Peers Effects on a Fertility Decision: an Application for Medellín Colombia

Por: Leonardo Fabio Morales

Borradores de ECONOMÍA

Núm. 777
2013

tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

Peers Effects on a Fertility Decision: an Application for Medellín Colombia*

Leonardo Fabio Morales†

lmoralzu@banrep.gov.co

Abstract

This paper addresses the estimation of peer group effects on a fertility decision. The peer group is composed of neighbors with similar socio-demographic characteristics. In order to deal with the endogeneity problem associated to the estimation of neighborhood effects, an instrumental variables procedure is performed. To control for the reflection problem, usual in linear effects models, this paper uses an identification strategy that relies on the definition of peer groups at the individual level. This paper provides evidence that peer effects explain the age at which poor women in Medellín (Colombia) decide to have their firstborn. These social forces are hazardous factors that may increase the incidence of adolescent pregnancy.

Keywords: Fertility, Family Planning, Demographic Economics, Social Interaction Models

JEL Codes: J13, J130, C310

*This paper is a preliminary work. The opinions expressed here are those of the author and not of the Banco de la República de Colombia nor of its Board. All errors are author's responsibility. I thank Carlos Medina and Jorge Tamayo for their comments and for providing the data used for the econometrical estimations of this paper. I thank Samuel Freije for detailed comments on successive drafts, Eleonora Dávalos and David Guilkey for their useful comments and all help provided in the realization of this research.

†Junior Researcher. Banco de la República de Colombia

1 Introduction

In the last 50 years there has been a substantial reduction in the world's total fertility rate (TFR). The Latin American and the Caribbean (LAC) region is not an exception to this rule, and it has experienced a deep process of demographic transition. The reduction in the total fertility rates is a homogeneous phenomenon throughout the entire region. An illustration of the state of demographic transition in the LAC region can be observed from the behavior of the TFR during the first decade of this century. The region's TFR was 2.67 children per woman in 1999, and by the end of 2010 the TFR was 2.12. Surprisingly close to the widely accepted replacement rate of 2.1.

The evolution of the fertility rate for young populations is especially important because of the negative consequences of teenage childbearing. This is a problem widely associated in the literature to low human development and poverty (Joshi and Schultz, 2007; Buvinic, 1998; Burt, 1998; Gage, 1995; Singh and Wulf, 1990; Hayes, 1987). During the last decade, there has been also a reduction in the Fertility Rate for women between 15 and 19 (FR15-19) in the LAC region. The FR15-19 decreased from 83.95 children per 1000 women in 1999 to 71.68 in 2010. Certainly, this is an important reduction in the FR15-19; nevertheless, in general terms it has been smaller that the reduction that the TFR has experienced during the same period. Between 1999 and 2010, the reduction in the TFR was 26% while the reduction in the FR15-19 was 17%. This is a very interesting phenomenon which, among other things, implies that adolescent fertility has become a more important component of total fertility in most of the LAC countries. This means that relative to adult fertility, adolescent fertility is becoming greater and greater in Latin America.

With some exception like Argentina and Perú, the contribution of adolescent fertility to the total fertility has increased continuously in almost all LAC Countries. As mentioned before, this is due to the faster reduction of the fertility rate for the adult population in comparison to the rate for teenagers' population (Florez and Soto, 2007B). A good illustration of this phenomenon can be observed from the evolution of the ratio adolescent fertility to total fertility (per 1000 women). In 1999, adolescent fertility was 15.72% of total fertility in developing LAC countries, and by 2010 this ratio had increased to 16.29%. There are some remarkable cases as Brazil and Ecuador where the ratio adolescent fertility to total fertility increased by more than 2 percentage points between 1999 and 2010. The result of this phenomenon is, on average, an earlier individual onset of childbearing.

In order to find explanations to this interesting phenomenon, it is important to study the factors that determine the age at which a mother decide having her firstborn. From an individual's point of view it may seems rational having a child early in life given her education, her household socioeconomic conditions, and the characteristics of the social group at which the individual belongs to. This paper will try to explain the mother's chosen timing for the onset of childbearing in an urban context in Colombia, emphasizing in the role that peer effects may play, and using longitudinal individual information that allows characterizing mothers before or at the time of the pregnancy.

Many social researchers in the last three decades have been interested in the phenomenon that takes place when an individual behavior is partly explained by the influence of other individuals' behavior. In economics this has been called social interactions (SI) or peer effects. There are several channels through which these effects may take place; for example, individuals may learn from peer's behavior (social learning), or they may embrace the norms of the community in regards to socially accepted practices (social influence) (Behrman and Watkins, 2001). The main purpose of this paper is testing the existence and measuring the magnitude of peer effects on a fertility decision. The fertility decision considered is the women's age at the onset of childbearing. The mothers studied are a big sample of poor mothers in the city of Medellin, which have their firstborn between 2001 and 2010.

Social Interactions could be a potential explanatory factor in the reduction of the average age of mothers at first birth observed in LAC region, and certainly they could be a cause for the high incidence of teenage pregnancy in some LAC countries like Colombia. There is evidence on the existence of geographic sorting patterns governing the spatial distribution of several fertility outcomes in LAC cities (Gaviria et all 2010). For instance, in poor neighborhoods women have more children and the onset of childbearing is earlier than in other neighborhoods. Nevertheless, it has not been explored the existence of peer effects in the literature about fertility in LAC countries. Several studies in this literature have suggested that there has been an underestimation of the importance of contextual and cultural factors. The influence of new social norms, like the general acceptation of an early beginning for sexual relationships, or the influence of peer pressure are factors that have not given enough relevance in the study of teen pregnancy in the LAC region (Florez and Soto, 2007B). In this paper, I use data from Medellín to evaluate the existence of peer effects that influence the age at which mothers decide having their firstborn. Medellin is an example of the urban context in Latin America; there are high levels of adolescent pregnancy concentrated in poor neighborhoods.

I order to achieve the purpose of this paper, I estimate a classical linear-in-means model of social interactions where the relevant peer effect groups are defined using weighting matrices with weights defined using spatial and social distance criteria. In order to deal with the reflection problem, typical in the estimation of endogenous peer effects models, I design peer groups varying at the individual level. This strategy has been recently proved to successfully overcome the reflection problem (Bramoullé et al, 2009; De Giorgi et al, 2010). In addition, the definition of non-perfectly overlapping groups is useful to overcome a second problem also typical in this kind of estimations, the endogeneity of the peer group. The definition of peer groups that are different for each individual implies the existence of excluded peers. They are peers of individual's peers who do not belong to the individual's peer group. Using information from excluded peers, I construct instruments to estimate the social interaction models by two stages least squares methods. For all specifications estimated in this paper the endogenous peer effects coefficient is positive and significant. In other words, an important factor explaining a woman's decision of having her firstborn at a specific age is the influence of her peer group. This influence is measured in terms of the average age for the onset of childbearing among the member of the peer group. Therefore, a woman has more probability of becoming a teenage mother if her peer group has an important composition of teenage mothers.

From a social policy perspective this topic is crucial, given that pregnancy at an early age has been widely associated with negative socio-economic outcomes for the mother and the child (i.e. Joshi and Schultz, 2007; Buvinic,1998; Gage, 1995; Singh and Wulf, 1990; Hayes, 1987; Case y Katz 1991; Grogger and Korenman, 1993; Hotz et al, 1999). These studies remark the fact that educational achievements, health markers, and measures of involvement in risky behaviors tend to be worse for teenage mothers and their children.

In section 2 of this paper I present a brief summary of the literature in which this paper fits. In section 3 the theoretical foundations in which this paper is based are described. In section 4 I make a description of the data. Section 5 describes the empirical strategy implemented to get reliable estimations. In section 6 the results of the estimations are provided. In section 7 some simulations and robustness checks are described and their results are presented. Finally in section 8, I present the main conclusions of this paper, and some relevant policy implications are commented.

2 Related Literature

The fertility outcome of interest for this paper is the women's age at the onset of childbearing. Most of the literature related with this topic has been focused on the teenage childbearing, which is an extreme case of early onset of childbearing. Several papers have been written on the negative consequences of teenage pregnancy and childbearing (Joshi and Schultz, 2007; Buvinic, 1998; Burt, 1998; Gage, 1995; Singh and Wulf, 1990; Hayes, 1987). Women who were mothers before their twenties are usually less educated and wealthier that the ones who became mothers after their twenties. In the case of Colombia for example, Florez and Soto (2007B) found that having a children as a teenager implies for that the mother will get 3.9 years of education less than if she would have had the child in her adulthood.

There are several papers that describe and analyze the adolescent childbearing phenomenon in Colombia and the LAC region (Florez and Nuñez, 2001; Florez and Soto, 2007A; Florez and Soto, 2007B). One important remark from some of these papers is that the contribution of the teenager's fertility rate to the TFR has increased for several of the LAC countries (Florez and Soto, 2007B). More recent data[1] on fertility rates confirm that trend for the LAC average. In terms of regional means, adolescent fertility became a more important component of total fertility during the first decade of this century. The average LAC proportion of teenage fertility in total fertility increased in almost one percentage point during the period between 1999 and 2010[2]. This is not a phenomenon that take place in all LAC countries, there are some important exceptions. For example, during the mentioned period the ratio adolescent fertility over total fertility decreased in Peru from 11.19% to 10.22%, and in Argentina from 13.02% to 12.5%. Nevertheless, for most of countries in the region this ratio increased or remained relatively constant. In cases like Brazil this ratio increased from 18.38% to 20.71%, and in Ecuador the increment was from 14.13% to 16.42%; for both countries in the period 1999-2010. In the specific case of Colombia the ratio adolescent fertility over total fertility in 2010 was 16.81%, close to what it was in 1999 (17.41%).

For the specific case of Colombia, several papers have used econometric models to explain the probability of teen pregnancy (Gaviria, 2000; Barrera and Higuera, 2004; Florez and Nuñez, 2001). These papers found evidence that low education, disadvantaged socio-economic conditions, and poor family backgrounds increases the probability of teenage pregnancy. None of these papers uses longitudinal information; therefore, it is not possible to use covariates before or at the time of the pregnancy to explain the fertility decision. This may derive in endogeneity bias, as long as important explanatory variables in these models are determined simultaneously with the pregnancy. In this paper I can know the characteristics of a future mother before the childbirth, in that way I can control better for the simultaneity of several covariates and the pregnancy. I addition, none of these papers emphasizes in the role that social interactions may play in the determination of fertility outcomes. At the LAC region level, very few papers consider seriously SI type or similar effects on fertility (Rosero-Bixby and Casterline,1994; Lindstrom and Muñoz, 2005).

There several studies in demography and sociology on the role of SI in fertility outcomes. One illustrative example of is the study of diffusion effects. A diffusion effect takes place when a behavior is adopted and reproduced through the social networks. The mechanisms through which social networks can perform an influence among its members include social influence and social learning. In order to

[1] Public data from World Bank (http://data.worldbank.org/)

[2] The average ratio teenage fertility over total fertility was 15.51% for all LAC countries, in 2010 this ratio was on average 16.29%

measure the diffusion effects, some papers in this literature have used, as explanatory variables in the estimations, variables that describe if the woman discusses about contraceptive practices with relatives or members of the community (Montgomery et al (2001)). Other papers have used aggregated fertility levels (Montgomery and Casterline (1993), Rosero and Casterline (1994)) or the proportion of family planning users in a woman's network (Kohler et all (2001)). The empirical techniques often rely on the use of longitudinal data at individual levels or aggregated data by geographical areas.

In economics, there is an important branch in the applied micro literature focused on the detection and identification of the group effects on many outcomes, including fertility. For example, the identification of peer effects in education has been widely debated in several papers. Part of the discussion was originated with the publication of the Colleman Report (1966); in which one of the most polemic findings was that students perform better if their fellow students are high achievers (Oates et al (1992)). There are several papers that seek the identification of SI when the main dependent variable is a fertility decision (i.e. Oates et al, 1992; Iyer and Weeks, 2009; Case and Katz, 1991). In some of these papers, the authors have realized that the peer group itself is an endogenous decision, and because of this the estimation results will be biased (e.g. Oates et al, 1992). As it will be discussed later on in this paper this self-selection issue and the reflection problem[3], pointed by Manski (1993), are the main identification threats for empirical models of social interactions.

3 Theoretical Foundations

Let's consider a representative woman i, who derives utility from consumption z_i and from children n_i (if she has). I will focus on the decision about starting a family or not; this fact will be represented by the term c_{1ia}. Therefore, any woman i, with no children until the previous period, can derive utility from her firstborn if she decides to become a mother in the current period; in which case, $c_{1ia} = 1$ where $c_{1ia} \in \{1, 0\}$. The subscript "a" represents the age of the woman, and the subscript 1 stands for firstborn. The total number of children is the summation of fertility choices (c_{ia})[4] up to age a, $n_i = \sum_{t=0}^{a} c_{ti}$. Additionally, if woman i decides to become a mother, she will also get utility from the quality of her child, q_i. The quality of the child is a household production function that takes as inputs education, parental time, or any other resources that can improve the child development. For simplicity it is assumed that this representative woman solves a static optimization problem every year,

[3] The reflection problem is originated because the individual is influenced by the reference group, but at the same time, an individual's decision also determines the group behavior.

[4] A general fertility choice is represented by c_{ia} where c_{ia} equal 1 if woman decide to have a childbirth (not necessarily the first one) at age a.

instead of assuming that she maximizes her lifetime utility. This assumption allows the development of a "demand for children" model similar to the one proposed by Becker (1981) in his book "A Treatise on the Family." For the year in which the woman i decided to have her firstborn or any year before, the utility function can be represented by the following equation:

$$U_{i,a} = U_{i,a}\left(c_{1ia}, z_{ia}, q_i\right) \tag{1}$$

At each age a, mother i decides if she will have her firstborn in this year or not and her consumption z_{ia}, subject to the standard budget constraint:

$$p_c.q.c_{1i,a} + \pi_z z_{i,a} = I_{ia} \tag{2}$$

Where I_{ia} denotes income, p_c is the unitary cost of quality, and $p_c.q.c_{1i,a}$ is the total amount spent on the child each year, which is zero if $c_{1i,a} = 0$ (no child in this period). In addition, π_z represent the price of the consumption good. Every year the woman i maximizes (1) subject to (2)[5]. The woman i will have her first childbirth at age a if:

$$U_{ia}\left(1, z^*_{i,a}, q^*_{ia}\right) > U_{ia}\left(0, z^*_{i,a}, q^*_{ia}\right) \quad | \quad p_c.q^*_i. + \pi_z z^*_{i,a} = I_{ia} \tag{3}$$

Where the utility function is evaluated in the optimal quantities $(z^*_{i,a}, q^*_{ia})$. Starting from this "Becker type" model I will introduce the possibility that the fertility variable can be explained by social interactions. If social interactions have some impact in the fertility decisions, then the fertility decisions of the peer group members should enter into the utility function of the woman i. Let's assume that there is perfect knowledge about the woman $i's$ social network, and that the level of interaction between a woman and a member of her peer group is perfectly measured by ω_j, where j is a generic member of the woman $i's$ peer group, and ω_j is a normalized "interaction index." For several reasons (social influence, social learning, etc.) one can assume that woman i gets utility from exhibiting a behavior similar to the one displayed by her peers. Therefore, optimization problem of the mother can be can be rewritten as:

$$\max \quad U_{i,a}\left(c_{1i,a}\left(m_{ia}\right),\ z_{ia}, q_i, \left| m_{ia} - \sum_{j\neq i} \omega_j.m_j \right| \right) \quad \text{s.t (2)} \tag{1a}$$

$$\text{with } \begin{cases} c_{1i,a} = 1 \text{ if } m_{ia} \geqq \bar{m}_i \\ c_{1i,a} = 0 \text{ if } m_{ia} < \bar{m}_i \end{cases}$$

[5] After her first child the decision of women i will be whether or not to have her second child $c_{2i,a} \in \{1, 0\}$. And her budget constraint will be:

$$p_c.q + p_c.q.c_{2i,a} + \pi_z Z_{i,a} = I.$$

The quality is assumed the same for each child.

Where m_{ia} is a continuous variable representing a fertility attitude or behavior of woman i, and m_j is a continuous variable representing a fertility attitude or behavior of each member of the women $i's$ peer group. The term $c_{1i,a}$ is a function of m_{ia}, and $\bar{m}_i$ is an arbitrary threshold beyond which a child is generated. One can think of m_i as some continuous index revealing for example attitudes toward sex or simply the desire of becoming mother; it is assumed that m_i is under the mother's control each period. The term $\sum_{j \neq i} \omega_j.m_j$ represents the weighted average of the fertility behavior among the peer group; the weights are the interaction indexes ω_j. This construction implies that the stronger is the relationship between i and j, the greater will be the weight than peer j has in the computation of the average. The main hypothesis of this paper is that in every period the woman i will get additional utility from mimicking the behavior of her peers. Therefore, the contribution of the term $\left| m_i - \sum_{j \neq i} \omega_j.m_j \right|$ is assumed to be negative.

3.1 Determination of the age at the first childbirth

The discrete framework explained before is useful to connect the theoretical foundations to the empirical approach. At every age a, woman i solve the optimization problem represented in (1a) subject to the budget constraint (2); therefore, the woman's age at the first childbirth can be defined as:

$$A_i = \min \left\{ a \in [12, 45] \quad \text{s.t.} \left[\begin{array}{c} U_{ia}\left(1,\ z^*_{ia},\ q^*_{ia},\ \left| m^*_i - \sum_{j \neq i} \omega_j.m_j \right| \right) > \\ U_{ia}\left(0,\ z^*_{ia},\ q^*_{ia},\ \left| m^*_i - \sum_{j \neq i} \omega_j.m_j \right| \right) \\ \text{given that}\ \ p_c.q.c^*_{1i,a} + \pi_z z^*_{i,a} = I_a \end{array} \right] \right\} \tag{4}$$

Where a can only take continuous values between 12 and 45 given the biological fertility period in the woman's life. If one is willing to assume that individials are followers of their peers' behavior[6], it would be interesting to ask what is the response of A_i given a exogenous increase in m_j. Let's assume that the individual already choose the optimal quantities of m^*_i, z^*_{ia}, and q^*_{ia} to maximize her utility. If individuals follow the behavior of their peers, an increase in m_j will produce an increase in m_i. This increase in m_i could cause a jump in $c^*_{1i,a}$ from zero to one if the increase is enough to overcome the threshold $\bar{m}_i$. Then through this mechanism the mother will chose to have her first child in the current period, at age a, and not latter; which can be interpreted as an effect of m_j on A. In this paper I will assume the existence of a continuous function $G(.)$ that maps each possible combination of the inputs in the utility function to single value A_i. Therefore, A_i can be written as:

[6] Assuming this is equivalent to say that $\frac{\partial m_i}{\partial m_j} > 0$. In a continuous and simplified version of the model it is possible to find an expression for $\frac{\partial m_i}{\partial m_j}$. Under minimal assumptions, it is not possible unambiguously determine the sign of this derivative. Under some circumstances individuals are peer followers, but the contrary case it is also possible. A valid interpretation of the main question of that paper is testing whether or not the individuals are followers of the peers' fertility behavior.

$$A_i = G\left(z_{ia},\ q_{ia},\ \left|m_i - \sum_{j\neq i}\omega_j.m_j\right|\right) \tag{5}$$

In forthcoming sections a linear approach to equation (5) will be proposed and estimated. The main interest of this paper is the peer effects on A_i, in other words, the influence that the behavior of the peer group has on the women's age at first childbirth. The prediction that can be derived from the main hypothesis of this paper is that $\frac{\partial m_i}{\partial m_j} > 0$. In terms of the dependent variable, this means that mother i may find optimal reduce her age at her first childbirth given a reduction in the expectation of this variable among her peer group.

4 Data

The System for Selecting Beneficiaries of Social Spending (SISBEN in Spanish) is a household targeting system that has been used in Colombia from the lately 90's to target social programs within the poor and vulnerable. This system is based on assessment of living conditions of individual families (Castañeda (2005)), and it is currently used to select beneficiaries of subsidized health insurance, educational subsidies and conditional cash transfers in Colombia. The information in the SISBEN databases could be considered as a census for poor populations; for example by 2002 60% of the total population in the country was registered in the system, and about 30% received benefit (Castañeda (2005)). The concise instrument to target the spending is a statistically derived "proxy means" index. The computation of the index relies on information about availability and quality of housing, basic public services, possession of durable goods, human capital endowments and current income. To collect this information, a questionnaire was implemented using mainly two tools for selecting the responders: using a geographical selection based on previous targeting systems, and by demand in hospitals, municipalities' offices, or other benefits suppliers.

The SISBEN was not initially conceived as a panel, but it is mandatory to update the information every 3 years. This means that in crossing the information of the original collection and updates a panel of the population surveyed could be constructed. For this paper the original collection and 2 updates have been used to construct a panel of 3 periods of information. Given that the survey was not initially conceived as a panel, there is not an official ID to following the same individuals through different periods. Nevertheless, there are mechanisms for matching individuals and households over time; unfortunately, great part of the observations will be missed, but given that there is a huge amount of observations in each period it is possible to preserve a satisfactory sample.

Given that the SISBEN is not a panel itself, the construction of the estimation sample is a process that involves several stages in which data from different SISBEN collections and external data are merged. As a result of this process the estimation sample is obtained. At this point is important to make clear that recent mothers included in the SISBEN system (Recent SISBEN Mothers, RSM) are the targeted population for this research. A recent mother in this paper, is one who had their first born any time between the first and the third SISBEN collections[7]. Roughly speaking this is between 2001 and 2010. The conclusions derived in this paper are only intended to be applicable to this population. Summarizing, the RSM are SISBEN surveyed women who became mothers or got pregnant, during the period in which SISBEN information is available. A recent SISBEN mother observation belongs to the estimation sample if it can be linked to a previous period. This is because I need to observed the covariates explaining the decision about the timing for the onset of childbearing before (or at) the pregnancy time. A summary statistics table comparing the RSM sample with the estimation sample is provided in Appendix A of this paper.

The total sample of Recent SISBEN Mothers that can be identified in any of the three SISBEN collection consist of 75768 individuals. Only a fraction of those individual may be linked to a previous SISBEN collection, which is crucial in order to know characteristics before pregnancy. At the end, the estimation sample consists of 11461 individuals. As mentioned before, the main reason for an observation being excluded of the estimation sample is that it cannot be linked to a previous SISBEN collection. Some covariates present differences between population and estimation sample, as the reader may see in Appendix A of this paper, nevertheless robustness checks show that this is not an issue driving the results of this research[8]. A whole subsection with more details on the construction of the estimation sample is available upon request[9].

5 Empirical Strategy

Empirically, the goal of this paper is the estimation of a single equation model where the dependent variable is the age in years of the mother at the first childbirth[10]. The main interest is the identification

[7]The average collection date of the SISBEN 3 is January 7, 2010 and the standard deviation are 57 days. The average SISBEN 2 collection date is May 22 of 2006, and the standard deviation are 316 days. The average collection date of the SISBEN 1 is September 9, 2002 and the standard deviation are 431 days.

[8]In order to make sure that the determination of the estimation sample is not a factor driving the results of this research, as a robustness check I estimate econometric models in which the process of selection into the estimation sample is modeled using Heckman selection procedures. The results of the endogenous peer effects coefficient do not show important variation after controlling for the selection.

[9]This section was included in the paper in previous versions of it, but following referees' suggestions in order to reduce the extension of the o paper I decided to separate this section from the paper and make it a supplemental section available upon request.

[10]This variable is replaced by the age of the woman if she is pregnant by the time of the survey

of endogenous social effects that could explain the dependent variable. Usually, the estimation of these effects is biased due to two fundamental problems, the reflection problem, and the group selection problem. In order to deal with the reflection problem, I implemented a strategy similar to the one in De Giorgi et al (2010). I am able to define peers groups that vary at the individual level. Once the reflection problem is controlled, there is still an endogeneity problem because the peer group effect is an individual's endogenous decision. In order to face this problem, I perform a standard instrumental variables (IV) methodology.

The instruments that I propose in this paper are based on the idea of using the expectation of outcomes and covariates computed only for the excluded peers[11]. These are peers of individual's peers who do not belong to the individual's peer group. These are good IVs because the covariates and the endogenous fertility variable of exclude peers explain the fertility outcome of individual's peers via social interactions. In addition, there is not a direct effect of these variables on the individual's fertility outcome because it is assumed that any effect that excluded peers' behavior may have on individual behavior works indirectly through the effect of peers' behavior on individual's behavior[12]. More details on this will provided later on. The following section starts by describing the main challenges that the estimation of endogenous peer effect estimation must face. These are the reflection problem and the endogenous nature of the peer group. After an introduction to each of these problems, I provide detail on the empirical strategy to overcome each one of them.

5.1 The Reflection Problem

5.1.1 Introduction to the reflection problem

Manski (1993) was the first paper to explain the concept of reflection in the literature about SI. Basically, the reflection is originated in the fact that, inside a social group, individuals are influencing their peer's behavior and being influenced by them. The term "reflection" comes because one cannot know if one's action is the cause or the effect of peers' influence (DiGiorgio et al, 2010). Consider the following equation where y_i is the woman's age at first childbirth, z_i represents a vector of individual and family characteristics of i, $E\left[y|N_i\right]$ stands for the mean of y in the neighbors group of individual i, and $E\left[z|N_i\right]$ is a vector that includes the mean of the exogenous variables z among those persons in the individual i's neighbors group.

$$y_i = \alpha + \beta E\left[y|N_i\right] + E\left[z|N_i\right]' \gamma + z_i' \eta + u_i \tag{7}$$

[11] Instruments of this nature have been proposed recently in the field (DiGiorgio et al, 2010; Bramoullé et al, 2009)

[12] In the case of exogenous covariates of excluded peers this statement can be formally proved in a system of equations framework.

This equation is the standard linear expression for the estimation of social interactions; it represents formally two types of social effects. The most important effect is represented in the coefficient β, formally known as the endogenous peer group effect. The endogenous effect is the response in the fertility behavior of mother i when the average fertility behavior of her reference group changes. The reader may notice that β is the main coefficient that I am trying to identify in this paper. Nevertheless, mother i may behave similar to her neighbors just because their socio-economic characteristics are similar and they share similar restrictions; that effect is usually known as contextual effect, and it will be captured by the vector of coefficients γ. Equation (7) is useful to illustrate the nature of the reflection problem. Taking the expectation conditional on the neighbors group N_i, solving for $E\left[y_i|N_i\right]$, and assuming $E\left[u_i|N_i\right]=0$, we get:

$$E\left[y_i|N_i\right]=\alpha+\beta E\left[y|N_i\right]+E\left[z|N_i\right]'\gamma+E\left[z_i|N_i\right]'\eta \tag{8}$$

$$E\left[y_i|N_i\right]=\frac{\alpha}{1-\beta}+E\left[z_i|N_i\right]'\left(\frac{\gamma+\eta}{1-\beta}\right) \tag{9}$$

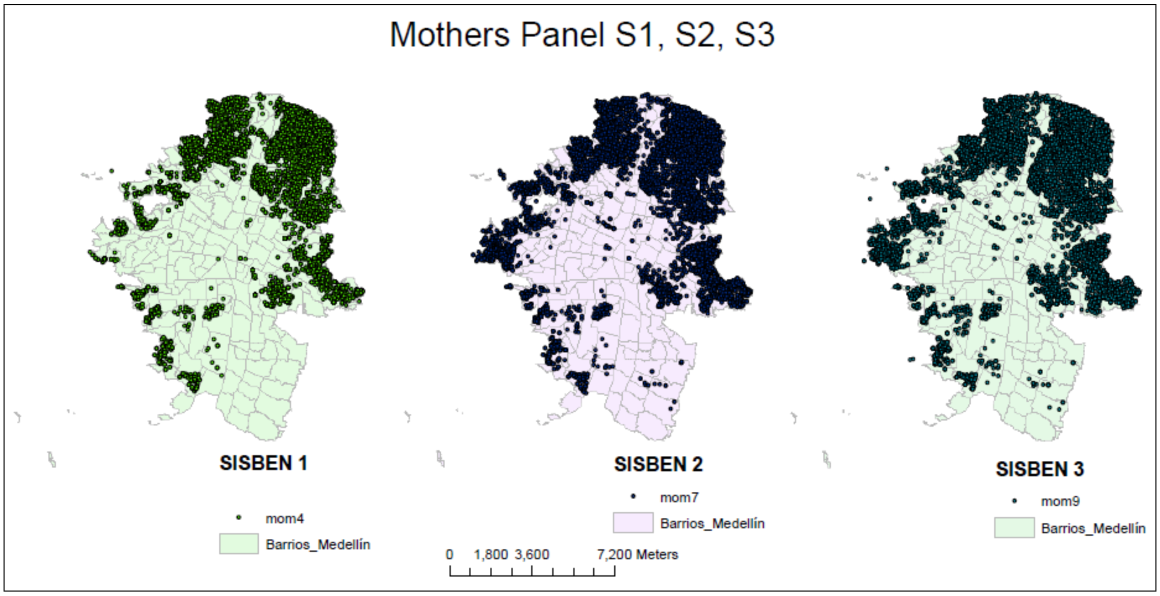
This straightforward algebra illustrates that in a standard setting, the parameters of interest cannot be identified separately. In this setting the peer groups are fixed across individuals; this means that if individual A is in the social group of individual B, and individual C is in the same social group that individual B is, it must be the case that individual A and C belongs to the same group. This characteristic causes that the term $E\left[y_i|N_i\right]$ appears in both sides of equation (2). In their 2010 paper, DeGiorgio et al show that identification can be achieved if instead of fixed, for every individual, peer groups are individual specific. If the neighbors groups are individual specific, the equation (3) can be re-written as:

$$E\left[y_i|N_i\right]=\alpha+\beta E\left[E\left(y|N_j\right)|N_i\right]+E\left[E\left[z_i|N_i\right]'|N_i\right]\gamma+E\left[z_i|N_i\right]'\eta \tag{10}$$

Where j represents a generic peer of the $i's$ neighbors group. Using a simple example in which the peer group are individual specific, DiGiorgio et al (2010) revealed that the identification is perfectly possible and that it relies only on observations with distinct peer groups.

5.1.2 Strategy

GIS information The relevant peer group for this research is a group of neighbors; these peers are defined according to some criteria of proximity and similarity, more technical details on this are provided in the next section. The relation with neighbors is by construction determined geographically. Therefore, GIS information is required in order to know who the neighbors of each individual are. To supply this necessity, the information provided by the administrative department of the city has the official codification for the census tract and block where the individuals' housing unit is located. Using this information, I merged the SISBEN data with an ARC-GIS shape file containing all the blocks in the city. As a result of this process, it is possible to know a closely approximated location for every household in the panel[13] in terms of their geographical coordinates[14]. In order to illustrate this procedure the following panel of maps shows the spatial location for all the SISBEN mothers in the panel (more specifically, the location of their block's centroid).



Map 1: SISBEN (recent) mothers.

Each point in the maps (mom4, mom7, mom9) represents the centroid of a block in the city, where one or more recent SISBEN mothers live. As the reader may remember, the estimation sample for this paper is a sample of SISBEN recent mothers. They are not necessarily beneficiaries of any social program, but they are registered in the system. The polygons in the map represent the most disaggregated geographical and political division in the city, "barrios" (neighborhoods). The maps show that the surveyed mothers are not located in every neighborhood in the city. Very highly valued neighborhoods, especially south-east, have low density of SISBEN households or nothing at all. SISBEN population is the set of potential beneficiaries of social programs; therefore, SISBEN households are usually poorer than the average household in the city. These maps at some extent show the economic spatial segregation, and the sorting patterns in the city. Disadvantaged (presumably) households are restricted to some areas of the city.

[13]The merge between the SISBEN panel and the GIS files for blocks allows, for most of the observations, to know the centroids' coordinates of the block where the household is located.

[14]Not for every observation in the panel was successfully matched with the GIS file of blocks, especially in S1 where the codification for census tracks and blocks was not available for all the observations. In those cases the centroids of the most disaggregated political division (barrios) is used.

Definition of the peer group The strategy used in this paper to overcome the reflection problem is similar to the one explained in the section 5.1.1. Therefore, the whole point is defining neighbor groups varying at the individual level. The basic idea in the definition of the reference group is to build a N by N matrix of weights (W); where N is the number of all SISBEN recent mothers in the estimation sample. The matrix operation $W.Y$, where Y is a vector containing the age at first childbirth for all SISBEN resent mothers in the estimation sample, is a nonparametric estimator of $E\left[Y\right]$. Therefore, a neighbor mother j can influence the fertility decision of the mother i, depending upon, the assigned weight she has in the computation of $\sum_{j\neq i} w_{ij}y_j$,where w_{ij} is an element of W and y_j is the $Y's$ element corresponding to the j^{th} neighbor of i.

The natural candidate for the weight w_{ij} is the inverse of the Euclidean distance between i and j. It is likely that mothers in contiguous blocks interact more than mothers separated by a considerable distance. Following the same logic, neighbors of the mother i, located farther away than a predetermined distance $\bar{d}$, should have no weight in the computation of the expectation; in other words, they should not belong to the peer group of the mother i. Furthermore, the distance between i and her neighbors is certainly not the unique criterion to exclude some mothers from the $i's$ peer group. There can be other social distances $\bar{s}^k$ (with $k = 1, 2, .., K$), such that, if neighbors are very different from i in any of the k characteristics, they also should be excluded from the $i's$ peer group. These characteristics can be socio-demographic variables such as education, age, or income. In the empirical work I defined several matrices using different criteria, and I estimate models using different matrices. Formally the construction of these matrices can be represented as:

$$W = \begin{bmatrix} 0 & S_{12}.\frac{1}{d_{12}} & \dots & S_{1N}.\frac{1}{d_{1N}} \\ S_{21}.\frac{1}{d_{21}} & 0 & \dots & S_{2N}.\frac{1}{d_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1}.\frac{1}{d_{N1}} & S_{N2}.\frac{1}{d_{N2}} & \dots & 0 \end{bmatrix} \tag{11}$$

$Where$

$$d_{ij} = \sqrt[2]{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{12}$$

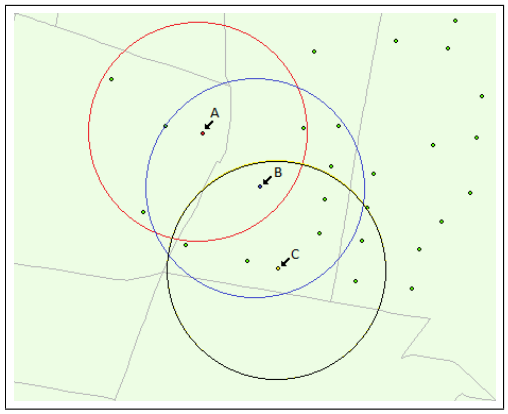
$$S_{ij} = 1\left\{d_{ij} < \bar{d}\right\} \times 1\left\{\left|s_i^1 - s_j^1\right| < \bar{d}^1\right\} \times 1\left\{\left|s_i^1 - s_j^1\right| < \bar{d}^2\right\} \times ... \times 1\left\{\left|s_i^K - s_j^K\right| < \bar{d}^K\right\} \tag{13}$$

In the empirical work a standardized[15] version of (11) is used. In expression (12), d_{ij} describes the Euclidean distance between mother i and mother j, where x, y stand for the geographical coordinates. In equation (13), S_{ij} describes a multiplication of indicators of functions; these indicators of functions

[15] The standardized version of W is a matrix such that the sum of every row or column is equal to one.

are equal to one when the condition inside of the parenthesis holds, and they are zero otherwise. The first condition is the distance condition; any neighbor beyond some radius $\bar{d}$ is excluded (i.e. a zero weight is assigned for that peer). The other conditions are the ones based on the socio-demographic variables, if a neighbor differs from i, such that, for any of these characteristics $|s_i - s_j| > \bar{d}$; then the neighbor is excluded. Age and education are the socio-demographic variables used in the construction of this matrix.

Figure 1: Peer group variation 1

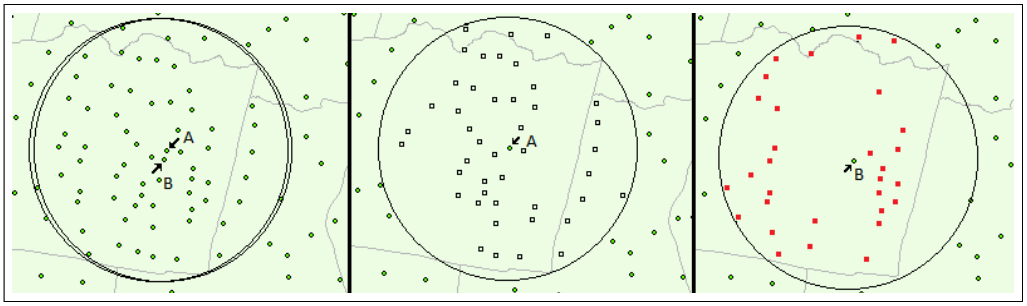


Peer groups varying at the individual level In figure (1) each point represents the centroid of a block, where a SISBEN mother resides; the polygons represent neighborhoods. Reader may notice that mother B belongs to the mother A's peer group, and she also belongs to the mother C's peer group, but A does not belong to C's group and neither C belongs to the A's group. When socio-demographic restrictions are operating the variation is greater. In the following figure (2), mother A and B are very close to each other; they live in contiguous blocks, but these mothers differ in some of the criteria used to form the social groups. Therefore, despite the fact that they live very close each other, none of them belongs to the social group of the other, and as it can be seen in figure (2) their peer groups are very different. The red filled squares represent peers of individual B, and the non-filled squares represent peers of individual A.

5.2 Fertility decision equation

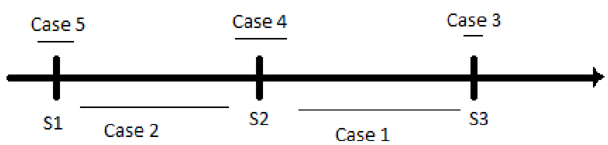
The fertility decision equation describes how family and personal characteristics, and the decisions of the peers, affect the age at which a mother decides to have her firstborn. The relevant information that explains this decision is the information previous to the pregnancy, or the information at the time of the pregnancy. The SISBEN information covers three different periods, based on those periods, I

Figure 2: Peer group variation 2



defined 5 different cases which describe the way how the information is used; the following timeline defines each case:

Figure 3: Sample Cases



The wide line represents the time line for a generic woman in the sample. The thin line represents different possibilities for the pregnancy time. In case 1, the woman got pregnant some time before she was surveyed in S3, but after she was surveyed in S2. In this case, the covariates and the expectation of the dependent variable among her peers are constructed with information from S2. In case 2 the woman got pregnant some time before she was surveyed in S2, but after she was surveyed in S1. In this case the covariates and the expectation of the dependent variable among her peers are constructed with information from the S1. The other three cases (3,4,5) represent the situation in which the woman was pregnant at the time of the survey. In these cases the covariates and the expectation of the dependent variable among her peers are constructed with information from the current survey at the time she was pregnant.

The fertility equation is specified as a linear social effects model, similar to the one represented in equation (7), which is estimated for recent mothers. In order to measure the endogenous peer effects, a nonparametric estimation of $E\left(y_i|G_j\right)$ is included in the fertility equation (where j represents a generic neighbor of the $i's$ neighbors group). The computation of this expectation[16] is $E\left(y_i|N_j\right) =$

[16] The reader may notice that in some of the cases described before, this expectation is not computed using the dependent variable in the period in which the women got pregnant but one period before. This helps to alleviate the simultaneity that generates the reflection problem and makes appears the expectation term in both sides of the equation

$\sum_{j\neq i} w_{ij}y_j$, where y_j is the fertility decision of neighbor j, and w_{ij} is the weight explained in section 5.1.2. In order to control for contextual effects a nonparametric estimation of $E\left(z_i|N_j\right)$ is also included in the fertility equation; similar to the previous case, this expectation is computed as $E\left(z_i|G_j\right) = \left[\sum_{j\neq i} w_{ij}z_{1j} \; ... \; \sum_{j\neq i} w_{ij}z_{kj}\right]$, where z_{kj} is the neighbor j's exogenous covariate z_k. The following equation is the one estimated in this paper.

$$y_i = \alpha + \beta.\sum_{i\neq j}\omega_{ij}y_j + \left[\sum_{j\neq i} w_{ij}z_{1j} \; ... \; \sum_{j\neq i} w_{ij}z_{kj}\right].\gamma + \sum_{b=1}^{B}\upsilon_b\delta_{bi} + z_i'\eta + u_i \tag{14}$$

Where y_i is the age of the mother i at first childbirth, z_i is a vector of individual and household characteristics. The parameter that represents the endogenous peer effects is β. As it was mentioned before, ω_{ij} represents an element of the weighting matrix W. The coefficient of interest in this paper is β; nevertheless, it is important to control for other non-endogenous effects. In the presence of these non-endogenous effects (contextual or correlated effects) the β coefficient could be overestimated. Contextual effects[17] have their roots in the fact that mothers in the same peer group have similar socio-economic composition, and fertility behavior may vary with different socio-economic characteristics of the group. Correlated effects may be important because individuals with the same characteristics or who face similar institutional environments tend to behave similar (Manski, 1993). In many situations the correlated and context effects are indistinguishable from each other because the characteristics of the reference group have to be defined in terms of averages of the exogenous variables. In order to control for these non-endogenous peer effects, I include in the regression the average of the exogenous covariates among those persons in the reference group of mother i. In addition, I include a set of dummy variables υ_b, where $b = 1...B$ is a index of the neighborhood and δ_{bi} is the coefficient measuring the fixed effect of neighborhood b.

5.3 Endogenous reference group problem and IVs.

In the estimation of SI effects there is an additional problem that has to do with the determination of the peer group. The problem arises because the peer group (neighbors or local residents) is often

(2).

[17]Exogenous or contextual effects are associated to the fact that individuals behavior vary with some exogenous characteristic of the group. The classical example is that in a classroom the achievement of a student may be explained by the socio-economic conditions in a school district. The reader may consider for example a school district where all student's parents have college degree, and another schooling district where no parent has college degree. One would expect that student achievement will go in the same direction of average parents' education in the district; but, this variation does not obey to any endogenous interaction between students. In equation (14), this is precisely what γ coefficients are capturing, the variation of individual behavior given changes in average exogenous characteristics of the group.

itself a matter of individual choice (Oates and Schwab (1992)). In other words, individuals self-select themselves in the peer group that best fits with their possibilities and preferences (unobserved factors). When individuals make their residential location decisions, choosing the neighborhood in which they want to live, they are at the same time choosing their peer group. Therefore, the expectation of the dependent variable, conditional on the social group, is an endogenous variable. Technically, there can be un-observables that are correlated with the location decisions, and they are also correlated with the fertility decision. Given that the location decision determines the woman's neighbors, an estimated expectation of the mother's age at her first childbirth (among a group of neighbors) is going to be correlated with the error term u_i in equation (14).

The estimation of (14) by OLS will be biased, even after applying the proposed procedure to correct for reflection. The direction of the bias depends on the unobservables driving the selection of the peer group, and the correlation of the later with the unobservables determining the fertility decision. In order to correct this bias, I performed an instrumental variable procedure. To obtain valid instruments I take advantage of the peers group structure proposed in the paper; I use information of the peers of individual's peers, which are not included in the individual's peer group. From now on this group will be referred as the excluded peers (EP). Using the information of EP is a practice that has been used very recently in papers that seek the identification of endogenous social effects (Bramoullé et al, 2009; De Giorgi et al, 2010).

Expectation of covariates and the fertility decision among the EP should have a strong correlation with the individual's fertility decision. This happens through a series of interconnected social interactions. EP exogenous covariates explain their fertility decisions and then via social interactions individual's peers fertility behaviors are explained by fertility behaviors of their peers (from the perspective of the individual they are excluded peers). By the same reason the fertility behavior of the individual is explained by fertility behavior of their peers. In addition, it can be formally proved[18] that expectation of covariates and the endogenous outcome among excluded peers[19] are exogenous to the individual's fertility decision. The intuition for this is that there is no direct effect of these variables on the individual fertility decision; all happens through the relationship between EP and individual's peers. Several instruments are used in the paper, all of them using the principle of the "excluded peers". I use the expectation among the EP of the fertility outcome, and the expectation of other

[18] In a system of equations framework, with one equation explaining the endogenous outcome for each individual it is easy to see that EP exogenous covariates are exogenous to the individual fertility decision. Reader may refer to De Giorgi (2010) for more details on the subject.

[19] In this case additional conditions are required. One of them is no correlation between EP unobservables and individual's unobservables.

covariates among the EP[20].

6 Results

The empirical strategy described in the previous section implies that several specifications are possible. This is because several matrices W can be used. The matrices are important because they define the peer group that is allowed to influence the fertility decision. In the previous section two types of criteria were mentioned as useful to construct the peer group: physical distance and social distance. In this paper the characteristics for the latter that are included are age and education. The following table defines the conditions inside the brackets of each indicator of function in the equation (13) that were used for the regressions presented in this paper[21].

Networking level	Distance (di)	Age (ai)	Education (ej)
low	$\lvert di-dj \rvert <=500$	$\lvert ai-aj \rvert <=5$	$\lvert ei-ej \rvert <=3$
medium		$\lvert ai-aj \rvert <=10$	$\lvert ei-ej \rvert <=5$
high	$\lvert di-dj \rvert <=1000$		$\lvert ei-ej \rvert <=7$

Table 1: Networking criteria

For each criterion (distance, age, and education) two or three levels of networking are used. In each level different sets of peers are allowed to affect the fertility decision (the peer receive a nonzero weight). Given the form of the restrictions these sets are nested. In the low level, the restriction in the criterion (distance, age, or education) is the strongest. Therefore, the number of neighbors included is smaller than in any other level; therefore, the number of peer excluded is bigger. In the medium or high levels, the restriction in the criterion is weaker. Therefore, the number of neighbors included is greater than in any other level. In other words, in the low level, a small distance (physical or social) is allowed between mothers to be included in a peer group; in the medium level, more distance is allowed between them. Let's consider for example the case of age. In the low networking level, a neighbor is included in the mother $i\prime s$ peer group if the age difference between the mother and this neighbor is less than or equal to five years. In the medium networking level, the neighbor is included if the difference is less than or equal to 10 years. Combinations of these restrictions will be used for determining the different specifications that are presented in this section.

[20]The expectation of exogenous covariates among the EP is a source of a big amount of instruments; to gain efficiency and make the specifications more parsimonuous, I selected the best of them in terms of the correlation with the endogenous variable, and using overidentifying restriction tests.

[21]Other criteria were used as well, results are comparable with the ones presented in the paper.

The following table presents a description of the covariates used in the regressions. For explaining the fertility variable, the covariates included are variables that intuitively may have an explanatory power in the decision about the age for the onset of childbearing. Taking advantage of the panel constructed with the SISBEN information, it is possible to know these covariates before the pregnancy time, or at the pregnancy time (in the case of mothers that were pregnant by the time they were surveyed). The covariates included are personal characteristics as educational attainment, marital status, employment status. Some household's characteristics as income or house's features were also included.

Variable Description	Obs	Mean	Std. Dev.	Min	Max
Age at first childbirth	11537	21.30164	5.037569	4	45
Education Attainment Incomplete elementary	11537	0.105833	0.307638	0	1
Education Attainment Complete elementary	11537	0.154113	0.361073	0	1
Education Attainment Incomplete high school	11537	0.453758	0.497879	0	1
Education Attainment Incomplete high school	11537	0.242611	0.42868	0	1
Education Attainment complete or incomplete high school	11537	0.02869	0.166942	0	1
Currently in school	11537	0.34203	0.47441	0	1
Currently unemployed	11537	0.081997	0.274372	0	1
Currently cohabitating with sentimental partner	11537	0.10878	0.311377	0	1
Currently widow	11537	0.0013	0.036036	0	1
Currently divorced	11537	0.012222	0.109878	0	1
Currently Single	11537	0.841987	0.364769	0	1
Women any sort physical or mental disability	11537	0.005634	0.074852	0	1
Monthly income of the household	11537	459963.9	408176.3	0	6588608
Dummy: living in house or apartment=1, other=0	11537	0.97105	0.167674	0	1
Dummy: Good or standard quality in walls material=1, other=0	11537	0.784693	0.411053	0	1
Dummy: Good or standard quality in floor material=1, other=0	11537	0.357979	0.479427	0	1
The house is own by the household	11537	0.480108	0.499626	0	1
Number of teenager mother in the household	11537	0.584727	0.77156	0	6
Dummy: stratum equal to 2=1	11537	0.331195	0.470664	0	1
Dummy: stratum equal to 2=1	11537	0.568952	0.495244	0	1
Dummy: stratum equal to 3=1	11537	0.096212	0.294895	0	1
Dummy: stratum equal to 4=1	11537	0.000173	0.013166	0	1

Table 2: Summary Statistics

6.1 OLS and IV Regressions

In this section the estimation results are presented. I used different restrictions for the maximum distance allowed between peers, and I conclude that the specifications that best fit the data are the ones using matrices with nonzero weights for peers inside a ball with 1000 meters of radius, and center in the mother i's residence. Therefore, the results presented in this section are of the specifications using this 1000 meter restriction by default. All regressions includes the set of neighborhood fixed effects $[\sum_b \upsilon_b \delta_{bi}]$ and the set of contextual effects $\left[\sum_{j\neq i} w_{ij} z_{1j} \ \ldots \ \sum_{j\neq i} w_{ij} z_{kj}\right]$.

As the reader may remember, the estimation sample used in the regressions is the sample of SISBEN recent mothers that can be followed throughout the different collections of the SISBEN (cases 1 to 5 in graph number three). The expectation of the endogenous variable among peers is therefore defined using that sample as well. Nevertheless, in order to reduce potential endogeneity of the contextual effects, all the expectations of the covariates among the peers are computed using the sample of SISBEN mothers which fulfill the conditions to be included as individual's peers; this regardless of the fact they are included or not in the estimation sample. The same strategy is used for the computation of the instruments. That strategy substantially improves the quality of the estimations in the paper.

In the following tables, I present the results of the estimation for different specifications of the W matrix and different set of instruments for a given W. Six different configurations for the peer group are presented in the tables; each configuration is a combination distance, age and education restrictions which define the peer group. All the restrictions of distance (d_i, d_j), education (e_i, e_j) and age (a_i, a_j) are indicated at the top of each table according with the criteria presented in table (1). In the first panel for example the configuration for the peer group includes peers located within a one-kilometer radius from the individual; in addition, the difference in age between the individual and the peer should be less than or equal to five years, and the disparity in years of education should be less than or equal to 3 years of education.

Inside each table's panel there are three subpanels; the first subpanel contains the OLS regression, and the second and third subpanels represents the 2SLS Regressions. I use two instruments in the regression in the second subpanel. The first instrument is the expectation of the fertility outcome among the Excluded Peers[22]. The second instrument is the expectation of the fertility outcome among the peers of excluded peers that at the same time do not belong to the peer group of individual's peer and neither to the individual's peer group (second level of excluded peers). The instruments used in the third panel are the expectation of some exogenous covariates among the Excluded Peers. I selected two of the best instruments for each regression[23]. The first stages for each one of the IV regressions are presented in the Appendix 2. The coefficients of the neighborhood fixed effects and the contextual effects have been omitted for presentational ease[24].

Table (3) and table (4) contain three different configurations for the peers group each. In all regressions the coefficient measuring the endogenous peer effects is positive, and less than one; in

[22] The process of identification of the excluded peers is developed in the following way. For each individual's peer, five of her closest peers are selected (in terms of the criteria discussed before) if they do not belong to the peer group of the original individual. The same process applies for the second level of excluded peers.

[23] More details about these instruments can be found in the table's footnotes.

[24] This information is available upon request

addition, it is highly significant in all 2SLS regressions and in most of the OLS regressions as well. In all specifications the OLS endogenous peer effects coefficients are smaller than the ones obtained by 2SLS. In this setting is impossible theoretically predicting the direction of the OLS bias, it would depend upon the correlation of individual factors driving the group selection with unobservables driving the fertility decision. For some specific unobserved factors the bias would be certainly negative. Let's say, for instance, there is an individual's restriction, unobserved to the econometrician, which have a positive effect in the fertility outcome; this is, the restriction delays the individual's onset of childbearing. Consider for example the presence of credit restrictions; if credit restriction is an unobserved component that influence the fertility outcome, probably, it will have a positive effect. The more restricted the less prone individuals will be to starting a family. Furthermore, the more restricted individuals are the more prone they are to living in bad quality neighborhoods, which are the ones with the smallest average of the fertility outcome. Therefore, in that case the correlation of unobservables with the endogenous variable would be negative and the OLS bias as well[25].

The endogenous peer effect coefficient is interpreted as the effect performed by the peer group on the individual's fertility decision. The coefficient describes the response in the mother's behavior given changes in the average behavior of the peer group; it is positive and significant in almost all specifications. Therefore, the mother's age at first childbirth is explained positively by the average age at first childbirth among mothers within her peer group. The endogenous peer effect coefficient is always positive and less than one in all 2SLS regressions; nevertheless, the coefficients are different for different definitions of the peer group, and different instruments as well. As it was already mentioned each set of instrument has their own advantages; the ones based on EP covariates are theoretically exogenous, and the ones based on EP outcomes are expected to be stronger. The preferred specification in this paper is the one described in panel [4], as reader may notice from the bottom of the table in this specification a mother is consider an individual's peer if she lives within a radius of 1000 meters of the individual's residential location, the age disparity with the individual is not greater than 10 years, and the educational disparity is not greater than 3 years of education. This specification is preferred for several reasons; the F test is remarkable high for the two sets of instruments used, the exogeneity test is strongly not rejected in both 2SLS regressions, and the difference between both 2SLS estimation is not very high. Nevertheless, the endogenous peer effects coefficients are specific for each definition of the peer group; therefore, other coefficients based on alternative specifications are also valid.

The quality of the instruments included in the regressions is tested using a F-test of the instruments

[25]The final direction of the OLS bias could be more complicated in the presence of correlation of some groups effects, if their effect is not ruled out by the introduction of neighborhood fixed effects and the contextual effects.

in the first stage regression[26].The F statistics is presented at the bottom of each table; in every regression presented in tables 3 and 4 the F-statistics is greater than 10, which is a minimal requirement for the consistency of the estimators. The exogeneity of the instruments cannot be strictly tested; nevertheless, over-identifying restriction test[27] are always informative. In none of the estimations presented in table 3 and 4 the null hypothesis of the exogeneity of the instruments in the over identifying restriction tests can be rejected.

The endogenous peer effects coefficient in the preferred specification and the best set of instruments is 0.70 (Panel 4, table 4). This means that an increase one of 1year in the peers group's average age at first childbirth implies an increment of the mother age at first childbirth of 0.7 years. In several of the alternative specifications the estimated coefficients are similar, but for some others, the magnitude is substantially lower; nevertheless, in any case the coefficients are strongly significant. In order to facilitate the interpretation of the endogenous peer effects coefficients, the results of a couple of simulations based on the estimated coefficients are presented at the bottom of table (3) and (4). In the first simulation, the incidence of teenage motherhood among the individual's peer group is increased in 10 percentage points. In order to do this, I randomly assign an age at the firstborn of 16 to members of the individual's peer group. I do this until the percentage of teenage mothers increase in 10 percentage points. In the second simulation, the age at the onset of childbearing is reduced in one standard deviation for each member of the peer group. In the case of the preferred specification an increase of 10 percentage points in the incidence of teenage motherhood would reduce the age at the onset of childbearing in 0.5 years. On the other hand, a reduction of 1 standard deviation in the onset of childbearing for each member of the peer group will reduce the individual's onset of childbearing in more than 3 years.

[26] The null hypothesis of the F test is $H_0 : \theta_1 = 0,\ \theta_2 = 0$ Where θ_1, θ_2 represent the coefficients of each instrument in the first stage regression.

[27] The over identifying restriction test is obtained as $N.R_u^2$; where N and R_u come from an auxiliary regression of $\hat{u}_i$ on $[X\ Z]$. In this auxiliary regression X stands for the matrix of exogenous covariates and Z stands for the matrix of instruments (Wooldridge, (2002)). $N.R_u^2$ is distributed χ^2 with freedom grades equal to the number of overidentifying restrictions. The null hypothesis of this test is the exogeneity of the instruments, mathematically $H_0 : E(Z'u) = 0$

Table 3: Estimation Results Specifications [1], [2], [3]

Variables	[1]			[2]			[3]		
	|di−dj|<1k, |ai−aj|<=5, |ei−ej|<=3			|di−dj|<1k, |ai−aj|<=5, |ei−ej|<=5			|di−dj|<1k, |ai−aj|<=5, |ei−ej|<=7		
	R-squared	0.840	0.830		0.861	0.850		0.866	0.862
	N	11465	11465		11465	11465		11465	11465
	OLS	2SLS: Wypp[1]	2SLS: Wxpp[2]	OLS	2SLS: Wypp[1]	2SLS: Wxpp[3]	OLS	2SLS: Wypp[1]	2SLS: Wxpp[3]
Wy	0.175***	0.403***	0.529***	0.144***	0.205**	0.513***	0.090**	0.283***	0.430**
	(0.028)	(0.113)	(0.164)	(0.037)	(0.081)	(0.179)	(0.041)	(0.085)	(0.201)
Some elementary	-0.305	0.339	0.231	1.663***	1.746***	1.663***	1.107***	0.930***	0.855***
	(0.391)	(0.395)	(0.406)	(0.293)	(0.287)	(0.279)	(0.208)	(0.209)	(0.208)
Elementary	-0.056	0.765*	0.662	1.913***	2.028***	1.937***	1.096***	0.836***	0.727***
	(0.425)	(0.416)	(0.428)	(0.303)	(0.295)	(0.286)	(0.205)	(0.206)	(0.222)
Some high school	-0.002	0.690	0.546	1.436***	1.649***	1.386***	0.945***	0.741***	0.613***
	(0.435)	(0.442)	(0.448)	(0.357)	(0.319)	(0.309)	(0.211)	(0.209)	(0.227)
High school	0.570	1.307***	1.172**	2.805***	2.924***	2.600***	1.620***	1.422***	1.295***
	(0.441)	(0.451)	(0.459)	(0.399)	(0.392)	(0.386)	(0.227)	(0.225)	(0.246)
College, some College	0.195	0.952**	0.786	2.545***	2.652***	2.272***	2.861***	2.492***	2.336***
	(0.425)	(0.457)	(0.477)	(0.389)	(0.384)	(0.391)	(0.291)	(0.286)	(0.293)
Assist to school	0.195***	0.202***	0.195***	0.198***	0.188***	0.160***	0.117*	0.159***	0.159***
	(0.058)	(0.058)	(0.058)	(0.057)	(0.058)	(0.061)	(0.061)	(0.056)	(0.056)
Unemployed	0.269***	0.260***	0.260***	0.301***	0.288***	0.300***	0.320***	0.315***	0.325***
	(0.068)	(0.069)	(0.070)	(0.063)	(0.062)	(0.064)	(0.062)	(0.060)	(0.063)
Cohabitating with sentimental partner	-0.480***	-0.447***	-0.498***	-0.375***	-0.318***	-0.420***	-0.337***	-0.294**	-0.340***
	(0.125)	(0.149)	(0.161)	(0.116)	(0.117)	(0.135)	(0.112)	(0.115)	(0.121)
Widow	-0.017	0.028	-0.024	0.015	0.143	0.124	0.149	0.215	0.195
	(0.459)	(0.498)	(0.535)	(0.395)	(0.409)	(0.449)	(0.347)	(0.367)	(0.396)
Divorced	-0.297	-0.223	-0.248	-0.164	-0.114	-0.124	-0.114	-0.016	-0.017
	(0.210)	(0.230)	(0.235)	(0.188)	(0.196)	(0.203)	(0.184)	(0.197)	(0.192)
Single	0.024	0.061	0.009	0.159	0.218**	0.123	0.173*	0.222**	0.172
	(0.109)	(0.134)	(0.145)	(0.104)	(0.107)	(0.120)	(0.099)	(0.103)	(0.106)
Physical or mental disability	0.109	0.414	0.429*	0.253	0.495**	0.529**	0.444*	0.408*	0.408*
	(0.291)	(0.255)	(0.258)	(0.275)	(0.240)	(0.251)	(0.229)	(0.239)	(0.242)

Table 3: Estimation Results (Continuationued from previous page)

Monthly household Income	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
living in apartment	-0.218*	-0.102	-0.093	-0.194	-0.125	-0.125	-0.165	-0.112	-0.118
	(0.122)	(0.123)	(0.132)	(0.120)	(0.108)	(0.131)	(0.106)	(0.103)	(0.113)
Good or standard quality in walls material	0.072	0.097**	0.088*	0.065	0.081*	0.069	0.048	0.058	0.049
	(0.045)	(0.047)	(0.049)	(0.043)	(0.044)	(0.046)	(0.044)	(0.043)	(0.044)
Good or standard quality in floor material	0.038	0.040	0.025	0.043	0.059	0.028	0.060*	0.059	0.044
	(0.039)	(0.040)	(0.041)	(0.038)	(0.038)	(0.042)	(0.036)	(0.037)	(0.040)
House is owned by the household	0.077*	0.061	0.062	0.079*	0.071	0.075	0.064	0.052	0.053
	(0.044)	(0.044)	(0.044)	(0.045)	(0.044)	(0.046)	(0.041)	(0.041)	(0.042)
# of teen mothers in the household	-0.107***	-0.098***	-0.095***	-0.114***	-0.109***	-0.099***	-0.108***	-0.102***	-0.099***
	(0.024)	(0.025)	(0.026)	(0.024)	(0.024)	(0.025)	(0.022)	(0.022)	(0.024)
Stratum[4] 2	-0.075	-0.043	-0.039	-0.033	-0.030	-0.025	-0.034	-0.015	-0.016
	(0.064)	(0.062)	(0.064)	(0.060)	(0.059)	(0.061)	(0.057)	(0.055)	(0.055)
Stratum[4] 3	-0.031	-0.028	-0.020	-0.013	-0.026	-0.016	-0.055	-0.026	-0.023
	(0.109)	(0.116)	(0.122)	(0.100)	(0.098)	(0.101)	(0.101)	(0.100)	(0.102)
Neigborhood Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Contextual Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
F		14.17	11.36		28.44	17.03		30.70	16.49
OID Pvalue[5]		0.55	0.66		0.98	0.99		0.98	0.77
First Simulation[6]		0.28	0.37		0.14	0.35		0.20	0.31
Second Simulation[7]		1.57	2.06		0.75	1.83		1.01	1.54

Notes: Robust standard errors in parentheses. All instruments use peer group definition |di-dj|<=1000,|ai-aj|<=5,|ei-ej|<=3; which is the one where more excluded peers are generated. All Instruments based on EP information presented here, impose a minimum distance of 250 meters between the individual and the Excluded Peer.

(1) Two instruments are used: the expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers.

(2) Two instruments are used; the expectation of the variables: Incomplete and Complete College and Standard quality in walls materials among the excluded peers

(3) Two instruments are used; the expectation of the variables: Widow and Standard quality in walls materials among the excluded peers

(4) Stratum is a socio-economic classification of the neighborhoods in Colombia as a strategy to target subsidies programs. The omitted alternative is stratum1.

(5) Overidentifying restriction Haussman Test (Pvalues)

(6) Change in individual fertility behavior caused by an increase of 10 percentage points in the incidence of teenage motherhood among the individual's peer group

(7) Change in individual fertility behavior caused by a reduction of 1 standard deviation in the onset of childbearing for each member of the peer group.

*** Significant at 1 percent level.

** Significant at 5 percent level.

* Significant at 10 percent level

Table 4, Estimation Results Specifications [4], [5], [6]

Variables	[4]			[5]			[6]		
	|di-dj|<1k, |ai-aj|<=5, |ei-ej|<=3			|di-dj|<1k, |ai-aj|<=5, |ei-ej|<=5			|di-dj|<1k, |ai-aj|<=5, |ei-ej|<=7		
	R-squared	0.794	0.800		0.831	0.819		0.847	0.839
	N	11465	11465		11465	11465		11465	11465
	OLS	2SLS: Wypp[1]	2SLS: Wxpp[3]	OLS	2SLS: Wypp[1]	2SLS: Wxpp[3]	OLS	2SLS: Wypp[1]	2SLS: Wxpp[3]
Wy	0.199***	0.696***	0.583***	0.065	0.285***	0.662***	0.033	0.293***	0.605***
	(0.041)	(0.114)	(0.135)	(0.050)	(0.085)	(0.165)	(0.044)	(0.091)	(0.175)
Some elementary	0.318	1.230***	1.251***	1.161***	1.082***	1.023***	1.433***	1.283***	1.172***
	(0.536)	(0.413)	(0.399)	(0.335)	(0.339)	(0.354)	(0.285)	(0.280)	(0.281)
Elementary	0.317	1.533***	1.528***	1.314***	1.203***	1.110***	1.488***	1.253***	1.075***
	(0.593)	(0.448)	(0.432)	(0.359)	(0.361)	(0.375)	(0.286)	(0.279)	(0.288)
Some high school	0.497	1.525***	1.538***	1.391***	1.232***	0.974**	1.414***	1.231***	1.041***
	(0.594)	(0.459)	(0.447)	(0.369)	(0.362)	(0.377)	(0.284)	(0.273)	(0.280)
High school	1.717***	2.782***	2.800***	2.783***	2.550***	2.319***	2.297***	2.168***	1.947***
	(0.628)	(0.468)	(0.459)	(0.449)	(0.454)	(0.471)	(0.326)	(0.311)	(0.316)
College, some College	1.597**	2.749***	2.798***	3.101***	2.834***	2.579***	3.999***	3.872***	3.704***
	(0.664)	(0.477)	(0.465)	(0.432)	(0.438)	(0.461)	(0.422)	(0.404)	(0.402)
Assist to school	-0.128*	-0.227***	-0.199***	-0.057	-0.122*	-0.246***	0.021	-0.026	-0.107
	(0.065)	(0.069)	(0.068)	(0.062)	(0.067)	(0.080)	(0.061)	(0.066)	(0.077)
Unemployed	0.370***	0.422***	0.406***	0.381***	0.406***	0.463***	0.396***	0.427***	0.472***
	(0.081)	(0.080)	(0.083)	(0.072)	(0.071)	(0.078)	(0.067)	(0.067)	(0.072)
Cohabitating with sentimental partner	-0.564***	-0.633***	-0.625***	-0.431***	-0.457***	-0.578***	-0.347***	-0.387***	-0.433***
	(0.143)	(0.173)	(0.163)	(0.133)	(0.134)	(0.148)	(0.124)	(0.124)	(0.132)
Widow	-0.043	-0.043	-0.049	0.117	0.224	0.194	0.200	0.193	0.215
	(0.507)	(0.600)	(0.575)	(0.405)	(0.441)	(0.510)	(0.370)	(0.421)	(0.476)
Divorced	-0.191	-0.132	-0.151	-0.124	-0.102	-0.116	-0.099	-0.036	0.015
	(0.251)	(0.261)	(0.253)	(0.209)	(0.210)	(0.214)	(0.195)	(0.192)	(0.188)
Single	-0.245**	-0.257*	-0.257*	-0.084	-0.095	-0.195	0.009	-0.010	-0.041
	(0.118)	(0.149)	(0.139)	(0.117)	(0.120)	(0.131)	(0.108)	(0.108)	(0.114)
Physical or mental disability	0.182	0.682**	0.662**	0.207	0.568*	0.661*	0.460*	0.396	0.419
	(0.391)	(0.325)	(0.315)	(0.374)	(0.306)	(0.340)	(0.274)	(0.290)	(0.311)

Table 4, Estimation Results (Continuationued from previous page)

Monthly household Income	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
living in apartment	-0.270**	-0.114	-0.123	-0.188*	-0.179	-0.152	-0.122	-0.134	-0.129
	(0.125)	(0.116)	(0.116)	(0.106)	(0.109)	(0.118)	(0.097)	(0.101)	(0.107)
Good or standard quality in walls material	0.058	0.116**	0.109*	0.073	0.085*	0.102*	0.051	0.061	0.071
	(0.053)	(0.057)	(0.056)	(0.046)	(0.047)	(0.052)	(0.047)	(0.047)	(0.050)
Good or standard quality in floor material	0.081*	0.075	0.079	0.092**	0.093**	0.075*	0.096**	0.084**	0.078*
	(0.049)	(0.048)	(0.049)	(0.045)	(0.043)	(0.045)	(0.042)	(0.042)	(0.043)
House is owned by the household	0.073	0.057	0.061	0.077	0.071	0.072	0.060	0.060	0.059
	(0.050)	(0.051)	(0.050)	(0.049)	(0.049)	(0.050)	(0.045)	(0.044)	(0.045)
# of teen mothers in the household	-0.140***	-0.125***	-0.130***	-0.136***	-0.128***	-0.117***	-0.124***	-0.117***	-0.108***
	(0.027)	(0.027)	(0.027)	(0.025)	(0.025)	(0.026)	(0.024)	(0.024)	(0.024)
Stratum[4] 2	-0.088	-0.074	-0.070	-0.006	-0.032	-0.036	-0.009	-0.005	-0.013
	(0.073)	(0.072)	(0.072)	(0.068)	(0.067)	(0.069)	(0.061)	(0.060)	(0.062)
Stratum[4] 3	-0.085	-0.048	-0.049	0.015	-0.006	-0.018	-0.027	0.015	-0.004
	(0.137)	(0.137)	(0.134)	(0.117)	(0.116)	(0.122)	(0.116)	(0.116)	(0.120)
Neigborhood Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Contextual Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
F		50.35	20.85		85.42	18.46		117.93	24.80
OID Pvalue[5]		1.00	0.98		0.30	1.00		0.27	0.98
First Simulation[6]		0.53	0.44		0.20	0.45		0.24	0.49
Second Simulation[7]		3.12	2.60		1.19	2.71		1.33	2.70

Notes: Robust standard errors in parentheses. All instruments use peer group definition |di-dj|<=1000,|ai-aj|<=5,|ei-ej|<=3; which is the one where more excluded peers are generated. All Instruments based on EP information presented here, impose a minimum distance of 250 meters between the individual and the Excluded Peer.

(1) Two instruments are used: the expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers.

(2) Two instruments are used; the expectation of the variables: Incomplete and Complete College and Standard quality in walls materials among the excluded peers

(3) Two instruments are used; the expectation of the variables: Widow and Standard quality in walls materials among the excluded peers

(4) Stratum is a socio-economic classification of the neighborhoods in Colombia as a strategy to target subsidies programs. The omitted alternative is stratum1.

(5) Overidentifying restriction Haussman Test (Pvalues)

(6) Change in individual fertility behavior caused by an increase of 10 percentage points in the incidence of teenage motherhood among the individual's peer group

(7) Change in individual fertility behavior caused by a reduction of 1 standard deviation in the onset of childbearing for each member of the peer group.

*** Significant at 1 percent level.

** Significant at 5 percent level.

* Significant at 10 percent level

Almost in all 2SLS specifications higher educational achievements, especially complete high school and some college, have positive and significant effects. These coefficients are smaller in the specifications where the educational disparity restriction between individuals and peers is strong (panel 1). When this restriction is relaxed the estimated coefficients for educational achievements are such that having some college or beyond implies at least two years of delay in individual's onset of childbearing. The positive effect of unemployment is significant in all specifications, which capture the fact that everything else constant unemployed women are less prone to starting a family. Family income also has a positive and significant coefficient in all specifications, which is consistent with the fact that poor mothers are usually younger. The dummy for cohabitation with their sentimental partner is always negative and significant; therefore, personal relationships out of marriage have a negative effect in the fertility outcome. The variable number of mothers who had their firstborn when they were teenagers, and that currently live together with the individual is significant in every specification and had a negative impact. In most of the cases these teen mothers are sisters or sisters in law of the mother. This variable is important because it is a proxy measure of fertility behavior in the individual's household.

In order to put in context the effect of the endogenous peer effects presented in the previous tables. It may be useful to compare the magnitude of the endogenous peer effect coefficient with the magnitude of other variables' coefficients. Using the results of specification number 4, one can see that the effect of a reduction of one year in the age at the firstborn for all individual's peer is a reduction of almost 0.69 years in the endogenous fertility outcome (using the WY set of instruments in the third column). This reduction is similar in magnitude to the one associated with the dummy variable for cohabitation (-0.63). In other words, the peer effects have almost the same effect on the endogenous fertility outcome as the important individual decision of cohabitating with a love partner. Another interesting comparison can be made using the results of specification 4 is the following. The effect of college or some college on the fertility outcome studied in this paper is a significant delay of almost 3 years in the age that women decide for the onset of childbearing. The response of the individual given an increase of one standard deviation (5 years) in the age at the firstborn of all members of her peer group, would be a delay in the age that women decide for the onset of childbearing of almost 3 years. In other words, a big perturbation in the peers' fertility outcome (one standard deviation) would have the same impact of having college or some college on individual decision for the onset of her childbirth. These comparisons give an idea of the importance of the endogenous peer effects.

7 Robustness

As the reader may remember, information before or at the time of the pregnancy was used in this paper to estimate the fertility equation. A potential drawback of this methodology is the following. If this information is much time before the pregnancy, it may not describe the characteristics of the mother at the real time of the pregnancy. The covariates and the expectation of the dependent variable used in the previous section estimations are constructed based on information that may be from several years before woman's pregnancy. For example, consider a woman who got pregnant before she was interviewed in S3 and after she was interviewed in S2; there can be more than three and a half years between these two events because the average date of collection for S3 is January 2010, and the average date of collection for S2 is May 2006.

The purpose in this section is testing if the peer effect coefficient is sensible to changes in the amount of time allowed between the pregnancy and the time when the information is collected. This robustness check consists of the estimation of the same specifications presented in previous section, but the sample is restricted to mothers who were pregnant not much time before they were interviewed in the SISBEN. The fertility equation specifications, [1] to [6], are the same as indicated in the tables' (3) and (4) headlines. Three sample restrictions are imposed, which are denoted by A, B, and C in table 5. Sample A only includes mothers who got pregnant at most two years before they were interviewed in the SISBEN, mothers who got pregnant before this 2 year threshold are excluded from the sample. Sample B only includes mothers who got pregnant, at most, one year before they were interviewed in the SISBEN. Finally, sample C only includes mothers who were pregnant at the time they were interviewed in the SISBEN.

Table 5, Robustness Checks

	[1]		[2]		[3]		[4]		[5]		[6]	
A: Sample restricted to mothers who got pregnat 2 years before the SISBEN Interview												
Sample size	8,163	8,164	8,163	8,164	8,163	8,164	8,163	8,164	8,163	8,164	8,163	8,164
R^2	0.879	0.867	0.901	0.873	0.903	0.879	0.842	0.843	0.877	0.863	0.892	0.884
Coefficient	0.491***	0.629**	0.313***	0.756***	0.411***	0.762***	0.615***	0.583***	0.216**	0.649***	0.229**	0.560***
Std dev	(0.105)	(0.249)	(0.086)	(0.233)	(0.109)	(0.280)	(0.102)	(0.171)	(0.096)	(0.200)	(0.098)	(0.179)
B: Sample restricted to mothers who got pregnat 1 year before the SISBEN Interview												
Sample size	5,995	5,996	5,995	5,996	5,995	5,996	5,995	5,996	5,995	5,996	5,995	5,996
R^2	0.876	0.868	0.909	0.865	0.907	0.871	0.844	0.858	0.891	0.878	0.908	0.901
Coefficient	0.629***	0.691**	0.440***	0.872***	0.563***	0.905**	0.747***	0.487**	0.325***	0.637***	0.327***	0.558***
Std dev	(0.142)	(0.292)	(0.116)	(0.328)	(0.159)	(0.437)	(0.116)	(0.187)	(0.102)	(0.227)	(0.102)	(0.210)
C: Sample restricted to mothers who were pregnat at the time of the SISBEN Interview												
Sample size	3,462	3,463	3,462	3,463	3,462	3,463	3,462	3,463	3,462	3,463	3,462	3,463
R^2	0.883	0.871	0.919	0.754	0.929	0.486	0.845	0.844	0.904	0.888	0.922	0.911
Coefficient	0.606***	0.706*	0.426***	1.664	0.359***	2.740	0.852***	0.857***	0.252*	0.699**	0.277*	0.685**
Std dev	(0.165)	(0.420)	(0.127)	(1.124)	(0.131)	(3.495)	(0.174)	(0.283)	(0.138)	(0.308)	(0.146)	(0.265)

Notes:

[1] Peer Group definition: $|di-dj|<1k$, $|ai-aj|<=5$, $|ei-ej|<=3$.

First Column Instruments: The expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers

Second Column Instruments: The expectaition of the variables Divorsed and Standard quality in walls materials among the excluded peers

[2] Peer Group definition: $|di-dj|<1k$, $|ai-aj|<=5$, $|ei-ej|<=5$.

First Column Instruments: The expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers

Second Column Instruments: The expectation of the variables divorced and widow among the excluded peers

[3] Peer Group definition: $|di-dj|<1k$, $|ai-aj|<=5$, $|ei-ej|<=7$.

First Column Instruments: The expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers

Second Column Instruments: The expectation of the variables divorced and widow among the excluded peers

[4] Peer Group definition: $|di-dj|<1k$, $|ai-aj|<=10$, $|ei-ej|<=3$.

First Column Instruments: The expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers

Second Column Instruments: The expectation of the variables Widow and Standard quality in walls materials among the excluded peers

[5] Peer Group definition: $|di-dj|<1k$, $|ai-aj|<=10$, $|ei-ej|<=5$.

First Column Instruments: The expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers

Second Column Instruments: The expectation of the variables Widow and Standard quality in walls materials among the excluded peers

[6] Peer Group definition: $|di-dj|<1k$, $|ai-aj|<=10$, $|ei-ej|<=7$.

First Column Instruments: The expectation of the outcome among the Excluded Peers, and the expectation of the outcome among the Excluded Peer of the Excluded Peers

Second Column Instruments: The expectation of the variables Widow and Standard quality in walls materials among the excluded peers

As a result of the restrictions, the sample estimation is significantly reduced. In the first restriction the sample is reduced in about 1/3, in the second the reduction is about 1/2, and in the last one the reduction is about 2/3 of the sample. In the case of sample reduction A and B, the peer effects coefficient remains positive and strongly significant; in most of the cases similar in magnitude or bigger, but always less that one. In the case of restriction C, where only mothers who were pregnant at the time of the interview are included in the sample, with the exception of a couple of specifications, in all the others the endogenous peer effect coefficient is positive and significant as well; furthermore, the magnitude of the coefficients remain similar or bigger and always less than one.

8 Conclusions

This paper presents evidence on the existence of endogenous peer effects that explain the SISBEN mothers' onset of childbearing. In the preferred model specification, the endogenous peer effect coefficient is 0.7; which means that a reduction of one year in the age at first childbirth for all members of the peer group will cause a reduction of 0.7 years in the individual's onset of childbearing. Using different definitions for the peers group the coefficient may vary, but in in all 2SLS specifications presented in section 6 the coefficient for β in equation (14) is positive and strongly significant. From simulations based on the estimated coefficients I obtain that, in the case of the preferred model specification, a reduction of one standard deviation in the onset of childbearing for each member of the peer group will cause a significant reduction of 3.1 years in the individual's onset of childbearing. This reduction of 3.1 in the individual's onset of childbearing for the sample of recent SISBEN mothers will correspond to an increase of the teenage-motherhood prevalence of 20 percentage points.

In many aspects Medellin is good representation of a standard LAC city; with high levels of teenage motherhood in poor neighborhoods. Therefore, peer effects are probably one of the factors explaining the generalized reduction in the average onset of childbearing observed for almost of LAC countries. In the case of Colombia this paper present evidence that social interactions play a crucial role in the determination of fertility outcomes, and is one of the explanations for the increase in the teenage fertility rates observed in the last 30 years.

From theoretical point of view, individuals are able to decide the best time for starting a family; nevertheless, there are factors that interfere with this process and individuals end up making inefficient decisions, as teenage pregnancy for instance. This paper provides evidence that peer effects is one of those factors. In the presence of peer effects high incidence of teenage pregnancies among the individual's peer group will cause reductions in the individual's age for the onset of childbearing. In simulations based on the estimated coefficients of the model I found that an increase of 10 percentage points in the adolescent fertility incidence among the individual's peers will cause a reduction of 0.5 years in the individual's age at the firstborn. It is not easy to think of a policy that can control a social force like peer effects. Nevertheless, is clear that socioeconomic segregation is a factor that exacerbates the negative effects of them. In segregated cities individuals in poor neighbors have as their peers mothers that had their onset of childbearing earlier than the population average. In a situation like this, peer effects can be seen as a factor contributing to the formation of poverty traps. Any social policy that contributes to a more random spatial distribution of households in the city, in terms of socioeconomic conditions, would help to reduce the negative consequences of social interactions in fertility decisions.

References

[1] Barrera Felipe and Higuera Lucas (2004). "Embarazo y Fedundudad Adolescente". *Documentos de Trabajo FEDESARROLLO* No. 24

[2] Becker, G. S. (1981): *A Treatise on the Family*. Harvard University Press, Cambridge, MA, USA.

[3] Case, Anne C., and Katz, Lawrence. "The Company You Keep: The Effects of Family and Neighborhoods on Disadvantaged Youth." Manuscript. Cambridge, Mass. NBER, 1990.

[4] Buvinic, M. 1998. "Costos de la Maternidad Adolescente en Barbados, Chile, Guatemala and México." Studies in Family Planning 29(2): 201-209.

[5] Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of Peer Effects through Social Networks." *Journal of Econometrics*, 150(1): 41–55.

[6] Castañeda Tarsicio (2005). "Targeting Social Spending To The Poor With Proxy–Means Testing: Colombia's SISBEN System". Social Protection Discussion Paper Series No. 0529, World Bank.

[7] Coleman, James S., et al. (1966). "Equality of Educational Opportunity" Washington: Government Printing Office, 1966.

[8] David M. Cutler and Edward L. Glaeser, (1997), "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics*, Vol. 112, No. 3 (Aug., 1997), pp. 827-872

[9] De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of Social Interactions through Partially Overlapping Peer Groups." *American Economic Journal: Applied Economics*, 2(2): 241–75.

[10] Florez and Nuñez (2001). "Teenage Childbearing in Latin American Countries" Inter-American Development Bank Banco Interamericano de Desarrollo. Research Network Working paper #R-434

[11] Flórez Carmen and Soto Eugenia (2007A). "Fecundidad adolescente y desigualdad en Colombia" *Notas de Población* No. 83

[12] Flórez Carmen and Soto Eugenia (2007B). "Fecundidad adolescente y pobreza Diagnóstico y lineamientos de política" Departamento Nacional de Planeación (Colombia), 2007

[13] Gaviria Alejandro, Medina Carlos, Morales Leonardo, Núñez Jairo (2010). "The Cost of Avoiding Crime: The Case of Bogotá" In Rafael Di Tella, Sebastian Edwards, and Ernesto Schargrodsky (Eds.), The Economics of Crime: Lessons for and from Latin America (pp. 101 - 132). Chapter 3. University of Chicago Press.

[14] Gaviria, A. (2000), "Decisiones: sexo y embarazo entre las jóvenes colombianas" Fedesarrollo, *Coyuntura Social* No. 23.

[15] Gage, A. 1995. "The Social Implications of Adolescent Fertility." Paper presented at the International Union for the Scientific Study of Population-United Nations Children's Fund Seminar on Demography and Poverty, Florence, Italy.

[16] Grogger, J. Y S. Korenman (1993), "The Socioeconomic Consequences of Child Bearing: Results from a NaturalExperiment". Mimeo, University of California, SantaBarbara.

[17] Hayes, C. 1987. Risking the Future: Adolescent Sexuality, Pregnancy and Childbearing. Washington: National Academic Press.

[18] Henderson, Vernon; Mieszkowski, Peter; and Sauvageau, Yvon. "Peer Group Effects and Educational Production Functions."J. *Public Econ.* 10 (August 1978): 97-106.

[19] Hotz, V. J., S. W. McElroy y S. G. Sanders (1999), "TeenageChildbearing and its Life Cycle Consequences: Exploitinga Natural Experiment". NBER, WP No. 7397.

[20] Iyer, Sriya and Weeks, Melvin. 2009. "Social Interactions, Ethnicity and Fertility in Kenya", working paper CWPE 0903

[21] Joshi, Shareen and Schultz, T. Paul. 2007. "Family planning as an investment in development and female human capital: Evaluating the long-run consequences in Matlab, Bangladesh", mimeo, Yale University.

[22] Hoxby Caroline M (2000). "Does Competition among Public Schools Benefit Students and Taxpayers?", *The American Economic Review*, Vol. 90, No. 5 (Dec., 2000), pp. 1209-1238

[23] Kohler, Hans-Peter, Behrman ,Jere R. and Watkins Susan C. (2001) "The Density of Social Networks and Fertility Decisions: Evidence from South Nyanza District, Kenya" *Demography*, Vol. 38, No. 1 (Feb., 2001), pp. 43-58

[24] Montgomery, M. R., G. Kiros, D. Agyeman, J. B. Casterline, P. Aglobitse, and P. C. Hewett (2001): "Social Networks and Contraceptive Dynamics in Southern Ghana". Population Council Research Division Working Paper No. 153, New York.

[25] Mark R. Montgomery and John B. Casterline (1993). "The Diffusion of Fertility Control in Taiwan: Evidence from Pooled Cross- Section Time-Series Models". *Population Studies*, 47, 457-479

[26] Manski, Charles, F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60, 531.542.Bloom et. all. (2008)

[27] Medina, Carlos and Morales Leonardo (2007). "Stratification and Public Utility Services in Colombia: Subsidies to Households or Distortions on Housing Prices", *Economia* vol.7 n.2, issue #14 spring 2007.

[28] Medina, Carlos,

[29] Meléndez, Marcela. 2004. *Subsidios al consumo de los servicios públicos en Colombia ¿ hacia donde movernos?*, Paper prepared for Misión de Servicios Públicos. Bogotá: La Fundación para la Educación Superior y el Desarrollo (FEDESARROLLO).

[30] Oates,Wallace E. , Evans, William N and Schwab Robert M. (1992). "Measuring Peer Group Effects: A Study of Teenage Behavior." *The Journal of Political Economy*, Vol. 100, No. 5 (Oct., 1992), pp. 966-991

[31] Rosero-Bixby, Luis and Casterline John B. (1994) "Interaction Diffusion and Fertility Transition in Costa Rica", *Social Forces*, Vol. 73, No. 2 (Dec., 1994), pp. 435-462

[32] Summers, Anita A., and Wolfe, Barbara L. "Do Schools Make a Difference?" *A.E.R.* 67 (September 1977): 639-52.

[33] Wooldridge, J. M. (2002), *Econometrics Analysis of Cross Section and Panel Data Models.* Massachusetts Institute of Technology.

Appendix

Comparison SISBEN Recent Mothers population with the Estimation Sample.

The Following table presents the summary statistics and a difference in means t-test, for the comparison of the population of SISBEN recent mothers and the estimation sample. Reader may remember that the main reason for a observation being excluded of the estimation sample is that it can not be linked to a previous SISBEN collection; in which case, information before the pregnancy is not available[28]. There is not a particular reason to think that the missing values generated in the estimation construction process follow a special endogenous pattern; nevertheless, some covariates show important differences between population and estimation sample. In order to make sure that the determination of the estimation sample is not a factor driving the results from this research, I estimate econometric models in which the process of selection into the estimation sample is modeled using Heckman selection procedures; the results of the endogenous peer effects coefficient do not show important variation after controlling for the selection[29].

Variable	SISBEN Recent Mothers			Estimation Sample			
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	T-test
Onset of Chilbearing Age	75768	22.18	5.43	11461	21.43	5.02	14.8
Some elementary	75768	0.07	0.25	11461	0.07	0.25	-1.0
Elementary	75768	0.11	0.31	11461	0.11	0.32	-1.6
Some high school	75768	0.31	0.46	11461	0.34	0.47	-5.7
High school	75768	0.41	0.49	11461	0.42	0.49	-0.6
College, some College	75768	0.09	0.29	11461	0.05	0.22	17.4
Assist to school	75768	0.12	0.33	11461	0.14	0.35	-5.9
Unemployed	75768	0.04	0.20	11461	0.06	0.24	-7.6
Cohabitating with partner	75768	0.34	0.48	11461	0.19	0.40	36.9
Widow	75768	0.00	0.06	11461	0.00	0.06	0.0
Divorced	75768	0.03	0.17	11461	0.03	0.17	-0.4
Single	75768	0.50	0.50	11461	0.71	0.46	-45.3
Physical or mental disability	75768	0.01	0.07	11461	0.01	0.08	-2.4
Monthly household Income	75768	664585	793465	11461	694033	727496	-4.0
living in apartment	75768	0.96	0.20	11461	0.97	0.17	-6.6
Good or standard quality in walls	75768	0.80	0.40	11461	0.79	0.41	3.2
Good or standard quality in floor	75768	0.49	0.50	11461	0.43	0.49	11.8
House is owned by the h/h	75768	0.36	0.48	11461	0.53	0.50	-32.2

[28] In order to compare the population with the estimation sample, this table presents the mother characteristics after the pregnancy. The reader may note taht not having information from a previous collection is the most important reason for missing an observation from the estimation sample.

[29] Following suggestions from referees, those results were ommitted from the final version of this paper.

First Stage Regressions.

The following table present the first stage regression of the 2SLS procedures presented in table (3) and (4). Instrumental variables changes for each specification, a description of the instruments can be found in section 6, and table (3) and (4) footnotes.

Appendix, First Stage Regression

	[1]		[2]		[3]		[4]		[5]		[6]	
Intrumental Variable 1	0.845***	2.855***	0.926***	1.926**	0.752***	1.803**	0.935***	1.329**	0.978***	1.317**	0.848***	1.467**
	(0.164)	(0.935)	(0.129)	(0.763)	(0.096)	(0.722)	(0.105)	(0.646)	(0.086)	(0.595)	(0.059)	(0.572)
Instrumental Variable 2	-0.021	8.034***	0.023	17.452***	0.079**	14.049***	0.129*	27.621***	0.146**	18.276***	0.164***	14.744***
	(0.087)	(2.137)	(0.070)	(3.627)	(0.037)	(3.009)	(0.071)	(4.601)	(0.059)	(3.570)	(0.036)	(2.185)
Some elementary	0.706	0.761*	0.025	0.125	0.352***	0.405***	0.105	0.324	-0.005	0.101	0.179	0.309**
	(0.449)	(0.446)	(0.216)	(0.235)	(0.126)	(0.140)	(0.321)	(0.333)	(0.197)	(0.215)	(0.124)	(0.139)
Elementary	0.406	0.620	-0.178	0.108	0.490***	0.621***	-0.335	0.074	-0.155	0.169	0.297**	0.496***
	(0.485)	(0.476)	(0.257)	(0.280)	(0.159)	(0.175)	(0.351)	(0.355)	(0.223)	(0.241)	(0.141)	(0.152)
Some high school	0.642	0.943**	0.388	0.678**	0.510***	0.745***	-0.328	0.205	0.326	0.588**	0.228	0.532***
	(0.481)	(0.463)	(0.251)	(0.273)	(0.148)	(0.168)	(0.349)	(0.352)	(0.236)	(0.258)	(0.141)	(0.152)
High school	0.665	0.904*	0.588*	0.835**	0.485***	0.767***	-0.249	0.232	0.365	0.521	0.179	0.614***
	(0.488)	(0.476)	(0.335)	(0.366)	(0.172)	(0.192)	(0.352)	(0.359)	(0.306)	(0.326)	(0.167)	(0.177)
College, some College	1.211**	1.216**	1.049***	1.038***	0.821***	0.903***	0.365	0.525	0.760**	0.589*	0.249	0.449*
	(0.503)	(0.506)	(0.354)	(0.387)	(0.241)	(0.272)	(0.358)	(0.382)	(0.322)	(0.345)	(0.213)	(0.238)
Assist to school	0.042	0.037	0.089***	0.089***	-0.017	-0.009	0.278***	0.254***	0.327***	0.325***	0.262***	0.264***
	(0.051)	(0.053)	(0.030)	(0.032)	(0.026)	(0.028)	(0.034)	(0.036)	(0.027)	(0.030)	(0.023)	(0.025)
Unemployed	0.037	0.011	-0.013	-0.029	-0.048	-0.062*	-0.105**	-0.124***	-0.137***	-0.147***	-0.136***	-0.139***
	(0.053)	(0.056)	(0.043)	(0.045)	(0.033)	(0.034)	(0.041)	(0.045)	(0.037)	(0.040)	(0.026)	(0.029)
Cohabitating with partner	0.113	0.177	0.178	0.289*	0.058	0.147	0.203	0.300**	0.244**	0.337***	0.119**	0.197***
	(0.205)	(0.201)	(0.160)	(0.163)	(0.131)	(0.137)	(0.128)	(0.132)	(0.108)	(0.116)	(0.057)	(0.072)
Widow	0.315	0.231	0.120	0.074	0.040	-0.012	0.335	0.296	0.156	0.115	0.030	0.016
	(0.501)	(0.526)	(0.415)	(0.535)	(0.326)	(0.432)	(0.358)	(0.433)	(0.299)	(0.378)	(0.240)	(0.324)
Divorced	-0.026	-0.044	-0.056	-0.046	-0.199	-0.187	0.045	0.045	0.022	0.015	-0.129	-0.132
	(0.237)	(0.236)	(0.189)	(0.189)	(0.158)	(0.160)	(0.156)	(0.156)	(0.131)	(0.134)	(0.087)	(0.099)
Single	0.127	0.156	0.173	0.255*	0.077	0.144	0.177	0.227*	0.208**	0.269**	0.085	0.141*
	(0.200)	(0.199)	(0.145)	(0.146)	(0.110)	(0.114)	(0.127)	(0.134)	(0.105)	(0.116)	(0.052)	(0.074)
Physical or mental disability	0.023	0.016	0.001	-0.025	0.050	0.036	-0.060	-0.094	-0.131	-0.174	-0.010	-0.042
	(0.219)	(0.235)	(0.179)	(0.192)	(0.132)	(0.142)	(0.175)	(0.176)	(0.163)	(0.176)	(0.125)	(0.137)
Monthly household Income	0.000	0.000	0.000	0.000	0.000	0.000	-0.000	-0.000**	-0.000	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
living in apartment	-0.005	-0.040	0.054	0.010	0.097	0.064	-0.026	-0.088	-0.016	-0.074	0.038	-0.013
	(0.138)	(0.138)	(0.137)	(0.136)	(0.139)	(0.138)	(0.063)	(0.073)	(0.066)	(0.071)	(0.054)	(0.058)
Good or standard quality in walls	0.019	0.043	0.011	0.032	0.012	0.029	-0.050*	-0.036	-0.044*	-0.034	-0.020	-0.017
	(0.039)	(0.039)	(0.033)	(0.035)	(0.030)	(0.032)	(0.030)	(0.031)	(0.024)	(0.024)	(0.020)	(0.020)
Good or standard quality in floor	0.077**	0.074**	0.079***	0.081***	0.062***	0.061***	0.061**	0.060**	0.048**	0.043*	0.026	0.020
	(0.038)	(0.037)	(0.027)	(0.028)	(0.022)	(0.023)	(0.029)	(0.030)	(0.022)	(0.025)	(0.018)	(0.021)
House is owned by the h/h	0.015	0.026	-0.014	-0.000	0.012	0.020	-0.005	0.010	-0.019	0.000	-0.008	0.004
	(0.031)	(0.032)	(0.026)	(0.028)	(0.022)	(0.025)	(0.022)	(0.025)	(0.017)	(0.020)	(0.014)	(0.017)
# of teen mothers in the h/h	-0.036**	-0.038**	-0.033**	-0.031**	-0.029**	-0.027**	-0.031***	-0.020*	-0.032***	-0.029***	-0.026***	-0.024***
	(0.016)	(0.017)	(0.013)	(0.013)	(0.012)	(0.012)	(0.010)	(0.011)	(0.010)	(0.010)	(0.008)	(0.009)
Stratum4	-0.029	-0.023	-0.010	0.009	0.005	0.022	0.039	0.037	0.044	0.039	0.046*	0.043
	(0.046)	(0.049)	(0.037)	(0.040)	(0.033)	(0.035)	(0.036)	(0.038)	(0.029)	(0.031)	(0.028)	(0.030)
Stratum4	-0.012	-0.015	-0.005	0.008	-0.001	0.025	0.020	0.019	0.053	0.058	0.060	0.079
	(0.126)	(0.127)	(0.068)	(0.078)	(0.067)	(0.070)	(0.091)	(0.090)	(0.064)	(0.066)	(0.058)	(0.062)
Constant	24.748***	0.869	23.530***	44.494***	24.339***	42.468***	11.828***	35.011***	14.137***	42.662***	15.321***	38.227***
	(5.495)	(2.600)	(4.442)	(3.009)	(4.667)	(3.059)	(4.169)	(3.163)	(3.790)	(3.382)	(2.750)	(2.530)
Neigborhood Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Contextual Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	11,591	11,592	11,591	11,592	11,591	11,592	11,591	11,592	11,591	11,592	11,591	11,592
R-squared	0.753	0.745	0.848	0.835	0.884	0.873	0.734	0.707	0.798	0.762	0.861	0.822