Reproductive selection and human evolution: An empirical analysis for Spain, 2010

Diego Varela
University of A Coruña
dvarela@udc.es

22 December 2011

Abstract
In this paper I investigate whether there is an economic bias in the reproduction of humans in Spain using data from the Encuesta de Presupuestos Familiares (Household Budget Survey) of 2010. The main result is that parents that are income earners tend to earn 16% more than non-parents income earners of equal age and gender. Parents also have greater probability of being income earners than the rest of the population of equal age and gender. I argue that the contribution of such reproductive selection to economic development can be as high as half a percentage point per year.

Keywords
Evolutionary economics, Sexual selection, reproduction, heritability, human capital, income, fertility

JEL codes
B52 - Institutional; Evolutionary
J13 - Fertility; Family Planning; Child Care; Children; Youth
J24 - Human Capital; Skills; Occupational Choice; Labor Productivity
O15 - Human Resources; Human Development; Income Distribution; Migration
1. Introduction

Sexual selection, a concept introduced by Charles Darwin in his book *On the Origin of Species* (1859), is a significant element of his theory of natural selection. The sexual form of selection “... depends, not on a struggle for existence, but on a struggle between the males for possession of the females; the result is not death to the unsuccessful competitor, but few or no offspring” (Darwin, 1859: 88). Darwin greatly expands his initial three-page treatment of sexual selection in *The Descent of Man and Selection in Relation to Sex* (Darwin, 1871). In summary, while natural selection results from the struggle to survive, sexual selection emerges from the struggle for sex.

The sexual struggle is of two kinds; in the one it is between individuals of the same sex, generally the males, in order to drive away or kill their rivals, the females remaining passive; whilst in the other, the struggle is likewise between the individuals of the same sex, in order to excite or charm those of the opposite sex, generally the females, which no longer remain passive, but select the more agreeable partners.

In this paper I investigate whether this sort of reproductive selection stands when applied to humans from an economic perspective. In other words, I want to investigate whether there is a reproductive selection of the fittest from an economic standpoint and in particular, whether people with higher incomes reproduce themselves to a greater extent than the rest of the population. This, if we assume a certain degree of heritability of income, would lead to a form of human evolution.

Applying the Darwinian idea of sexual selection to human evolution is not straightforward. The issue arises when the extension of contraceptive use breaks the linkage between sex and reproduction and, as a consequence, the struggle for sex does not necessarily have a corresponding effect on human reproduction. It is possible that, even if the fittest have more success in driving away their rivals or charming those of the opposite sex, they will not necessarily have more children.

When sex and reproduction are separate issues, the issue is whether parenthood is a normal good whose demand increases with income, and this is not straightforward either, as there are theoretical and empirical arguments in both directions, as I will show in the next section.

The rest of the paper is divided in three sections. In section two, I undertake a review of the theoretical and empirical literature on socioeconomic selection in human reproduction. In section three I present the data from the Spanish Household Budget Survey and I carry out the econometric analysis of the data. Finally, in the fourth section I present the main conclusions on the existence of reproductive selection and its effect on human evolution.

2. Literature review

If we are to apply a Darwinian concept of reproductive selection as a means of economic evolution of the human species, we should be able to verify that two requirements are met. First, the fittest in economic terms, i.e. the most productive members of society, should reproduce to a greater extent than the rest of the population. Secondly, the economic ability of parent should be
inherited to a certain degree by their descendants. I will review the literature on these two issues in turn.

2.1. The relationship between income and fertility

The most famous and influential theory of population change is that of Malthus, who assumed that populations grow at a rapid rate unless checked by limited supplies of food and other subsistence goods. When incomes fall because the growth in population exceeds the growth in subsistence goods, marriages are delayed, the frequency of coition within marriage is reduced, and fewer children survive to adulthood. The first two factors are “moral restraints” and the last produces “misery” (Malthus, 1826). Malthusian theory thus suggests the existence of a positive relationship between income and fertility.

But as Becker (1960) later pointed out, children are not purchased but self-produced by each family, using market goods and services and the own time of parents, especially of mothers. Thus, the relative cost of children is significantly affected by changes in the value of time of parents, and especially mothers, because the cost of the mother’s time is a major part of the total cost of producing and rearing children. As the value of time and the income of parents are closely related, this price effect would point to a negative indirect relationship between income and fertility. Thus, the net effect of income on fertility would depend on the relative importance of the income and price effects and is a matter of empirical investigation.

Empirical evidence on the relationship between income and fertility is also mixed. On the one hand, several studies document the existence of a positive relationship between income and fertility in several societies early on in the development process. Most of the studies that document such a positive relationship are set in agrarian economies, and often income is proxied by farm size. Examples include Simon (1977), who documents a positive relationship between farm size in hectares and the average numbers of children born for rural areas in Poland in 1948, and Clark and Hamilton (2006), who document a positive relationship between occupational status and the number of surviving children in England in the late 16th and early 17th century (see also Clark, 2005; Clark, 2007). Weir (1995) finds a weakly positive relationship between economic status and fertility in 18th century France, while Wrigley (1961) and Haines (Haines, 1976) document higher fertility in the coalmining areas of France and Prussia than in surrounding agricultural areas during the end of the 19th century. Also, Lee (1987) documents a similar finding using data from the U.S. and Canada.

On the other hand, it is sometimes argued that the fundamental forces determining the demand for children might be different in areas where agriculture is the primary economic activity, and in fact the relationship between income and fertility has reversed with industrialization (Jones et al., 2008). As Becker (1960: 217) points out, ‘most data tend to show a negative relationship between income and fertility.’ This is true of the Census data for 1910, 1940 and 1950, where income is represented by father’s occupation, mother’s education or monthly rental; the data from the Indianapolis survey, the data for nineteenth century Providence families, and several other studies as well." The studies
Becker is referring to are Grabill and U.S. Bureau of the Census (1947), U.S. Census (1956), Whelpton and Kiser (1951), and Jaffe (1940).

Many other studies have documented this kind of negative relationship between income and fertility, typically for a specific geographic area at a particular point in time. For example, Borg (1989) finds a negative relationship using panel data from South Korea in 1976, and Docquier (2004) documents a similar relationship for the U.S. using data from the PSID in 1994. Westoff (1954) finds a negative relationship between fertility and occupational status for the years 1900-1952 using U.S. Census data. In a recent study, Jones and Tertilt (2008) use U.S. Census Data on lifetime fertility and occupations to document this negative cross-sectional relationship in the United States. Income is based on the median annual income for a given occupation in 1950 and adjusted for TFP growth. They find a robust negative cross-sectional relationship between husband’s income and fertility for all cohorts for which data is available, that is for women born between 1826 and 1960.

Part of the literature argues that a negative income-fertility relationship is primarily a statistical fluke, i.e. that it is due to a problem of missing variables. The idea is that once enough variables are controlled for, one would actually find a positive income-fertility relation. Indeed, this was Becker’s original view on the topic. He went into great detail focusing on knowledge of the proper use of contraceptives as the important missing variable. He showed that, in his sample, in those households that were actively engaged in family planning, fertility and income were positively related while the opposite was true for families not engaged in family planning. Other early papers along this line are cited by Becker in his original piece. They include Edin and Hutchinson (1935) and Bash (1955).

Similarly, many authors have argued that a distinction between male and female income is crucial and that the relationship between male income and fertility is indeed (weakly) positive once one correctly controls for female income. Empirical studies distinguishing explicitly between husbands and wives include Cho (1968), Fleischer and Rhodes (1979), Freedman and Thorton (1982), Schultz (1986), Heckman and Walker (1990), Merrigan and Pierre (1998), Blau and van der Klaauw (2007), and Jones and Tertilt (2008). Authors of studies that find a positive relationship after controlling for women’s wages, often interpret such finding as having resolved the “puzzle” but Jones et al. (2008) do not agree. They argue that even though the finding reconciles the conditional correlations in the data with the simplest model of fertility, the question remains of what kind of theories would explain the unconditional negative correlation of men’s wages and fertility.

In any case, it can be argued that there is no reason why the fertility-income relationship should not change over time or vary in different cross sections. It may be that in some subgroups of the population, fertility increases in income once all other relevant correlates are controlled for, while in other subgroups the primary change across the income distribution is in the price of a child and, because of this, fertility is lower at higher income levels (Jones et al., 2008). That is one of the reasons for this empirical investigation.
2.2. The heritability of economic ability

As far as the heritability of economic traits in concerned, this can be due to genetic or social factors. Although the heritability of the IQ, i.e. the portion of the variability in IQs of a population attributable to the effects of genes, is still a controversial question, several Studies estimate that between 50 and 80 percent of the variation can be explained on the basis of genetic factors (Devlin et al., 1994; Bouchard, 1998). Intelligence has proven to have a significant influence as an explanatory variable of class structure in the US, affecting a number of variables related to productivity, such as poverty, schooling, unemployment, family, dependency, parenthood, criminality and citizenship (Herrnstein and Murray, 1994). However, other studies argue that the inheritance of IQ is a relatively minor factor for the intergenerational transmission of economic and social status. For these authors, socioeconomic circumstances such as parents’ income, the father’s years of schooling or occupational status play a more important role in the explanation of the educational performance, income, and socioeconomic status of the child (Bowles and Nelson, 1974). In any case, there is a positive correlation between the incomes of parents and their children.

3. Data and analysis

In order to test the relationship between income and fertility, and its consequences for economic development in Spain, I will use data from the Spanish Encuesta de Presupuestos Familiares (Household Budget Survey). This survey is carried out annually by the Spanish Instituto Nacional de Estadística (National Statistics Institute) and includes over 20 thousand households in its sample. The survey analyzes the expenditure of households resident in Spain, as well as its distribution by different areas of consumption, and offers indispensable information for estimations of household consumption expenditure and for the update of the weights in the consumer price index. The EPF obtains its information by means of personal interviews in randomly selected homes, which cooperate during two consecutive weeks in each of the two years they remain in the sample.

The survey includes information about the personal characteristics of each household member. One of those characteristics is age, which allows identifying the babies born in the year immediately before the survey. There are 641 such babies in the 2010 sample. The information about those babies allows us to identify their parents when they are members of the same household, which happens in most of the cases.
Table 1. Some preliminary evidence on reproductive selection based on income

<table>
<thead>
<tr>
<th></th>
<th>Household members</th>
<th>Average monthly income</th>
<th>Adults (&gt;16-year old)</th>
<th>Average monthly income</th>
<th>Parents</th>
<th>Average monthly income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>30,392</td>
<td>881.25</td>
<td>24,805</td>
<td>1,085.10</td>
<td>591</td>
<td>1,329.10</td>
</tr>
<tr>
<td>Female</td>
<td>31,906</td>
<td>520.75</td>
<td>26,570</td>
<td>627.05</td>
<td>627</td>
<td>803.42</td>
</tr>
<tr>
<td>Total</td>
<td>62,298</td>
<td>696.14</td>
<td>51,375</td>
<td>847.41</td>
<td>1218</td>
<td>1,058.50</td>
</tr>
</tbody>
</table>

Source: Elaborated with data from Encuesta de Presupuestos Familiares 2010, INE.

Table 1 presents some data about the size of the sample that I will use in the next section. There are a total of 22,203 households, with an average of 2.8 members per household, which makes a total of 62,298 members. Out of these, 30,392 are male and 31,906 are female. The sample contains 51,375 members older than 16, out of which 24,895 are male and 26,570 are female. As far as recent parents are concerned, the sample contains a total of 1218, including 591 fathers and 627 mothers.

The survey provides us with information about those parents, including their monthly income, which will allow us to compare them with the rest of the population in order to test whether there is some sort of reproductive selection based on income. Preliminary evidence in Table 1 seems to point in the direction of higher incomes for parents than for the rest of the adult population. Thus, if an average adult earns 847.41 euros per month, an average parent earned 1,058.50, i.e. 24.91% more. The differences seem to be similar for fathers and mothers. Thus, if an average male adult makes 1,085.10 euros per month, an average father makes 1,329.10, i.e. 22.49% more. Similarly, if a female adult earns 627.05 euros per month, an average mother earns 803.42, i.e. 28.13% more. But what if those differences are due to the fact that parents usually have their children in the middle of their lifecycles when their incomes tend to be higher?

Jones and Tertilt (2008) argue that a measure of income based on occupation is a better indicator of lifetime income than income in any particular year. See Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2004) for a description of how occupational income scores (OIS) are constructed as well as its robustness as a proxy for income. For Jones and Tertilt (2008), the focus on husband’s income allows a consistent analysis over time. In particular, it allows the analysis of periods for which data on wife’s income is practically nonexistent. I do not use occupational proxies of income but income itself, which has its advantages, but also requires an extra of caution when making income comparisons between people that may be at different stages of their lifecycles.
Figure 1. Mean income by gender, parenthood and age groups

Source: Elaborated with data from *Encuesta de Presupuestos Familiares 2010*, INE.

Figure 1 shows mean income by gender, parenthood and age groups. The first thing to note is that men seem to have higher incomes than women for all age groups. For both men and women, income tends to increase until it reaches a maximum somewhat in the middle of their lives, and declines thereafter. It also seems that this maximum is reached earlier for women than men. The existence of such a lifecycle confirms the need to control for the stage in the lifecycle in order to make sensible comparisons between parents and the rest of the population.
Preliminary evidence seems to indicate that parents have higher incomes than infertile people of the same gender and age group. This would point in the direction of a sort of income-based reproductive selection with potential implications for human evolution from an economic perspective. The data on Figure 1 show few exceptions to this rule, namely men between 50 and 55 and women between 20 and 25, which are age groups with relatively low numbers of parents (see Figure 2 and Figure 3). Still, we need to undertake further tests if we are to claim that there is a significant difference between parents and the general population as far as income is concerned.

I will opt for a semi-logarithmic specification because it is common in human capital theory, and also because the resulting coefficient for the effect of
Parenthood will be easier to interpret in the context of an evolution model. The equation of the model is as follows:

$$\text{LN(INCOME)} = C + \text{AGE} + \text{AGE}^2 + \text{MALE} + \text{MALE}^*\text{AGE} + \text{MALE}^*\text{AGE}^2 + \text{PARENT} + \epsilon$$  (1)

Where \(C\) is a constant, \(\text{AGE}\) is the age in years. \(\text{MALE}\) and \(\text{PARENT}\) are dummy variables indicating male gender and parenthood in the year immediately before the survey, respectively. The coefficient for \(\text{AGE}\) is expected to be positive because of the general increasing trend in income, whereas the coefficient for \(\text{AGE}^2\) is expected to be negative, due to the concavity of the age-income curve. The dummy variable \(\text{MALE}\) affects does not have a direct interpretation. An interaction term \(\text{MALE}^*\text{AGE}\) is expected to have a positive sign due to the steeper slope of the age-income curve for men than for women. The interaction term \(\text{MALE}^*\text{AGE}^2\) is expected to have a negative coefficient due to the greater concavity of the age-income curve for men. Finally, the coefficient for the dummy variable \(\text{PARENT}\) is expected to be positive.

Table 2. Log-linear regression of income by age, gender and parenthood

<table>
<thead>
<tr>
<th>Dependent Variable: LN(INCOME)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method: Least Squares</td>
</tr>
<tr>
<td>Included observations: 36212</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C)</td>
<td>5.796889</td>
<td>0.043732</td>
<td>132.5561</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\text{AGE})</td>
<td>0.041709</td>
<td>0.001776</td>
<td>23.48158</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\text{AGE}^2)</td>
<td>-0.000434</td>
<td>1.67E-05</td>
<td>-25.95257</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\text{MALE})</td>
<td>-0.300247</td>
<td>0.059705</td>
<td>-5.028800</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\text{MALE}^*\text{AGE})</td>
<td>0.022402</td>
<td>0.002432</td>
<td>9.210649</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\text{MALE}^*\text{AGE}^2)</td>
<td>-0.000182</td>
<td>2.31E-05</td>
<td>-7.888207</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\text{PARENT})</td>
<td>0.162745</td>
<td>0.020040</td>
<td>8.121124</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.124148 Mean dependent var 6.858914
Adjusted R-squared 0.124003 S.D. dependent var 0.647984
S.E. of regression 0.606479 Akaike info criterion 1.837899
Sum squared resid 13316.79 Schwarz criterion 1.839542
Log likelihood -33270.00 F-statistic 855.3132
Durbin-Watson stat 1.627022 Prob(F-statistic) 0.000000

Source: Elaborated with data from Encuesta de Presupuestos Familiares 2010, INE.

Table 2 shows the results of a log-linear regression of income as a function of age, gender and parenthood. The results of the regression confirm the hypotheses, as all the coefficients have the expected sign and are significant at very high levels (p-value = 0.0000). The coefficient for \(\text{AGE}\) (0.04) means that income is expected to grow at a 4% annually. The coefficient for \(\text{AGE}^2\) is negative and significant, as expected, indicating the convexity of the age-income curve. The coefficient for the \(\text{MALE}\) dummy (-0.30) is also highly significant. The coefficient for the interaction term \(\text{MALE}^*\text{AGE}\) (0.02) indicates that men's income grows faster than women's, at round 6% annually. The coefficient for the interaction term \(\text{MALE}^*\text{AGE}^2\) is negative and significant, indicating that the age-income curve is more concave for men than women. Finally, and most importantly, the coefficient for the parenthood variable
PARENT (0.16) is positive and highly significant, and can be interpreted in the sense that parents tend to have 16% higher incomes than the rest of the population of equal gender and age.

The results so far are applicable to those adults that are income earners, i.e. a sample of 36,212 out of 51,375 (70.49% of household members who are 16 or older). In order to test the robustness of the findings, it would also be interesting to know if parenthood is also related to whether a household member is an income earner or not.

Table 3. Logit regression of being an income earner by age, gender and parenthood

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-2.391484</td>
<td>0.089319</td>
<td>-26.77477</td>
<td>0.0000</td>
</tr>
<tr>
<td>AGE</td>
<td>0.112260</td>
<td>0.003908</td>
<td>28.72522</td>
<td>0.0000</td>
</tr>
<tr>
<td>AGE^2</td>
<td>-0.000958</td>
<td>3.89E-05</td>
<td>-24.59836</td>
<td>0.0000</td>
</tr>
<tr>
<td>MALE</td>
<td>-3.095588</td>
<td>0.148848</td>
<td>-20.79695</td>
<td>0.0000</td>
</tr>
<tr>
<td>MALE*AGE</td>
<td>0.161924</td>
<td>0.007017</td>
<td>23.07579</td>
<td>0.0000</td>
</tr>
<tr>
<td>MALE*AGE^2</td>
<td>-0.001250</td>
<td>7.34E-05</td>
<td>-17.02418</td>
<td>0.0000</td>
</tr>
<tr>
<td>PARENT</td>
<td>0.721923</td>
<td>0.074641</td>
<td>9.671946</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Mean dependent var 0.704856  S.D. dependent var 0.456112
S.E. of regression 0.409460  Akaike info criterion 1.014599
Sum squared resid 8612.232  Schwarz criterion 1.015805
Log likelihood -26055.52  Hannan-Quinn criter. 1.014976
Restr. log likelihood -31168.86  Avg. log likelihood -0.507163
LR statistic (6 df) 10226.68  McFadden R-squared 0.164053
Probability(LR stat) 0.000000

Obs with Dep=0 15163  Total obs 51375
Obs with Dep=1 36212

Source: Elaborated with data from Encuesta de Presupuestos Familiares 2010, INE.

Table 3 shows the results of a logit regression of the dummy variable INCOME_EARNERS, which takes the value of 1 for income earners and 0 for the rest. The results of this regression are all consistent with those in Table 2. All the variables included in the model are highly significant and their estimated coefficients have the same sign as in the previous model. According to the model, a typical male who has recently become a father (35.48 years old) has a probability of being an income earner of 89.88% as compared to 81.18% for the rest of males of equal age. Similarly, an average female who has recently become a mother (32.49 years old) has a probability of being an income earner of 72.44% as compared to 56.09% for the rest of females of equal age.
The findings relative to the difference between parents and the rest of the population have potential implications for human evolution and economic development. If children are assumed to inherit the economic ability of their parents, be it because of genetic or social factors, then the next generation will be more productive than their parents. If the average age of parents in the sample is 33.94 years, then an intergenerational increase in productivity of 16% would be equivalent to an annual increase of 0.45 percentage points.

4. Conclusions

The decision to have a child is conditioned by income. In principle, the higher the income, the higher the demand for children. However, higher-income parents (and especially mothers) also tend to face a higher opportunity cost of parenthood. This indirect price effect tends to work in the opposite direction as the income effect. The result is that the final relationship between income and fertility may be positive or negative depending on which factor prevails.

As it is apparent from the evidence presented above, there may be differences across countries and throughout time. It is that a matter for empirical investigation and in this paper I have tried to find out what the relationship is in Spain in 2010.

The results are interesting enough, as they suggest that there is a sort of reproductive selection that makes that the income of parents tends to be around 16% higher than for the rest of the population of the same gender and age. The potential implications of this finding for economic development can be considerable, because if children are assumed to inherit the characteristics of their parents, the contribution of this evolution effect could be up to half a percentage point of annual growth. Apparently, this is a positive result.

However, there is a caveat, and it is that differences in the income-fertility relationship may be due to differences not only in economic conditions but also in public policies. In Europe, for instance, there public provisions for a paid maternity leave based on the mother’s income, which in practice means that higher-income mothers will receive more from the public budget in absolute terms. Conversely, in the US such schemes are not universal. If the reproductive selection that is taking place in Spain is a result of such policies, its economic development gains must be weighed against its cost in budgetary terms. But this is a matter for further investigation.

References


