 comes from the Income and Wealth Register[9], which is a register based on taxation. Additional background information, as well as connections between children and social parents, comes from the Population and Housing Census[10], which is a nationwide census that took place 1965, 1970, 1975, 1980, 1985 and 1990. Social parents refer to parents living with the child, even though they are not necessarily biological or adoptive parents. The data are deflated to the 2001 price level by using the consumer price index.

The complete data set consists of 153 005 individuals. It is divided into two subsamples, consisting of 79 007 males and 73 998 females. Individuals who died during the period 1994-99 are excluded from the sample as are those whose parents died before 1985. If the analysis is performed using the fathers' incomes, only individuals who have the same social father identified as the social father for each of the years 1970, 1975 and 1980, are included in the sample. With these restrictions, 45 699 males

[8]In Swedish, Flergenerationsregistret.
[9]In Swedish, Inkomst och förmögenhetsregistret.
[10]In Swedish, Folk och bostadsräkningen, FoB.

and 42 918 females are included with their fathers. If the income of both parents is analyzed, the same social parents had to be identified for each of the years 1970, 1975 and 1980, in order for the individual to be included. 43 301 males and 41 027 females remain in the sample when families with both parents present are analyzed.

It is obvious that, as a consequence of using these criteria for selecting the samples, the final data sets consist, to a large extent, of nuclear families. Descriptive statistics of the family characteristics, as well as information about the neighborhood, can be found in *Table 1*. The variables are measured several times during the individual's childhood and averages are used in the estimations.[11] Factors that are connected to the neighborhood are all measured at the parish level.[12] On average, a parish includes about 3200 individuals, although the variation is substantial. At this level, the average income among 30-60 years old is used as well as the percentage of individuals aged 30-60 years who earned less than 50 % of the median income. Additional variables describing circumstances are, for example, the percentage of individuals in employment, the percentage of female headed households with children and the percentage of individuals who own their own home living in the parish. Similar variables to those used in this analysis are included in other studies where the outcome of a child is analyzed.[13]

[11]Most of the early empirical literature on intergenerational income mobility used the income for a specific year, but if interest is focused on estimating the effects of long-run income status, or permanent income, this yields an error-in-variables bias toward zero (Zimmerman, 1992). Traditionally, averaging the income over several years or using out of equation instruments are the methods used to reduce this bias. In this study, the income of the parents is averaged over 1971, 1974, 1977, 1980 and 1983.

[12]A parish is originally a geographical division related to the church, but it is a suitable level of analysis as it is much smaller than a municipality. Using the average of data for a municipality could, for example, hide important differences within the area.

[13]See for example Ginther *et al*. (2000) for a survey of the effects of neighborhoods.

Table 1. Descriptive statistics

Sample	Male		Female	
Statistics	Mean	Std.	Mean	Std.
Variables		errors		errors
Labor income (ln), average 1994-99	11.9140	1.0331	11.3720	1.0655
Disposable income (ln), average 1994-99	11.8690	0.4291	11.7380	0.3721
Combined annual income (ln), father and mother average 1971, 1974, 1977, 1980, 1983	12.5990	0.3844	12.6030	0.3856
Combined annual income (ln), father average 1971, 1974, 1977, 1980, 1983	12.2720	0.4150	12.2750	0.4193
Social mother (1970, 1975, 1980) is biological mother (dummy = 1)	0.9427	0.2325	0.9490	0.2200
Mother is foreign born (dummy = 1)	0.0775	0.2674	0.0815	0.2736
Age of mother 1971	32.8170	5.8353	32.754	5.8146
Social father (1970, 1975, 1980) is biological father (dummy = 1)	0.8002	0.3999	0.7827	0.4124
Father is foreign born (dummy = 1)	0.0801	0.2714	0.0799	0.2711
Age of father 1971	36.182	6.6366	36.226	6.6829
Child did not stay with social parent 1985 (dummy = 1)	0.1859	0.3890	0.3211	0.4669
Number of social siblings, average over 1970, 1975 and 1980	1.3815	0.8704	1.3848	0.8793
Number of social sisters, average over 1970, 1975 and 1980	0.6497	0.6688	0.6575	0.6969
Mother divorced (was married 1971, 1974, 1977 or 1980, and later divorced) (dummy = 1)	0.1345	0.3412	0.1440	0.3511
Father divorced (was married 1971, 1974, 1977 or 1980, and later divorced) (dummy = 1)	0.0651	0.2467	0.0581	0.2340
Owner (dummy = 1 if the owner of the home lives in household)	0.3847	0.4865	0.3721	0.4834
Mother received social assistance, 1983 (dummy = 1)	0.0442	0.2056	0.0441	0.2053
Father received social assistance, 1983 (dummy = 1)	0.0252	0.1567	0.0244	0.1543
Mother unemployed (dummy = 1 if mother received money due to unemployment 1974, 1977, 1980 or 1983)	0.0950	0.2932	0.0957	0.2942
Father unemployed (dummy = 1 if father received money due to unemployment 1974, 1977, 1980 or 1983)	0.0920	0.2890	0.0921	0.2891
Combined annual income (ln), average of 30-60 years old in parish (that mother lived), average 1971, 1974, 1977, 1980, 1983	12.0070	0.1326	12.0070	0.1318
Combined annual income (ln), average of 30-60 years old in parish (that father lived), average 1971, 1974, 1977, 1980, 1983	12.0020	0.1352	12.0010	0.1350
Share working, percent in parish, average 1970, 1975 and 1980	46.5140	3.3883	46.5090	3.3829
Female headed household with a child, percent in parish average 1970, 1975, 1980	4.0475	2.0452	4.0422	2.0335
Share of population in parish that was foreign citizen average 1970, 1975, 1980	5.0584	4.3825	5.0753	4.3966
Owner, percent in parish where the owner of the home lives in household, average 1970, 1975, 1980	49.7210	22.1340	49.4950	22.0830
Poverty measure; share with less than 50% of median income in parish (mothers), average (1971, 1974, 1977, 1980, 1983), percent	24.5340	3.1901	24.5170	3.1832
Poverty measure; share with less than 50% of median income in parish (fathers), average (1971, 1974, 1977, 1980, 1983), percent	24.7180	3.2486	24.7170	3.2552

Two different measures of income for the individual are used in the empirical part. The first measure is the average of the labor income for the years 1994-99. Equality of opportunity refers, in this case, to the possibility to earn a high labor income. Here opportunities indicate a more direct connection to productivity. The second income measure that is used is the average disposable income for the same years. In this case, the opportunities refer to reaching a high living standard after tax.

4. Econometric model

This econometric section describes how to illustrate the indirect opportunity sets. Following the theory of equality of opportunities, the first step consists of modelling the outcome variable as a function of circumstances and individual effort. However, in order to illustrate the opportunity sets, it is necessary to reduce the dimension along which the opportunities are compared. This is done within the econometric model, by estimating $y_i = y(e_i, C_{type(i)})$.

Both the theoretical and empirical literature concerning intergenerational income mobility suggest that there may be a nonlinear relationship between the income of the father and the income of the child as an adult. (Becker & Tomes, 1986 and Corak & Heisz, 1999). This nonlinearity is allowed for by estimating a semiparametric model. In addition, a parametric model is also estimated. The semiparametric model will be described first. Then, based on the semiparametric model, an explanation of the illustration of the indirect opportunity sets will follow.

4.1 Semiparametric model

Becker & Tomes (1986) suggest that there is a higher correlation between the incomes of parents and adult children for low-income families. This arises from borrowing constraints when it comes to the parents' human capital investments in their children. Using a nonparametric technique and Canadian data, Corak & Heisz (1999) find empirically an inverted V-pattern for the income elasticity. They suggest that this can be explained by a borrowing constraint and a positive relation between the ability of the children and the income of the parents.[14] As mentioned in the introduction, in the Swedish case, where higher education is free from tuition fees and there is a system of student grants and government loans, there seems to be a relatively lower risk for families to be constrained in their investment behavior in their children's human capital. However, for Swedish data, Österberg (2000) indicates nonlinearities by estimating the probability for the son to end up in a certain decile of the income distribution given the father's status. The highest income classes are reported as the least socially mobile. The method used is, however, different from estimating a nonlinear elasticity.

Bearing the previously mentioned literature in mind, it is appropriate to estimate the effects of parental income without assuming a particular functional form and allowing the income elasticity to vary according to the parental income.

[14]The argument is that low income parents would have low ability children, who would get an optimal investment in human capital. The elasticity should thus be low. For families with higher incomes, the children would have a higher ability. Even though the income is higher, the extra income would not fully cover the extra funds needed to invest in human capital, to produce an optimal investment. The elasticity would thus be higher. Families with even higher incomes would, however, have enough money to invest optimally, even though the children would have an even higher ability. The elasticity would, accordingly, be lower again.

A nonlinear specification is,

$$y_i = m(C_i) + e_i \tag{5}$$

where C_i is a vector of the circumstances that the individual cannot influence. e_i is the effort in the model. As individuals only receive a positive labor income when they participate in the labor market and actually have a job, there are individuals with no labor income at all in a specific year. This occurs both when the individuals decide not to participate in the labor market or when they do, but are unemployed. All income variables are measured in logarithms and, accordingly, only observations with positive incomes for both the adult child and the father/parents are included.[15] As (5) applies for the full sample, excluding observations with no income, should require some adjustments to the estimation procedure. An option would be to estimate a Tobit-model. However, as the income is averaged over a period of years, it is assumed that the potential problems using a standard procedure are limited. Inference is, of course, only possible with respect to the population with a positive income.

The curse of dimensionality refers to the slow convergence of nonparametric estimators when additional regressors are included. Due to this problem, the C_i variables are separated into a linear part, consisting of variables with parameters that are estimated parametrically, while a nonlinear part is estimated nonparametrically. The semiparametric model is described in Robinson (1988).

[15]The problem of zero income is, of course, more common when labor income is used as the output variable than when disposable income is used. Of course, parents with zero income also exclude observations, when the logarithm of income is used.

The model separates the nonlinear effect of (5), into a linear and a nonlinear part,

$$y_i = x_i\beta + m(z_i) + e_i \tag{6}$$

where z_i and x_i are circumstances beyond the control of individuals. z_i is the income of the parents, and x_i is other circumstances for which it is important to control. e_i is the individually specific effect, which corresponds to effort in theory. e_i, in equation (6) is assumed to be independent of the explanatory variables.[16] β is a vector of parameters and the functional form of $m(\cdot)$ is unknown. The semiparametric estimation of β in (6) is carried out in the following manner. First (6) is transformed as

$$R_i^{yz} = R_i^{xz}\beta + u_i$$

where $R_i^{yz} = y_i - E(y_i \mid z_i)$ and $R_i^{xz} = x_i - E(x_i \mid z_i)$. With this transformation the semiparametric estimator is,

$$\hat{\beta}_{SP} = \left(\sum_{i=1}^{n} R_i^{xz'} R_i^{xz} \hat{I}_i \right)^{-1} \sum_{i=1}^{n} R_i^{xz'} R_i^{yz} \hat{I}_i \tag{7}$$

To make the estimation of $\hat{\beta}_{SP}$ feasible, $E(y_i \mid z_i)$ and $E(x_i \mid z_i)$ are replaced by nonparametric estimates. This is done by using local linear

[16]Abul Naga, R. (2002) pays regard to the potential problem caused if this assumption is invalid. For example, if the ability and effort of the parents are transferred, through genes and/or behavior, to the ability and effort of the child, this could imply a correlation. In this study, a fairly large vector of control variables, in addition to the parental income, is used. A sensitivity analysis is also included in the results section.

regressions. $\hat{I}_i$ is an indicator function to avoid division by zero. ($\hat{I}_i = 1(|f_i| > b)$, where b is a small value, here set to 0.001 and f_i is the density function of z.) The nonparametric part of (6), $m(z)$ and its derivative, is estimated by means of a local linear regression of $y^* = y_i - \hat{y}_i$ on $m(z_i)$, where $\hat{y}_i = x_i \hat{\beta}_{SP}$. The procedure can be summarized in three steps.

1. Estimate $E(y_i \mid z_i)$ and $E(x_i \mid z_i)$ nonparametrically for all x_i, and make the transformation to R_i^{yz} and R_i^{xz}.

2. Estimate equation (7) and calculate $y^* = y_i - \hat{y}_i$.

3. Estimate $y^* = m(z_i)$ nonparametrically.

The semiparamtric estimator of the derivative of $m(z_i)$ is,

$$\gamma_{SP}(z) = \frac{\sum_{i=1}^{n}(z_i)(y_i^*)K((z_i - z)/h)}{\sum_{i=1}^{n}(z_i)^2 K((z_i - z)/h)} \tag{8}$$

A gaussian kernel is used with the bandwidth, h, which is estimated using Scott's rule of thumb (Scott, 1992, page 152).[17] In this case, the rule of thumb yields $\hat{h} = \hat{\sigma}\,(n)^{-1/5}$, where $\hat{\sigma}$ is the estimated standard deviation of z, and n is the number of individuals.[18] This semiparametric model results in estimates for each coefficient of the parametric part. Further, the

[17]See the appendix for a specification of the gaussian kernel.

[18]As seen in the formula, Scott's rule of thumb increases the bandwidth with the standard deviation, and decreases with the number of observations. The bandwidth is fixed over the whole sample, which can give a too small bandwidth in the ends of the distribution where there are few observations.

model results in estimates in each z. Together these constitute the derivative of the nonlinear part in (6). The indirect opportunity sets can easily be calculated based on this model.

4.2 Estimation of indirect opportunity sets

A central part of the theory of equality of opportunity is the distinction between choices and circumstances that restrict the individual's opportunities. To compare the income of individuals living in different circumstances, but making a similar effort, it is, therefore, important to estimate conditional indirect opportunity sets. O'Neill *et al.* (2000) illustrate opportunity sets using two different methods. The first uses a bivariate kernel density of the fathers' incomes on the incomes of the children as adults. The second method is to extract those fathers who have similar incomes and then rank the adult children according to their own income. The extraction could, for example, be fathers belonging to percentile 25, 50 and 75. Within each of these groups, the adult children would be ranked based on their own income.[19] To ease the interpretation, the outcome variable is divided by the mean for the sample. With opportunity sets, it is easy to compare the outcome of individuals with different circumstances but with the same rank within their type.

As the income distribution of parents and the adult children is rather compressed in Sweden, a bivariate kernel density did not give an illustrative figure of the opportunity sets.[20] The second method is, thus, applied instead. The difference compared with O'Neill, *et al.* (2000) is, however, that the illustration is done with z and y^*, instead of just the

[19]The ranking is made by ordering the observations, starting with the lowest income. The rank is also normalized to go from 0 to 100.

[20]The density is very high in a small area, and the figure of the density is a very sharp peak rather than a hill. This is the case even when a large proportion of the density is cut away to scale the figure into something interpretable.

incomes of the adult child (y). Since $y^* = y_i - x_i \hat{\beta}_{SP}$, the indirect opportunity sets are conditional on all the variables in $\mathbf{x}$. As y^* is measured in logarithmic form, it is transformed back before being analyzed. $(y^{**} = e^{y^*})$ Note that these indirect opportunity sets are conditional on the variables that are included parametrically and also on the income of the parents.[21] This makes the assumption that all relevant circumstances are included a little bit more convincing. In O'Neill *et al.* (2000) only the income of the father was included as a circumstance, and the rest of the variation in income was labelled as effort. Accordingly, this was interpreted as within the scope of the choice of the individual. The model used in this study has permitted the opportunity sets to be estimated without sorting the population into a tremendous number of groups which, even with a huge data set, could contain too few observations.

With this method, it is of course not possible to test all the circumstances at once. The method is instead a way of illustrating different opportunities arising from differences in one specific circumstance at each time, while keeping all the other circumstances constant. The researcher has, thus, to decide which circumstance to illustrate. The regression described in the previous section tests whether circumstances affect income. However, it does not illustrate the differences in incomes, while keeping the degree of effort constant, as in the theory. This is an addition to the analysis obtained by estimating the indirect opportunity sets.

[21]To get the indirect opportunity set for children with parents in the 25th percentile, the incomes of the parents are ranked and those who are included in the 25th percentile are extracted. Then *y***/mean is ranked for the children, where the mean refers to the mean of *y*** for the total population. The opportunity set for individuals with parents in the 75th percentile is done in the same way. Finally, the results are plotted together. The variables in *x* are kept constant in *y***, and the extraction takes care of the differences in the parents' income.

5. Results

This section presents the results from the estimations. It is, however, important to remember that the individual specific effect is estimated using a particular set of circumstances. It is, of course, debatable whether enough circumstances, or even the correct circumstances, are included in the analysis. In the final part of this section, this issue is investigated briefly by using information on siblings. If, for example, innate ability is important for income, this should be included in the vector of circumstances, since the individuals cannot choose their genes. Innate ability and other unobserved circumstances that are not shared by the siblings are not included in the analysis. All these unobserved factors are, accordingly, sorted into effort.

The analysis is performed in three different steps. First the hypothesis that equality of opportunity exists is tested by including the circumstances in an OLS regression of the labor income. The next step is to include the parents' income nonparametrically to allow for a possible nonlinear relation. Finally, indirect opportunity sets are calculated and different circumstances are compared keeping the rank constant. These three steps are also performed using disposable income as the outcome variable.

5.1 Labor income as dependent variable

Table 2 illustrates semiparametric estimation as well as the OLS estimation using labor income as the dependent variable for both the male and female samples. The different circumstances influence the income of the adult child and equality of opportunity clearly does not exist.

At the same time, it can be noted from the fairly small $R^2 - adj$ that the circumstances only explain a small part of the variation in income in adulthood.

As the choice of where the parents decide to live is potentially endogenous, the coefficient of neighborhood variables should be interpreted with caution.[22] A significant coefficient should not be interpreted as having a causal impact on the labor income of the adult child. It should rather be seen as correlation.

Variables indicating a stable family relationship, for example, if both social parents are the child's biological parents, seem to be important for labor income. Having a divorced mother appears to significantly influence the child's labor income in adulthood. The magnitude of the coefficient should, however, be interpreted keeping in mind the selection criteria of the samples. Social fathers (and mothers) are in this empirical part included if the child was living with them during each of the years 1970, 1975 and 1980. This implies, of course, that most of the families did not experience a divorce.

[22]It is possible that parents who pay a lot of attention in the upbringing of their child, also value a good neighborhood highly. Whether it is the good neighborhood, or the attention, that is important is, however, not separated.

Table 2. Estimates. Dependent variable: Labor income (ln) (average, 1994-99)

Sample	Male		Female	
Model	OLS	Robinson	OLS	Robinson
Variables				
Combined annual income (ln), father and mother	0.2979***	Fig. 1a	0.2248***	Fig. 2a
	(0.0149)		(0.0157)	
Social mother (1970, 1975, 1980) is biological mother	0.1502**	0.1479**	0.1744**	0.1764**
	(0.0663)	(0.0677)	(0.0690)	(0.0706)
Mother is foreign born	-0.0412*	-0.0435	-0.0394*	-0.0402*
	(0.0211)	(0.0216)	(0.0219)	(0.0224)
Age of mother 1971	0.0012	0.0011	0.0045***	0.0041*
	(0.0015)	(0.0016)	(0.0016)	(0.0017)
Social father (1970, 1975, 1980) is biological father	0.2035***	0.2010***	0.1736***	0.1699***
	(0.0432)	(0.0441)	(0.0458)	(0.0469)
Father is foreign born	-0.0427*	-0.0431**	-0.0458**	-0.0462**
	(0.0212)	(0.0217)	(0.0225)	(0.0231)
Age of father 1971	-0.0056***	-0.0053***	-0.0022	-0.0023
	(0.0013)	(0.0013)	(0.0014)	(0.0014)
Child did not stay with social parent 1985	-0.1132***	-0.1129***	-0.1111***	-0.1096***
	(0.0139)	(0.0142)	(0.0116)	(0.0119)
Number of social siblings	-0.0178**	-0.0147**	-0.0593***	-0.0590***
	(0.0071)	(0.0073)	(0.0077)	(0.0080)
Number of social sisters	-0.0043	-0.0041	0.0344***	0.0346***
	(0.0088)	(0.0090)	(0.0092)	(0.0095)
Mother divorced	-0.1625***	-0.1648**	-0.0670**	-0.0682**
	(0.0254)	(0.0260)	(0.0267)	(0.0273)
Owner (dummy = 1 if the owner of the home lives in household)	0.0274**	0.0287**	0.0834***	0.0798***
	(0.0112)	(0.0115)	(0.0118)	(0.0122)
Mother received social assistance, 1983	-0.2969***	-0.2869***	-0.3107***	-0.3102***
	(0.0341)	(0.0350)	(0.0360)	(0.0370)
Mother experienced unemployment (dummy = 1)	0.0050	0.0015	-0.0505***	-0.0472**
	(0.0179)	(0.0184)	(0.0190)	(0.0195)
Father experienced unemployment (dummy = 1)	-0.0727***	-0.0695***	-0.0404**	-0.0330*
	(0.0177)	(0.0182)	(0.0188)	(0.0193)
Combined annual income (ln), average in parish	0.2048***	0.2097***	0.0969*	0.0756
	(0.0534)	(0.0548)	(0.0567)	(0.0584)
Share working, percent in parish	0.0086***	0.0087***	0.0029	0.0026
	(0.0023)	(0.0024)	(0.0025)	(0.0025)
Female headed household with a child, percent in parish	-0.0072*	-0.0080*	-0.0038	-0.0045
	(0.0043)	(0.0044)	(0.0046)	(0.0047)
Share of population in parish that was foreign citizen	0.0026*	0.0026	0.0015	0.0014
	(0.0016)	(0.0016)	(0.0016)	(0.0017)
Owner of the home, percent in parish	0.0004	0.0004	0.0002	0.0001
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Poverty measure	0.0070**	0.0078***	-0.0019	-0.0024
	(0.0028)	(0.0028)	(0.0029)	(0.0030)
Constant	5.0503***		7.0090***	
	(0.6084)		0.6480	
N	39057	39019	36501	36477
R-sq, adj.	0.0294		0.0250	

Note: Coefficients that are significantly different from zero at 10, 5 and 1%-level are marked with *, ** and ***. Standard errors are included in the second row for each estimated parameter.

In the semiparametric model, parental income is the nonparametric part. Accordingly, it does not give a single measure of the elasticity, but rather a continuous measure of elasticities. The elasticities displayed in *Figures 1a* and *2a* correspond to columns 2 and 4 in *Table 2*. All figures can be found on the last pages.

Note that 98 percent of the observations of the logarithm of the combined annual income of the parents are included between 11.6 and 13.6 in *Figures 1a* and *2a*.[23] The confidence interval, for each estimated elasticity, is estimated as plus two standard deviations and minus two standard deviations. The confidence bands are, however, underestimated and interpretations should be made with caution. This is because the variance for each elasticity is estimated without taking into account that y_i^* also is estimated. See the appendix for details concerning this issue. There is no pattern of elasticities with regard to the incomes of the parents. Accordingly, the Swedish case does not appear to support the hypothesis in Becker & Tomes (1986) that there is a constraint on investment behavior in human capital formation for low income parents. The inverted V-pattern found by Corak & Heisz (1999) for Canadian data, does not apply for Swedish data. The pattern of immobility found by Österberg (2000) does not appear in the estimation of nonlinear elasticities. The method, however, answers different questions. Österberg (2000) investigated the probability of a child ending up in a certain decile given the parents' income decile. The nonparametric estimations of the elasticity indicate the expected consequence of a one-percent increase/decrease in the parents' income, given the level of the parental income. Since no nonlinearity of parent's incomes is found, the OLS estimates can be used

[23]The reason that the elasticity seems to vary more at the ends of the distribution is because of the constant bandwidth chosen with Scott's rule of thumb. The bandwidth is probably too small at the ends of the distribution, and the estimates accordingly appear to be under smoothed.

for interpreting how different characteristics are connected to labor income in adulthood.

The analysis presented above is carried out using the parents' combined annual income. The literature on intergenerational income correlation has, to a large extent, focused on estimating elasticities with respect to the father's income. In order to make a comparison with this literature, the models are estimated using the fathers' income instead of the parental income. The results can be found in *Table A1* in the appendix. The elasticity of the fathers' income is higher than the elasticity found previously by Österberg (2000) (0.125-0.185) for the father-son relationship in Sweden.[24] This is expected as the social parents are used in this study rather than the formal guardianship that was registered at the birth of the child.[25] However, adding more background variables typically reduces the elasticity of the fathers' income.

5.2 Disposable income as the dependent variable

Using labor income as the dependent variable answers questions about whether, and to what extent, circumstances constrain the individual's opportunities to acquire a high labor income in adulthood. As always, the results are estimated for a particular period and for particular public policies. Public policies can, thus, make different circumstances more or less important. The system of taxes and subsidies does, of course, affect the income distribution after taxation. Using disposable income instead of

[24]*Table A1* and *Figures 5a* and *6a* are difficult to compare with Österberg (2000) as no single measure of the elasticity is presented. The OLS regressions corresponding to columns two and four give, however, the elasticity 0.282 for the male sample and 0.190 for the female sample.

[25]Björklund & Chadwick (2003) study this issue for Swedish data. The elasticity is found to be "generally insignificantly different from zero" for sons who never lived with their biological fathers.

labor income, allows the after tax effects to be investigated. However, using disposable income also allows for inequality of opportunity arising from incomes other than those from labor. Incomes from capital are, for example, included in the income measure.

With the change of the dependent variable from labor income to disposable income, the coefficient of the different circumstances shrinks in magnitude. This is natural since the distribution of disposable income is more compressed than the distribution of labor income. Some of the variables are, however, not even significantly different from zero. The tax system does not only reduce after tax inequality, it also reduces the connection between circumstances and later income. *Table 3* shows the OLS estimates, as well as the results from the semiparametrical model. *Figures 3a* and *4a* show the nonparametric part. Estimates using the fathers' income instead of the family income can be found in *Table A1* in the appendix.[26] No particular nonlinear pattern can be found either with the parents' income or with the fathers' income.

It is notable that some of the circumstances that matter for labor income do not seem to matter at all when it comes to disposable income. For females, the coefficients for the biological father, number of siblings, divorced mother, whether the mother received social assistance, and unemployment of either parent are not significantly different from zero. In the case of women, not living with their social parents at the age of 20, actually means that they have a significantly higher disposable income. When labor income was analyzed, the corresponding coefficient was negative. For the male sample, the corresponding coefficient is significantly negative for both income measures.

[26]The OLS counterpart estimates 0.161 as the elasticity for the male sample, and 0.07 for the female sample.

Table 3. Estimates. Dependent variable: Disposable income (ln) (average, 1994-99)

Sample	Male		Female	
Model	OLS	Robinson	OLS	Robinson
Variables				
Combined annual income (ln), father and mother	0.1739***	Fig. 3a	0.0955***	Fig. 4a
	(0.0061)		(0.0057)	
Social mother (1970, 1975, 1980) is biological mother	0.1042***	0.1028***	0.0406	0.0406
	(0.0270)	(0.0275)	(0.0248)	(0.0250)
Mother is foreign born	-0.0088	-0.0087	-0.0037	-0.0032
	(0.0086)	(0.0088)	(0.0080)	(0.0080)
Age of mother 1971	-0.0005	-0.0004	-0.0007	-0.0008
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Social father (1970, 1975, 1980) is biological father	0.0533***	0.0540***	-0.0119	-0.0127
	(0.0176)	(0.0179)	(0.0165)	(0.0166)
Father is foreign born	-0.0324***	-0.0316***	-0.0222***	-0.0216***
	(0.0087)	(0.0088)	(0.0082)	(0.0082)
Age of father 1971	-0.0019***	-0.0019***	-0.0009*	-0.001**
	(0.0005)	(0.0006)	(0.0005)	(0.0005)
Child did not stay with social parent 1985	-0.0232***	-0.0237***	0.0355***	0.0360***
	(0.0057)	(0.0058)	(0.0042)	(0.0043)
Number of social siblings	-0.0074***	-0.0077**	0.0018*	0.0002
	(0.0029)	(0.0030)	(0.0028)	(0.0029)
Number of social sisters	-0.0034	-0.0033**	0.0056	0.0056
	(0.0036)	(0.0037)	(0.0034)	(0.0034)
Mother divorced	-0.0603***	-0.0594***	-0.0009	-0.0004
	(0.0104)	(0.0106)	(0.0097)	(0.0098)
Owner (dummy = 1 if the owner of the home lives in household)	0.0202***	0.0197***	0.0174***	0.0151***
	(0.0046)	(0.0047)	(0.0043)	(0.0044)
Mother received social assistance, 1983	-0.0977***	-0.1001***	0.0144	0.0103
	(0.0136)	(0.0139)	(0.0128)	(0.0130)
Mother experienced unemployment (dummy = 1)	-0.0079	-0.0077	0.0128	0.0155**
	(0.0073)	(0.0075)	(0.0069)	(0.0070)
Father experienced unemployment (dummy = 1)	-0.0235***	-0.0240***	0.0025	0.0045
	(0.0073)	(0.0074)	(0.0068)	(0.0069)
Combined annual income (ln), average in parish	0.1220***	0.1203***	0.0938***	0.0798***
	(0.0218)	(0.0223)	(0.0206)	(0.0209)
Share working, percent in parish	0.0035***	0.0035***	0.0002	-2E-05
	(0.0009)	(0.0010)	(0.0009)	(0.0009)
Female headed household with a child, percent in parish	-0.0034*	-0.0033*	0.0034**	0.0036**
	(0.0017)	(0.0018)	(0.0017)	(0.0017)
Share of population in parish that was foreign citizen	0.0020***	0.0021***	0.0010	0.0010
	(0.0006)	(0.0007)	(0.0006)	(0.0006)
Owner of the home, percent in parish	0.0004***	0.0004***	0.0002	0.0002*
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Poverty measure	0.0033***	0.0032***	0.0005	-0.0001
	(0.0011)	(0.0012)	(0.0011)	(0.0011)
Constant	7.9218***		9.3627***	
	(0.2490)		(0.2355)	
N	40211	40171	37492	37468
R-sq, adj.	0.0426		0.0165	

Note: Coefficients that are significantly different from zero at 10, 5 and 1%-level are marked with *, ** and ***. Standard errors are included in the second row for each estimated parameter.

Some of the background factors seem to be connected in different ways for men and women. It could, therefore, be interesting to test whether the differences are indeed significant. Simple F-tests reveal that if the mother is divorced, received social assistance, and if the child was living with a social parent at the age of 20, are clearly connected to the disposable income in different ways for the male and female samples. The background factors seem to be less important for the female sample's disposable income.

5.3 Illustration of Indirect Opportunity Sets

Estimating the indirect opportunity sets is an attempt to illustrate inequality of opportunities, and also to incorporate the individual's effort in the analysis.

Figure 1b (and *1c*)[27] is the indirect opportunity sets for men whose parents belong to the 25th and the 75th percentiles of the income distribution. *Figure 2b* (and *3c*) presents the female counterpart. These figures can be analyzed in two different ways. The figures show how much effort is required to achieve the average labor income, (i.e. y**/mean = 1), when the individual has parents belonging to 25th or 75th percentile of the income distribution. To reach the average income, a man with parents in the 25th percentile has to be ranked approximately 52.3 out of 100. For men with parents in the 75th percentile the rank of 42.6 is sufficient to achieve the mean income. [28] The necessary rankings for the female sample are 56.8 and 50.8 respectively.[29]

[27]*Figure 1b* illustrates all individuals, while *Figure 1c* is a close-up, in which individuals at either end of the distribution are cut away. This is done in order to make the figure easier to read. The same applies for *Figure 3c* and all the other indirect opportunity sets. The complete figures of the opportunity sets can be requested from the author.
[28]The results from the estimated models are summarized in *Tables A2* and A3. The figures, however, include much more information.
[29]Note that the male and female samples are analyzed separately. The necessary rankings

O'Neill *et al.* (2000) estimated, for a small male sample, a necessary rank of 70 for children with fathers belonging to the 25th percentile. On the other hand, children with fathers belonging to the 75th percentile only needed a rank of 40 to reach the average income.

Another way to use the opportunity sets in the analysis is to investigate how different circumstances affect the income, when the level of effort is held constant. If the rank is fixed at 50, the male with parents in the 75th percentile has 9.6 % higher income.[30] The counterpart for the female sample is 6.0 %.[31] O'Neill *et al.* (2000) calculate this relationship when the level of effort is kept at 10. In this case, children from families in the 75th percentile were expected to earn 56 % more. These calculated differences, between the different types, obviously vary depending on which level of effort is investigated. Even though there are substantial differences, both in terms of data and method, compared with the study by O'Neill *et al.* (2000), the Swedish case seems to be relatively closer to equality of opportunity.

A Kolmogorov-Smirnov test is performed to see whether the distributions are significantly different depending on the different opportunities. For the male sample, the test finds that the zero hypothesis of equality of the distribution can be rejected at a 1 percent significance level. If the test is performed separately for above and below y**/mean = 1, the results indicate that the difference only applies above the average labor income.

For the female sample, the test cannot, at any reasonable significance level, reject that the distributions are equal. This indicates that the difference in opportunities for earning a high labor income, for the female

for the female sample are, accordingly, to achieve the mean income among woman. These results are also included in *Table A2* in the appendix.

[30]((1.0701/0.9760)-1)*100 ≈ 9.6%. These numbers are also found in *Table A3* in the appendix.

sample, is not that different for individuals with parents in the 25th percentile and the 75th percentile.[32]

The opportunity sets illustrate inequality of opportunities with regard to parental income when all the circumstances in the regression are taken into account. This makes the assumption, required by the theory, that an individual can freely choose how much effort he/she exerts, more reliable than if the opportunity sets for the parents' incomes were analyzed without these variables. Note that the opportunity sets only analyze parental incomes. If opportunities with respect to other variables are to be investigated, then these variables would have to be treated differently.

The indirect opportunity sets presented so far have illustrated different opportunities arising from differences in parental income. It is possible to illustrate indirect opportunity sets with respect to other circumstances as well. For example, *Figure 9*, illustrates indirect opportunity sets for individuals in the male sample, who had parents within the first decile of the income distribution and had, or had not, a mother receiving social assistance in 1983. As social assistance is the focus of the analysis, y** is adjusted so that the effect of social assistance is not deducted from the income of the child. Accordingly y^* is estimated as $y^* = y_i - x_i \hat{\beta}_{SP}$, where neither the dummy variable nor the coefficient for whether or not the mother received social assistance is included in $x_i \hat{\beta}_{SP}$.

Figure 9 illustrates that whether or not the mother received social assistance is a more important factor when the individual's level of effort is

[31]((0.9937/0.9377)-1)*100 ≈ 6.0%

[32]Kolmogorov-Smirnov tests, corresponding to the other figures, give significant difference at the 1 percent level for the following cases; 3b, 5b, 7b, 8b and 9. The opportunity sets in fig. 6b, is not significantly different, at any reasonable significance level. The opportunity sets in figure 4b is significantly different at the 5 percent significance level.

low than when he/she has a high level of effort. A Kolmogorov-Smirnov test confirms a significant difference in the distributions, at the 1 percent significance level, depending on whether the mother received social assistance. Below the average income, the equality of the distributions is rejected at the 1 percent significance level. Above the average, the equality of the distributions is rejected at the 10 percent significance level.

Other circumstances can be analyzed in the same manner. If only a few individuals have the same circumstances, then this can cause problems as there will be too few individuals to illustrate indirect opportunity sets. This is also the reason why the whole of the first decile is included instead of just a percentile.

5.4 Are unobserved circumstances included in effort?

The opportunity sets are estimated using a finite number of observable circumstances and it is, of course, possible that important circumstances are left out. If this is the case, the analysis would then assign too much of the difference in income to differences in effort. To get an idea of whether important circumstances are missing in the analysis, it is possible to use information on siblings. Siblings have, in the literature, been used to capture whether shared family and community factors are important for a socioeconomic outcome.[33] A high correlation for the outcome variable between the siblings would suggest that shared factors are important.

To see whether important circumstances are missing, wage regressions are estimated for both siblings, using the observed circumstances. If all important circumstances are included in the analysis, the correlation of the residuals for two same sex siblings would be very close to zero. The

[33]See Solon, G. (1999) for a survey.

reason is that the observed circumstances would remove the grounds for correlation. To investigate this issue, each individual is matched with the eldest of his/her social siblings. The reason for matching the eldest siblings is to, as far as possible, exclude siblings who are still in education instead of in the labor market. To be identified as a social sibling, the individual and the sibling have to be living in the family home in 1970, and 1975. Further, only brothers are matched to the male sample, and only sisters are matched to the female sample.

The OLS model above is estimated as a seemingly unrelated regression (SUR) with almost the same equation for the brother and the sister. The only differences in the regressions are the addition of age and age-squared, to remove the effects of different ages, for the siblings. The residuals from both the estimations are then compared.

Table 4 illustrates that substantial correlation[34] persists even after eliminating variation due to the observed circumstances. This applies for both the male and female samples, and for both labor income and disposable income. Accordingly, the opportunity sets, based on the observed circumstances, are sorting too much of the variation in incomes into the effort category. The reason is that the siblings have unobserved shared characteristics that influence the income. This means that the opportunity sets give too optimistic pictures of the true possibilities to influence the income with different degrees of effort.

[34]This is the simple correlation coefficient. In the literature on sibling correlation, the estimates are usually components of income variation that, within a model, can be interpreted as correlation.

Table 4. Brothers and sisters correlations

Sample	Measure	Number of sibling pairs	Correlation	Correlation in different quintiles of parents' incomes				
				1	2	3	4	5
Male	Labor income	15257	0.1396					
	Residual	15170	0.1222	0.1106	0.1063	0.1394	0.1343	0.1305
	Disposable inc.	16154	0.2008					
	Residual	16055	0.1771	0.1672	0.1340	0.1323	0.2189	0.2051
Female	Labor income	12851	0.1286					
	Residual	12741	0.1130	0.1067	0.1469	0.0996	0.0848	0.1284
	Disposable inc.	13527	0.1453					
	Residual	13401	0.1406	0.1179	0.097	0.0928	0.1076	0.2101

Table 4 also reveals that the correlation between the residuals of the labor income of the brothers does not seem to vary when the sample is sorted into five quintiles according to the parents' income. This indicates, in the context of the model, that different opportunity sets are missing important circumstances in a similar fashion. In the case of disposable income, the correlation in some quintiles seems to be lower, which indicates that the opportunity set is less affected by unobserved circumstances, and is closer to being based on preferences. It is, however, important to note that none of the quintiles is particularly close to indicating that the circumstances included are sufficient. For all quintiles, there seem to be important circumstances that are missing in the analysis.

6. Concluding remarks

Theories of equality of opportunities distinguish between preferences and circumstances that the child is unable to influence. Roemer (1998) suggests in his theory, that income differences due to preferences that have developed out of circumstances should also be considered as inequality of opportunity. The theory argues that it is possible to control for circumstances and the residual variation is assigned to a "catch-all" term

called effort, which is based on preferences. Investigating this empirically, by sorting the population into discrete groups would, however, quickly result in a problem of groups with too few individuals when controlling for a greater number of circumstances. At the same time, it is important to take into account a broad range of circumstances, since the rest of the variation of income would be sorted into the "catch-all" term called effort. If circumstances that are not included are important, this "catch-all" term would not be based on preferences. In this study, this practical problem is handled within the model, and the individuals are analyzed as if they had the same circumstances, apart from the one that is being analyzed. Indirect opportunity sets are illustrated and analyzed for both a male and a female sample in Sweden. The opportunity sets illustrate how different circumstances require different amounts of effort to reach the average income. An additional analysis of siblings does, however, reveal that the opportunity sets are incomplete and too much of the variation in income is sorted into effort. With a limited set of circumstances the "catch-all" term does indeed catch too much. At the same time, the opportunity sets also illustrate how different circumstances reward individuals making the same degree of effort differently.

With a semiparametric model, it is possible to investigate whether the relation between the income of the parents and that of the adult child is nonlinear. In this study, the elasticity does not vary in a systematic pattern that would support nonlinearity. One interpretation is that low income parents can invest optimally in the human capital formation of their children, and the observed relation has to be explained by other hypotheses. It is, however, possible that there are investment constraints, and other explanations exist for the correlation of incomes for middle and high income parents. With respect to Swedish ambitions with, for example, a public education system without school fees or tuition charges, it seems that the institutions that allow access to education have reduced

the importance of the household budget for human capital formation. Despite this, the correlation between the parents' income and that of the adult children is still present. Investigating the disposable income, instead of the labor income, reduces the connection for the circumstances. For the female sample, some of the circumstances do not seem to matter at all for the disposable income in adulthood.

One way to extend the analysis would, of course, be to add more circumstances and see whether the opportunity sets for parents' incomes change. There are, however, practical difficulties. These extra circumstances, for example, may not be included in the data set, or they may be difficult to measure. One such circumstance, that has been briefly mentioned but not investigated in the empirical part, is innate abilities. The income correlation among siblings only reflects shared factors. Since innate abilities are only partly shared, this most likely means that the opportunity sets are even more incomplete than indicated. One way to investigate this issue would be to include twins in the analysis.

7. References

Abul Naga, R. (2002), Estimating the Intergenerational Correlation of Incomes: An Error-in-Variables Framework, *Economica 69*, 69-91.

Becker, G. & N. Tomes, (1986), Human Capital and the Rise and Fall of Families, *Journal of Labor Economics*, 4, 1-39.

Björklund, A. & L. Chadwick, (2003), Intergenerational income mobility in permanent and separated families, *Economic Letters*, 80, 239-246.

Björklund, A. & M. Jäntti, (1997), Intergenerational Mobility in Sweden Compared to the United States, *American Economic Review*, 87, 1009-1018.

Björklund, A. & M. Jäntti, (2000), Intergenerational Mobility of Socio-economic Status in a Comparative Perspective, *Nordic Journal of Political Economy*, 26, 3-32.

Corak, M. & A. Heisz, (1999), The intergenerational Earnings and Income Mobility of Canadian Men, *Journal of Human Resources*, Vol. 34, Issue 3, 504-533.

Dworkin R. (1981), What is equality? Part 1: Equality of Welfare. Part 2: Equality of Resources, *Philosophy & public affairs*, 10, 185-246 and 283-345.

Ginther, D., R. Haveman & B. Wolfe, (2000), Neighborhood Attributes as Determinants of Children's Outcomes, *The Journal of Human Resources*, Vol. 35, Number 4, 603-642.

O'Neill, D., O. Sweetman, & D. Van de Gaer, (2000), Equality of Opportunity, Kernel Density and Intergenerational Mobility, *Advances in Econometrics*, Vol 14, 259-274.

Robinson, P. M. (1988), Root-N Consistent Semiparametric Regression, *Econometrica*, 56, 931-954.

Roemer, J. E. (1998), *Equality of Opportunity*, Cambridge: Harvard University Press.

Roemer, J. E. (2002), Equality of Opportunity: A Progress Report, *Social Choice and Welfare*, 19, 455-471.

Scott, D.W. (1992), *Multivariate Density Estimation*, New York: Wiley.

Solon, G. (1999), Intergenerational Mobility in the Labor Market, In Ashenfelter, O. and Card, D. (eds.), *Handbook of Labor Economics*, Vol. 3, 1761-1800, Amsterdam: Elsevier.

Ullah, A. & N. Roy, (1998), Nonparametric and Semiparametric Econometrics of Panel Data, In Ullah, A. and D. E. A. Giles (eds.), *Handbook of Applied Economic Statistics*, Volume 155, 579-603, New York: Dekker.

Zimmerman, D. J. (1992), Regression Toward Mediocrity in Economic Stature, *American Economic Review*, Vol. 82, No. 3, *409-429.*

Österberg, T. (2000), Intergenerational Income Mobility in Sweden: What do Tax-data Show?, *Review of Income and Wealth*, Series 46, Number 4, 421-436.

8. Appendix

8.1 Econometrical details

The description of the econometric model is not explained in detail. Robinson (1988) is recommended for a more comprehensive presentation of the method. However, since the $V(\gamma_{SP}(z))$ is used to estimate the confidence bands for the elasticity, it is reasonable to explain this part. The standard error for the coefficients of the parametric part are estimated as follows,

$\hat{\sigma}^2 = V(\varepsilon|X,Z)$ is estimated as;

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} R_i^{yz\prime} R_i^{yz} \hat{I}_i + \frac{2}{n}\sum_{i=1}^{n} R_i^{yz\prime} R_i^{xz} \hat{I}_i \hat{\beta}_{SP} +$$

$$\hat{\beta}_{SP} \frac{1}{n}\sum_{i=1}^{n} R_i^{xz\prime} R_i^{xz} \hat{I}_i \hat{\beta}_{SP}$$

and the covariance matrix is estimated as,

$$Cov(\hat{\beta}_{SP}) = \hat{\sigma}^2 \left[\sum_{i=1}^{n} R_i^{xz'} R_i^{xz} \hat{I}_i\right]^{-1}$$

The variances that are used to estimated the confidence band for the elasticities of the nonparametric part are estimated as follows, (Ullah & Roy, 1998),

$$V(\gamma_{SP}(z)|z_i) = (Z'(z)K(z)Z(z))^{-1} Z'(z)\Omega(z)Z(z)(Z'(z)K(z)Z(z))^{-1}$$

where, $Z(z)$ is a $n * (q+1)(2)$ matrix $[1 \quad z_i - z]$ and $\Omega(z) = K(z)\Sigma K(z)$, where Σ is a diagonal matrix with, $\hat{\sigma}_u^2(z)$ estimated through local linear estimation.

$$(1 \;\; 0)(Z'(z)K(z)Z(z))^{-1}Z'(z)K(z)\,\tilde{\dot{u}}$$

where $\tilde{\dot{u}}$ is a vector of local linear squared residuals, $y_i^* = z_i\gamma_{SP}(z) + u_i$.

$V(\gamma_{SP}(z))$ is thus estimated as if y_i^* is correct, even though y_i^* is an estimate in itself. The confidence bands in the figures are estimated as +/- two standard errors from each estimated $\gamma_{SP}(z)$. These confidence bands are accordingly underestimated. Interpretations should, thus, be done with caution.

Through this paper a gaussian kernel is used, i.e.

$$K(z) = K\left(\frac{z_i - z}{h}\right) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{z_i - z}{h}\right)^2\right\}$$

8.2 Tables and figures

Table A1. Estimates. Dependent variables in (ln) (average, 1994-99)

Sample	Male		Female	
Dependent variable	Disposable	Labor	Disposable	Labor
Variables	income	income	income	income
Combined annual income (ln), father	fig. 7a	fig. 5a	fig. 8a	fig. 6a
Social mother (1970, 1975, 1980) is biological mother	0.0391***	0.0807***	0.0158*	0.1093***
	(0.0096)	(0.0238)	(0.0093)	(0.0266)
Mother is foreign born	-0.0067	-0.0386*	-0.0017	-0.0469**
	(0.0085)	(0.0209)	(0.0078)	(0.0220)
Social father (1970, 1975, 1980) is biological father	0.0723***	0.2109***	-0.0090	0.1572***
	(0.0138)	(0.0340)	(0.0127)	(0.0365)
Father is foreign born	-0.0229***	-0.0198	-0.0192**	-0.0274
	(0.0086)	(0.0211)	(0.0080)	(0.0227)
Age of father 1971	-0.0020***	-0.0041***	-0.0016***	0.0010
	(0.0003)	(0.0008)	(0.0003)	(0.0008)
Child did not stay with social parent 1985	-0.0219***	-0.1046***	0.0356***	-0.1174***
	(0.0056)	(0.0136)	(0.0041)	(0.0117)
Number of social siblings	-0.0109***	-0.0201***	-0.0023	-0.0674***
	(0.0029)	(0.0071)	(0.0028)	(0.0078)
Number of social sisters	-0.0026	-0.0052	0.0054	0.03313***
	(0.0029)	(0.0088)	(0.0033)	(0.0094)
Father divorced	-0.0401***	0.0217	0.0032	-0.0727***
	(0.0090)	(0.0222)	(0.0085)	(0.0242)
Owner (dummy = 1 if the owner of the home lives in household)	0.0182***	0.0291***	0.0152***	0.0810***
	(0.0046)	(0.0112)	(0.0043)	(0.0120)
Father received social assistance, 1983	-0.0941***	-0.2953***	0.0179	-0.2508***
	(0.0135)	(0.0339)	(0.0128)	(0.0372)
Father experienced unemployment (dummy = 1)	-0.0176**	-0.0573***	0.0036	-0.0328*
	(0.0072)	(0.0178)	(0.0067)	(0.0190)
Combined annual income (ln), average in parish	0.1029***	0.1703***	0.0890***	0.0866
	(0.0219)	(0.0539)	(0.0205)	(0.0582)
Share working, percent in parish	0.0019**	0.0086***	-0.0006	0.0024
	(0.0009)	(0.0023)	(0.0009)	(0.0025)
Female headed household with a child, percent in parish	-0.0028*	-0.0085*	0.0035*	-0.0069
	(0.0017)	(0.0043)	(0.0016)	(0.0046)
Share of population in parish that was foreign citizen	0.0030***	0.0027*	0.0011**	0.0015
	(0.0006)	(0.0015)	(0.0006)	(0.0016)
Owner of the home, percent in parish	0.0005***	0.0004	0.0002	0.0003
	(0.0002)	(0.0004)	(0.0001)	(0.0004)
Poverty measure.	0.0014	0.0041	-0.0010	-0.0059
	(0.0011)	(0.0028)	(0.0011)	(0.0030)
N	42527	41293	39328	38261

Note: Coefficients that are significantly different from zero at 10, 5 and 1%-level are marked with *, ** and ***. Standard errors are included in the second row for each estimated parameter.

Table A2. Necessary effort to reach average income with different opportunities

	Labor income		Disposable income	
	Male	Female	Male	Female
Income percentile 25, father	52.8	54.7	64.7	62.6
Income percentile 75, father	40.8	52.0	47.6	55.9
Income percentile 25, parents	52.3	56.8	59.7	61.9
Income percentile 75, parents	42.6	50.8	47.9	52.8

*Table A3. y**/mean at effort = 50 with different opportunities*

	Labor income		Disposable income	
	Male	Female	Male	Female
Income percentile 25, father	0.980	0.939	0.918	0.937
Income percentile 75, father	1.088	0.979	1.013	0.962
Income percentile 25, parents	0.976	0.938	0.949	0.940
Income percentile 75, parents	1.070	0.994	1.012	0.987

Figure 1a. Nonparametric part

Figure 2a. Nonparametric part

Figure 1b. Indirect Opportunity Sets

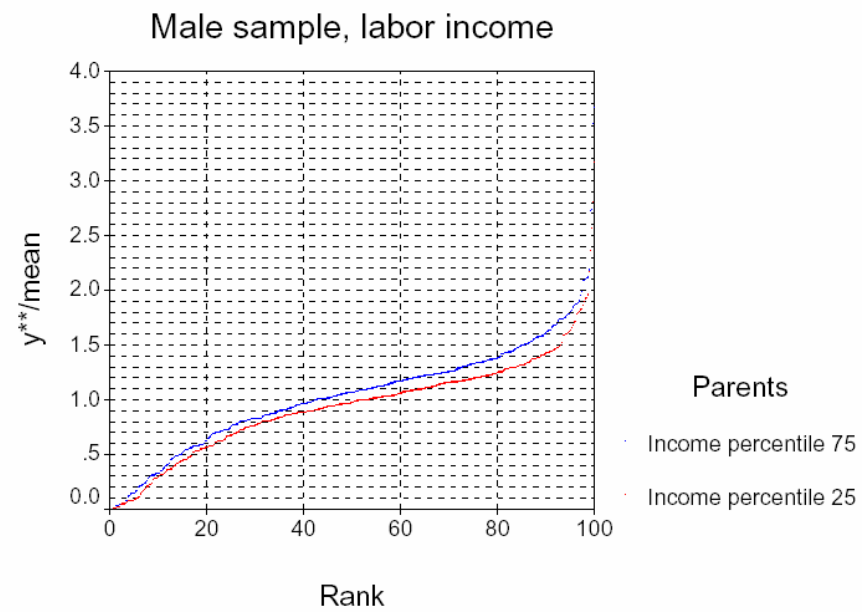

Figure 1c. Indirect Opportunity Sets

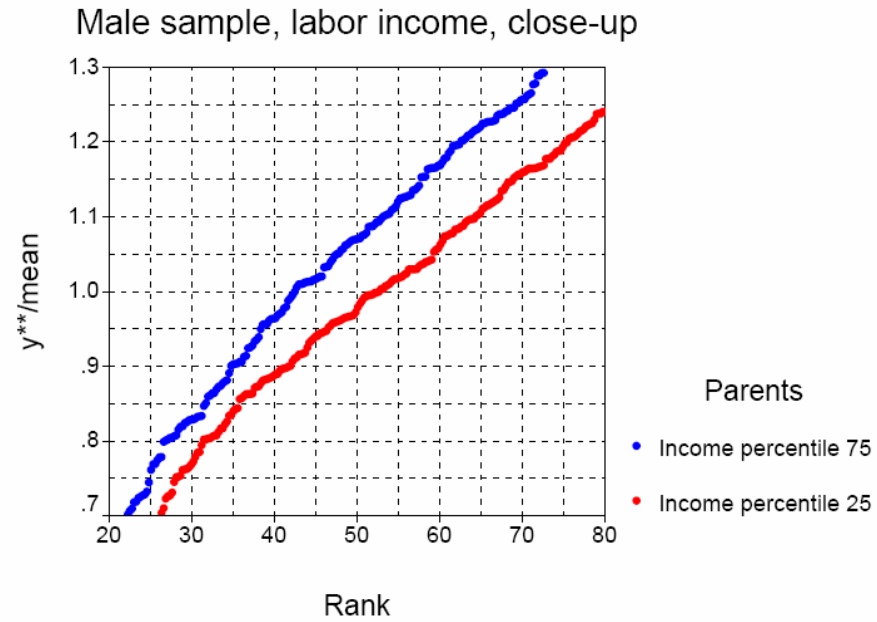

Figure 2b. Indirect Opportunity Sets

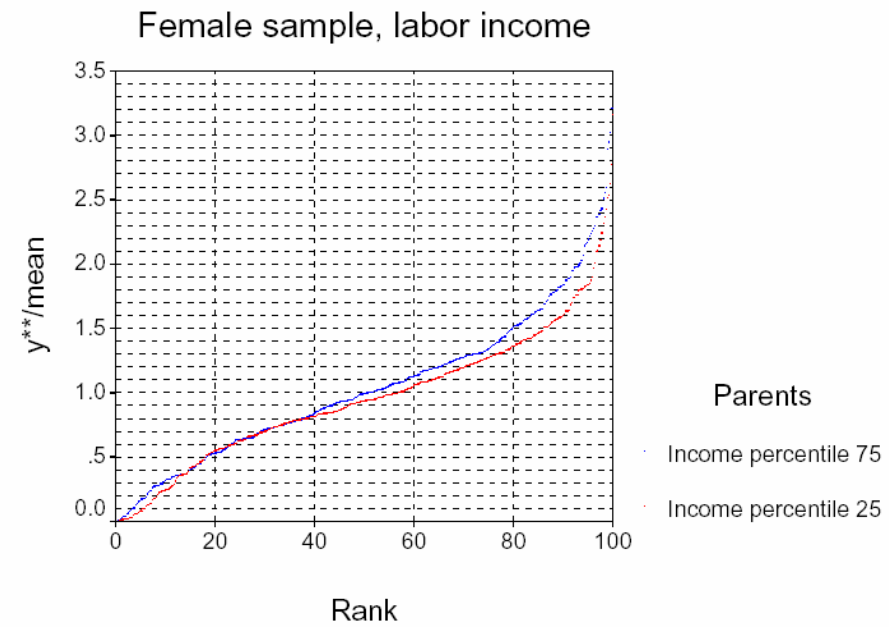

Figure 2c. Indirect Opportunity Sets

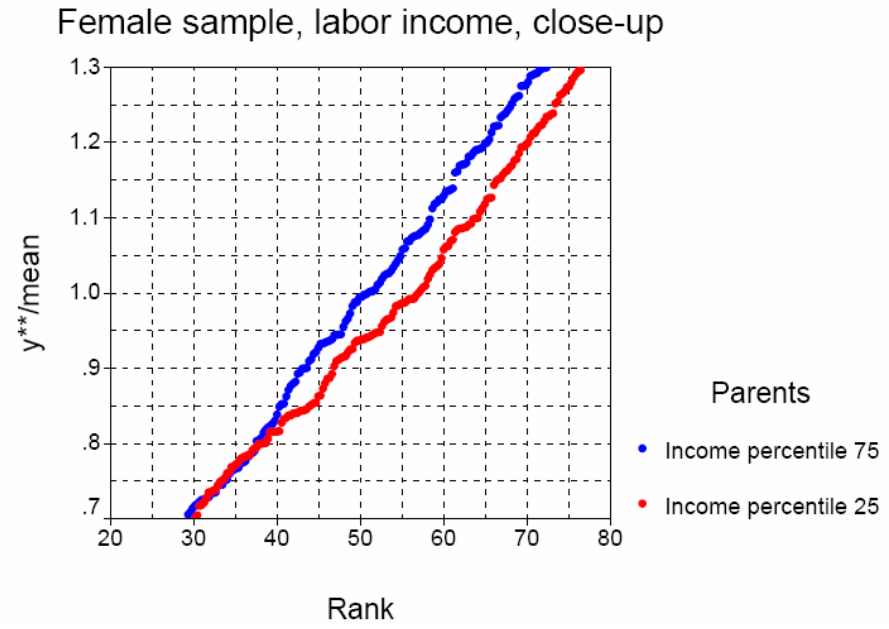

Figure 3a. Nonparametric part

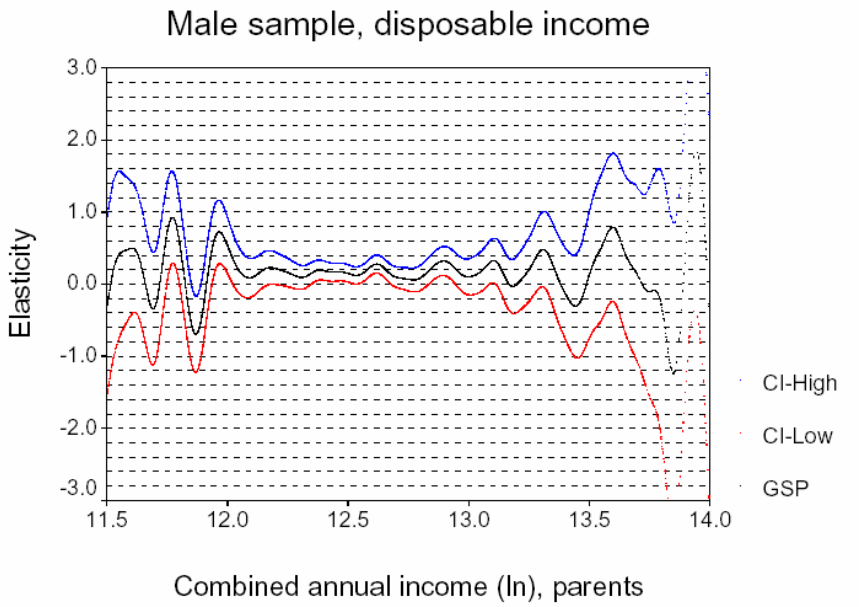

Figure 4a. Nonparametric part

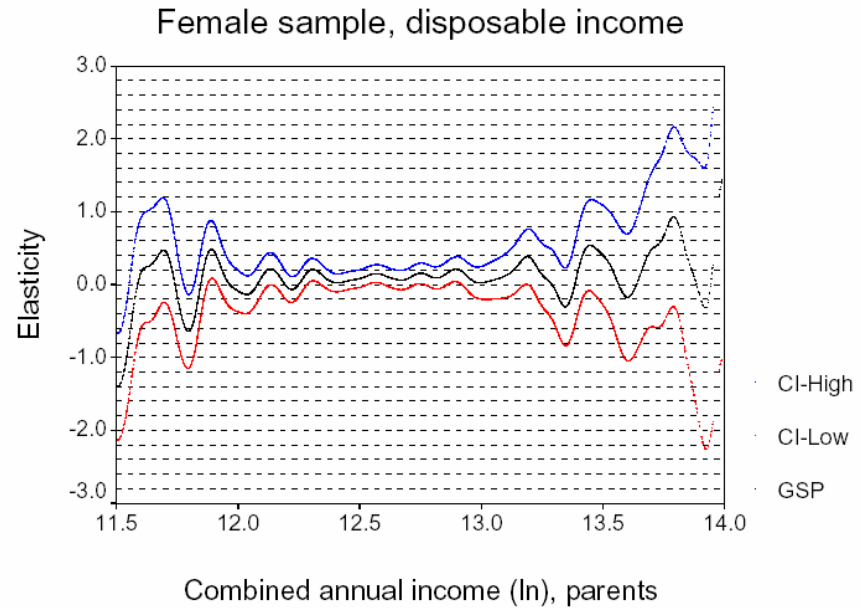

Figure 3b. Indirect Opportunity Sets

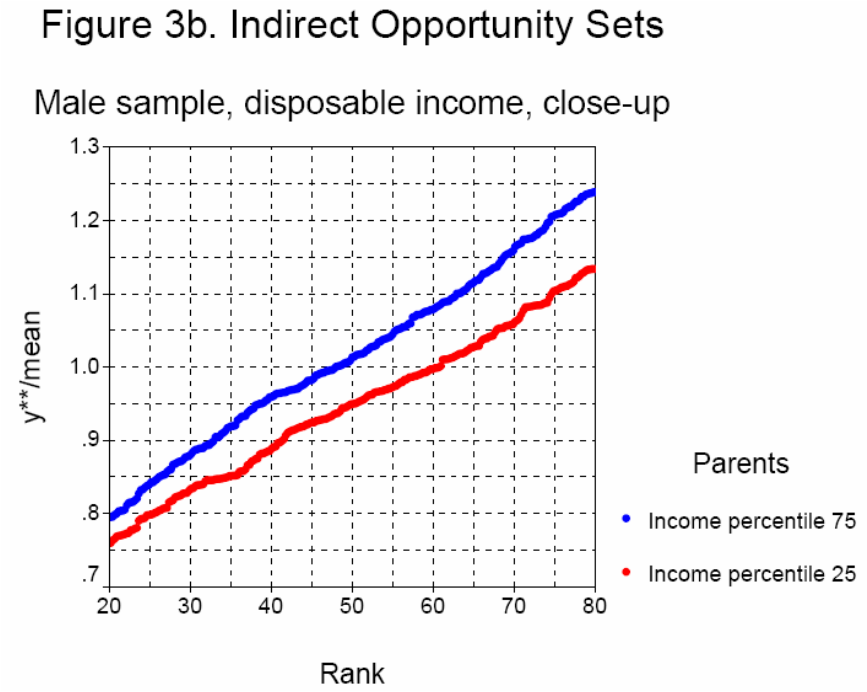

Figure 4b. Indirect Opportunity Sets

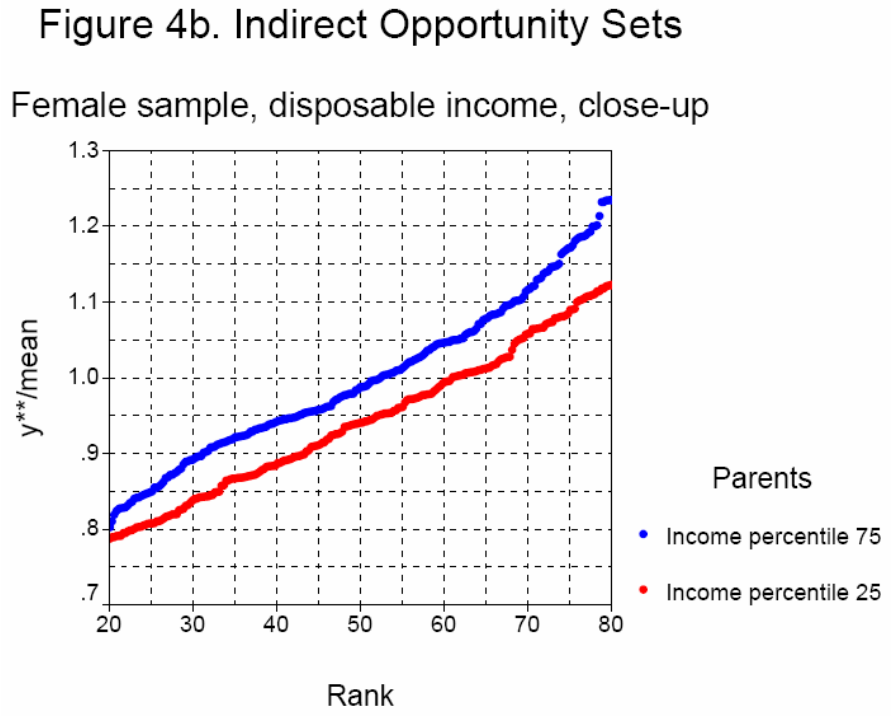

Figure 5a. Nonparametric part

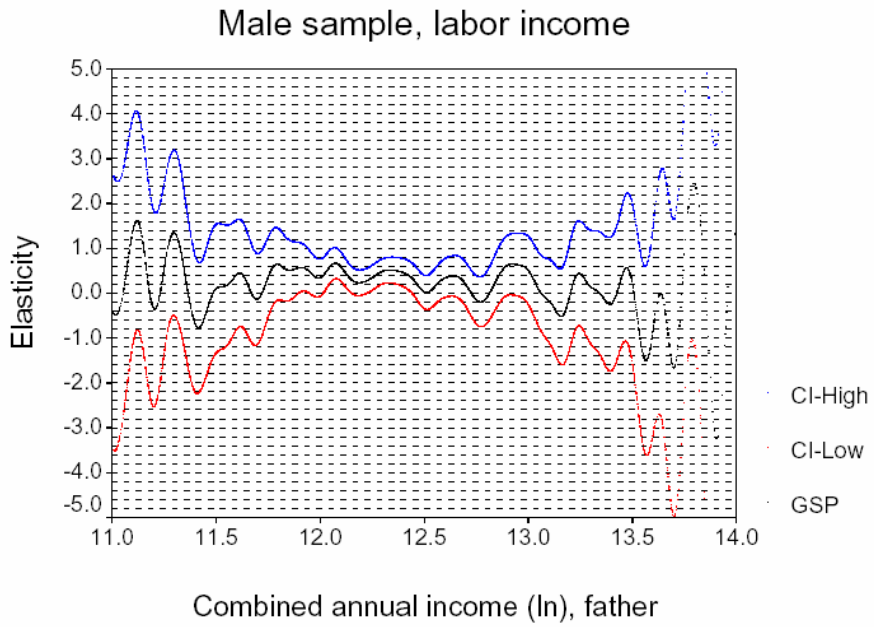

Figure 6a. Nonparametric part

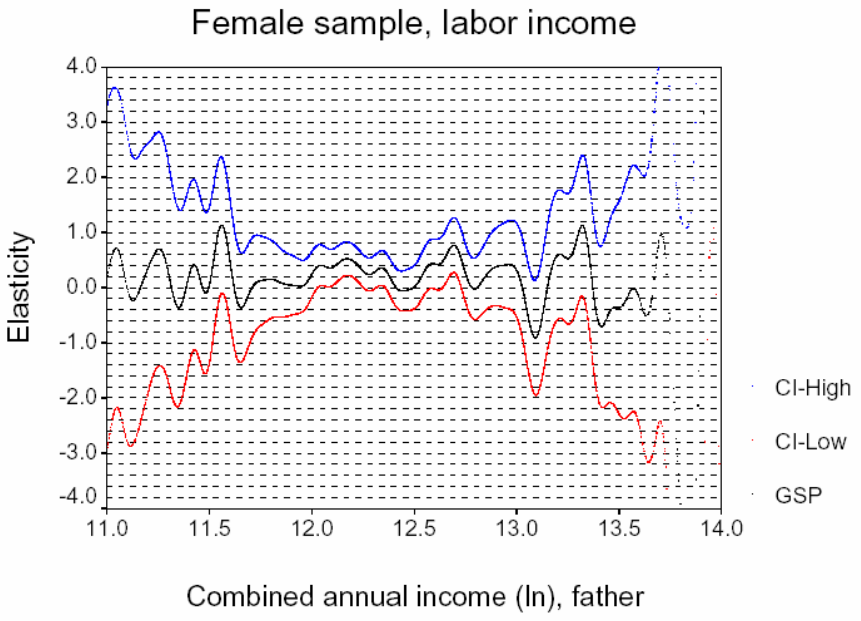

Figure 5b. Indirect Opportunity Sets

Figure 6b. Indirect Opportunity Sets

Figure 7a. Nonparametric part

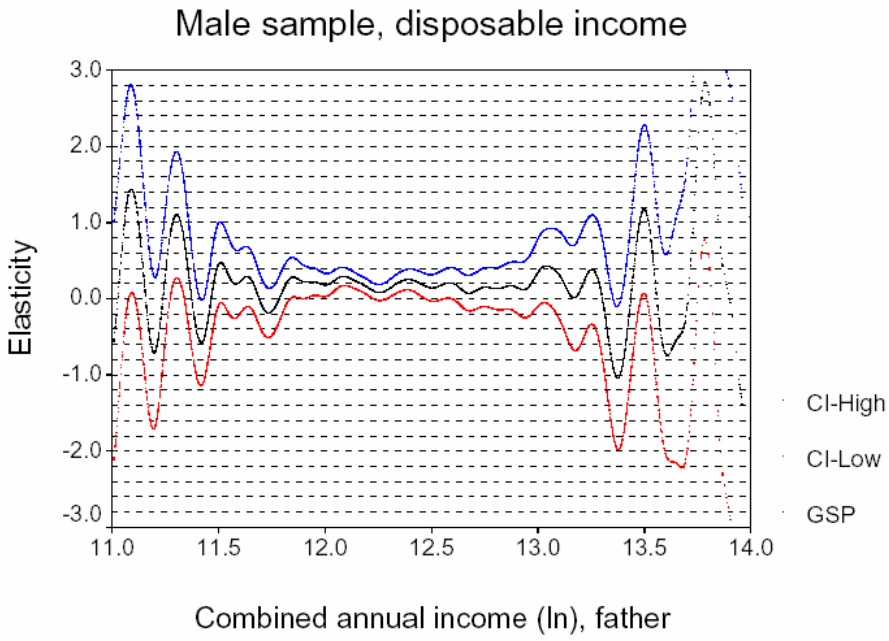

Figure 8a. Nonparametric part

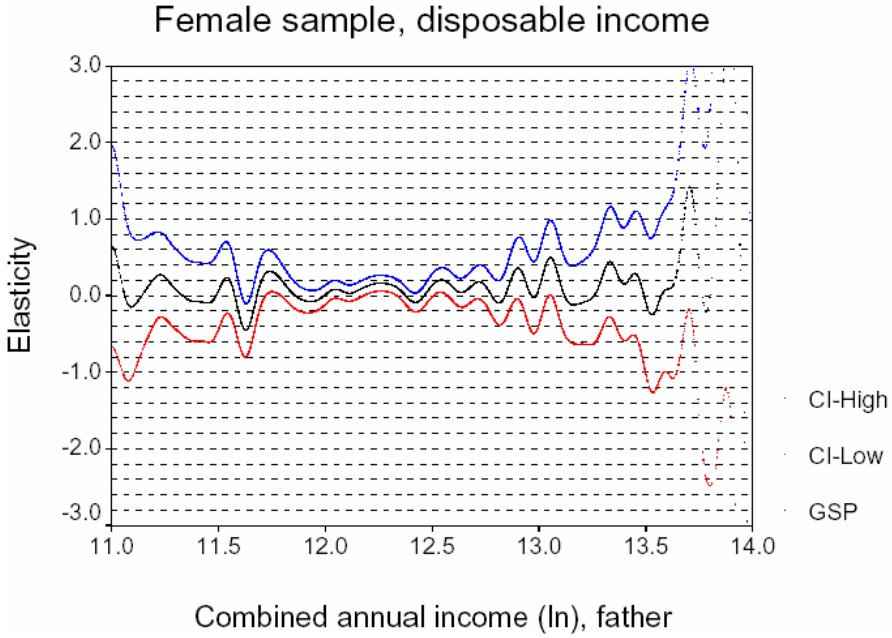

Figure 7b. Indirect Opportunity Sets

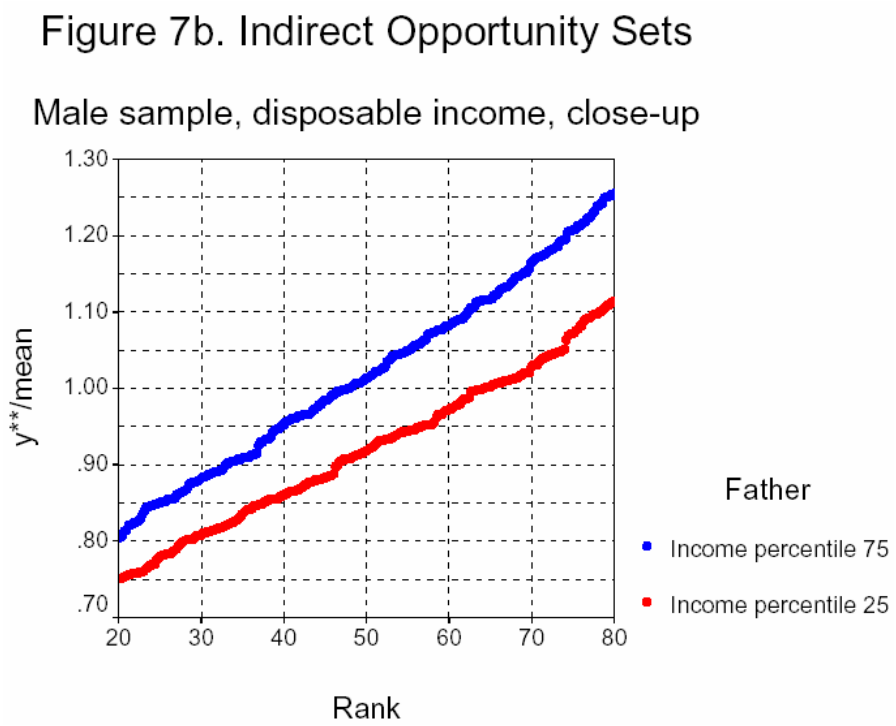

Figure 8b. Indirect Opportunity Sets

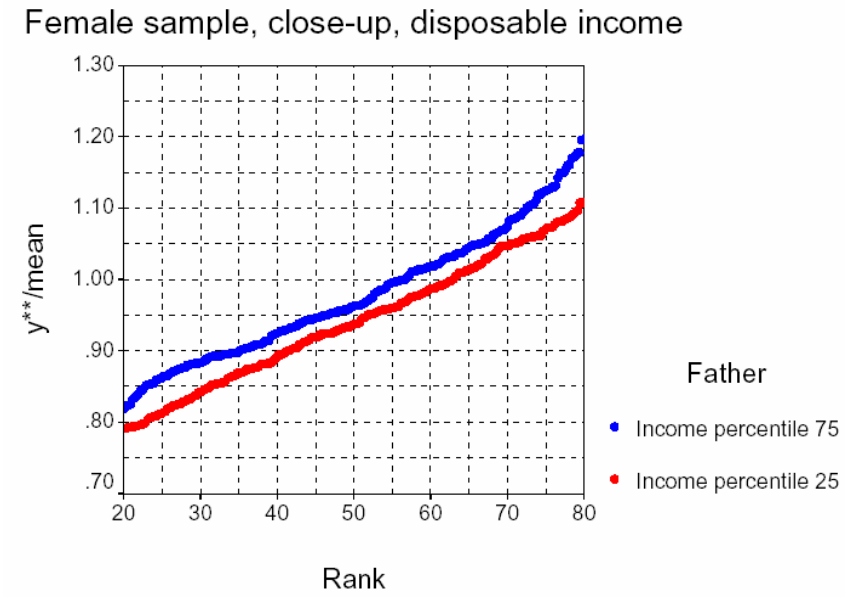

Figure 9. Indirect Opportunity Sets