Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2006/13

Assessing the Causal Effect of Childbearing on Economic Wellbeing in Albania

Francesca Francavilla,
Alessandra Mattei

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Assessing the Causal Effect of Childbearing on Economic Wellbeing in Albania

Francesca Francavilla and Alessandra Mattei

July 18, 2006

Abstract

In this paper we analyze to what extent births may lead to changes in economic wellbeing. In contrast to most previous studies on this issue we apply appropriate econometric techniques based on longitudinal micro data in order to identify the causal effects of child bearing events on income. We perform our analysis on longitudinal data from the Albanian Living Standard Measurement Survey. We take a quasi experimental approach, that is, we consider the experience of a childbearing event as the treatment variable, and our measure of wellbeing as the outcome variable. In order to deal with the confounding due to the presence of systematic differences in background characteristics between the treatment groups, we first fit a multiple linear regression model that includes relevant background characteristics as well as an indicator variable for the treatment (i.e., childbearing). This estimation is then compared and contrasted with a matching approach, based on the bias-corrected matching estimator introduced by Abadie and Imbens (2002). Our analysis suggests that there is some evidence that childbearing events can in fact increase household wellbeing in Albania. In addition, the treatment effect is highly heterogeneous with respect to observable characteristics such as the woman's working status and the woman's parity. All the results appear to be robust with respect to the estimated equivalence scale: changing the equivalence scale leaves the childbearing effect on income positive and non-significant.

KEYWORDS: Matching Estimators, Average Treatment Effects, Unconfoundedness, Longitudinal data, ALSMS, Childbearing, Economic Wellbeing, Equivalence Scale, Sensitivity Analysis.

1 Introduction

The relationship between demographic developments and economic performance has been the subject of rather intense debate in the economics literature for nearly two centuries. Whereas the view of the literature during the 1960s and 1970s seemed to favour the neo-Malthusian view that fertility decline would slow population growth in developing countries and thus reduce poverty, this view came in for a great deal of criticism during the 1980s. The former view argues that high birth rates prevent countries and families from making appropriate long-term investments which may lift them out of poverty. That is, instead of investing in the country's infrastructure and development, resources are diverted to meet immediate needs. Such pessimism spawned many publicly funded family-planning programmes, which were touted as a panacea for poor countries' economic ills (Coale and Hoover, 1958). While many critics have argued that demographic considerations are largely irrelevant to poverty reduction (Kelley, 2001), economists have recently reverted to the view that demographic trends are indeed important. At the same time there has been greater acknowledgements of the importance of timing and intensity of demographic change, the economic and social status of women, and the type and focus of economic policies in countries undergoing demographic change (Birdsall and Sinding, 2001).

However, general theoretical assertions are not sufficient for our understanding of the relationship between demographic and economic processes. As McNicoll (1997) points out, the link between poverty and demography is "an institutionally contingent relationship", depending on the social and institutional environment, as well as policy instruments - such as education, health services and family planning policies - that impinge on both processes.

Until recently limitations on both data sources and statistical techniques have prevented clear insights into the relationship between population growth and economic wellbeing (Birdsall et al., 2001), and most of the existing studies have relied on either cross sectional or aggregate level data. Cross sectional data, no matter what techniques applied, is unlikely to provide robust causal information about the relationship between the occurrence of life events (such as a childbearing event) and economic wellbeing. Past empirical studies concerning the relationship between economic wellbeing and fertility have consequently showed mixed results, indicating that the relationship does not appear to be unidirectional. Some studies suggest a positive relationship, others find it to be negative, and yet others find it to have an inverse J-shaped relationship. Other studies find very little evidence of any relationship at all (see Schoumaker and Tabutin, 1999 for further details).

The focus of this paper is primarily on the relationship between fertility and poverty. Fertility is measured in terms of child bearing events and we analyze to what extent births may lead to changes in economic wellbeing or poverty. In contrast to most previous studies on this issue we apply appropriate econometric techniques based on longitudinal micro data in order to identify the causal effects of child bearing events on poverty. We perform our analysis on longitudinal data from the Albanian Living Standard Measurement Survey (ALSMS). Albania is interesting for a range of reasons. Since 1992, when democracy was re-installed in Albania, the country has experienced rapid political, social and economic changes. However, the country is by far the poorest in Europe, and in terms of the human development indicator, only ranked at 72th out of 177 countries.

We take a quasi experimental approach, that is, we consider the variable of interest (the experience of a childbearing event) as the treatment variable, and our measure of wellbeing as outcome variable. Individuals experiencing a childbearing event might be self selected, generating systematic differences in background characteristics between the treatment groups. In order to deal with this confounding factor, we first fit a multiple linear regression model that includes relevant background characteristics as well as an indicator variable for the treatment (i.e. childbearing). This estimation is then compared and contrasted with a matching approach, which is specifically designed to deal with the problem of confounding in observational studies. We apply the bias-corrected matching estimator introduced by Abadie and Imbens (2002), which allows to regression-adjust the difference within matches for the difference in covariate values. Our analysis suggests that there is some evidence that childbearing events can in fact increase household wellbeing in Albania, although the causal parameter estimate is not significant. In addition, the treatment effect is highly heterogeneous with respect to observable characteristics such as the woman's working status and the woman's parity. All the results appear to be robust with respect to the estimated equivalence scale: changing the equivalence scale leaves the childbearing effect on income positive and non-significant.

The structure of the paper is as follows. Section 2 briefly describes the Albanian context. Section 3 gives a short description of the ALSMS data. Section 4 explains how we define wellbeing putting particular emphasis on the choice of the equivalence scale. Using this wellbeing definition, section 5 provides interesting descriptive patterns of wellbeing for different family types. Section 6 explains the methodological strategy for the causal analysis and Section 7 presents the results along with a dissertation on the robustness of our estimates with respect to the selected equivalence scale. Section 8 draws some conclusions.

2 The Albanian Background

Given the socialist background, Albania has a history of strong social protection. Before the collapse of communism, guaranteed employment schemes protected most families from poverty ensuring them income from earnings. Wages were low but prices and rents were controlled and the state invested extensively in maternal and early child health. Since 1992, when democracy was reinstalled, Albania has enjoyed strong economic growth and its economic progress is rapidly transforming Albania to become a middle income country. From the mid 1990s, the Albania's GNP started to grow and surpassed the so-called Lower Income countries, and currently the GNP is moving toward the levels of the Middle Income countries. Despite the impressive performance of the economy over the last years, Albania continues to have one of the lowest level of per capita income in Europe and the incidence of poverty in Albania is large compared with countries in the region[1]. According to the World Bank Poverty Assessment in 2003 one-quarter of the Albanian population - about 780,000 persons - fell below the poverty line, and around 5 percent of the population - 150,000 persons - are extreme poor. The modernization that the country experienced in

[1] According to the World Development Indicator database and the Country Poverty Assessment Reports by World Bank, Albania is the eighth poorest countries among the transition economies in Europe and Central Asia. Herzegovina (19 percent), FYR of Macedonia (16 percent), Bulgaria (13 percent) and Croatia (8 percent); whereas some recent studies show that the poverty dimension in Albania is near to some countries of the Commonwealth of Independent States such as Uzbekistan and Moldova. See World Bank, Making Transition Work for Everyone: Poverty and Inequality in Europe and Central Asia, 2004.

the last decade has benefited Tirana and other urban areas more than rural areas. Poor individuals in rural areas comprise nearly 35 percent of the population and almost half of residents in the most remote districts in the North and North-East Mountain regions are poor.

Amongst the Southeast European countries, Albania performs badly in many health indicators, education attainment, and dependency ratio (see International Monetary Found, 2005 for further details). The official statistics suggests that in 2003 Albania experienced infant and maternal mortality rates equal to 18 per 1000 births and 21 per 100,000 births, respectively, - which appear to be the highest levels of the Southeast European area. However in the last years the two indicators have reported an encouraging downward trend. Despite these pictures the life expectancy at birth, currently 74 years, is comparable with European countries.

The strong economic growth following the transformation to a market economy, obviously produced rapid and dramatic social changes. Several structural reforms have been carried out involving banking, land reforms and privatization. Almost all the small and medium enterprisers and the strategic sectors (such as telecommunication) have been privatized.

In 1993 the Social Insurance System existing since 1946 was completely reorganized. The new law introduced in 1993 and the following amendments brought substantial changes in the Albanian social assistance program which included old age, disability and survivor pensions, sickness and maternity benefits, work injury, as well as unemployment benefits and family allowances, which were introduced for the first time. There is however, no specific child benefit, but general “economic assistance” is allocated on a means-tested basis for families with low earned income. Employees with at least twelve moths of contributions are entitled to 365 days of paid maternity leave. The benefit is 80 percent of the average daily wage in the last calendar year for the leave period taken before childbirth and for 150 days after, whereas the benefit is paid at 50 percent of the average daily wage for the remainder of the entitlement period. For more children extensions are provided. Compensation is payable for changes of employment due to pregnancy. A lump-sum payment is payable to either insured parents with a minimum of 1 year’s contributions.

Moreover, the Albanian social system provides a child supplement for each dependent child under age 15. It is clear therefore, that there is still reasonably good support available for mothers with young children. Whereas, the total fertility rate has declined steadily over the years, it seems to have stabilized in recent years, and there is little indication that Albania will experience lowest low fertility as experienced in Italy and other Mediterranean countries.

3 The Albania Living Standards Measurement Study

Our analysis is based on data from the Albania Living Standards Measurement Survey (ALSMS), a periodic study carried out by the Albanian Institute of Statistics (INSTAT) with the technical and financial assistance of the World Bank. The first survey was conducted in 2002, and provided individual level and household level socio-economic data from $3,599$ households drawn from urban and rural areas in Albania. The sample was designed to be representative of Albania as a whole, Tirana, other urban/rural locations, and the three main agro-ecological areas (Coastal, Central, and Mountain). The 2002 ALSMS was followed by two panel surveys (in 2003 and 2004) on a sub-sample of the original households. The sample size for the panel took approximately half the ALSMS households and has re-interviewed these households annually in 2003 and 2004. The ALSMS data collected in 2002 therefore constitute "wave 1" of the panel survey and giving three waves of panel data altogether.

The sample selected from the ALSMS for the panel was designed to provide a nationally representative sample of households and individuals within Albania. This differs from the original ALSMS where the sample was designed to be representative of each strata which broadly represented the main regions in Albania so that regional level statistics could be generated (Mountain, Central, Coastal, Tirana). The panel is essentially an individual level survey as individuals are followed over time regardless of the household they live in at any given interview point.

The 2002 survey contains a wealth of information collected at the individual and household levels. Information collected at the household level includes housing, subjective poverty, consumption expenditures, agriculture, non-farm enterprises, and other income. Information collected at the individual level includes demographics, migration, education, health, fertility, labor, transfers and social assistance, and anthropometrics (for children under 6 years of age). The ALSMS also collects community level information on the basic characteristics of the community, access to public services such as education, health, and transportation, community services, community organizations, community safety, migration, child labor and problems related to the environment. Finally, the ALSMS has information concerning price which can be used to adjust for regional price differences.

The two following panel waves provide updated individual level and household level socio-economic data for household members 15 years of age and older. It is important to note that we have no panel information on consumption expenditure. In addition, whereas the first wave contains complete fertility histories, waves 2 and 3 only provide additional information on any new births. All the analyses in this paper are based on a sub-sample of women of child-bearing age (15-49 years) with complete information on the relevant variables drawn from the Albanian panel survey.

4 A Measure of Wellbeing

The focus of our study is on the extent to which childbearing events lead to changes in wellbeing. In order to address this issue we first have to define a measure of wellbeing. As a multidimensional phenomenon, wellbeing can be defined and measured in a multitude of ways. One approach is to think of one's wellbeing as the command over commodities in general, so people are better off if they have a greater command over resources. In this view, the main focus is on whether households or individuals have enough resources to meet their needs, and wellbeing is typically seen in monetary terms. The most common welfare-monetary indicators for poverty measurement are expenditure on household consumption and household income. In our study we use an income-based measure for poverty analysis. This choice was mainly driven by the availability of data. As previously noted, in the ALSM study information on consumption expenditure is only available for the first wave; whereas we have data on income for all the three waves of the Albanian panel survey.

Our measure of monetary wellbeing is constructed using the monthly total household income, which comprises income from dependent work (wages, in-kind salaries, bonuses) as well as non-dependent work, earnings transfer (only incoming), public transfers and other income (such as rental income, inheritance, lottery/gambling winnings and other).

When assessing economic wellbeing it is paramount to adjust for price variability across space and time and household heterogeneity. Conceptually, variability of prices also includes variability in quality and in quantity constraints. Failure to account for such variability can distort comparisons of wellbeing across time and space. Microeconomic theory suggests that we may wish to account for price variability by comparing real as opposed to nominal income. Several procedures can be followed to enable such comparisons. Here we deflate the level of total nominal income by a cost-of-living index. Specifically, we convert income in 2004 to be real with respect to 2002 Leks prices, using the aggregate consumption price index reported by the International Monetary Found (2004).

Household size and demographic composition vary across households, as do the prices they face, including wage rates. As a result, it takes different resources to make ends meet for different households. In other words, differences in household size and composition can be expected to create differences in household "needs", so a simple comparison of aggregate household income can be quite misleading about the wellbeing of individuals in a given household. Most researchers recognize this problem and use some form of normalization. The most straightforward method is to convert from household income to individual income by dividing household income by the number of people in the household. Then, total household income per capita is the measure of welfare assigned to each member of the household. Although this is by far the most common procedure, it is not very satisfactory, for two reasons. First, different individuals have different needs. A young child typically needs less food than an adult, and a manual laborer requires more food than an office worker. Second, there are economies of scale in consumption (at least for such items as housing). It costs less to house a couple than to house two single individuals.

In principle, the solution to this problem is to apply a system of weights. For a household of any given size and demographic composition (such as one male adult, one female adult, and two children), an equivalence scale measures the number of adults to which that household is deemed to be equivalent. So each member of the household counts as some fraction of an adult. Effectively, household size is the sum of these fractions and is not measured in numbers of persons but in numbers of adult equivalents. Economies of scale can be allowed for by transforming the number of adult equivalents into "effective" adult equivalents. Formally, the number of adult equivalents, n_e, in each household is given by

$$n_e = (A + \alpha \cdot K)^{\theta}, \tag{4.1}$$

where A and K stands for number of adults and children, respectively. Both α and θ take a value between 0 and 1. The parameter α is the adult-equivalence of a child, and the parameter θ reflects possible economies of scale favoring larger households, due to the allocation of fixed costs (such as heat and light) over a greater number of people.

The notion of equivalence scale is compelling. It is much less persuasive in practice, because of the problem of picking an appropriate scale. How the parameters α and θ should be calculated and whether it makes sense to even try is still subject to debate, and there is no consensus on the matter. There are two possible solutions to this problem: either pick a scale that seems reasonable on the grounds that even a bad equivalence scale is better than none at all, or try to estimate a scale typically based on observed consumption behavior from household surveys. In our study, preliminary analyses suggested that standard equivalence scales does not work very well. Looking at the cases where α and θ take values of 0.5 or 1, we found that our results were highly sensitive to both the choice of a weight of a child relative to an adult and economies of scale. Therefore, we decided to estimate the equivalence scale from the data.

Following Lanjouw and Ravallion (1995), we focus on the class of equivalence scales whereby the money metric of an individual's welfare has an elasticity θ with respect to household size. As in Lanjouw and Ravallion (1995), the parameter θ is often termed the "size elasticity". The welfare of a typical member of any household is then measured in monetary terms by x/n^θ, where x denotes total household consumption expenditure, and n denotes household size; n^θ can be interpreted as the equivalent number of single-persons.

It is well known that empirical data alone cannot reveal equivalence scales. Additional assumptions are needed to identify equivalence scales from observed data on household consumption patterns. The approach we follow is based on what is sometimes called Engel's second law, which asserts that the food share is an inverse indicator of welfare across households of different sizes and compositions, namely, the higher the share of non-food spending the better off members of the household are deemed to be. Generally, an Engel curve measures the relationship between the expenditure on a particular good and the total expenditure of the household. In our study, as in Lanjouw and Ravallion (1995), we estimate size elasticity by regressing the food share on the log expenditure per person and a set of demographic variables. The basic specification is the following:

$$\begin{aligned} \omega_{ij} &= \mu + \beta \ln(x_{ij}/n_{ij}^\theta) + X_{ij}'\gamma + \nu_j + \epsilon_{ij} \\ &= \mu + \beta \ln(x_{ij}) - \beta\theta \ln(n_{ij}) + X_{ij}'\gamma + \nu_j + \epsilon_{ij}, \end{aligned} \tag{4.2}$$

where ω_{ij} is a food share of household i in village j, x_{ij} is total household expenditure, n_{ij} denotes the number of household members, θ is the size elasticity, X_{ij} is a set of demographic variables, ν_j is community specific characteristics including prices in village j, and ϵ_{ij} represents an error term. We consider a community fixed effect regression in order to control for relative prices across regions. The estimate of size elasticity, θ, is obtained by taking the ratio of the coefficient on log of household size to that of log of household expenditure in equation (4.2).

Recall that, in our application information on consumption expenditure is only available for the first wave; so we estimate size elasticity using observed panel data from wave 1 of

ALSM survey, and apply the estimated equivalence scale both to income in 2002 and income in 2004.

Table 1 shows the results. We consider different specifications of the Engel curve, both imposing the homogeneity restriction, that is $\theta = 1$, (models (5), (6), and (7)) and do not (models (1)-(4)). Column 1 is the simple community fixed effect regression of the food share on the logarithm of the household size. There is a slight tendency for larger households to have higher food shares, but the correlation is not strong (the correlation coefficient is 0.108). When expenditures are added (column 2) the estimated size elasticity of the money metric of welfare is 0.415. The homogeneity restriction is rejected ($t - value = -4.428$). In column 3 we give the augmented model including both household size and household composition (represented by the numbers of people in each demographic group) as independent variables. We obtain a value for θ of 0.221, with a standard error of 0.196. The homogeneity restriction is again rejected ($t - value = -3.975$). For this model the demographic composition parameters are not significant; only if the homogeneity restriction is imposed (column 6), we observe significant even if not strong differences in food shares among households with a different number of adult members. As alternative, model in column 4 includes the demographics as proportion of children in household. This specification gives an elasticity of 0.338, and leads to rejected the homogeneity restriction ($t - value = -5.053$). In addition, the model suggests that there exists a positive although not strong relationship between demographic composition and food share in the Engel curve (the regression coefficient on proportion of children appears to be significant according to a standard two-sided t-test at the 10% level). Therefore, once relaxed, the equivalence scale implied by the Engel curve appears to be approximated well by n^{θ} with adjustment for the proportion of children in household.

Thus, we estimate θ to be 0.338. This size elasticity implies surprisingly large falls in food spending per head for consumers. According to these estimated size economies, ten individuals, each spending, say, 1 Lek per day in separate single-dweller households could achieve the same welfare level living as a 10-person single household with total expenditures of just 2 Leks per day ($10^{0.338} = 2.18$).

[Table 1 about here.]

5 Descriptive Statistics

Table 2 presents some descriptive statistics of the empirical distributions of the (real) total household income in wave 2002 and wave 2004, equivalised using as number of adult equivalents $n_e = n^{\hat{\theta}}$, where $\hat{\theta} = 0.338$ is the estimated size elasticity from model (4) (see section 4). The same descriptive statistics are also presented for the difference between equivalised income in wave 2004 and equivalised income in wave 2002. The descriptive statistics are shown for the sample of 1698 panel women grouped by the number of household members and a binary variable, Z_i, equal to 1 if woman i experienced a childbearing event between the time of the first wave and 31st, December 2003, and 0 if she did not.

[Table 2 about here.]

Table 2 suggests that within the group of women belonging to households with more than two members, there exists negative correlation between the number of household members and the income level both in wave 2002 and in wave 2004. Specifically, this relationship appears to be strong when we compare income level of three-member households with that of four-member households.

Concerning the childbearing status, Table 2 shows that the mean income level in 2004 is higher for women who experienced a childbearing event than for those who did not. The same figure appears comparing income level between the two groups of women in 2002, although the difference is quite small in this case. The summary statistics also suggest that the trend in time of income for women who gave birth differs from that of the other group of women. Looking at the difference between 2004 and 2002 income level, we observe a strong growth in wellbeing within the group of women who experienced a childbearing event, whereas the mean income level for women who did not have a new child decreases, although the difference does not seem relevant.

Table 3 shows the mean values of the components of the (real) total household income in wave 2002 and wave 2004. All the income components are equivalised and real with respect to 2002. Table 3 also presents the average number of workers by household and the percentage of women belonging to a household where at least a member received maternity benefits in the last 12 months. Finally, the last two columns in the table show the mean difference between waves as percentage of the mean in wave 2002. All the descriptive statistics in the table are presented for the sample of 1698 panel women classified by childbearing status.

[Table 3 about here.]

Table 3 suggests that women who experienced a new birth belong to household with a higher number of workers with respect to women who did not. This result characterizes both wave 2002 and wave 2004. However, while the higher number of household workers in 2002 appears to be a consequence of a large number of female and male workers, in 2004 the lower number of female workers is compensated by a higher number of male workers. Looking at the trend in the period 2002-2004 Table 3 suggests that the number of household workers decreases for the two groups of women in the time. However, the reduction in the number of workers in households where there are women who gave birth to a new child is four percent points greater than the reduction in the number of workers experienced by the other households. This is probably due to the fact that households who experienced a childbearing event are affected by a high reduction in the number of female workers ($21,5\%$ with respect to 2002), which is not sufficiently compensated by the increasing in the number of male workers (8% with respect to 2002). This result strongly suggests that there exists a reorganization of labor supply in households who experienced a new birth. Between 2002 and 2004 households who experienced a childbearing event tend to decrease the supply of female work, and increase the supply of male work. On the contrary, households who did not experience a childbearing event seem to reduce their labor supply at all. Table 3 also shows some differences in the income composition of the two groups of women defined by the childbearing status. Both in 2002 and 2003 women who experienced a childbearing event belong to households with a higher self-employed labor income and a lower wage income. This results is partially explained by the fact that women who gave birth to a new child received lower bonuses (which are a component of household income). A further explanation could be related with the higher capacity of household members who work as self-employed to react to a childbearing event modifying their labor supply.

As discussed in section 2, public and private transfers are a crucial component of household income in Albania. The descriptive statistics suggest that women who experienced a new birth belong to households with a slightly higher level of public transfers but a substantial low level of private transfers. This evidence appears in both the waves. Concerning the trend of the different components of income, Table 3 shows an increase in the household wages for both groups of women, although we observe a higher growth, of about six percentage points, among women who experienced a new birth. On the contrary, income of self-employed workers appear to increase for household who experienced a new birth and decrease for households who did not. Table 3 shows that public and private transfers move in opposite directions: the former increases while the latter decreases for both groups of women, even if public transfers increase more for women who experienced a new birth, and private transfer have a heavier reduction for women who did not experience a new birth.

6 Identifying the Causal Effect of a New Birth

6.1 The Quasi-Experimental Approach

The aim of our analysis is to assess whether in Albania a childbearing event leads to changes in wellbeing. We face such a problem using a quasi-experimental approach, that is, we consider the endogenous variable of interest as the treatment variable and a measure of wellbeing as the outcome variable. In our study, the treatment is given by the childbearing status, Z, that is, our binary treatment variable is equal to 1 if a woman experiences a childbearing event between the time of the first wave (t_0) and December 31, 2003, and 0 otherwise. The outcome of interest is the income-based measure of wellbeing at the time of the third wave (t_1) defined in section 4.

More formally, consider a set of N individuals, and denote each of them by subscript i, $i = 1, \ldots, N$. At time t_i ($t_0 < t_i < t_1$), subject i is "treated", i.e., she gives birth to a new child, or "untreated"; in this latter case she will also named "control". The treatment indicator is $Z \in \{0, 1\}$. Interest lies in the continuous scalar outcome representing the equivalised income at the time of the third wave t_1: $Y \in \mathbb{R}_+$. Note that the distance between the treatment assignment - that is, the birth of a new child - and the time at which we observe the outcome variable ($t_1 - t_i$) varies among women. For each individual i, $i = 1, \ldots, N$, with all units exchangeable, let $(Y_i(0), Y_i(1))$ denote the two potential outcomes, that is, $Y_i(0)$ is the income level for individual i when she is not exposed to the treatment, and $Y_i(1)$ is the income level for individual i when she is exposed to the treatment. If both $Y_i(0)$ and $Y_i(1)$ could be observed, then the effect of the treatment on i would be $Y_i(1) - Y_i(0)$. The root of the problem is that only one of the two outcomes is observed. Let the observed outcome be denoted by Y_i:

$$Y_i \equiv Y_i(Z_i) = Z_i \cdot Y_i(1) + (1 - Z_i) \cdot Y_i(0).$$

In this study, we are interested in the estimation of the average treatment effect for the subpopulation of women who experience a childbearing event, usually called, the Average effect of Treatment on the Treated (ATT):

$$\tau = \mathrm{E}\left(Y_i(1) - Y_i(0) | Z = 1\right).$$

If we could observe both outcomes, we could estimate this causal effect using the estimator

$$\frac{1}{N_1} \sum_i Z_i \cdot \left(Y_i(1) - Y_i(0)\right),$$

where $N_1 = \sum_i Z_i$ is the number of treated units in the sample. In practice, for each treated unit i we observe only the income level under treatment, $Y_i(1)$; the untreated income level $Y_i(0)$ have to be estimated.

If the decision to give birth to a new child was "purely random", we could expect that the background characteristics in the treatment groups to be similar, so that comparisons of the groups' outcome variables would measure the effect of the treatment.

However, it is reasonable to believe that subjects who experience childbearing events might be self-selected, and so large differences may exist between women experiencing a new birth and those who do not on observable as well as unobservable covariates, which can lead to severe bias in the estimates of treatment effects.

Tables 4 and 5 show some descriptive statistics for the observed background variables separately for women who experience a childbearing event and women who do not. Table 4 presents, for each continuous covariate, the mean, the standard deviation, and the standardized percentage difference, defined as the mean difference between women who experience a childbearing event and women who do not, as a percentage of the standard deviation: $[100(\bar{x}(1) - \bar{x}(0))]/\sqrt{(s^2(1) + s^2(0))/2}$, where $\bar{x}(1)$ and $\bar{x}(0)$ are the sample means in the childbearing and no-childbearing groups, and $s^2(1)$ and $s^2(0)$ are the corresponding sample variances.

[Table 4 about here.]

Table 5 shows, for categorical covariates, the proportion of women in each category in the two groups defined by the childbearing status, Z, as well as the absolute differences in percentage (third column).

[Table 5 about here.]

As we can seen in Tables 4 and 5, there exists considerable differences between women who experience a childbearing event and women who do not: sixteen of the continuous covariates have standardized differences larger than 10%; and the distributions of most of the categorical variables appear to be substantially different in the two groups of women. These differences indicate the possible extent of the bias when comparing outcomes between the two groups of women due to the different distributions of observed covariates. Therefore, before estimating the causal effect of interest we have to think clearly about the correct way to adjust for the systematic differences in background characteristics.

6.2 Econometric Framework

In our non-experimental context, because treatment and outcome can be endogenous, an identifying assumption is needed to consistently estimate the treatment effects of interest. We assume that assignment to treatment, Z, is independent of the outcome for untreated units, $Y(0)$, conditional on the covariates, X; and that the probability of assignment is bounded away from one. Formally, for all x in the support of X,

(i) (Unconfoundedness) Z is independent of $Y(0)$ conditional on $X = x$;

(ii) (Overlap) $\Pr(Z = 1|X = x) < 1 - \eta$*, for some* $\eta > 0$.

The combination of these two conditions represents a relaxed form of strong ignorability introduced by Rosenbaum and Rubin, 1983 (e.g., Abadie and Imbens, 2002). The first assumption requires that all variables that affect the unobserved outcome and the likelihood of receiving the treatment are observed, and the second one requires that there is sufficient overlap in the probability of receiving the treatment among treated and controls. These conditions are strong, and in many cases may not be satisfied. In many studies, however, researchers have found it useful to consider estimators based on these or similar conditions (see, for example, Rosenbaum and Rubin, 1983; Heckman, Ichimura, and Todd, 1997; Dehejia and Wahba, 1999; Becker and Ichino, 2002).

Under this strong ignorability assumption, the average treatment effect for the subpopulation with $Z = 1$ is equal to:

$$\begin{aligned} \tau &= \mathrm{E}\left(Y(1) - Y(0)|Z = 1\right) \\ &= \mathrm{E}\left(Y(1) - Y(0)|Z = 1, X = x\right) \\ &= \mathrm{E}\left[\mathrm{E}\left(Y|Z = 1, X = x\right) - \mathrm{E}\left(Y|Z = 0, X = x\right)|Z = 1\right] \\ &= \mathrm{E}\left(\tau(x)|Z = 1\right), \end{aligned} \tag{6.1}$$

where the outer expectation is over the distribution of X conditional on $Z = 1$, and $\tau(x)$ is the average treatment effect for the subpopulation with $X = x$ and $Z = 1$. Therefore, under Assumptions (i) and (ii), the τ effect can be estimated by first estimating $\tau(x)$, for all x in the support of X for the treated (say $\mathcal{X}_1$), and then averaging over the distribution of X conditional on $Z = 1$.

A usual way to control for differences in the groups' background variables is to specify a multiple regression of the outcome variable on the covariates, including an indicator variable for treatment status. When the model is well specified, the resulting estimated coefficient of the treatment indicator is a consistent estimate of the average causal effect of the treatment. Hahn (1998) showed that under the unconfoundedness assumption the use of non-parametric series regression adjusting for all covariates can achieve efficiency bounds of the treatment effect.

However, the estimate can be badly biased when the model is not well specified as, for example, when the treatment is assumed constant, but instead it varies depending on the covariate values. In addition, when the data in the treated and comparison groups have different multivariate distributions of the covariates, the fitted regression involves extrapolations over much of the multidimensional covariate space (Rubin, 1997). Such violations of model assumptions can be difficult to detect.

As an alternative to multiple linear regression, we can use matching methods to create groups of treated and control units that have similar background characteristics so that comparisons can be made within these matched groups. For each subject i, matching estimators impute the missing outcome by finding other individuals in the data whose covariates are similar but who were exposed to the other treatment. Specifically, the matching estimator we consider imputes the missing potential outcome, $Y_i(0)$, by using average outcomes for individuals with "similar" values for the covariates. We use matching with replacement, allowing each unit to be used as a match more than once.

A simple way to do this is imputing $Y_i(0)$ for a treated individual ($Z_i = 1$) with covariate values $X = x$ as the average of the outcomes we observe among controls with the similar covariate values $X = x$. When the available covariates for predicting acceptance of treatment are plentiful and/or continuous, such as in our study, the resulting matching estimator can be biased, since it may not be possible to come up with exact matches. Abadie and Imbens (2002) show that subject to some regularity assumptions, the simple matching estimator defined above is inconsistent if the number of (continuous) covariates available for matching exceeds two. In order to address this problem, they develop a bias-corrected matching estimator where the difference within the matches is regression-adjusted for the difference in covariate values.

In our study we apply their bias-corrected matching estimator. Let $\mathcal{J}_M(i)$ be the set of indices for the matches for treated unit i that are at least as close as the Mth match; i.e., for the set $\{j : Z_j = 0\}$, find the M nearest neighbors of i in the predictor space $\mathcal{X}$, using a metric. The missing potential outcome, $Y_i(0)$, is then imputed as

$$\widetilde{Y}_i(0) = \frac{1}{\sharp\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} \Big(Y_l + \widehat{\mu}_0(X_i) - \widehat{\mu}_0(X_l) \Big),$$

where $\sharp\mathcal{J}_M(i)$ is the number of elements of $\mathcal{J}_M(i)$, and $\widehat{\mu}_0(x)$ denotes the estimated regression function for the controls with covariate values $X = x$. The corresponding estimator for τ is

$$\tau_M^{bcm} = \frac{1}{N_T} \sum_{i: Z_i = 1} \Big(Y_i - \widetilde{Y}_i(0) \Big), \tag{6.2}$$

where *bcm* stands for bias-corrected matching.

Our motivation for using this bias-corrected matching estimator is twofold. First, it has better statistical properties than the simple matching estimator. Abadie and Imbens (2002) show that their bias-corrected matching estimator is consistent and has a sampling distribution that is asymptotically normal. In addition, they provide expressions for computing the variance of the bias-corrected estimator making it possible to test the significance of the treatment effect without relying on bootstrapping. Second, in our study, the bias-corrected matching estimator perform much better. It allows us to improve the balancing in the covariates after matching, and to obtain better results in terms of efficiency and robustness[2].

7 Results

In this section, we apply both the regression and the Abadie-Imbens bias-corrected matching approaches to our subsample of panel women from Albania Living Standards Measurement Study (ALSMS) in the attempt to assess the impact of childbearing on economic wellbeing in Albania. Both the regression and matching approaches produce consistent estimates of the treatment effect only when we have controlled for all confounding covariates. When there are important confounding variables that have not been controlled for, either method can lead to biased estimates of treatment effects. It is important to keep in mind, however, that the two methods estimate the ATT effect under different assumptions. The simple linear regression model estimates the average treatment effect assuming that the treatment effect is constant across the subpopulation defined by the covariate values. Therefore, when the treatment effect is a non-constant function of the covariates, the regression model and the matching approach can achieve different estimates of the treatment effect even if each method produces unbiased estimates.

7.1 Regression results

We first estimate the causal effect of interest using a multiple linear regression model of the form $Y|X, Z \sim N(\alpha + X\beta + \gamma Z, \sigma^2)$, where X denote the matrix of background covariates. We control for the geographic characteristics, the socio-demographic and economic variables and the pregnancy history. The regression model also contains a quadratic term for woman's age. Table 6 presents our regression results.

[2]The choice of estimating the causal effect of interest using the bias-corrected matching estimator proposed by Abadie and Imbens (2002) is the result of a lots of preliminary analyses, concerning the selection of an appropriate set of pre-treatment matching variables, which allows us to consider the unconfoundedness assumption reasonable, and the comparison among different matching methods and matching estimators. Specifically, the goals of this preliminary work were: (1) investigating which variables were most likely to confound any comparison between treated and control units in such a way that the assumption that all relevant variables were observed might be a reasonable approximation; (2) choosing the matching method and the matching estimator which gave the best results in term of efficiency and robustness of the estimated effects.

[Table 6 about here.]

The results in Table 6 show that there is a statistically significant shift in the regression equation for women who give birth to a new child relative for women who do not: the birth of a new child causes an increase of living standard by 8.838 Leks by month (with a standard error of 4.056 Leks)[3]. As a reference, note that the observed average monthly income for treated units is $28,632$ Leks. Therefore, for the treated the estimated "counterfactual" average monthly income in the case of no-childbearing is $19,794$ Leks (i.e., $28,632 - 8,838$). This means that having a new child would increase the average monthly income level by 44.6 percentage points (i.e., $100 \cdot 8,838/19,794$) with respect to the "counterfactual" situation of not having a new child. This result is surprising and puzzling. We worry about the scientific validity of the inference drawn from the regression model, which relies heavily on the correct specification of the functional form of the relationship (e.g., linearity) between the outcome and the covariates. In particular, the regression results might be driven by the specific way of extrapolating outcome values from the model (Dehejia and Wahba, 1999; Rubin, 1997). In our data, the observed average monthly income for controls is $21,447$ Leks, which is higher than the estimated "counterfactual" average monthly income in the case of no-treatment for treated women ($17,658$ Leks), so that there is some sign that the regression results can be affected by the specific form of the model to extrapolate estimates of childbearing differences. In addition, the goodness of fit of our model appears to be very poor: the adjusted-R^2 is 8.5%. We could fit different specifications of the model, but we prefer to relax model assumptions by focussing on the matching approach.

7.2 Matching results

The main purpose of matching is to re-establish the conditions of an experiment when no randomized control group is available. The matching method aims to construct the correct sample counterpart for the missing information on the outcome treated for individuals, had they not been treated by pairing each childbearing woman with women of the control group. Also matching estimators depend on the unconfoundedness assumption, but the diagnostics for matching analysis (checking for balance in the covariates) are much more straightforward than those for regression analysis and, enable the researcher to easily determine the range over which comparisons can be supported. Furthermore, the matching approach is more objective in the sense that the comparison group can be constructed without ever looking at the outcome variables. These two aspects of the analysis are inextricably linked in the linear regression analysis.

The literature on matching methods is vast and growing. We apply the Abadie-Imbens bias-corrected matching estimator described in the previous section[4].

[3] US$1.00 equals 105.6 Leks.

[4] All of the analysis is implemented by the use of the `nnmatch` module in STATA (Imbens et al., 2001). This program estimates the average treatment effects either for the overall sample or for the subsample of treated or control units using nearest neighbor matching estimators. The `nnmatch` command implements the specific matching estimators developed in Abadie and Imbens (2002), including their bias-corrected

Here, the biased-corrected matching estimator uses one match and the weighted Euclidean norm to measure the distance between different values for the covariates, with weights given by the inverse of the sample standard errors of the pre-treatment variables used in matching. The bias adjustment uses linear regression on all the pre-treatment covariates in Table 4 and 5, but not higher order terms or interactions. The bias correction is estimated using only the matched units in the comparison group.

7.2.1 Covariate Balance After Matching

To see how well the bias-corrected matching estimator performs in terms of balancing the covariates, Figure 1 and Figure 2 evaluate balance on observed continuous and categorical covariates, respectively, in the matched sample derived from the model.

The matching performs very well in reducing the bias of the background covariates with moderate-large initial standardized differences. For instance, the initial standardized bias for "Age" is 86%, and the matching reduces it to 15%. In addition, exact matches have been obtained for three covariates, "Number of children between 3 and 6 years", "Number of male workers", and "Head's grade level", which have initial standardized differences equal to 34%, 20%, and 33%, respectively.

[Figure 1 about here.]

For the indicator variables "Region" and "Area" we specified exact matching, and for "Religion" exact matching is obtained. The other categorical variables are not matched exactly, but the quality of the matches appears very high: the average difference within the pairs is very small compared to the average difference between treated and comparison units before the matching.

[Figure 2 about here.]

These results suggests that the matched units can be considered sufficiently similar to the treated units. Therefore, provided the unconfoundedness assumption holds, one may proceed to estimate the causal effect of interest.

7.2.2 Estimated Causal Effects

Table 7 presents the estimated average causal effect of childbearing on income for the subpopulation of childbearing women using the Abadie-Imbens bias-corrected matching estimator. The estimate of the ATT effect is equal to $10,416$ Leks, with a standard error of $9,441$ Leks. Thus, as the linear regression model, the matching analysis shows some evidence that giving birth to a new child increases living standard in Albania. In contrast with the regression analysis, however, the matching-based estimate of the ATT effect does not appear to be statistically significant.

---

matching estimator. The procedure `nnmatch` allows individuals to be used as a match more than once. Compared to matching without replacement this generally lowers the bias but increases the variance.

[Table 7 about here.]

There are several considerations behind the positive but negligible effect. First note that, using the Abadie-Imbens bias-corrected matching method, we estimate the "counterfactual" average monthly income for treated women in the case of no-treatment being equal to $17,658$ Leks; this value is lower than the observed pre-treatment income level for the treated, which is equal to $21,826$ Leks. Between wave 2002 and wave 2004 the observed average monthly income level for treated women decreases, but the difference does not appear to be relevant (see Table 2). On the contrary, it seems that whether the treated women did not have experienced a childbearing event, their average monthly income level would be decreased heavily from $21,826$ to $17,658$ Leks. Thus, childbearing appears to have a positive effect for women who would have suffered from a stronger reduction in their monthly income level in absence of childbearing.

Our estimated effect appears consistent with the descriptive statistics shown in Table 2. As we argued in section 5, households where treated women reside seem to undertake a reorganization of the labor supply by increasing the number of male workers and decreasing the number of female workers. This descriptive result is in line with the positive ATT effect in the sense that treated households try to compensate the additional cost of a new member (the newly born child) and the possible loss of an active labor member (the woman who gives birth to the new child) increasing the number of male workers. In fact, it is reasonable that mothers will be completely inundated by the child bearing event whereas the other women of the family assist them with housework, while men focus on the market work. However, as we noted in section 5, the increase of the number of male workers is insufficient to compensate the reduction in the number of female workers. This fact can plausibly affect the significance of our estimated ATT effect.

The positive sign of our estimated effect appears to be also fairly consistent with the Albanian welfare system. According to the Albanian Labor Code, "a woman is entitled to maternity leave provided she has been included in the social insurance scheme for the last 12 months and has been employed with an employment contract from the initial moment of pregnancy until the beginning of maternity leave. Maternity leave benefits are provided for one year, including a minimum of 35 days before delivery and 42 days after delivery. Women carrying more than one child during pregnancy are entitled to 390 days leave, including a minimum of 60 days before delivery and 42 days after delivery. Women in employment receive during maternity leave 80 percent of the average daily payment for the period before delivery and 50 percent of the average daily payment for 150 days after delivery, based on previous year's average salary". For women who are employers or the self-employed, the maternity benefit is equal to the basic old-age pension.

The Albanian Social Insurance System also offers birth grants to an insured person who is the mother or father of a newborn child, provided one of them has contributed for one year prior to the childbirth. The grant is however payable only once and the mother have priority in eligibility, if insured. Birth grant is a lump sum of one-half of the minimum wage.

This system enables working mothers to make informed choices concerning the number and timing of their children. Specifically, maternity benefits and birth grants allow working mothers to recover from childbirth and to care for their newborn infants, providing them for comprehensive protection against income loss due to childbirth and maternity.

These law-based arguments also tally with the descriptive statistics shown in Table 2. As discussed previously in section 5, between wave 2002 and wave 2004 the average self-employed income level and the percentage of household members who received maternity benefits in the last 12 months increase among treated women and decrease among controls. The other income components follow the same trend for treated and control women, but the differences in time appear to be generally more relevant for women who gave birth to a new child (see the last two columns of Table 2).

In spite of these theoretical and practical positive aspects of the Albanian Social Insurance System, we have to keep in mind that parental leave and child support policies are mainly addressed to working women. In our sample, about half of the treated women worked in the week before the first interview, and the other half did not. This distribution of treated women by working status can at least partially explain the fact that the estimated effect is statistically non significant. In other words, we expect that the treatment effect is heterogeneous with respect to woman working status with a stronger and more significant effect for working women.

In addition, recent work on fertility behaviour in Albania during the nineties suggests that "traditionalism" or "norms" persist for the onset of family formation, whereas "modernity" and economic constraints impacts the number of children, especially for third births and higher parities. For instance, using data from the Albanian Living Standard Measurement Survey family, Aassve et al. (2006) show that formation in Albania is still traditional and having (at least) one child is still the norm.

These remarks suggest us to investigate the heterogeneity of the treatment effect along observable characteristics such as "woman's working status" and "number of children". Table 8 shows some sources of heterogeneity in the treatment effect[5]. The most of the estimated effects are statistically negligible, confirming the global analysis, and the corresponding standard errors are sometimes fairly large (e.g., the estimated ATT effect for women with one child - equal to $28,480$ Leks - has a standard error of $33,751$ Leks). This result can be partially due to the small number of observations belonging to each subgroup.

Due to the small size of each subsample and the high sample variability, it is unlike that we can draw robust inference on the size of the childbearing effect in each subgroup of women from our heterogeneity analysis. However, keeping in mind this caveat, we can look at the results in order to obtain some insight on the possible presence of treatment-effect

[5]All the ATT effects are estimated using the Abadie-Imbens matching estimator in its simple form. We do not regression-adjust the results because of the small size of each subsample defined by the marginal and joint values of the two covariates, "woman's working status" and "number of children", which we suspect being source of treatment-effect heterogeneity. For each subsample we first find one match for each treated woman using the weight Euclidean norm to measure the distance between units, with weights given by the inverse of the subsample standard errors of the matching variables. Then, we estimate the ATT effects separately in each subsample.

heterogeneity.

As we can see in Table 8, there appears to be a somewhat strong even if not much significant positive effect of a newly born child on income for working women, whereas this effect becomes small and totally negligible when we focus on women who do not work. This result appears to be consistent with the Albanian Social Insurance system.

Concerning the heterogeneity of the treatment effect with respect to the number of children, we find a significant positive childbearing effect for women who give birth to the first child. This effect appears to be larger for women who have the second child, but in this case it looses much of its significance. Finally, the effect of childbearing for women who already have at least two children is negative and much lower in absolute value. The heterogeneity of the childbearing effect with respect to the initial parity can be linked to the Albanian traditionalism in the family formation in the sense that the birth of the first infant is expected, and so the family is able to prevent income loss due to it. The birth of the second child is still quite normal in Albania, and the cost of the newly born child can be at least partially cushion re-using baby accessories and nursery equipment purchased for the first child. The negative impact of the third birth and higher parities can be due to the fact that the average income level of treated women with two or more children is significantly lower than the income level for the other two groups of treated women.

In order to understand better how treatment effect heterogeneity occurs, we also investigate the differences in the ATT effect across subgroups of women defined by the joint value of "woman's working status" and "number of children". Not unexpected we find a quite strong and highly significant childbearing effect for working women who give birth to the first child in the treatment spell (see Table 8). For women with a child, the working status seems to heavily affect the size of the positive effect, although a standard two-sided t-test suggests the two effects are not significant. For women with at least two children at the time of the first wave, there appears to be no relevant difference in the treatment effect with respect to the working status: we find a negative and barely significant effect of childbearing on living standards.

[Table 8 about here.]

These results suggest that the treatment effect is highly heterogeneous with respect to "woman's working status" and "number of children". Therefore, it may be of substantive interest to investigate whether this heterogeneity in average treatment effects by "woman's working status" and "number of children" is statistically significant or whether it is simply due to the sampling variability. We check whether the observed heterogeneity in the average treatment effects is statistically not negligible by regressing the average effect conditional on "woman's working status" and "number of children" on the two covariates:

$$\tau(x_1, x_2) = \gamma_0 + \gamma_1 \cdot x_1 + \gamma_2 \cdot x_2 + \epsilon,$$

where we denote with x_1 and x_2 "woman's working status", and "woman's parity", respectively. Note that we consider "number of children" as a continuous covariate in this regression model. In order to allow for heteroscedasticity of the average treatment effects, we use a variance-weighted least squares model, where the variance-weights are given by the square of the estimated standard errors of the ATT effects we computed in each subsamples. As we can seen in Table 9, the regression model confirms that there exists relevant heterogeneity in the treatment effect along "woman's working status" and "number of children": all the estimated regression coefficients are statistically significant.

[Table 9 about here.]

7.2.3 Sensitivity of the Estimated Causal Effects to the Equivalence Scale

All the previous estimates rest on the plausibility of our income-based measure of well-being as proxy for poverty, which has been adjusted for differences in household size and composition using an equivalence scale.

We estimated the equivalent scale implied by the data using a variation on the well-known Engel method as described in Lanjouw and Ravallion (1995). Unfortunately, this method has some limitations. Gibson (2002) showed that Engel estimates of size economies are large when household expenditures are obtained by respondent recall but small when expenditures are obtained by daily recording in diaries. This results suggest that the Engel method could not give robust empirical estimates of scale economies, which should not depend on the method used to gather expenditure data. In our study, food consumption was collected by means of a 14-day diary, so we could expect that our estimate of size elasticity ($\theta = 0.338$) is biased downwards.

In addition, the assumption that the food share is an inverse welfare measure across household types, underlying the Engel method, does not always make sense. For instance, consider a larger household with the same per capita expenditures as a smaller household. If there are scale economies, the larger household is better off. Thus, according to Engel's second law, the larger household should have a lower food share. But a decline in the food share with constant per capita expenditures can occur only if there is a decline in food spending per person. It is very unlikely that people who are better off would spend less on food, especially in mid-low income countries where nutritional needs are not being met.

Given these conceptual and empirical problems with the Engel method, it seems important to carry out sensitivity analyses to see whether any conclusions reached previously using our measure of wellbeing are overturned. Our sensitivity analyses is based on equation (4.1), trying different values of α and θ. Specifically, we approximated the continuous function (4.1) with a discrete function on a grid of points: we computed the equivalence scale (4.1) at a set of 20×20 evenly spaced values, (α_j, θ_j), that cover the range of the parameter space of α and θ - that is $[0, 1] \times [0, 1]$. Then, for each $j = 1, \ldots, 400$, we equivalised the household total income using $n_{e,j} = (A + \alpha_j \cdot K)^{\theta_j}$ as equivalence scale, and re-estimated the ATT effect of interest.

As we can see in Figure 3, the estimates of the average treatment effect appear to decrease almost monotonically with respect to the relative cost of a child, α, and the size elasticity, θ, ranging from 3,064 Leks (with a standard error of 3,267 Leks) - which corresponds to $\alpha = \theta = 1$ - to 18,113 (with a standard error of 15,313 Leks) - which corresponds to $\alpha = \theta = 0.05$.

[Figure 3 about here.]

This descending trend also appears looking at the two marginal functions in Figure 4. Examining the trend of the ATT effects with respect to the relative cost of the child when the size elasticity is fixed at its estimated value $\widehat{\theta} = 0.338$ (Figure 4(a)), we see that our estimated causal effect, equals to 10,416 Leks, is the lowest. This means that if the assumption that adults and children have the same weight (equal to 1) does not hold, our estimated average treatment effect would underestimate the real treatment effect. Finally, Figure 4(b) - which shows the distribution of the ATT effect as function of the size elasticity, θ, when the relative cost of the child, α, is fixed to 1 - suggests that our estimated size elasticity could be actually biased downwards, implying an enlargement of the real causal effect.

[Figure 4 about here.]

Our sensitivity analysis allows us to make clear two important remarks. First, all the estimates of the ATT effect we obtain ranging α and θ between 0 and 1 appear to be positive and statistically negligible[6] - confirming the result reached previously; therefore, we are safe to say that our poverty estimates are not heavily affected by the adult equivalence weights that we chose. Second, the sensitivity analysis supports the conclusion that having an additional child has a non-negative effect on the living standards in Albania, although our data seem to be unable to identify the size of this effect.

8 Conclusions

This paper evaluates whether and to what extent a childbearing event changes economic wellbeing for Albanian women. We use a panel sample of women drawn from the Albania Living Standard Measurement Study. Studying the causal relationships between poverty and fertility involves several crucial issues. First, a suitable measure of economic wellbeing is developed. Second, an appropriate econometric methodology is chosen, which works correctly with longitudinal information and takes into account that variation in fertility can be endogenous with respect to wellbeing. We use an income-based measure of wellbeing adjusted for household heterogeneity applying an equivalence scale. We estimate the equivalent scale from the data assuming that the number of adult equivalents in a household is given by the household size to the power of the size elasticity. Following Lanjouw and Ravallion (1995), the implied size elasticity from Engel curve estimation in the ALSMS is 0.338. We then identify the causal effect of a childbearing event on our measure of monetary wellbeing applying both a linear regression model and the Abadie-Imbens bias-corrected matching estimator. Both approaches lead to a positive effect of childbearing on living standards, but whereas the regression model suggests that this effect is highly significant, the Abadie-Imbens bias-corrected matching approach shows a negligible and insignificant

[6] The standard errors are omitted. However, their values along with further details are available on request from the authors.

effect. The regression results are most likely driven by the specific way of extrapolating outcome values from the model, thus preference is given to the results drawn from the Abadie-Imbens bias-corrected matching estimator, which leads to an average causal effect of $10,416$ Leks (s.e. $= 9,441$) for childbearing women.

We find that the treatment effect is fairly heterogeneous along observable characteristics such as woman's working status and woman's parity. Because of the high sample variability and the small number of observations of each subgroup of women defined by the marginal and/or joint values of the two covariates, it is difficult to draw clear insights on the size of the effects in each subsamaple. However, our heterogeneity analysis casts considerable doubt on the hypothesis that the average effect conditional on the covariates is identical for all subpopulations.

All these results rest on the plausibility of our income-based measure of wellbeing as a proxy for poverty, which depends on the estimated equivalence scale. In order to investigate the sensitivity of our results depending on the way in which household size and household composition is taken care of, we re-estimated the ATT effect using different equivalence scales, that is, different values of the parameters α, the weight for a child relative to an adult, and θ, the size elasticity. This sensitivity analysis finds that in Albania the estimated ATT effect is robust with respect to the estimated equivalence scale: all the estimates of the ATT effect appear to be positive and not significant.

There are two main directions for future research. The first is to extend this study by using other measures of wellbeing including multidimensional measures (such as deprivation indices) and subjective measures. Secondly, it is of considerable interest to analyze the conditional distribution of the difference between the two potential outcomes $(Y(1)-Y(0))$ given a childbearing event $(Z = 1)$ as a whole, instead of focussing on its expected value as we have done in this paper.

Acknowledgements

We are grateful to all the participants at the project "Poverty Dynamics and Fertility in Developing Countries" for their support and encouragement. Special thanks are due to Fabrizia Mealli and Steve Pudney for their insightful suggestions and discussions. We also thank Arnstein Aassve and Letizia Mencarini for their detailed comments which have improved the paper.

References

[1] Aassve, A., Burgess, S., and Propper, C. Modeling poverty transitions as the outcome of employment, family unions, and childbearing decisions in the United Kingdom. *Mimeo*, Department of Economics, University of Leicester, 2003.

[2] Aassve, A., Gjonca, A., and Mencarini, L. The highest fertility in Europe for how long? the analysis of fertility change in Albania based on individual data. *European Population Conference* (2006).

[3] Aassve, A., Mencarini, L., and Mazzucco, S. Poverty and household dynamics in the European Union: An application of the difference-in-difference estimator with matching. *Mimeo*, Department of Economics, University of Leicester, 2003.

[4] Abadie, A. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics 113*, 2 (2003), 231–263.

[5] Abadie, A., Drukker, D., Herr Leber, J., and Imbens, G. W. Implementing matching estimators for average treatment effects in Stata. *The Stata Journal 1*, 1 (2001), 1–18.

[6] Abadie, A., and Imbens, G. W. Simple and bias-corrected matching estimators. *Mimeo*, Department of Economics, UC Berkeley, 2002.

[7] Angrist, J. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica 66* (1998), 249–289.

[8] Angrist, J., Imbens, G. W., and Rubin, D. B. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91* (1998), 444–472.

[9] Angrist, J. D., and Krueger, A. B. Empirical strategies in labor economics. In *Handbook of Labor Economics*, A. Ashenfelter and D. Card, Eds., vol. 3. New York, Elsevier Science, 2000.

[10] Ashenfelter, O. Estimating the effect of training programs on earnings. *Review of Economics and Statistics 60* (1978), 47–57.

[11] Bane, M. J., and Ellwood, D. Slipping into and out poverty: The dynamics of spells. *Journal of Human Resources 21* (1986), 1–23.

[12] Barnow, B. S., Cain, G. G., and Goldberger, A. S. Issues in the analysis of selectivity bias. In *Evaluation Studies*, E. Stromsdorfer and G. Farkas, Eds., vol. 5. San Francisco, Sage, 1980.

[13] Becker, S., and Ichino, A. Estimation of average treatment effects based on propensity scores. *The Stata Journal 2*, 4 (2002), 358–377.

[14] Bigsten, A., Kebede, B., and Shimeles, A. Growth and poverty reduction in Ethiopia: Evidence from household panel surveys. *World Development 31*, 1 (2003), 87–106.

[15] Birdsall, N., and Sinding, S. W. How and why population matters: New findings, new issues. In *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*, N. Birdsall, A. C. Kelley, and S. S. W., Eds. New York: Oxford University Press, 2001.

[16] Birdsall, N. M., Kelley, A. C., and Sinding, S. W. *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World.* Oxford University Press, Oxford, Part III. Fertility, Poverty, and the Family, 2001.

[17] Blundell, R., and Costa Dias, M. Alternative approaches to evaluation in empirical microeconomics. *Institute for Fiscal Studies, Cemmap working paper cwp10/02* (2002).

[18] Blundell, R., Dias, M. C., Meghir., C., and Van Reenen, J. Evaluating the employment impact of a mandatory job search assistance program. *IFS Working Paper WP01/20* (2001).

[19] Card, D., and Sullivan, D. G. Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica 56*, 3 (1988), 497–530.

[20] Carletto, G., Azzari, C., Davis, D., and Zezza, A. Monitoring poverty without consumption data: An application using the albania panel survey. *Eastern European Economics 44*, 1 (2006).

[21] Coale, A. J., and Hoover, E. M. *Population Growth and Economic Development in Low-Income Countries: A Case Study of Indias Prospect.* Princeton University Press, 1958, pp. 304–320.

[22] Cochran, W. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics 24* (1968), 295–314.

[23] Dehejia, R., and Wahba, S. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association 94* (1999), 1053–1062.

[24] Dercon, S., and Krishnan, P. Vulnerability, seasonality and poverty in Ethiopia. *Journal of Development Studies 36*, 6 (2000).

[25] Firpo, S. *Efficient Semiparametric Estimation of Quantile Treatment Effects.* PhD thesis, Chapter 2, Department of Economics, University of California, Berkeley, 2003.

[26] Gibson, J. Why does the Engel method work? Food demand, economies of size and household survey methods. *Oxford Bulletin of Economics and Statistics 64*, 4 (2002), 341–360.

[27] Hahn, J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*, 2 (1989), 315–331.

[28] Heckman, J., and Hotz, J. Alternative methods for evaluating the impact of training programs (with discussion). *Journal of the American Statistical Association* (1989).

[29] Heckman, J., Ichimura, H., Smith, J., and Todd, P. Characterizing selection bias using experimental data. *Econometrica 66* (1998), 1017–1098.

[30] Heckman, J., Ichimura, H., and Todd, P. Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies 64* (1997), 605–654.

[31] Heckman, J., Ichimura, H., and Todd, P. Matching as an econometric evaluation estimator. *Review of Economic Studies 65* (1998), 261–294.

[32] Heckman, J., and Robb, R. Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data*, Heckman and Singer, Eds. Cambridge, Cambridge University Press, 1984.

[33] Heckman, J. J., Lalonde, R. J., and Smith, J. A. The economics and econometrics of active labor markets programs. In *Handbook of Labor Economics*, A. Ashenfelter and D. Card, Eds., vol. 3. New York, Elsevier Science, 2000.

[34] Hirano, K., Imbens, G. W., and Ridder, G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* (2003).

[35] Hotz, J., Imbens, G. W., and Mortimer, J. Predicting the efficacy of future training programs using past experiences. *NBER Working Paper* (1999).

[36] International Monetary Fund. The world economic outlook. The global demographic transition. Tech. rep., International Monetary Fund, 2004.

[37] International Monetary Fund. Albania: Selected issues and statistical appendix. Tech. rep., International Monetary Fund, 2005.

[38] Jalan, J., and Ravallion, M. Is transient poverty different? Evidence for rural China. *Journal of Development Studies 36*, 5 (2000).

[39] Justino, P., and Litchfield, J. Welfare in Vietnam during the 1990s: Poverty, inequality and poverty dynamics. Working paper 8, PRUS, 2003.

[40] Kedir, A., and McKay, A. Chronic poverty in urban Ethiopia: Panel data evidence. Nimeo, University of Leicester, UK, 2003.

[41] Kelley, A. C. The population debate in historical perspective: Revisionism revised. In *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*, N. Birdsall, A. C. Kelley, and S. W. Sinding, Eds. New York: Oxford University Press, 2001.

[42] Kim, J., Engelhardt, H., Prskawetz, A., and Aassve, A. Does fertility decrease the welfare of households? An analysis of poverty dynamics and fertility in Indonesia. Working Paper 5, Vienna Institute of Demography, 2005.

[43] Lalonde, R. J. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76* (1986), 604–620.

[44] Lanjouw, P., and Ravallion, M. Poverty and household size. *Economic Journal 105*, 433 (1995), 1415–1434.

[45] Lechner, M. Earnings and employment effects of continuous off-the-job training in East Germany after unification. *Journal of Business and Economic Statistics 17*, 1 (1999), 74–90.

[46] Manski, C. Nonparametric bounds on treatment effects. *American Economic Review Papers and Proceedings 80* (1990), 319–323.

[47] Manski, C. *Identification Problems in the Social Sciences.* Harvard University Press, Cambridge, MA., 1995.

[48] Mcculloch, N., and Baulch, B. Stimulating the impact of policy upon chronic and transitory poverty in rural Pakistan. *Journal of Development Studies 36*, 6 (2000).

[49] McNicoll, G. Population and poverty: A review and restatement. Policy Research Division Working Paper 105, New York: Population Council, 1997.

[50] Neyman, J. On the application of probability theory to agricultural experiments: Essay on principles, section 9, 1923. *Translated in Statistical Science 5* (1990), 465–480.

[51] Ravillion, M. Poverty comparisons. a guide to concepts and methods. *Living Standards Measurement Study, Working Paper 88* (1994).

[52] Rosenbaum, P. T., and Rubin, D. B. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association 79* (1983), 516–524.

[53] Rubin, D. B. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology 66*, 5 (1974), 688–701.

[54] Rubin, D. B. Assignment to a treatment group on the basis of a covariate. *Journal of Educational Statistics 2* (1977), 1–26.

[55] Rubin, D. B. Bayesian inference for causal effects. *Annals of Statistics 6* (1978), 34–58.

[56] Rubin, D. B. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine 127* (1997), 757–763.

[57] Schoumaker, B., and Tabutin, D. Relationship between poverty and fertility in southern countries. Knowledge, methodology and cases. Working paper 2, Department of science of population and development, Universit catholique de Louvain, 1999.

[58] United Nations Development Programme. Human development report. Report, UNDP, 2003.

[59] World Bank. Making transition work for everyone: Poverty and inequality in Europe and Central Asia. Report, World Bank.

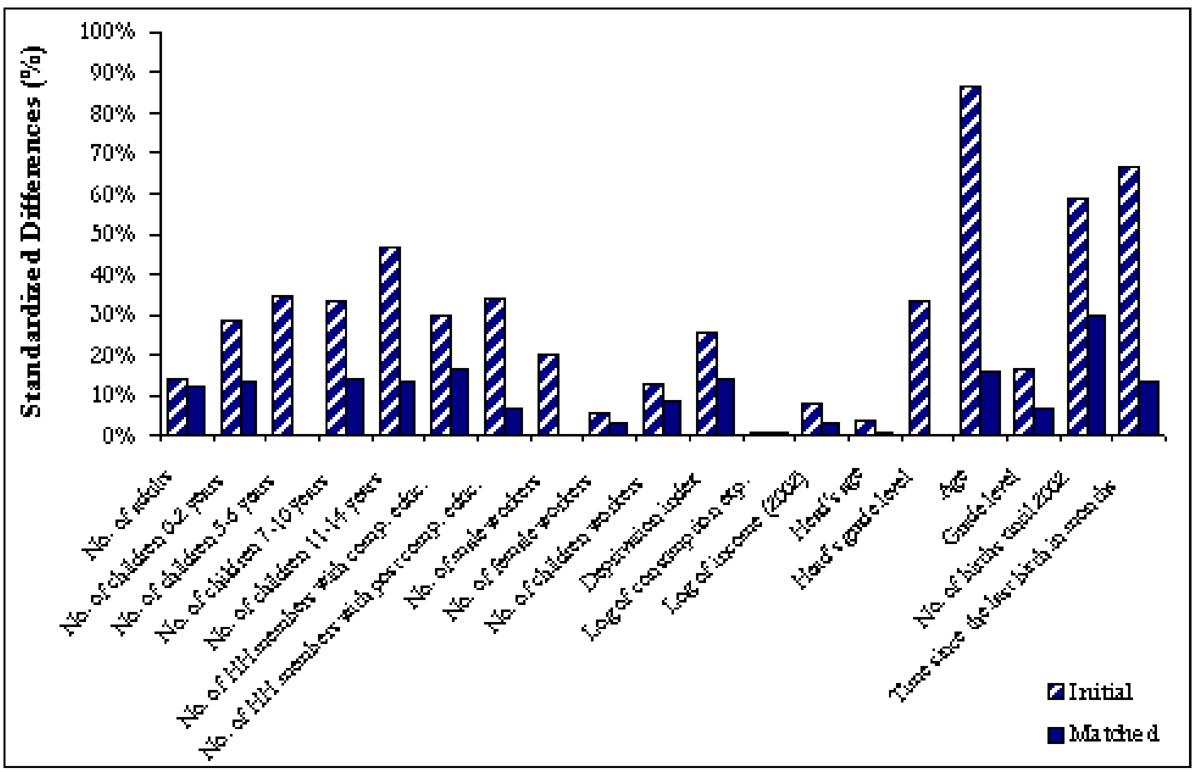

Figure 1: Comparison of standardized differences (in %) for covariates between childbearing and no-childbearing women

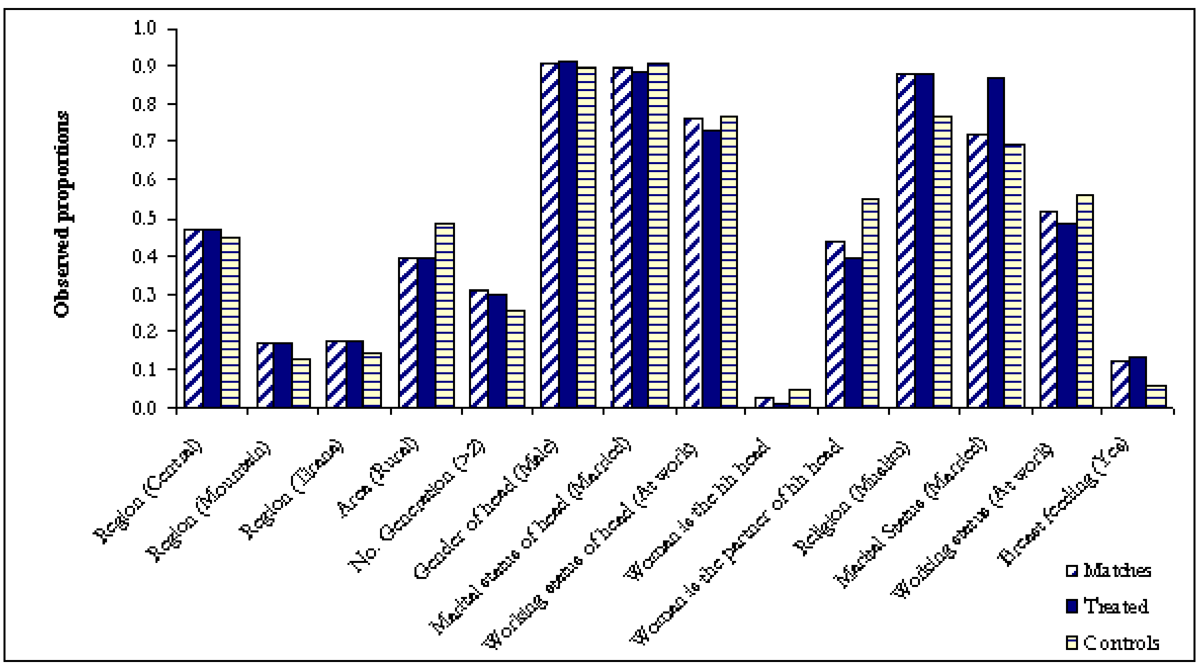

Figure 2: Comparison of observed proportions for categorical covariates between childbearing and no-childbearing women

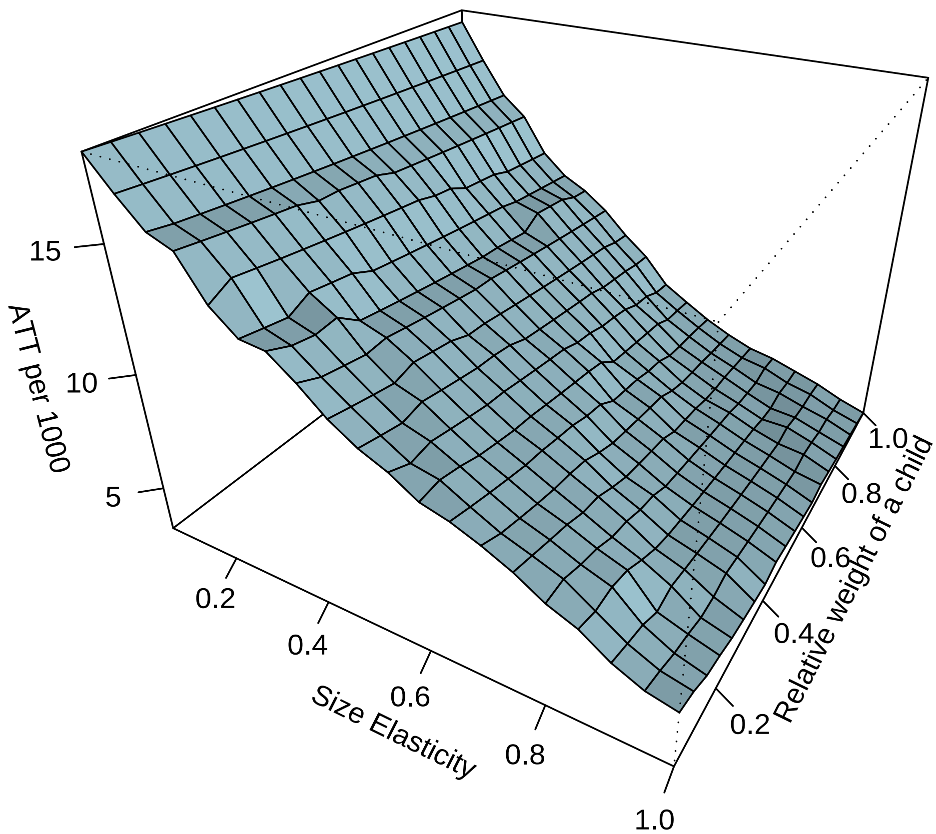

Figure 3: Average Treatment Effect on Treated by Relative Weight of a Child and Size Elasticity

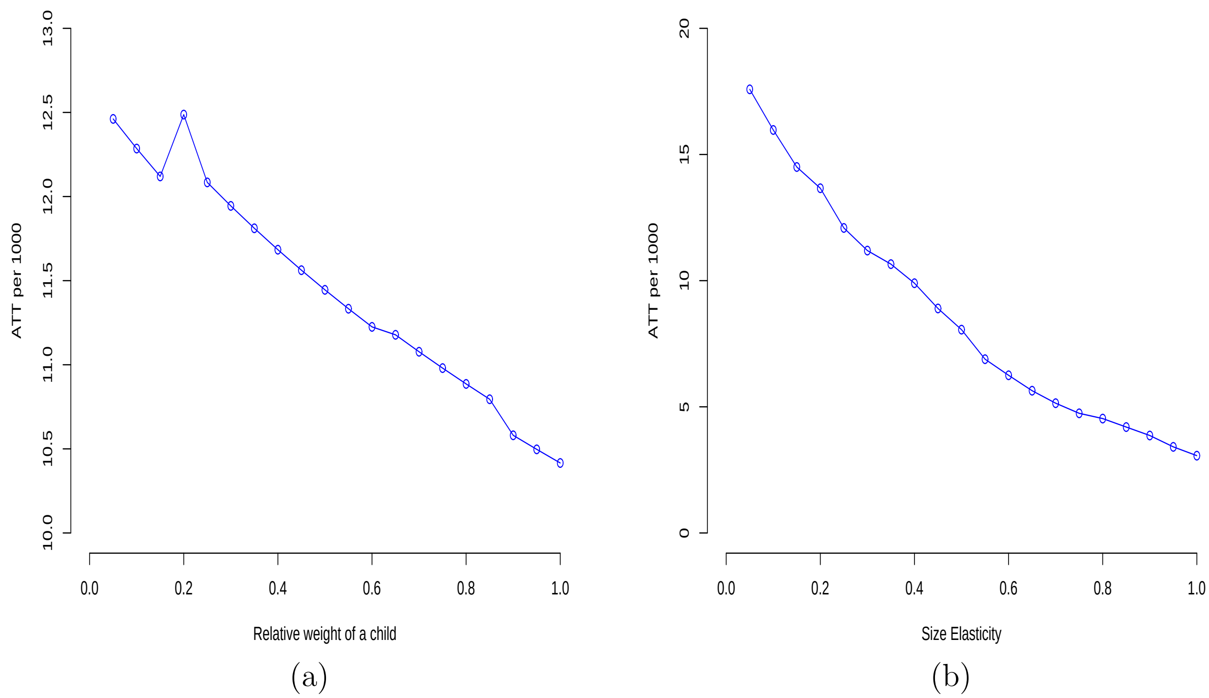

Figure 4: Average Childbearing Effect on Treated: (a) by Relative Weight of a Child (Size Elasticity, θ, equals to 0.338); (b) by Size Elasticity (Relative Weight of a Child, α, equals to 1)

Table 1: Engel Curve Estimation of the Size Elasticity using the first wave of ALSMS. Community Fixed Effect Regression. (Standard errors in parentheses)

	Models						
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log total expenditure		-0.064	-0.062	-0.063			
		(0.008)	(0.008)	(0.008)			
Log household size	0.007	0.026	-0.014	0.021			
	(0.011)	(0.011)	(0.037)	(0.011)			
Log expenditure per person					-0.051	0.060	0.050
					(0.007)	(0.008)	(0.007)
No. of adults			0.005			0.010	
			(0.008)			(0.003)	
No. of children			0.013			0.003	
			(0.009)			(0.003)	
Proportion of children				0.022			0.006
				(0.017)			(0.016)
Constant	0.631	1.262	1.268	1.255	1.092	1.208	1.084
	(0.016)	(0.084)	(0.085)	(0.084)	(0.065)	(0.080)	(0.068)
Observations[a]	1301	1301	1301	1301	1301	1301	1301
No. Communities	283	283	283	283	283	283	283
R-squared	0.0117	0.0498	0.0527	0.0501	0.0568	0.0503	0.0571
Implied Size Elasticity (θ)[b]		0.415	0.221	0.338	1	1	1
		(0.132)	(0.196)	(0.131)			

[a] The number of observations is given by the number of households which our 1698 panel women belonged to at the time of the first wave.

[b] The estimate of size elasticity, θ, is obtained by taking the ratio of the coefficient on log of household size to that of log of household expenditure. The standard error for θ is computed using the Delta method.

Table 2: Means of the real net equivalised income (Leks) by number of household members in 2002 and childbearing status

	OBS	MEANS		
		Wave		
		2002	2004	2004-2002
No of HH Members in 2002				
1	6	29,722	17,015	-12,708
2	53	24,471	21,374	-3,097
3	236	30,027	27,288	-2,739
4	508	21,721	21,735	14
5	390	22,912	23,483	572
6	225	17,691	18,942	1,251
more than 6	280	14,550	18,035	3,485
Childbearing				
Yes	107	21,826	28,632	6,807
No	1,591	21,528	21,447	-80
Total	1,698	21,547	21,900	354

Table 3: Means in wave 2002 and wave 2004, and relative mean differences between waves by childbearing status for income variables and some demographic variables[(a)].

	MEANS				Rel. mean difference (%)[(b)]	
	Wave 2002		Wave 2004			
	Childbearing		Childbearing		Childbearing	
	Yes	No	Yes	No	Yes	No
Woman's bonuses	7	67	25	109	242.6%	62.2%
Wage	7,713	9,304	9,144	10,385	18.5%	11.6%
Income for self-employed	14,069	8,440	16,289	7,383	15.8%	−12.5%
Private Transfer	661	5,562	574	1,307	−13.1%	−76.5%
Public Transfer	1,807	1,736	2,296	2,087	27.1%	20.2%
Total Income	21,826	21,528	28,632	21,447	31.2%	−0.4%
Maternity benefits (Yes)[(c)]	0.9%	0.8%	4.7%	0.5%	3.7%	−0.3%
No. of HH workers	2.20	2.01	2.07	1.98	−5.5%	−1.3%
No. of HH male workers	1.20	1.06	1.29	1.04	7.8%	−1.8%
No. of HH female workers	1.00	0.95	0.79	0.94	−21.5%	−0.7%

[(a)] All the income variables are equivalised using as equivalence scale $n_e = n^{\hat{\theta}}$, where $\hat{\theta} = 0.338$.

[(b)] The relative mean difference is the mean difference between waves as percentage of the mean in wave 2002: $[100(\bar{x}_{2004}(z) - \bar{x}_{2002}(z))]/\bar{x}_{2002}(z)$, where for each variable $\bar{x}_{2004}(z)$ and $\bar{x}_{2002}(z)$ are the sample mean in wave 2002 and wave 2004 in the group of women with $Z = z$, $z = 0, 1$.

[(c)] "Maternity benefits" is a binary variable equal to 1 if at least a household member received maternity benefits in the last 12 months. Therefore, the means are proportions and the relative difference in percent is the percent difference between waves for each group of women defined by childbearing status.

Table 4: Means (standard deviations), and standardized differences in percent for continuous covariate in both treatment groups before matching.

Covariate	Childbearing No mean	(s.d.)	Childbearing Yes mean	(s.d.)	Standardized Difference (%)[b]
Demographic Variables					
No. of adults	3.473	(1.399)	3.673	(1.484)	13.9
No. of children under 2 years	0.217	(0.477)	0.364	(0.573)	28.0
No. of children 3 to 6 years old	0.334	(0.576)	0.561	(0.742)	34.2
No. of children 7 to 10 years old	0.407	(0.612)	0.224	(0.501)	-32.6
No. of children 11 to 14 years old	0.504	(0.666)	0.224	(0.537)	-46.2
Educational Attainment					
No. of household members with:					
sub compulsory education	2.043	(1.562)	2.168	(1.850)	7.3
compulsory education	1.559	(1.410)	1.963	(1.359)	29.1
post compulsory education	1.332	(1.285)	0.916	(1.167)	-34.0
Working Status					
No. of male workers	1.059	(0.700)	1.196	(0.679)	20.0
No. of female workers	0.952	(0.897)	1.000	(1.019)	5.0
No. of children workers	0.089	(0.406)	0.047	(0.253)	-12.4
Measures of Welfare					
Deprivation index	0.358	(0.186)	0.403	(0.176)	25.1
Log of consumption expenditure[a]	9.844	(0.454)	9.841	(0.414)	-0.7
Log of income in wave 2002[a]	9.007	(2.257)	8.827	(2.362)	-7.8
Household Head Characteristics					
Age of the household head	48.223	(11.667)	47.701	(15.252)	-3.8
Grade level of household head	9.724	(3.466)	8.643	(3.140)	-32.7
Woman Characteristics					
Age	31.985	(10.365)	24.832	(5.509)	-86.2
Grade level	9.771	(2.686)	9.362	(2.473)	-15.8
No.of births until 2002	1.926	(1.720)	1.056	(1.204)	-58.6
Time since the last birth in months	96.998	78.526	55.548	41.011	-66.2

[a] The consumption expenditure and income variables are equivalised using as equivalence scale $n_e = n^{\hat{\theta}}$,where $\hat{\theta} = 0.338$.

[b] The standardized difference is the mean difference as a percentage of the average standard deviation: $[100(\bar{x}(1) - \bar{x}(0))]/\sqrt{(s^2(1) + s^2(0))/2}$, where for each covariate $\bar{x}(1)$ and $\bar{x}(0)$ are the samples means in the childbearing and no-childbearing groups and $s^2(1)$ and $s^2(0)$ are the corresponding sample variances.

Table 5: Table of observed proportions and percent differences for categorical covariates.

Covariate		Childbearing No	Childbearing Yes	Difference (%)
Demographic Variables				
Region				
	Costal	0.280	0.187	9.3
	Central	0.449	0.467	1.9
	Mountain	0.129	0.168	3.9
	Tirana	0.142	0.178	3.6
Area				
	Urban	0.518	0.607	
	Rural	0.482	0.393	9.0
No. of generations				
	≤ 2	0.748	0.701	
	> 2	0.252	0.299	4.7
Household Head Characteristics				
Gender				
	Female	0.105	0.084	
	Male	0.895	0.916	2.1
Marital status				
	Unmarried	0.093	0.112	
	Married	0.907	0.888	1.9
Working status				
	Head does not work	0.232	0.271	
	Head works	0.768	0.729	3.9
Woman Characteristics				
Relation to household head				
	Household head	0.043	0.009	3.4
	Partner of the household head	0.549	0.393	15.7
	Other	0.407	0.598	19.1
Religion				
	No Muslim	0.232	0.121	
	Muslim	0.768	0.879	11.0
Marital status				
	Unmarried	0.309	0.131	
	Married	0.691	0.869	17.8
Working status				
	Woman does not work	0.439	0.514	
	Woman works	0.561	0.486	7.5
Currently Breast feeding				
	No	0.940	0.869	
	Yes	0.060	0.131	7.1

Table 6: Regression results (Y = Equivalised household real (with respect to 2002) monthly income at the time of the third wave). Standard errors in parentheses*.

Adjusted R^2	0.085
Overall F-statistic	5.490
Sample size	1698

Covariates	Coef.	(s.e.)
Intercept	0.074	(16.552)
Childbearing status		
No childbearing		
Childbearing	8.838	(4.056)
Household variables		
Region		
Costal		
Central	1.144	(2.278)
Mountain	-4.121	(3.223)
Tirana	9.078	(3.232)
Area		
Rural		
Urban	2.526	(2.600)
Deprivation index	-16.215	(6.090)
Income in wave 1 (per 1000)	-0.013	(0.017)
Consumption expenditure (per 1000)	0.563	(0.097)
No. of generations		
No more than 2		
More than2	1.668	(2.787)
No. of adults	0.789	(1.571)
No. of children under 2 years	2.936	(2.556)
No. of children between 3 and 6 years	-0.030	(1.892)
No. of children between 7 and 10 years	-0.762	(1.791)
No. of children between 11 and 14 years	-0.449	(1.596)
No. of HH members with compulsory education	-0.856	(1.291)
No. of HH members with post compulsory education	-1.909	(1.683)
No. of men who work in Household	2.974	(1.942)
No. of women who work in Household	-2.198	(1.739)
No. of children who work in Household	-1.081	(2.469)

*For the categorical variables, the level which no coefficient value corresponds to, represents the baseline group

Table 6 continued: Regression results (Y = Equivalised household real (with respect to 2002) monthly income at the time of the third wave). Standard errors in parentheses*.

Covariates	Coef.	(s.e.)
<u>*Household head variables*</u>		
Gender		
Female		
Male	0.472	(6.291)
Age	0.260	(0.153)
Marital status		
Unmarried		
Married	1.524	(5.498)
Grade level	0.642	(0.360)
Activity status		
Household head does't work		
Household head works	-1.375	(3.279)
<u>*Woman variables*</u>		
Age	-0.967	(0.899)
Square of Age	0.017	(0.013)
Relation to household head		
Head		
Partner of household head	-1.559	(6.974)
Other	-3.819	(7.996)
Religion		
No Muslim		
Muslim	1.266	(2.269)
Marital status		
Unmarried		
Married	1.786	(4.189)
Grade level	0.588	(0.478)
Working status		
Woman doesn't work		
Woman works	3.086	(2.818)
Number of births until 2002	-0.912	(1.145)
Time since the last birth in months	0.002	(0.026)
Currently Breast feeding		
No		
Yes	-3.831	(4.618)

*For the categorical variables, the level which no coefficient value corresponds to, represents the baseline group

Table 7: Means (standard deviations) for income (Leks) in both treatment groups after matching, and Average Causal Effect of childbearing on income (Leks) for the subpopulation of childbearing women in Albania.

Estimand	Mean	(s.e.)
Income for Childbearing Women	28,632	(90,470)
Income for Matched No-Childbearing Women	17,658	(13,466)
Average Treatment Effect on Childbearing Women	10,416	(9,441)

Table 8: Heterogeneity of the treatment effect (Standard errors in parentheses)

Heterogeneity of the treatment effect with respect to "woman's working status"

Covariate	Treated	Controls	Average Income Controls	Average Income Treated	Average Income Matched	ATT	(s.e.)
Woman does not work	55	698	22,008	22,307	20,500	1,807	(2,592)
Woman works	52	893	21,009	35,322	12,762	22,560	(19,758)

Heterogeneity of the treatment effect with respect to "number of children"

Covariate	Treated	Controls	Average Income Controls	Average Income Treated	Average Income Matched	ATT	(s.e.)
0 children	45	481	19,388	23,689	16,139	7,550	(3,709)
1 child	32	137	22,847	50,006	21,526	28,480	(33,751)
More than 1 child	30	973	22,269	13,250	16,541	-3,291	(2,772)

Heterogeneity of the treatment effect with respect to "woman's working status" and "number of children"

Covariate	Treated	Controls	Average Income Controls	Average Income Treated	Average Income Matched	ATT	(s.e.)
Woman does not work and she has							
0 children	23	264	20,093	25,342	19,489	5,852	(5,570)
1 child	28	64	20,721	22,541	21,046	1,495	(3,631)
more than 1 child	14	370	23,597	17,022	22,371	-5,349	(4,014)
Woman works and she has							
0 children	22	217	18,530	21,960	12,363	9,597	(3,182)
1 child	14	73	24,711	85,318	24,915	60,403	(71,362)
more than 1 child	16	603	21,454	9,949	15,562	-5,613	(4,104)

Table 9: Heterogeneity of the treatment effect: variance-weighted least squares results

Goodness of fit χ^2	1539.150
Model χ^2	2849.030
Sample size	1698

Covariates	Coef.	(s.e.)
Intercept	4.230	(0.195)
Woman's working status*		
Not at work		
At work	1.765	(0.205)
Number of children	-3.038	(0.058)

* Women who do not work, whom no coefficient value corresponds to, represent the baseline group