UNIVERSITÀ
DI SIENA
1240

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Baris Ucar
Gianni Betti

The effect of a newborn on household poverty: a multi-indicator analysis

n. 742 – Dicembre 2016

The effect of a newborn on household poverty: a multi-indicator analysis

Baris Ucar and Gianni Betti

University of Siena

Abstract

The causal relationship between fertility and poverty has not been thoroughly discussed in the existing literature. There are only a few articles directly addressing the issue. Studies which analyze the effect of fertility on overall poverty measures from a dynamic perspective are even fewer.

This study aims to analyze the causal relationship between fertility and poverty. The main issue is to reveal the effect of fertility on poverty. In literature, there are various methods, which attempt to analyze this causal relationship. This study basically focuses on micro level analysis. The relationship is analyzed at household level.

Such a study requires monitoring households throughout time to analyze the differences in their well-being occurring after the birth of a child. Their well-being is examined by using various indicators. In addition to consumption expenditure and income, conventional poverty indicators based on consumption expenditure and income are used along with fuzzy measures of poverty and deprivation index in a comparative way.

The analysis throughout time is possible by making use of data from a panel survey where households are interviewed regularly at different times. For this study, Turkish SILC (Statistics on Income and Living Conditions) Survey data is used. From

Baris Ucar, Institute of Population Studies, Hacettepe University; email: ucarbaris1@hotmail.com;

Gianni Betti, Department of Economics and Statistics, University of Siena; email: gianni.betti@unisi.it;

* This study has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK) through its International Research Fellowship Program for Doctorate Students.

SILC survey it is possible to monitor households within a four years of time span. Since SILC lacks consumption expenditure variable, this is made available by statistical matching from the Household Budget Survey where such information exists.

There are different suggested methods to analyze the causal effect of fertility on household well-being. The most widely used approach depends on using propensity scores, either running a regression or applying propensity score matching (PSM). The causal relationship analyses in this study employs mainly PSM method.

Keywords: Fertility, Childbirth, Poverty, Fuzzy Poverty, SILC, Treatment Effect, Propensity Score Matching

JEL Classification: C21, I32, J13

1. Introduction

Ageing is a recent demographic issue that is seen as a threat and some countries implement pro-natalist policies in order to struggle with it. In this respect, the effect of fertility on household well-being is still an issue that should be in the research agenda. Having information on this effect would enable policy makers to implement better policies targeting fertility and poverty. In this way, necessary support mechanisms could be developed and households having new children could be prevented from falling into poverty. This study uses Turkish data, so the context will be limited to Turkey for some situations.

There are several dimensions of a discussion regarding effects of high fertility and well-being relationship. Starting with a framework of these dimensions would be profitable for putting the question right.

First, it should be discussed that whether it is logical and possible to increase the population continuously. An increased population will obviously be less likely to experience sustainable growth, which indicates an intergenerational dimension of the issue. Therefore, this first dimension for discussion can be defined as the sustainability of ever-increasing population and intergenerational aspects.

The second dimension is the direct negative effects of high fertility on the economy with regard to high child dependency rate. If the total fertility rate rises up to 3 in 2050 and stays constant throughout 2050 to 2075, the population of Turkey would increase to 140.7 million by 2075 (Turkstat, 2013). Considering that Turkey's current population is below 80 million, even such a great increase in population will not be enough to offset ageing, where the old age dependency ratio will increase to 31 per cent in this scenario. In a more probable scenario where total fertility rate decreases in its natural flow and reaches to its lowest value 1.65 in 2050, and then increases after this year and reaches the value 1.85 in 2075, the old age dependency ratio will increase to 48 per cent.

The third dimension is the negative effects of high fertility at household level. As well as having a direct effect on the well-being of households, the effects on the households also have an indirect effect on the economy as a whole via smaller amount of human capital. So this dimension can be studied under two sub-dimensions. First is the direct effects of high fertility on the well-being of households and the second is the

indirect effects of high fertility on the economy depicted as the human capital accumulation.

This study focuses on the third dimension in the defined framework, and analyzes the direct effects of fertility on household well-being in Turkey, as a case study. How a newborn affects the well-being of households and its relationship with household poverty is studied with different indicators.

The increased fertility would lead to diminishing financial means provided to the extra children, which in turn would lead to lower quantity and quality of nutrition, education, etc. provided to those children.

It is necessary to understand the potential negative impacts of higher fertility especially at household level. Whether higher fertility leads to higher poverty and the degree of the lessened economic well-being of the households resulting from more children are important issues which must be dealt with. The analysis of these issues would contribute to the implementations of policies regarding both fertility and poverty which are two of the core policy targets.

This study will analyze the relevance whether families would be poorer because of having more children and the extent of the situation which is to say, the extent of change in average income and consumption expenditure, how many households would be pushed to poverty and for those who are already poor how would exit rate from poverty is different between households with or without a newborn. How the well-being of households is affected as a result of having children will be investigated.

The contribution of this study to the literature is that this subject is put forward for the first time in Turkey. There is no study explicitly addressing the relationship between fertility and poverty at household level in Turkey. Also, the causal relationship between fertility and poverty has not been thoroughly discussed in the existing literature. There are only a few articles directly addressing the issue. The effect on income has not been studied, either. The study is unique in the variety of the well-being indicators employed. This study uses consumption expenditure and income which is not used in previous studies as well as different conventional poverty measures along with monetary and supplementary fuzzy poverty measures and deprivation index which enable comparison with regard to the chosen indicator. Moreover, supplementary fuzzy measures and deprivation index are equivalence scale-free indicators which gives the opportunity for analyses in this regard.

After this introduction section, in section 2, literature review is demonstrated. Section 3 presents the methodology and data. In section 4 the application of the analyses and the results are demonstrated and finally section 5 concludes.

2. Literature review

The relationship between population dynamics and economic well-being has been a matter of interest since old times. Malthus was one of the first to bring out the issue and his theories have been discussed extensively by many researchers. Malthus (1798) suggested that food production increases arithmetically where population increases geometrically resulting in over increasing of population. Malthus was criticized because he didn't take into consideration the pace of technological improvements. Unlike Malthus' views, technological improvements in food production exceeded the population growth in the 20th century and although there was an unprecedented population growth in this century the economy was growing ceaselessly. In spite of this growth some researchers such as Simon (1981); Kahn et al. (1976) indicated that this increase is not limitless, but unlike Malthus' opinion there were expandable limits on population growth.

Most of the literature on population and economic well-being are at macro level. For instance, Eastwood and Lipton (1999) found negative impact of fertility on well-being measured by consumption based poverty. They carried out a cross-national study and the results of the study indicated that higher fertility increased poverty by retarding economic growth and skewing distribution against the poor.

Micro level research at household level is relatively recent and most of this literature depends on the relationship between household size as the indicator of population. Sinding (2009) attribute this scarce literature to scarcity of longitudinal datasets that would enable such research.

Household size has been seen as an important factor of well-being of households. Besides household size, number of children in the household is also used in various studies. Most of the time, these studies are bound to be static analyses at one specific time and are not capable of indicating the change in household composition.

In one such study Desta (2014) found that the effect of a large number of children on consumption expenditure of households is negative for rural households, whereas results for urban households are not as clear.

Among studies that use child birth as fertility indicator, Gupta and Dubey (2003) found that fertility has a significant positive effect on poverty, which is halved when endogeneity of fertility in poverty is taken into account. Mussa (2014) analyzed the relationship between fertility with objective and subjective poverty using IV method based on son preference and concluded that fertility has a positive effect on objective poverty and a negative effect on subjective poverty in Malawi.

Aassve et al. (2005) analyzed the effect of fertility on overall poverty levels for Albania, Ethiopia, Indonesia and Vietnam. In their study, they used dynamic models besides static ones. The dynamic model for fertility aimed to understand how the rate of childbearing differed by background characteristics using longitudinal data. Their model used a probit regression of entry into and exit from poverty between the two waves of the longitudinal data. Their findings suggest that households with many children (i.e. high fertility) tend to have a higher rate of entering poverty and lower rate of exiting poverty.

Aassve et al. (2006) used simultaneous and dynamic random effects models to analyze the causal relationship between fertility and poverty in both directions. Their study indicate that poverty itself has little effect on fertility, whereas there is evidence of state dependence in poverty and important feedback from fertility on future poverty. In the study they use both consumption expenditure and poverty status as indicators of economic well-being.

Recently, there are studies which measure the effect of fertility through analyzing the difference before and after the birth of a child in a household which facilitate dynamic analyses, but the literature is not rich in studies targeting to examine the effect of fertility on poverty in a dynamic perspective. Kim et al. (2009) and Arpino and Aassve (2013) are among these studies in the very rare literature in this respect.

Kim et al. (2009), "Does Fertility Decrease Household Consumption: An Analysis of Poverty Dynamics and Fertility in Indonesia", analyzes the effect of fertility on household consumption by using propensity matching method. The findings assert that when per capita consumption measure is used, household consumption decreases 20 per cent. On the other hand, the magnitude of the effect is rather sensitive to the equivalence scale in use. When different equivalence scales are used instead of per capita measures the magnitude of the effect of fertility on household consumption is found 20 to 65 per cent of the effect found by using per capita measures.

Arpino and Aassve (2013) used different methods for analyzing the causal relationship between fertility and consumption expenditure in Vietnam and found that

those households having children between the recorded waves have considerably worse outcomes in terms of changes in consumption expenditure by using methods based on the unconfoundedness assumption. With the instrumental variable approach which relies on son preference, they estimated a negative impact of fertility on poverty with a magnitude not dramatically different from that obtained by methods based on the unconfoundedness assumption.

3. Methodology

This study particularly aims to analyze the causal relationship between fertility and poverty. The main issue is to reveal the effect of fertility on poverty.

Most of the studies in literature focus on a relationship at one point in time mostly due to lack of data availability. This static analysis limits the quality of the findings of the studies. As Aassve et al. (2005) indicates studies concerned with the dynamic side of poverty are few and none of these have explicitly considered the link with the fertility behaviour. Only after the study of Aassve et al. (2005) and Kim et al. (2005) there are dynamic analyses focusing on the causal relationship between fertility and poverty, but only a few studies are realized. The main target of this study is to explore the effect in a dynamic perspective. Such a study requires monitoring households throughout time to analyze the differences in their well-being.

For analyzing causal relationships there are various methods in literature. In this study, mainly propensity score matching method is used. Some analyses are also made by regression methods based on propensity scores. Instrumental variable (IV) methods for analyzing causal relationships are also abundant in literature, but this method cannot be used in this study due to the small sample size. Moreover, it should mentioned that the outcome measured by PSM and IV methods are different. In PSM methods the outcome is either average treatment effect on the treated (ATT) or average treatment effect (ATE). On the other hand, the outcome is local average treatment effect (LATE) for IV methods.

This study utilizes many indicators of well-being in order to demonstrate the relationship in a multidimensional perspective. Among the well-being indicators that are used are consumption expenditure; income; poverty indicators based on consumption expenditure and income; fuzzy measures of poverty and material deprivation index. The variety of analyzed indicators will enable more robust inferences along with variety in the acquired results. Supplementary fuzzy measures and material deprivation index are

independent from equivalence size effects. Therefore the analyses with these indicators will enable further robust results in this respect.

Regarding the fertility measure, in literature there are studies which use household size and number of children. This is mostly due to lack of data. This study uses the newborn as an indicator of fertility throughout time. Multiple births between waves could create complications in the measurement of the effect, therefore the sample is restricted to households which have only one child between the waves.

3.1. Measuring the causal effect

As mentioned above, there are several methods to analyze the causal effect of fertility on household well-being. Aassve and Arpino (2013) have summarized these methods. There are two main approaches. The first approach depends on unconfoundedness assumption (UNA). Unconfoundedness assumption suggests that there are enough controls in the analysis of causal inference. There aren't unobservable variables in the model used for estimation, which indicates that there is no selection bias. It is also known as Conditional Independence Assumption (CIA), selection on observables or the exogeneity assumption.

Under this assumption either a regression can be run or propensity score (PS) matching method can be applied. Using this method with sensitivity analysis relaxes the UNA to a certain level. The estimates obtained by these methods are called Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT).

This study will use methods that rely on UNA with supplementary sensitivity analysis. Rubin's Causal Model presents a useful framework for the application of analyses for causal inferences, which will be the main approach in this study.

3.2. Rubin's Causal Model

Holland (1986) draws a valuable framework for the understanding of the terminology on causal effects in statistics. He explains Rubin's Causal Model and suggests using terminology of an experiment setting which is the basic framework of the method. First of all, he mentions that the effect of a cause is always relative to another cause. When it is mentioned that A causes B, this means that A causes B relative to other causes. In this regard, the experiment terminology is introduced by saying that these causes are separated as treatment and control indicating the one cause versus another cause.

The population of units under analysis is denoted by U and units in the population are denoted by u. There are only two causes in this framework for simplicity, namely the treatment and control, as mentioned above. Treatment is denoted by t and control by c. A variable S, which indicates the cause is exposed to either t or c. In a controlled study the scientist sets the values for S. In an observational study this is not the case. Y is the response variable, for which the effect of the cause is to be explained. A denotes the characteristics of u.

Variables are divided into two groups with regard to exposure to cause. Before exposure they are called pre-exposure, and after the exposure they are post-exposure. Y is required to be a post-exposure variable to measure its effect. Potentially there are two values for Y for each unit; Yt(u), when the unit is exposed to treatment and Yc(u) when the unit is exposed to control. The difference between the two gives the effect that is looked for. Unfortunately, these two outcomes cannot be observed at the same time for the same unit. This is addressed as the fundamental problem of causal inference by Holland (1986). Fortunately, this can be overcome in a statistical context, by measuring the average causal effect T, which is expected value of the difference between Yt(u)-Yc(u). This can be denoted as,

$E(Yt - Yc) = T$ (1), which can also be written as,

$T = E(Yt) - E(Yc)$ (2)

Although Yt and Yc cannot be observed for the same unit, information can be derived from observed values from different units. Hereby, the statistical solution replaces the impossible-to-observe causal effect of t on a specific unit with the possible-to-estimate average causal effect of t over a population of units (Holland, 1986).

When this is the case, the causal indicator variable S determines whether Yc or Yt is observed for a given unit.

if S(u)=t then Yt(u) is observed,

if S(u)=c then Yc(u) is observed,

The model contains three variables, S, Yt and Yc, on the other hand there are only two observable variables, S and Ys. Hereby,

$E(Ys \mid S = t) = E(Yt \mid S = t)$, (3)

$E(Ys \mid S = c) = E(Yc \mid S = c)$ (4)

The first part of the equation refers to the average treatment effect (ATE) and the second part refers to the average treatment effect on the treated (ATT). According to unconfoundedness assumption the process is randomized which means that the treatment is independent of all other variables. In this case ATE is equal to ATT. So, under this assumption the effect of treatment can be computed by the following equation:

$$E(Ys \mid S=t) - E(Ys \mid S=c) \qquad (5)$$

The model relies also on other assumptions. Stable Unit Treatment Value Assumption (SUTVA) implies that the potential outcomes for a given unit do not vary with the treatments assigned to any other unit, and that there are no different versions of treatment Rubin (1980). In our case, the implication is that fertility behaviours in other households do not have an effect on the outcome of other households. The "different versions" refer to different characteristics of the newborn. The effect of differences in sex, weight and other characteristics of the newborn on the outcome is assumed to be null.

Temporal stability assumes that the effect of the treatment does not depend on when it is applied, and causal transience, assumes that the exposure to control at time t does not affect the result of the exposure to treatment at succeeding times, or vice versa (Holland, 1986).

3.3. Propensity scores

For computation of the treatment effect there is need for specifying the treated and control groups. In order to ensure the plausibility of UNA the treatment and control groups in the process are selected so that their pretreatment characteristics are the same and thus the experiment is fully randomized. When this is accomplished UNA is no more a strong assumption. The treatment effect now depends also on the pretreatment variables (X).

$$E(Ys \mid S=t,X) - E(Ys \mid S=c,X) \qquad (6)$$

If a model that guarantees the two groups are matched according to pretreatment characteristics can be provided then UNA would be more plausible. When

this is done by using many variables the so called curse of dimensionality is faced. There are computational problems while trying to match treatments with controls. A solution to this problem was found by creating propensity scores based on the characteristics of the observations. Propensity score is the conditional probability of receiving a treatment given pretreatment characteristics (Rosenbaum and Rubin, 1983). By making use of propensity scores the dimensions are reduced to a scalar therefore the treatment and control groups can be matched on this.

There are various ways to use propensity scores for estimating treatment effects. In literature there are applications with propensity score matching, regression methods and stratification.

In propensity score matching method treated and control units are matched on the propensity scores. Since in most of the cases it is not possible to find a perfect match between the treated and the controls an interval or a distance is used instead. There are various ways to apply propensity score matching (PSM). Such matching methods will be clarified in the following sections.

PSM method is mainly based on a five steps process (Caliendo and Kopeinig, 2008).

1. Propensity score estimation
2. Choice of matching algorithm
3. Checking overlap / common support
4. Matching quality / effect estimation
5. Sensitivity analysis

In the following paragraphs, these steps are explained.

The propensity scores are derived through a regression model application. While constructing the model to be used in the estimation of propensity scores, there are two main choices to be made. The type of model and the variables to be used in the model should be considered.

The purpose of a model is classification rather than estimating structural coefficients. Therefore any discrete choice model could be used for obtaining propensity scores. In general logit and probit models are used for the estimation of propensity scores when the treatment is binary. Both models produce similar results (Caliendo and Kopeinig, 2008).

The basic idea of propensity score estimation is that because the matching is based on UNA, the outcome variable should be independent of the treatment conditional on propensity score. Therefore the variables to be chosen for the construction of the model and computation of propensity scores should serve to this end.

Omitting important variables increase the bias (Heckman, Ichimura and Todd, 1997). Therefore all important variables should be included in the model. Variables that influence simultaneously the treatment and the outcome variable should be included in the model. Moreover, knowledge on the theory of the research area and literature and also information about the institutional settings should guide the researcher in the construction of the model (Caliendo and Kopeinig, 2008).

Another important issue is that the variables that are unaffected by treatment assignment should be chosen. In order to to ensure this, variables should either be fixed over time or measured before treatment. If the variable is a pretreatment measure also the anticipation effect should be considered and it should be made sure that the variable is not affected by an anticipation of treatment (Caliendo and Kopeinig, 2008).

The number of the variables included in the model is also a matter of concern. Bryson et al. (2002) Augurzky and Schmidt (2000) argue that over-parameterized models should be avoided. Such models could lead to problem in the common support and increase the variance of the estimates. All the same, they wouldn't lead to bias. On the other hand, Rubin and Thomas (1996) recommend against decreasing the number of variables. They argue that a variable should only be excluded from analysis if the variable is either unrelated to the outcome or not a proper covariate. If there are doubts about these two points, they explicitly advise to include the relevant variables in the propensity score estimation. Black and Smith (2003) indicate that when the model is constructed in a minimal setting there are problems with regard to the plausibility of UNA. Austin, (2011) also argues more covariates are better. Because propensity score reduces information from many dimensions into a single score it is less sensitive to model misspecification Therefore concerns about collinearity and model fit do not apply in the context of the propensity score model. Accuracy of propensity score model is less important than the balance on covariates obtained and model misspecification is an iterative process with balance checking.

Arpino (2013) has used community level variables as well as regional dummies in addition to household level variables, in the model construction. Since, our data set

lacks such variables, the variables of the model comprise of household level variables of which some are derived from individual level.

In literature, the taxonomy of matching algorithms for propensity score matching have an intertwined characteristic.

The most commonly used and straightforward matching estimator is nearest neighbor (NN) matching. An observation from the treatment group is matched to an observation in the control group that is closest in terms of propensity score. This can be implemented with or without replacement. In the former case, a control can be used more than once as a match, whereas in the latter case it is considered only once. Matching with replacement involves a trade-off between bias and variance. If replacement is allowed, the average quality of matching will increase and the bias will decrease (Caliendo and Kopeinig, 2008). Also more than one nearest neighbour could be used. Such matching algorithms are called 1:N instead of 1:1 matching, where N is the decided number of neighbours to be selected for each treatment. This also involves a trade-off between variance and bias. As N increases the matches are less quality, hence the bias increases. On the other hand, variance decreases.

Nearest neighbour matching could be conducted by making use of a caliper. The magnitude of the caliper decides the distance of the match and hence the quality of each match. Radius matching is also a type of caliper matching where all of the comparison members within the caliper are matched.

Kernel matching (KM) and local linear matching (LLM) are non-parametric matching estimators that use weighted averages of all observations in the control group to construct the counterfactual outcome. One major advantage of these approaches is the lower variance achieved with more information and the drawback of these methods is that possibly bad matches are generated (Caliendo and Kopeinig, 2008).

Overlap or common support is an important issue in PSM. By common support it is ensured that for every treated unit there are control units with similar propensity scores. Various methods are suggested in the literature. Visual analysis of the density distribution of the propensity score in both groups, comparing the minima and maxima of the propensity score in both groups are among the most straightforward ones. Minima and maxima analysis is based on deleting all observations whose propensity score is smaller than the minimum and larger than the maximum in the opposite group (Caliendo and Kopeinig, 2008). A different method for checking common support is

suggested by Smith and Todd (2005), which is trimming to determine the common support

After the matching is realized in the common support the quality of the matching should be checked. The distributions of the matching variables in the model should be similar. In this study, we use the methodology in Psmatch2 module in STATA. The common support is determined and balancing hypothesis is tested through this algorithm: 1. Split the sample in k equally spaced intervals of e(x) 2. Within each interval test that the average e(x) of treated and untreated do not differ 3. If the test fails, split the interval and test again 4. Continue until, in all intervals, the average e(x) of treated and untreated units do not differ 5. Within each interval, test that the means of each characteristic do not differ between treated and untreated (Leuven and Sianesi, 2003).

Besides this, in a similar way, quality is checked also by calculating the absolute standardized bias (ASB). ASB is defined as the absolute difference of sample means in the treated and matched control subsamples as a percentage of the square root of the average of sample variances in both groups (Rosenbaum and Rubin, 1985).

Regression and matching methods based on propensity scores involve Unconfoundedness Assumption which is also known as the Conditional Independence Assumption or Selection on Observables. It assumes that the model for estimation of propensity scores, includes every variable that explains the treatment and the outcome variables and no nonobservable variables are left. UNA is not informed by the data which means it is untestable by the data set. Since it is a strong assumption, sensitivity tests were generated for the inspection of the assumption. Such sensitivity analyses provide valuable information to draw conclusions on reliability of the estimates. By realizing these tests inference could be made to what degree this assumption holds. There are some indirect and direct tests.

One of the ways to assess the UNA is using multiple control groups. In this method another control group is created and the causal effect of belonging to one of the control groups which is expected to have a zero effect, is estimated (Rosenbaum, 1987). If the effect is found to be different than zero this means at least one of the control groups is invalid (Arpino, 2008).

Another indirect method suggests estimating the causal effect of the treatment on variables which are expected to be unaffected by it. In this case pretreatment

variables are useful for testing. If the effect is not different from zero, it means that UNA is more likely to hold (Imbens, 2004).

In this respect a major contribution was made by Ichino et al. (2008) who suggested a more direct method. The sensitivity analysis proposed by them builds on Rosenbaum and Rubin (1983) and Rosenbaum (1987), and it is based on a simple idea. Parameterization is not necessary in this sensitivity analysis. They suggest that when unconfoundedness is not satisfied given observables but would be satisfied if we could observe an additional binary variable, U.

$$Ys \perp S \mid (X, U).$$

This binary variable can be simulated in the data and used as an additional matching factor. A comparison of the estimates obtained with and without matching on this simulated variable shows the extent of robustness of the estimator to this specific source of failure of the UNA. This simulation is done in two ways in practice. In the first one a binary variable which is significant in the model is used by mimicking this variable. In the second method no such variable is required, but some parameter values are attached which simulate the unobserved variable's (U) distribution and the sensitivity analysis is carried out accordingly. The analysis tests to what extent unobservables impose a threat to the baseline ATT.

$$Pr (Y=1 \mid S, X, U) \neq Pr (Y=1 \mid S, X) \qquad (7)$$

U values that would drive the treatment effects to zero are looked for. These are called "killer confounders" or "dangerous confounders". After this the plausibility of the configuration of parameters that generate such a U value is assessed. If it is not found to be likely, then the PSM is considered to be robust (Nannicini, 2007). For the sensitivity analysis, Sensatt module of STATA is used.

3.4. Well-being indicators for analyses (outcomes)

Consumption expenditure is seen as one of the most important indicators of household well-being in literature. For this reason, an extra effort was realized to make use of this indicator.

Information on consumption expenditures is not available in SILC data. Detailed information on consumption expenditures is collected with another survey, the Household Budget Survey. In this case, to make use of consumption expenditure in this study, the most suitable method is to incorporate information on consumption expenditures in HBS to SILC data. The incorporation of the two surveys is executed by using statistical matching method. and consumption expenditure is fused into SILC data set to be used in this study.

The effect on income is analyzed with two indicators. The first is the total household income and the other one is the income of the mother. By this way, it will be possible to observe a more direct effect of fertility by observing the income forgone by the mother.

For the analyses with regard to poverty the choice should be made for the method of poverty measurement, indicator of poverty, and construction of poverty line. To this end, when conventional poverty measures are calculated this study will utilize the previous poverty measurement methodology of Turkey which was used until 2009. This allows for a fixed poverty line is calculated according to Cost of Basic Needs Approach (Ravallion and Bidani, 1994).

In this methodology, a minimum level of food consumption expenditure is calculated based on a minimum calorie level. By multiplying this value with the food consumption expenditure share in total consumption expenditure of a reference household type, a poverty line is calculated. Poverty measures will be calculated with this methodology at household level. The ratio of the poor refers to the percentage of households that are poor.

For the conventional measures of poverty, a poverty line for the distinction of poor from non-poor is necessary for the analyses. This is a crucial point since, whether a household is poor or non-poor will be decided according to the comparison of its consumption expenditure and income level with the poverty line. The poverty line for each year will be the adjusted values of poverty line from 2009, which is the last year when official conventional poverty measures were presented, using CPI.

For comparing consumption expenditures and incomes with the poverty line there is need for adjustment according to household characteristics, so that the households with different characteristics can be made comparable with the same poverty line. Most widely used is the household size for such adjustment. The adjustment is realized by using equivalence scales. This study will utilize recently computed

equivalence scales for Turkey (Betti et al., 2017). These equivalence scales are the most recent measures derived from Turkish data. Each additional adult corresponds to 0.65 of the first adult and each child correspond to 0.35 of the first adult.

Conventional poverty measures are based on both consumption expenditure and income. The measures are produced for both indicators.

In the conventional approach poverty is characterized by a dichotomization of the population into poor and non poor defined in relation to some chosen poverty line. This approach presents two main limitations: firstly, it is unidimensional, i.e. it refers to only one proxy of poverty, namely low income or consumption expenditure, and secondly it divides the population into a simple dichotomy. However, poverty is a complex phenomenon that cannot be reduced solely to monetary dimension but it must also take account of non–monetary indicators of living conditions; moreover it is not an attribute that characterises an individual in terms of presence or absence, but is rather a vague predicate that manifests itself in different shades and degrees.

The fuzzy approach considers poverty as a matter of degree rather than an attribute that is simply present or absent for individuals in the population.

The Fuzzy Monetary (FM) indicator is defined as a combination of the $(1-F_{(M),i})$ indicator, the proportion of individuals less poor than the person concerned, proposed by Cheli and Lemmi (1995), and of the $(1-L_{(M),i})$ indicator, the share of the total equivalised income received by all individuals less poor than the person concerned, proposed by Betti and Verma (2008). Formally:

$$\mu_i = FM_i = (1-F_{(M),i})^{\alpha-1}(1-L_{(M),i}) = \left(\frac{\sum_{\gamma=i+1}^{n} w_\gamma \mid y_\gamma > y_i}{\sum_{\gamma=2}^{n} w_\gamma \mid y_\gamma > y_1} \right)^{\alpha-1} \left(\frac{\sum_{\gamma=i+1}^{n} w_\gamma y_\gamma \mid y_\gamma > y_i}{\sum_{\gamma=2}^{n} w_\gamma y_\gamma \mid y_\gamma > y_1} \right) \quad (8)$$

where, y_γ is the equivalised income, $F_{(M),i}$ is the income distribution function, w_γ is the sample weight of individual of rank γ $(\gamma = 1,...,n)$ in the ascending income distribution, $L_{(M),i}$ represent the value of the Lorenz curve of income for individual *i*. The parameter α is estimated so that the mean of the FM indicator is equal to the head count ratio computed for the poverty line, which is 60% of the median income.

In addition to the level of monetary income, the standard of living of households and individuals can be described by indicators, such as housing conditions, possession of durable goods, perception of hardship, expectations, norms and values.

According to Betti et al. (2015), aggregation over a group of items in a particular dimension h (h = 1, 2, …, m) is given by a weighted mean taken over j items: $s_{hi} = \sum w_{hj} \cdot s_{hj,i} / w_{hj}$ where w_{hj} is the weight of the j-th deprivation variable in the h-th dimension.

An overall score for the i-th individual is calculated as the unweighted mean:

$$s_i = \frac{\sum_{h=1}^{m} s_{hi}}{m} \tag{9}$$

Then, the Fuzzy Supplementary (FS) indicator for the i-th individual over all dimensions is calculated as:

$$FS_i = \left(1 - F_{(S),i}\right)^{\alpha - 1}\left(1 - L_{(S),i}\right) \tag{10}$$

As for FM indicator, the parameter α is determined so as to make the overall non-monetary deprivation rate numerically identical to the head count ratio computed for the official poverty line (60% of the median). The parameter α estimated is used to calculate the FS indicator for each dimension of deprivation separately.

The FS indicator for the h-th deprivation dimension for the i-th individual is defined as combination of the $(1 - F_{(S),hi})$ indicator the $(1 - L_{(S),hi})$ indicator.

$$\mu_i = FS_{hi} = \left(1 - F_{(S),hi}\right)^{\alpha - 1}\left(1 - L_{(S),hi}\right) = \left[\frac{\sum_{\gamma=i+1}^{n} w_{h\gamma} \mid s_{h\gamma} > s_{hi}}{\sum_{\gamma=2}^{n} w_{h\gamma} \mid s_{h\gamma} > s_{h1}}\right]^{\alpha - 1}\left[\frac{\sum_{\gamma=i+1}^{n} w_{h\gamma} s_{h\gamma} \mid s_{h\gamma} > s_{hi}}{\sum_{\gamma=2}^{n} w_{h\gamma} s_{h\gamma} \mid s_{h\gamma} > s_{h1}}\right], \tag{11}$$

$h = 1, 2, ..., m; i = 1, 2, ..., n; \mu_{hn} = 0$

The $(1 - F_{(S),hi})$ indicator for the i-th individual is the proportion of individuals who are less deprived, in the h-th dimension, than the individual concerned. $F_{(S),hi}$ is the value of the score distribution function evaluated for individual i in dimension h and $w_{h\gamma}$

is the sample weight of the i-th individual of rank γ in the ascending score distribution in the h-th dimension.

The $(1-L_{(S),hi})$ indicator is the share of the total lack of deprivation score assigned to all individuals less deprived than the person concerned. $L_{(S),hi}$ is the value of the Lorenz curve of score in the h-th dimension for the i-th individual. The list of variables used to derive the fuzzy supplementary poverty measure is presented in the appendix.

The material deprivation rate is an indicator in EU-SILC that expresses the inability to afford some items considered by most people to be desirable or even necessary to lead an adequate life. The indicator distinguishes between individuals who cannot afford a certain good or service, and those who do not have this good or service for another reason, e.g. because they do not want or do not need it. The indicator adopted by the Social protection committee measures the percentage of the population that cannot afford at least three of the following nine items (Eurostat, 2016):

- to pay their rent, mortgage or utility bills;
- to keep their home adequately warm;
- to face unexpected expenses;
- to eat meat or proteins regularly;
- to go on holiday;
- a television set;
- a washing machine;
- a car;
- a telephone.

There are three main reasons for the change in consumption expenditure and income in consecutive years as formulized in this study. The first is economic growth in the country which increases the well-being of the average household. The second is inflation which increases the monetary elements in nominal terms. And in our case the third is the birth of a child. Since our target is to single out the effect of child birth we make use of the dynamic approach. The monetary terms are all equivalized with regard to inflation and all monetary values refer to 2013 level. For comparison of treated and control households we take differences between the first period and the following or the last period. Because we are comparing the differences, the outcome under investigation

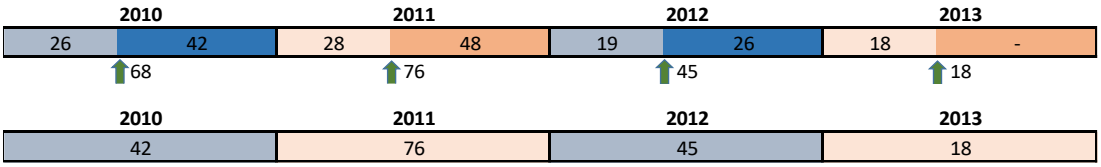
will be the comparison of differences in change of well-being (which could also be defined as economic growth for solely monetary indicators) between two periods.

3.5. Timing of the birth

In causal inference the role of time is profoundly important (Holland, 1986). Contrary to this fact, the literature analyzing the causal effect of fertility on well-being lacks the discussion on timing of the birth and compare births between two periods without considering when the birth has occurred. In general, births between panels waves of four or five years are analyzed with regard to their effect on well-being. On the other hand, all births are considered the same disregarding the time of the birth. Doubtlessly these births have different effects according to the time when the birth took place. This study will try to distinguish the difference that depends on the time of the birth. The shortcoming for this analysis is that, as the births are stratified according to time periods, there will be less observations left for examination. All the same, an effort will be to distinguish time effects to the possible point.

Because there is no available pre-treatment information for the 26 households that had a child before the interview in 2010, we exclude these households from the analyses. They are right at the border of analyses, son they are not kept in the controls either.

The household size is based on the household roaster and determined at the time of the interview. There are two reasons why this is accepted as it is and no action is taken. Firstly, because almost all other collected variables regarding household and individuals refer to the time of interview and secondly the births are spread within the year so their effect does not apply directly to the year after the birth, so the interview date also provides the average effect. In Turkey, recently infant mortality rate is around as low as 10 per thousand, so any such case that would exist in the data is ignored.



The timing of changes in consumption expenditure and income with regard to child birth also require a closer look. It is hard to tell the exact time a change would commence. The household could be starting to increase its consumption before the birth

and this would mean a smaller amount of change after the birth which is hard to tell with the data at hand. Income as well as consumption expenditure could be affected by pregnancy rather than birth. Such analyses in depth, requires longer panels with bigger samples. All the same, in this study we will conduct some analyses taking into consideration the timing as so long as the data sets enables such analyses.

3.6. Unit of analysis, sample restrictions and weights

The measurements of well-being are at household level and the measured effect will be at household level. Since the analyses will be made at household level the unit of analysis will be households. The SILC longitudinal survey follows individuals and provides weights at individual level. But in this case, because there is need for household level measurements, individual level weights are not very convenient for use. Household weights can be derived by making use of individual weights, but since there entries to and exits from households which lead to household splits and formation of new households there is no way to follow all households throughout the panel waves. Also, the child birth, which is the main point of issue in this study requires a change in the household structure between the waves.

The analyses will be restricted to households which have only one married woman of age between 15 and 49, and in which there were no other changes, with regard to its formation between the waves, with the exception of a child birth. This way the households are restricted to those having exactly one woman that is subject to risk of child bearing.

Also the treated households will be restricted to those which have only one child between the waves. When such restrictions are made, household weights could be derived for the last wave of the panel and the analyses could be made accordingly. But the weights will no more be representative of the whole population, therefore such effort would be unnecessary. Taking all these issues into consideration, the use of weights will be dropped. Restricting the sample and dropping of weights cause bias which can't be measured. This will be one of the shortcomings of the study. On the other hand, dealing with complexity of household formation and deriving household weights from individual weights for a sample which is no more representative of the whole population will be more problematic an issue than the otherwise.

Another restriction was made at this point to reduce the outlier effect and households with either consumption expenditure or income over 100 000 Turkish liras

(TL) were excluded from the subsample. This choice is arbitrary and decreasing the level would lead to further cuts in the sample which is not desired. After this restriction, total number of households were 947 and households with a childbirth between 2010 and 2013 were 181. The number of the treated in the analyses is 181 and the control is 766.

3.7. Data

The main data source for this study is the Turkish Income and Living Conditions (SILC) Survey. This survey has been conducted every year since 2006. It is carried out yearly by using panel survey technique for displaying the income distribution between individuals and households, measuring the living conditions of the people, social exclusion and poverty with the income dimension. The aim of the survey is to produce comparable data with the EU Countries, on income distribution, relative poverty, living conditions and social exclusion. It is applied according to the European Union Compliance Programme.

Respondents in the sample are monitored for four years in this survey. Panel survey technique is used and field application is carried out regularly every year. 25 per cent of the households are changed every year and this way, cross-sectional and panel data are obtained from the survey for each year.

4. Application of the analyses and results

In order to demonstrate the effect of fertility there is need for selecting households that are more likely to have children. In this respect, the households with married women aged between 15-49 are selected. As also suggested by Aassve and Arpino (2007) this is a useful selection as it eliminates units that are unlikely to experience child birth.

After the restrictions are applied in the data set, the sample is decreased to 947 observations. Among these 181 are treatments, meaning that there was exactly one birth in these households after the first interview in 2010 and before the last interview in 2013. When the differences of income and consumption expenditure between the first and last year are compared, it can be seen that there is a loss for the treatment group and a gain for the control group (Table 4.1). The analysis at a yearly basis does not provide meaningful results, neither for all births (Table 4.1), nor for births for each year (Table 4.2). This is probably due to the very low sample size when there are further

splits. In this regard, it is not possible to pursue analysis on a yearly basis, therefore the subsequent analyses will be made for all births that occurred during the four-year panel.

Table 4.1. Differences in income and consumption expenditure for treatment and control groups

		Income					Consumption				
	n	2009	2010	2011	2012	diff	2010	2011	2012	2013	diff
Treatment (all births)	181	12 218	12 225	12 106	11 177	-1 041	11 618	12 564	10 968	10 639	-980
Control	766	12 228	12 709	13 373	13 415	1 187	11 756	12 199	12 694	12 967	1 211

Table 4.2. Differences in income and consumption expenditure for treatment and control groups, by year of birth

		Income			Consumption		
	n	tb-1	tb	diff	tb	tb+1	diff
Births in 2010	42	10 435	9 308	-1 127	10 130	11 971	1 841
Births in 2011	76	13 172	12 488	-684	12 403	10 222	-2 181
Births in 2012	45	12 872	11 072	-1 801	10 708	10 190	-518
Control in 2010	766	12 228	12 709	480	11 756	12 199	443
Control in 2011	766	12 709	13 373	664	12 199	12 694	495
Control in 2012	766	13 373	13 415	42	12 694	12 967	273

Note: tb refers to time (year) of birth

4.1. Analysis with equivalence scale modification

In this subsection the effect of the newborn is analyzed by modifying the equivalence scale in a retrospective "what if" approach perspective. The assumption here is that when the consumption expenditure and the income level of the household is preserved how would the poverty status of the households change if the newborns were not added to the household. By altering the equivalence scale, the value of adult equivalent consumption expenditure and income values are increased and some of the households will rise above the poverty line and the aim is to observe the number of such households.

The consumption expenditure and income variables from Household Budget Survey (HBS) are used because there are more observations available in HBS and it is

better to use the original consumption expenditure variable instead of the estimated one in SILC whenever it is possible. The households that are under poverty line are assumed to be using all their available resources in order to reach a consumption level as high as possible since they are under the poverty line. These households might have obtained transfers related to the newborn. However, this will be ignored because this means there are resources available to these households which are under the poverty line and these transfers could have been made to pull them out of poverty instead. Therefore the analysis is made with this approach.

As mentioned before the national equivalence scale values computed by Betti et al. (2017) are used. In this scale, the value for children is 0.35. The equivalence scale value are revised according to the number of children of age 0, 1, 2 and 3 in the household. Accordingly, equivalized consumption expenditure and income are revised and these revised values are compared with the poverty line.

The results indicate that both consumption poverty and income poverty decrease 0.4 points when children of age 0 are excluded from the household. The difference is almost 2 points when children up to age 3 are excluded. The number of the poor would have been more than 1 million lower than the actual number.

Table 4.3. Number and percentage of poor when the newborn are excluded

		All	- 0 year	- 0,1 year	- 0,1,2 year	- 0,1,2,3 year
Number of people	Consumption Poor	15 176 552	14 878 766	14 553 781	14 158 900	13 909 196
	Income Poor	13 759 151	13 486 387	13 080 990	12 724 028	12 361 213
%	Consumption Poor	20.4	20.0	19.5	19.0	18.7
	Income Poor	18.5	18.1	17.6	17.1	16.6

The analyses at household size level show that as household size increases, also the difference between the actual poverty rate and the poverty rate that would be when the children are excluded, increases. This is an expected result since larger households include more children.

Table 4.4. Percentage of consumption poor when the newborn are excluded, by household size (HHB)

	Consumption Poverty				
HHB	**All**	**- 0 year**	**- 0,1 year**	**- 0,1,2 year**	**- 0,1,2,3 year**
1	13.0	13.0	13.0	13.0	13.0
2	10.3	10.3	10.3	10.3	10.3
3	7.8	7.6	7.4	7.3	7.2
4	11.9	11.8	11.4	11.1	10.6
5	21.9	21.1	20.8	20.0	19.7
6	35.0	34.7	33.5	32.1	31.4
7+	54.2	53.1	52.1	50.8	50.2

Table 4.5. Percentage of income poor when the newborn are excluded, by household size

	Income Poverty				
HHB	**All**	**- 0 year**	**- 0,1 year**	**- 0,1,2 year**	**- 0,1,2,3 year**
1	9.4	9.4	9.4	9.4	9.4
2	7.1	7.1	7.1	7.1	7.1
3	6.7	6.6	6.4	6.3	6.1
4	11.2	10.8	10.2	9.6	9.2
5	20.6	20.2	19.4	18.5	18.3
6	32.3	31.4	31.1	29.8	28.6
7+	49.6	49.0	47.7	47.5	46.3

4.2. Analyses with PSM

This subsection presents the propensity score estimation and analyses with PSM. A logit model is used for the propensity score estimation. The dependent variable is the treatment which means occurrence of a birth during the four-year panel. Treatment variable takes a value of one when there is a birth and zero in the absence of a birth.

Table 4.6. Model for propensity score estimation

	coeff.	s.e.	
Intercept	1.471	1.846	
Gender of reference person	0.010	0.387	
Age of reference person	-0.698	0.131	***
Household size	0.066	0.103	
Children of age 0-4 dummy	0.307	0.221	
Children of age 5-9 dummy	-0.200	0.215	
Children of age 10-14 dummy	-0.793	0.260	***
Percentage of household members with high school education or more * Percentage of household members that worked last week	2.444	0.697	***
Men with high school education or more dummy	0.243	0.220	
Women with high school education or more dummy	-0.675	0.270	**
Women that worked last week dummy	-0.302	0.230	
Log of consumption expenditure in 2010 (with 2013 prices)	-0.049	0.209	
Log of income in 2009 - measured in 2010 (with 2013 prices)	-0.097	0.216	

The graphical analysis shows that the overlap is sufficient (Figure 4.1). The balancing property is satisfied and the region of common support is [0.0406, 0.764]. At the end of the common support analysis with minima-maxima criteria 2 observations from the treated sample is dropped from the matching sample (Table 4.7). Further trimming is not desired in order not to lose observations.

Figure 4.1. Common support / overlap

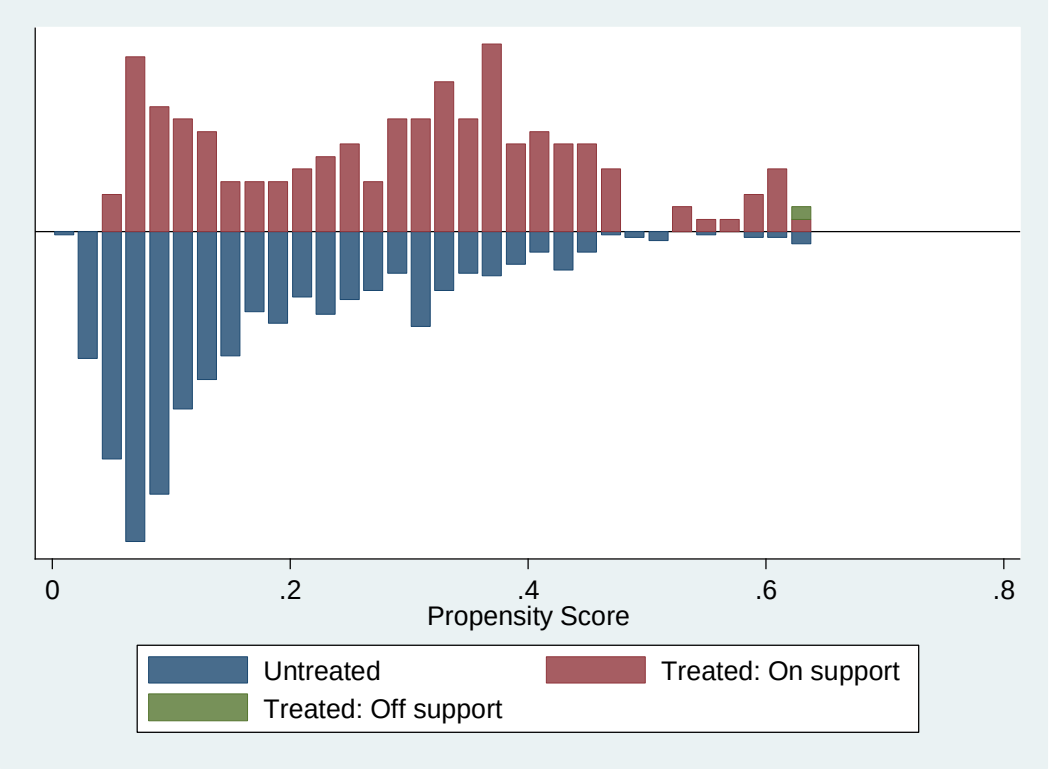


Table 4.7. Common support

	Off support	On support	Total
Untreated	0	766	766
Treated	2	179	181
Total	2	945	947

Matching quality is assessed by the standardized bias. There is only one variable (child_0_4_dummy) which has an absolute value slightly over 5 per cent. The matching quality is considered to be high.

Table 4.8. Matching quality

Variable	Sample	Mean		% bias
		Treated	Control	
ref_sex_1_0	**Unmatched**	0.94	0.93	2.9
	Matched	0.94	0.94	-2.3
ref_age	**Unmatched**	2.29	2.84	-70.1
	Matched	2.30	2.32	-2.1
hsize_orj	**Unmatched**	3.77	4.08	-22.1
	Matched	3.79	3.82	-2.0
child_0_4_dummy	**Unmatched**	0.60	0.41	39.4
	Matched	0.61	0.64	-5.7
child_5_9_dummy	**Unmatched**	0.39	0.48	-18.1
	Matched	0.40	0.42	-4.5
child_10_14_dummy	**Unmatched**	0.17	0.42	-55.8
	Matched	0.17	0.16	2.5
perc_edu_worked	**Unmatched**	0.15	0.09	31.7
	Matched	0.14	0.14	1.7
men_edu_high_dummy	**Unmatched**	0.54	0.40	27.0
	Matched	0.53	0.52	2.3
women_edu_high_dummy	**Unmatched**	0.31	0.29	6.0
	Matched	0.31	0.30	1.2
women_worked_dummy	**Unmatched**	0.29	0.33	-8.4
	Matched	0.28	0.27	2.4
log_cons	**Unmatched**	9.14	9.14	0.5
	Matched	9.14	9.14	0.4
log_inc	**Unmatched**	9.22	9.21	1.3
	Matched	9.21	9.21	0.2

The results of the PSM analyses for different types of outcomes are presented in the following paragraphs. Both income and consumption expenditure behave similarly for treatment and control groups. There is a decrease in the treated and an increase in the controls. The difference between the two is more than 2 300 TL per year which is

nearly half of the poverty line. Fuzzy monetary poverty measure decrease for both groups, but the decrease is 0.05 points greater for the controls which supports the finding that control group is better off compared to the treatment group. Fuzzy supplementary poverty measure increases for the treatment group showing that they are worse than they used to be in absolute terms, without even comparing to the control group.

It can also be seen that the average treatment effect is not significantly different from average treatment effect on the treated.

Regarding the standard errors, there isn't a consensus in literature. Bootstrapping is used in some studies. Abadie and Imbens (2008), on the other hand, argue that bootstrapping is inappropriate for the matched data. In this study, unadjusted standard errors are presented. Only specific to the sensitivity analyses tables, are between-imputation standard errors are used as these are provided by the Sensatt package.

Table 4.9. ATT and ATE for income, consumption expenditure and fuzzy poverty

		ATT				
	n	**Treated**	**Controls**	**Difference**	**s.e.**	**ATE**
income	179	-979	1 347	-2 326	786	-2 516
consumption expenditure	179	-944	2 326	-3 270	1 267	-2 654
fuzzy monetary	179	-0.013	-0.062	0.049	0.022	0.032
fuzzy supplementary	179	0.009	-0.026	0.035	0.029	0.032

For those households which were not income poor in the beginning of the panel there is not considerable difference with regard to their entrance into poverty. Around 9 per cent of the treatment group and 8 per cent of the control group enter into poverty at the end of the panel. On the other hand, consumption poverty demonstrate a difference among the groups indicating the control group is better off in general. The situation is reversed for entrance into deprivation. All the same, it is seen that the difference is not statistically significant for deprivation.

Table 4.10. Entrance into poverty (for those who were not poor or deprived in 2010)

		ATT			
	n	Treated	Controls	Difference	s.e.
income poverty	130	0.085	0.077	0.008	0.043
consumption poverty	144	0.174	0.069	0.104	0.041
deprivation	60	0.283	0.367	-0.083	0.105

When exit from poverty is considered the results for income poverty is significant. The results indicate that 60 per cent of income poor in the beginning of the panel among the treatment group, was still poor at the end of the panel. This percentage was 46 per cent for the control group. The result is not significant for consumption poverty. The analysis of deprivation also shows that the treatment group is worse-off.

Table 4.11. Exit from poverty (for those who were poor or deprived in 2010)

		ATT			
	n	Treated	Controls	Difference	s.e.
income poverty	48	0.604	0.458	0.146	0.116
consumption poverty	32	0.344	0.469	-0.125	0.141
deprivation	119	0.773	0.639	0.134	0.067

The sensitivity analysis shows that choice among logit and probit models and matching algorithm does not have a significant impact on the estimated treatment effect.

Table 4.12. Sensitivity to matching algorithm

Model type and matching algorithm	n	ATT	s.e.
logit / nnd / with replacement	179	-2 326	786
probit / nnd / with replacement	179	-2 559	746
logit / nnd / without replacement	179	-2 297	634
logit / nnd / without replacement / caliper (0.1)	174	-2 209	645
logit / nnd / without replacement / caliper (0.01)	167	-2 475	663
logit /radius	179	-2 166	466
logit /kernel	179	-2 482	544

When PSM method, which relies on UNA is applied it is utmost important to carry out sensitivity analysis to account for robustness. In this respect indirect and direct tests are carried out for analysis.

The indirect tests with different outcomes show that these outcomes are not significantly different for the treatment and control groups.

Table 4.13. Indirect Tests

Name of variable	ATT	s.e.
income in 2010	-721	1 321
consumption expenditure in 2010	-434	1 084
hot water	0.04	0.04
internet	0.07	0.06
household size	-0.03	0.16

The direct tests for the change in income are presented in this section. Using variables from the model to mimic the unobserved variable is the first direct test. The baseline ATT value is -2 326 and none of the ATT with tested variables are significantly different from the baseline ATT. G refers to outcome effect which is the effect of unobserved variable on the outcome variable, that is the difference in income in this case and A refers to selection effect, which refers to the effect of unobserved variable on the treatment variable. p_{ij} refers to the distribution of the variable where both i and j take values of 0 and 1. i refers to the treatment variable and j refers to a transformed form of the outcome variable which takes a value of 1 if the outcome is greater than the median and 0 otherwise.

Table 4.14. Direct Tests 1

Unobserved variable simulated according to	p11	p10	p01	p00	Outcome effect (G)	Selection effect (A)	ATT	s.e.
none							-2 326	786
child_0_4_dummy	0.71	0.54	0.38	0.44	0.79	2.28	-2 534	901
child_5_9_dummy	0.35	0.41	0.47	0.49	0.91	0.70	-2 632	829
child_10_14_dummy	0.20	0.16	0.39	0.45	0.77	0.30	-2 696	908
women_edu_high_dummy	0.32	0.31	0.31	0.27	1.23	1.16	-2 551	822
women_worked_dummy	0.31	0.28	0.35	0.31	1.23	0.82	-2 551	817
men_edu_high_dummy	0.58	0.51	0.45	0.35	1.48	1.80	-2 581	853

Note: Standard errors are between-imputation standard errors

The second direct test is presented below. At some extremes the ATT value is different from the baseline value or close to zero. Even so, in general the ATT's obtained with the test are not significantly different from the baseline ATT and the they always have a value which is negative and significantly different from zero. The results indicate that the PSM is not robust only in the case of an unreasonably effective unobserved variable. In this test, $d = p_{01} - p_{00}$ is a measure of the effect of on the untreated outcome and $s = p_{1.} - p_{0.}$ is a measure of the effect of unobserved variable on treatment assignment (Ichino et al., 2008).

Table 4.15. Direct Tests 2

					G < 1 and A < 1							
		s = -0.1				**s = -0.3**				**s = -0.5**		
	ATT	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**
d = -0.1	-2 661	450	0.67	0.68	-2 808	512	0.67	0.21	-3 007	609	0.59	0.11
d = -0.3	-2 756	500	0.28	0.74	-3 153	538	0.29	0.31	-3 669	576	0.25	0.11
d = -0.5	-2 810	498	0.07	0.76	-3 530	521	0.07	0.33	-4 222	616	0.07	0.11
					G > 1 and A < 1							
		s = -0.1				**s = -0.3**				**s = -0.5**		
	ATT	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**
d = +0.1	-2 570	447	1.54	0.70	-2 399	468	1.52	0.23	-2 082	560	1.75	0.12
d = +0.3	-2 385	489	3.65	0.68	-1 986	500	3.61	0.28	-1 511	504	4.17	0.10
d = +0.5	-2 185	489	15.20	0.64	-1 524	491	15.03	0.28	-789	446	14.70	0.09
					G < 1 and A > 1							
		s = +0.1				**s = +0.3**				**s = +0.5**		
	ATT	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**
d = -0.1	-2 468	432	0.58	1.82	-2 347	543	0.59	4.16	-2 173	589	0.58	10.86
d = -0.3	-2 297	463	0.24	1.76	-1 820	474	0.25	4.73	-1 550	554	0.25	13.87
d = -0.5	-2 123	478	0.07	1.80	-1 326	469	0.07	4.94	-734	506	0.07	14.57
					G > 1 and A > 1							
		s = +0.1				**s = +0.3**				**s = +0.5**		
	ATT	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**	**ATT**	**s.e.**	**G**	**A**
d = +0.1	-2 721	476	2.39	2.02	-2 964	558	2.42	5.02	-3 169	718	2.38	10.42
d = +0.3	-2 760	483	6.41	1.49	-3 388	555	6.29	3.90	-3 814	622	6.29	8.67
d = +0.5	-3 028	519	14.73	1.55	-3 631	516	14.50	3.62	-4 397	576	15.04	12.48

Note: Standard errors are between-imputation standard errors

The analyses with stratification and regression methods for the difference in income return similar results with PSM. These analyses are conducted to be sure that the results by other PS methods are in line with PSM method. In the stratification, the observations are divided into stratas according to their propensity scores. The results show that there is a difference between the control and treatment groups with regard to the difference in income between 2010 and 2013. This difference is in line with findings from PSM method. Only in the first strata we don't observe any difference. In this strata, there are only 11 treatment observations, which is probably the cause of this situation.

Table 4.16. Stratification

	1		2		3		4		5	
	mean	**s.e.**	**mean**	**s.e.**	**mean**	**s.e.**	**mean**	**s.e.**	**mean**	**s.e.**
Control	232	565	776	523	2 294	487	1 224	548	1 678	460
Treatment	278	721	-3 080	1 993	-1 396	950	-2 411	805	33	589
Difference	46	2 285	-3 856	1 702	-3 690	1 357	-3 635	1 104	-1 645	736

Table 4.17. ANCOVA

Control	1 274
Treatment	-1 408
Difference	-2 681

Table 4.18. Regression with inverse weights

Parameter	**Estimate**	**s.e.**
Intercept	-1 399	287
Control	2 660	407
Treatment	0	.

5. Conclusion

Economic growth by itself wouldn't be adequate for increasing well-being of all households and decreasing poverty. In order to struggle with poverty, specific groups should be targeted. Especially when there is a pro-natal policy in action, the extent of the effect of childbirth on household well-being should be well analyzed and understood. Only by this way, anti-poverty policies could be put in action and poverty could be avoided.

Establishing policy targets requires information on the issues to be dealt with. Information on causality is particularly important in this regard. Knowing causal relationships enables better understanding of socio-economic structures and in this way provides a basis for better targeting for policy makers. In the case of the relationship between fertility and poverty, the acquired information would provide useful tools for policy makers to compensate for the households if there is a risk of decreased well-being depending on having an extra child.

The results indicate the negative effect of having a newborn in the household on well-being. Analyses using propensity scores either with PSM method, stratification or regression methods bring out similar results.

The sensitivity analyses demonstrate the unconfoundedness assumption is not a strong one. Even if this was the case the information provided by this study claims policy implication since causal or not, households that have a child between the waves are worse off compared to the control group. There is negative effect of a newborn on household well-being with most of the measures used to analyze. This finding calls for further support for households having a child, in the context of pro-natal policies. Such support, by providing the necessary funds for households, would encourage pro-natal policies as well if it is deemed necessary.

In this study, the shortcoming with regard to use of consumption expenditure is that it is a derived variable via statistical matching. The procedure to obtain this variable could have caused a variation which cannot be measured. Although using income which is closely related to consumption expenditure in its estimation, the results associated with this variable should be treated with caution.

On the other hand, the income variable in SILC has low variability between successive years. High imputation process in SILC links one year to another decreasing the variability between them. In this regard, use of relatively new poverty measures such as fuzzy measures of poverty and deprivation index, demonstrates itself as a dominant alternative over conventional measures. Therefore, besides expenditure, income and the conventional measures of poverty based on these, multidimensional poverty are be calculated and compared between waves.

One target of the study was to carry out the analyses by considering the year of birth. Due to the low number of observations, this exercise didn't return results that are consistent and interpretable. However, since this is an important issue, it should be considered in further research agenda whenever bigger samples are available.

References

Aassve, A., Arpino, B. (2007) “Estimation of Causal Effects of Fertility on Economic Well-being: Evidence from Rural Vietnam”, ISER Working Paper 2007-24. Colchester: University of Essex.

Aassve, A., Engelhardt, H., Francavilla, F., Kedir, A., Kim, J. Mealli, F., Mencarini, L., Pudney, S., Prskawetz, A. (2005) "Poverty and Fertility in Less Developed Countries: A Comparative Analysis," Discussion Papers in Economics 05/28, Department of Economics, University of Leicester.

Aassve, A., Kedir, A., Weldegebriel, H. T. (2006) “State Dependence and Causal Feedback of Poverty and Fertility in Ethiopia”, ISER Working Paper 2006-30. Colchester: University of Essex.

Abadie, A., Imbens, G. W. (2008) “On the failure of the bootstrap for matching estimators”, Econometrica, 76, 1537–1557.

Arpino, B. (2008) “Causal inference for observational studies extended to a multilevel setting. The impact of fertility on poverty in Vietnam”, PhD Thesis, University of Florence.

Arpino, B., Aassve, A. (2013) “Estimating the causal effect of fertility on economic well-being: data requirements, identifying assumptions and estimation methods”, Empirical Economics, 44, 355-385.

Augurzky, B., Schmidt, C. (2001) “The Propensity Score: A Means to an End”, Discussion Paper No. 271, IZA.

Austin, P. C. (2011) “An introduction to propensity score methods for reducing the effects of confounding in observational studies”, Multivariate Behavioral Research, 46(1), 399-424.

Betti G., Gagliardi F., Lemmi A., Verma V. (2015) “Comparative Measures of Multidimensional Deprivation in the European Union”, Empirical Economics, 49(3), 1071-1100.

Betti, G., Karadag, M. A., Sarica, O., Ucar, B. (2017) "Regional Differences in Equivalence Scales in Turkey", Economy of Region. (forthcoming)

Betti G., Verma V. (2008) “Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective”, Statistical Methods and Applications, 12(2), 225-250.

Black, D., Smith, J. (2003) “How Robust is the Evidence on the Effects of the College Quality? Evidence from Matching", Working Paper, Syracuse University, University of Maryland, NBER, IZA.

Bryson, A., Dorsett, R., Purdon S. (2002) “The Use of Propensity Score Matching in the Evaluation of Labour Market Policies”, Working Paper No. 4, Department for Work and Pensions.

Caliendo, M., Kopeinig, S. (2008) "Some Practical Guidance For The Implementation Of Propensity Score Matching", Journal of Economic Surveys, 22(1), 31-72.

Cheli, B., Lemmi A. (1995) “A ‘Totally’ Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty”, Economic Notes, 24 (1), 115-134.

Desta, C. G. (2014) "Fertility and Household Consumption Expenditure in Ethiopia: A Study in the Amhara Region”, Journal of Population and Social Studies, 22(2), 202-218.

Eastwood, R., Lipton, M. (1999) “The impact of changes in human fertility on poverty”, The Journal of Development Studies, 36(1), 1-30.

Eurostat (2016) http://ec.europa.eu/eurostat/statistics-explained/index.php/Material_deprivation_statistics_-_early_results, accessed: 24.11.2016.

Gupta, N. D., Dubey, A. (2003) “Poverty and Fertility - An Instrumental Variables Analysis on Indian Micro Data”, Scandinavian Working Papers in Economics. No 03-11: November, 2003.

Heckman J.J., Ichimura H., Todd P. (1997), “Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, Review of Economic Studies, 64, 605-654.

Holland, P. (1986) “Statistics and causal inference”, Journal of American Statistical Association, 81, 945–970.

Ichino, A., Mealli, F., Nannicini, T. (2008) “From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity?”, Journal of Applied Econometrics, 23(3), 305-327.

Imbens, G. W. (2004) “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review”, Review of Economics and Statistics, 86, 4-30.

Kahn, H., Brown, W., Martel, L. (1976) “The Next 200 Years: A Scenario for America and the World”, New York: William Morrow and Company, ISBN 0688030297

Kim, J., Engelhardt, H., Prskawetz, A., Aassve, A. (2005), "Does fertility decrease the welfare of households?: An analysis of poverty dynamics and fertility in Indonesia", Vienna Institute of Demography Working Paper, 06 / 2005.

Kim, J., Engelhardt, H., Prskawetz, A., Aassve, A. (2009) "Does fertility decrease household consumption?: An analysis of poverty dynamics and fertility in Indonesia" Demographic Research, 20, 623-656.

Leuven, E., Sianesi, B. (2003) "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing".

Malthus, T.R. (1798) "An Essay on the Principle of Population". Oxford World's Classics reprint.

Mussa, R. (2014) "Impact of fertility on objective and subjective poverty in Malawi", Development Studies Research, 1(1), 202-222.

Nannicini, T. (2007) "Simulation-Based Sensitivity Analysis for Matching Estimators," Stata Journal, 7, 334-350

Ravallion, M., Bidani, B. (1994) "How Robust Is a Poverty Profile?", World Bank Economic Review, Oxford University Press, 8(1), 75-102.

Rosenbaum, P. R. (1987) "The Role of a Second Control Group in an Observational Study". Statistical Science, 2(3), 292-306.

Rosenbaum, P. R., Rubin, D. B. (1983), "The central role of the propensity score in observational studies for causal effects", Biometrika, 70 (1): 41-55.

Rosenbaum, P. R., Rubin, D. B. (1985) "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score", The American Statistician, 39(1), 33-38.

Rubin, D. (1980) "Discussion of Randomization Analysis of Experimental Data: The Fisher Randomization Test", Journal of the American Statistical Association, 75, 591-93.

Rubin, D. B., Thomas, N. (1996) "Matching Using Estimated Propensity Scores: Relating Theory to Practice", Biometrics, 52, 249-264.

Simon, J. (1981) "The Ultimate Resource". Princeton, N. J.: Princeton University Press, ISBN 0-85520-563-6

Sinding, S. W. (2009) "Population, poverty and economic development", Philosophical Transactions of the Royal Society B, 364(1532), 3023-3030.

Smith, J., and Todd, P. (2005) “Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," Journal of Econometrics, 125(1-2), 305-353.

Turkstat (2013) “Press Release on Population Projections, 2013-2075” : http://www.turkstat.gov.tr/PreHaberBultenleri.do?id=15844.

Appendix 1. List of variables used to derive the fuzzy supplementary poverty measure

Dimension	Items of deprivation
1 Basic lifestyle	Meals with meat, fish or chicken
	Household adequately warm
	Holiday away from home
	Ability to make ends meet
2 Consumer durables	Car
	PC
	Telephone
	Washing Machine
	TV
3 Housing amenities	Bath or Shower
	Indoor flushing toilet
	Leaking roof and damp
4 Financial situation	Inability to cope with unexpected expenses
	Arrears on mortgage or rent payments
	Arrears on utility bills
	Arrears on hire purchase instalments
5 Work & Education	Low education
	Worklessness
6 Health related	General health
	Chronic illness
	Mobility restriction