LEONARDO FABIO MORALES

Peer Effects on a Fertility Decision: An Application for Medellín, Colombia

The world's total fertility rate has fallen substantially in the last fifty years. The Latin American and Caribbean region is not an exception to this rule, as a deep process of demographic transition has swept through the entire region. In the first decade of this century, the region's total fertility rate (TFR) fell from 2.67 children per woman in 1999 to 2.12 by the end of 2010. The current rate is 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, which is widely associated with low human development and poverty.[1] In the last decade, the Latin American and Caribbean region also saw a reduction in the fertility rate for women between fifteen and nineteen years of age (FR15-19), from 83.95 children per 1,000 women in 1999 to 71.68 in 2010. While this reduction in the FR15-19 is substantial, it is smaller than the reduction in the TFR in the same period. Between 1999 and 2010, the TFR dropped 26 percent, while the FR15-19 fell 17 percent. This implies that adolescent fertility has become a more important component of total fertility in most of the countries in the region.[2] Relative to adult fertility, adolescent fertility is becoming greater and greater in Latin America.

Morales is with the Central Bank of Colombia.

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.

1. Joshi and Schultz (2007); Buvinic (1998); Gage (1999); Hayes (1987).
2. Flórez and Soto (2007b).

With some exceptions like Argentina and Peru, the contribution of adolescent fertility to total fertility has increased continuously in almost all Latin American and Caribbean countries. In 1999, adolescent fertility was 15.72 percent of total fertility in developing Latin American and Caribbean countries; by 2010, this ratio had increased to 16.29 percent. There are some remarkable cases, such as Brazil and Ecuador, where the ratio of adolescent fertility to total fertility increased by more than two percentage points between 1999 and 2010. The result of this phenomenon is, on average, an earlier individual onset of childbearing.

To explain this interesting phenomenon, it is important to study the factors that determine the age at which a mother decides to have her first child. From an individual's point of view, it may seem rational to have a child early in life given her education, her household socioeconomic conditions, and the characteristics of her social group. This paper explores the mother's chosen timing for the onset of childbearing in an urban context in Colombia, emphasizing the role of peer effects 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's behavior is partly explained by the influence of other individuals' behavior. In economics this has been called social interactions or peer effects. There are several channels through which these effects may take place: for example, individuals may learn from peers' behavior (social learning), or they may embrace the norms of the community with regard to socially accepted practices (social influence).[3] The main purpose of this paper is to test the existence and measure the magnitude of peer effects on a fertility decision. The fertility decision considered is the woman's age at the onset of childbearing. The study draws on a large sample of poor mothers in the city of Medellín, who had their first child 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 Latin America and the Caribbean, and they could certainly be a cause for the high incidence of teenage pregnancy in countries such as Colombia. There is evidence on the existence of geographic sorting patterns governing the spatial distribution of several fertility outcomes in Latin American cities.[4] For instance, in poor

3. Kohler, Behrman, and Watkins (2001).
4. Gaviria and others (2010).

neighborhoods women have more children and the onset of childbearing is earlier than in other neighborhoods. Nevertheless, the literature on fertility in the region has not explored the existence of peer effects. Several studies in this literature suggest that the importance of contextual and cultural factors has been underestimated. The influence of new social norms, like the general acceptance of early initiation of sexual relationships, and the influence of peer pressure have not been given enough attention in the study of teen pregnancy in the region.[5] In this paper, I use data from Medellín to evaluate the existence of peer effects that influence the age at which mothers decide to have their first child. Medellín is an example of the urban context in Latin America, where high levels of adolescent pregnancy are concentrated in poor neighborhoods.

To explore these issues, I estimate a linear-in-means model of social interactions in which the relevant peer effect groups are defined using weighting matrices, with weights defined using spatial and social distance criteria. To deal with the reflection problem, which is typical in the estimation of endogenous peer effects models, I design peer groups varying at the individual level. This strategy has recently been shown to successfully overcome the reflection problem.[6] In addition, the definition of non-perfectly-overlapping groups is useful to overcome a second problem that is also typical in this kind of estimation: namely, the endogeneity of the peer group. The definition of peer groups that are different for each individual implies the existence of excluded peers. These are peers of an 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-stage 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 to have her first child 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 members of the peer group. Therefore, a woman has more probability of becoming a teenage mother if her peer group has a significant composition of teenage mothers.

From a social policy perspective, this topic is crucial, given that pregnancy at an early age is widely associated with negative socioeconomic outcomes for the mother and the child. Studies show that educational achievements, health markers, and measures of involvement in risky behaviors all tend to be worse for teenage mothers and their children.[7]

5. Flórez and Soto (2007b).
6. Bramoullé, Djebbari, and Fortin (2009); de Giorgi, Pellizzari, and Redaelli (2010).

I start by presenting a brief summary of the literature in which this paper fits. The next section then describes the theoretical foundations on which the paper is based. After describing the data, I outline the empirical strategy implemented to get reliable estimations and present the results of the estimations. A subsequent section describes some simulations and robustness checks, together with their results. Finally, I present the main conclusions of the paper and discuss some relevant policy implications.

Related Literature

The fertility outcome of interest for this paper is the woman's age at the onset of childbearing. Most of the related literature focuses on teenage childbearing, which is an extreme case of early onset of childbearing. Several papers address the negative consequences of teenage pregnancy and childbearing.[8] Women who were mothers before the age of twenty are usually poorer and less educated than women who became mothers after the age of twenty. In the case of Colombia, for example, Flórez and Soto find that having a child as a teenager implies that the mother will get 3.9 years of education less than if she would have had the child as an adult.[9]

Several papers describe and analyze adolescent childbearing in Colombia and the region.[10] One important finding in these papers is that the contribution of the teenage fertility rate to the TFR has increased for several Latin American and Caribbean countries.[11] More recent data on fertility rates confirm that trend for the regional average.[12] For the region as a whole, the average share of teenage fertility in total fertility increased by almost one percentage point between 1999 and 2010.[13] This phenomenon is not generalized across all Latin American and Caribbean countries, however, as there are some important exceptions. For example, the ratio of adolescent fertility to total fertility decreased in Peru

7. See, for example, Joshi and Schultz (2007); Buvinic (1998); Gage (1999); Hayes (1987); Case and Katz (1991); Grogger and Bronars (1993); Hotz, McElroy, and Sanders (1999).

8. Joshi and Schultz (2007); Buvinic (1998); Gage (1999); Hayes (1987).

9. Flórez and Soto (2007b).

10. For example, Flórez and Núñez (2001); Flórez and Soto (2007a, 2007b).

11. Flórez and Soto (2007b).

12. Public data from the World Bank (available online at data.worldbank.org).

13. The average ratio of teenage fertility to total fertility was 15.51 percent for all Latin American and Caribbean countries in 1999 versus 16.29 percent in 2010.

from 11.19 percent in 1999 to 10.22 percent in 2010 and in Argentina from 13.02 percent to 12.5 percent. Nevertheless, for most of the countries in the region, this ratio increased or remained relatively constant. In Brazil it rose from 18.38 percent to 20.71 percent in the period; in Ecuador, from 14.13 percent to 16.42 percent. In Colombia the ratio was 16.81 percent in 2010, close to what it was in 1999 (17.41 percent).

For the specific case of Colombia, several papers use econometric models to explain the probability of teen pregnancy.[14] These papers find evidence that low education, disadvantaged socioeconomic conditions, and poor family backgrounds increase the probability of teenage pregnancy. None of these papers uses longitudinal information, so they cannot use covariates before or at the time of the pregnancy to explain the fertility decision. This may result in endogeneity bias, as long as important explanatory variables in these models are determined simultaneously with the pregnancy. In this paper, I have information on the characteristics of a future mother before the childbirth, so I can better control for the simultaneity of several covariates and the pregnancy. In addition, none of these papers emphasizes the role that social interactions may play in the determination of fertility outcomes. At the regional level, very few papers seriously consider social interactions or similar effects on fertility.[15]

Several studies in demography and sociology explore the role of social interactions in fertility outcomes. One example is the study of diffusion effects. A diffusion effect takes place when a behavior is adopted and reproduced through social networks. The mechanisms through which social networks can affect their members include social influence and social learning. To measure diffusion effects, some papers in this literature use, as explanatory variables in the estimations, variables that describe whether the woman discusses contraceptive practices with relatives or members of the community.[16] Other papers use aggregate fertility levels or the proportion of family planning users in a woman's network.[17] The empirical techniques often rely on the use of longitudinal data at individual levels or aggregate data by geographical areas.

14. Gaviria (2000); Barrera and Higuera (2004); Flórez and Núñez (2001).
15. Rosero-Bixby and Casterline (1994).
16. Montgomery and others (2001).
17. On aggregate fertility levels, see Montgomery and Casterline (1993) or Rosero-Bixby and Casterline (1994); on the share of family planning users, see Kohler, Behrman, and Watkins (2001).

In economics, an important branch of the applied microeconomic literature focuses on the detection and identification of group effects on many outcomes, including fertility. For example, the identification of peer effects in education has been widely debated. Part of the discussion originated with the publication of the Coleman Report in 1966, in which one of the most polemic findings was that students perform better if their fellow students are high achievers.[18] Several papers seek the identification of social interactions when the main dependent variable is a fertility decision.[19] Some of the authors realize that the peer group itself is an endogenous decision, and because of this the estimation results will be biased.[20] As discussed later in this paper, the self-selection issue and the reflection problem are the main identification threats for empirical models of social interactions.[21]

Theoretical Foundations

The model starts with a representative woman *i,* who derives utility from consumption z_i and from children n_i (if she has any). I focus on the decision to start a family or not; this fact is represented by the term c_{1ia}. Therefore, any woman *i,* with no children until the previous period, can derive utility from her first child 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 woman's age, and the subscript 1 stands for the first child. The total number of children is the summation of fertility choices (c_{1ia}) up to age a: [22]

$$n_i = \sum_{t=0}^{a} c_{ti}.$$

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

18. Coleman (1966); Oates, Evans, and Schwab (1992).
19. For example, Oates, Evans, and Schwab (1992); Iyer and Weeks (2009); Case and Katz (1991).
20. For example, Oates, Evans, and Schwab (1992).
21. The reflection problem arises because the individual is influenced by the reference group, but at the same time, an individual's decision also determines the group behavior. See Manski (1993).
22. A general fertility choice is represented by c_{1ia}, where $c_{1ia} = 1$ if the woman decides to have a child (not necessarily the first one) at age a.

production function that takes as inputs education, parental time, and any other resources that can improve the child's development. For simplicity, it is assumed that this representative woman solves a static optimization problem every year, instead of maximizing her lifetime utility. This assumption allows me to develop a model for the demand for children similar to the one proposed by Becker.[23] For the year in which woman *i* decided to have her first child or any prior year, the utility function can be represented by the following equation:

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

At each age *a,* mother *i* decides whether she will have her first child this year and her consumption, z_{ia}, subject to the standard budget constraint:

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

where I_{ia} denotes income, p_c is the unitary cost of quality, and $p_c \cdot q \cdot c_{1ia}$ is the total amount spent on the child each year, which is zero if $c_{1ia} = 0$ (no child in this period). In addition, π_z represents the price of the consumption good. Every year, woman *i* maximizes equation 1 subject to equation 2.[24] Woman *i* will have her first child at age *a* if:

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

where the utility function is evaluated in the optimal quantities ($z^*_{i,a}$, $q^*_{i,a}$).

Starting from this Becker-type model, I introduce the possibility that the fertility variable can be explained by social interactions. If social interactions have some impact in fertility decisions, then the fertility decisions of the peer group members should enter into the utility function of woman *i*. I assume that there is perfect knowledge about the woman'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's peer group and ω_j is a normalized interaction index. For many sociological reasons

23. Becker (1981).

24. After her first child, the decision of woman *i* will be whether or not to have her second child, $c_{2ia} \in \{1, 0\}$. Her budget constraint will be $p_c \cdot q + p_c \cdot q \cdot c_{2ia} + \pi_z z_{ia} = I$. The quality is assumed to be the same for each child.

(social influence, social learning, and so on), one can assume that woman i gets utility from exhibiting a behavior similar to her peers. Therefore, the mother's optimization problem can be rewritten as follws:

$$\text{(1a)} \quad \max U_{ia}\left[c_{1ia}\left(m_{ia}\right), z_{ia}, q_i \left| m_{ia} - \sum_{j \neq 1} \omega_j \cdot m_j \right|\right], \text{ subject to equation 2,}$$

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

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 woman's peer group. The term c_{1ia} 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 to become a mother; it is assumed that m_i is under the mother's control in each period. The term

$$\sum_{j \neq i} \omega_j \cdot m_j$$

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 the relationship between i and j, the greater will be the weight that peer j has in the computation of the average. The main hypothesis of this paper is that in every period, woman i will get additional utility from mimicking the behavior of her peers. Therefore, the contribution of the term

$$\left| m_{ia} - \sum_{j \neq i} \omega_j \cdot m_j \right|$$

is assumed to be negative.

Determination of the Age at the First Childbirth

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

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

where a only takes continuous values between twelve and forty-five given the biological fertility period in the woman's life. If one is willing to assume that individuals are followers of their peers' behavior, it would be interesting to ask what is the response of A_i given an exogenous increase in m_j.[25] I assume that the individual already chooses 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 $c^*_{1,ia}$ from zero to one if the increase is enough to overcome the threshold $\bar{m}_i$. Through this mechanism, the mother will choose to have her first child in the current period, at age a and not later, which can be interpreted as an effect of m_j on A. In this paper I assume the existence of a continuous function, $G(\cdot)$, that maps each possible combination of the inputs in the utility function to a single value, A_i. Therefore, A_i can be written as

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

In later sections, I propose and estimate a linear approach to equation 5. The main interest of this paper is the peer effects on A_i, or the influence that

25. Assuming this is equivalent to saying that $(\partial m_i / \partial m_j) > 0$. In a continuous and simplified version of the model, it is possible to find an expression for $(\partial m_i / \partial m_j)$. Under minimal assumptions, it is not possible to unambiguously determine the sign of this derivative, as individuals are peer followers under some circumstances but not others. A valid interpretation of the main question of this paper is testing whether or not the individuals are followers of the peers' fertility behavior.

the behavior of the peer group has on a woman's age at first childbirth. The prediction that can be derived from the main hypothesis of this paper is that $(\partial m_i/\partial m_j) > 0$. In terms of the dependent variable, this means that mother *i* may find it optimal to reduce her age at her first childbirth given a reduction in the expectation of this variable among her peer group.

Data

The System for Selecting Beneficiaries of Social Spending (SISBEN in Spanish) is a household targeting system that has been used in Colombia since the late 1990s to target social programs within the poor and vulnerable population. This system is based on an assessment of the living conditions of individual families, and it is currently used to select beneficiaries of subsidized health insurance, educational subsidies, and conditional cash transfers in Colombia.[26] The information in the SISBEN databases could be considered as a census for poor populations; for example, by 2002, 60 percent of the total population in the country was registered in the system, and about 30 percent received benefits.[27] The instrument used to target spending is a statistically derived proxy means index. The computation of the index relies on information about the 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 two main tools for selecting the responders: a geographical selection based on previous targeting systems; and demand in hospitals, municipal offices, and other benefit providers.

The SISBEN was not initially conceived as a panel, but it is mandatory to update the information every three years. Consequently, a panel of the surveyed population can be constructed by crossing the information from the original collection and the updates. For this paper, the original collection and two updates were used to construct a panel of three periods of information. Given that the survey was not initially conceived as a panel, the surveys do not include an official identification number for following the same individuals through different periods, but there are mechanisms for matching individuals and households over time. Although this means that a large share of the

26. Castañeda (2005).
27. Castañeda (2005).

observations will be missed, the huge number of observations in each period makes it possible to preserve a satisfactory sample.

The construction of the estimation sample thus involves several stages in which data from different SISBEN collections and external data are merged. The target population for the study is recent mothers included in the SISBEN system. A recent mother is here defined as a woman who had her first child any time between the first and third SISBEN collections (roughly speaking, between 2001 and 2010).[28] The conclusions derived in this paper are only intended to be applicable to this population. A SISBEN recent mother can only be included in the estimation sample if the observation can be linked to a previous period. This is because I need to observe the covariates explaining the decision about the timing of the onset of childbearing before (or at) the time of pregnancy. Table A1 in appendix A provides summary statistics comparing the sample of SISBEN recent mothers and the estimation sample.

The total sample of SISBEN recent mothers that can be identified in any of the three SISBEN collections consists of 75,768 individuals. Only a fraction of those individuals can be linked to a previous SISBEN collection, which is crucial for identifying characteristics before pregnancy. Consequently, the estimation sample consists of 11,461 individuals. Some covariates present differences between population and estimation sample, as evident in table A1, but robustness checks show that this issue is not driving the results.[29]

Empirical Strategy

Empirically, the goal of this paper is the estimation of a single-equation model where the dependent variable is the mother's age in years at her first childbirth.[30] The main interest is the identification of endogenous social effects that could explain the dependent variable. The estimation of these effects is usually

28. The average collection date of the first SISBEN is 9 September 2002, with a standard deviation of 431 days. The average collection date for the second SISBEN is 22 May 2006, with a standard deviation of 316 days. The average collection date of the third SISBEN is 7 January 2010, with a standard deviation of 57 days.

29. To verify that the determination of the estimation sample is not driving the results, 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 change significantly after controlling for selection. More details on the construction of the estimation sample are available on request.

30. This variable is replaced by the woman's age if she is pregnant at the time of the survey.

biased due to two fundamental problems: the reflection problem and the group selection problem. To deal with the reflection problem, I implemented a strategy similar to the one developed by De Giorgi, Pellizzari, and Redaelli, which allows me to define peer groups that vary at the individual level.[31] Once the reflection problem is controlled, there is still an endogeneity problem because the peer group effect is an individual's endogenous decision. To address this problem, I perform a standard instrumental variables (IV) methodology.

The instruments proposed in this paper are based on the idea of using the expectation of outcomes and covariates computed only for the excluded peers.[32] These are peers of an 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 excluded peers explain the fertility outcome of the individual's peers via social interactions. In addition, these variables do not have a direct effect on the individual's fertility outcome because it is assumed that any effect that excluded peers' behavior may have on the individual's behavior works indirectly through the effect of peers' behavior on the individual's behavior.[33] More details are provided later in the paper. The following subsections describe 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 them.

The Reflection Problem

Manski was the first to explain the concept of reflection in the literature on social interactions.[34] Basically, inside a social group, individuals are both influencing their peers' behavior and being influenced by them. The term reflection refers to the fact that one cannot know if one's action is the cause or the effect of peers' influence.[35] Consider the following equation:

$$y_i = \alpha + \beta E\left[y \middle| N_i \right] + \mathbf{E}\left[\mathbf{z} \middle| N_i \right]' \boldsymbol{\gamma} + \mathbf{z}_i'\eta + u_i, \tag{6}$$

where y_i is the woman's age at first childbirth, $\mathbf{z}_i$ represents a vector of individual and family characteristics of individual *i*, $E[y \mid N_i]$ stands for the mean

31. De Giorgi, Pellizzari, and Redaelli (2010).

32. Instruments of this nature have been proposed recently in the field (de Giorgi, Pellizzari, and Redaelli 2010; Bramoullé, Djebbari, and Fortin 2009).

33. In the case of exogenous covariates of excluded peers, this statement can be formally proved in a system-of-equations framework.

34. Manski (1993).

35. De Giorgi, Pellizzari, and Redaelli (2010).

of y in the neighbor group of individual *i,* and $\mathbf{E}[\mathbf{z} \mid N_i]$ is a vector that includes the mean of the exogenous variables z among those people in the individual *i*'s neighbor group.

This equation is the standard linear expression for the estimation of social interactions; it formally represents two types of social effect. 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. In this paper, β is the main coefficient that I am trying to identify. Nevertheless, mother *i* may behave similarly to her neighbors just because they have similar socioeconomic characteristics and share similar restrictions; that effect is usually known as a contextual effect, and it will be captured by the vector of coefficients $\boldsymbol{\gamma}$. Equation 6 is useful for illustrating the nature of the reflection problem. Taking the expectation conditional on the neighbor group N_i, solving for $E[y \mid N_i]$, and assuming $E[u_i \mid N_i] = 0$ yields

$$(7) \qquad E\left[y \middle| N_i\right] = \alpha + \beta E\left[y \middle| N_i\right] + \mathbf{E}\left[\mathbf{z} \middle| N_i\right]' \boldsymbol{\gamma} + \mathbf{E}\left[\mathbf{z} \middle| N_i\right]' \eta;$$

$$(8) \qquad E\left[y \middle| N_i\right] = \left(\frac{\alpha}{1-\beta}\right) + \mathbf{E}\left[\mathbf{z} \middle| N_i\right]' \left(\frac{\boldsymbol{\gamma}+\eta}{1-\beta}\right).$$

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 as individual B, then it must be the case that individual A and C belong to the same group. This characteristic causes the term $E[y \mid N_i]$ to appear on both sides of equation 2. De Giorgi, Pellizzari, and Redaelli show that identification can be achieved if instead of being fixed, peer groups are specific to every individual.[36] If the neighbor groups are individual specific, then equation 3 can be rewritten as follows:

$$(9) \qquad E\left[y \middle| N_i\right] = \alpha + \beta E E y \middle| N_j \middle| N_i$$
$$+ E\left[\mathbf{E}\left[\mathbf{z} \middle| N_i\right]' \middle| N_i\right] \boldsymbol{\gamma} + \mathbf{E}\left[\mathbf{z} \middle| N_i\right]' \eta,$$

36. De Giorgi, Pellizzari, and Redaelli (2010).

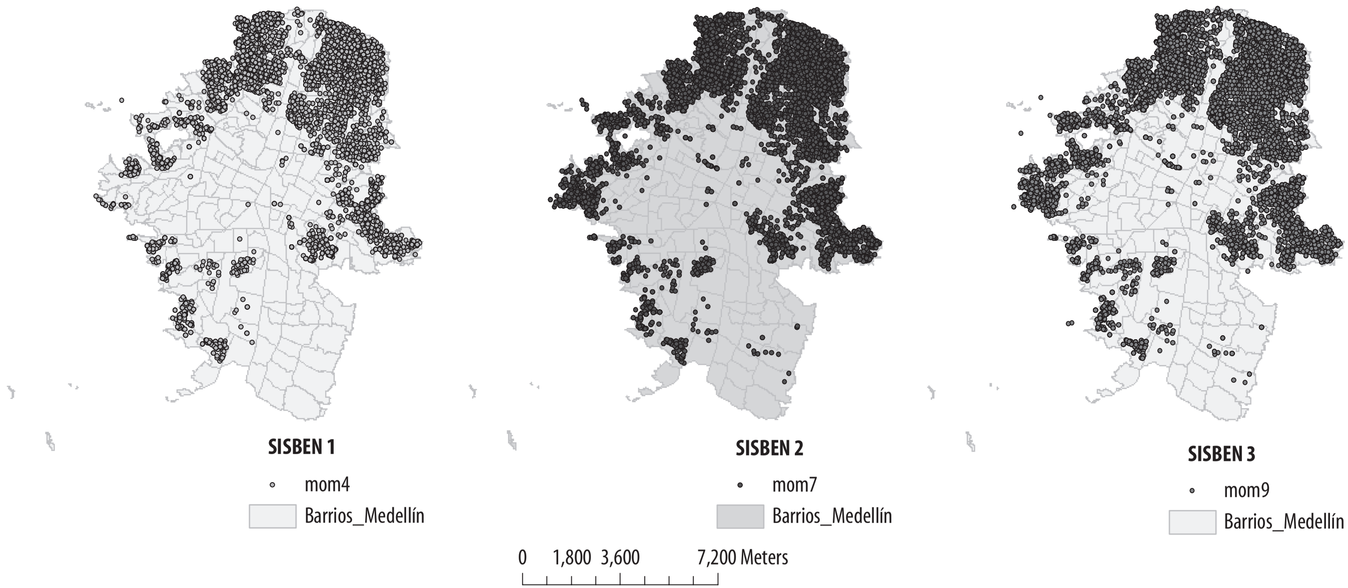
where j represents a generic peer of individual i's neighbor group. Using a simple example in which the peer groups are individual specific, de Giorgi, Pellizzari, and Redaelli show that identification relies only on observations with distinct peer groups.

STRATEGY. The relevant peer group for this research is a group of neighbors, who are defined according to some criteria of proximity and similarity. The relationship with neighbors is, by construction, determined geographically. Therefore, geographic information system (GIS) data are required to identify the neighbors of each individual. The information provided by the administrative department of the city includes the official codification for the census tract and block where the individual's housing unit is located. Using this information, I merged the SISBEN data with an ArcGIS file containing all the blocks in the city. This process generates a closely approximated location for every household in the panel in terms of its geographical coordinates.[37] To illustrate this procedure, figure 1 presents a panel of maps showing the spatial location for all the SISBEN mothers in the panel (more specifically, the location of their block's centroid).

Each point on the maps (mom4, mom7, mom9) represents the centroid of a block in the city, where one or more SISBEN recent mothers live. These women are not necessarily beneficiaries of any social program, but they are registered in the system. The polygons on the maps 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 expensive neighborhoods, especially in the southeast, have a low density of SISBEN households or none at all. The 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 thus show the economic spatial segregation and the sorting patterns in the city. Disadvantaged (presumably) households are restricted to some areas of the city.

37. Merging the SISBEN panel and the GIS files for blocks identified the centroids' coordinates of the block where the household is located, for most of the observations. Not every observation in the panel was successfully matched with the GIS file of blocks, especially in the first SISBEN, where the codification for census tracts and blocks was not available for all the observations. In those cases, the centroids of the most disaggregated political division (*barrios*) are used.

FIGURE 1. SISBEN Recent Mothers[a]



a. Each point on the maps represents the centroid of a block in the city, where one or more recent SISBEN mothers live. The polygons represent neighborhoods.

The strategy used in this paper to overcome the reflection problem is based on defining neighbor groups that vary at the individual level. The basic idea in the definition of the reference group is to build an N-by-N matrix of weights ($\mathbf{W}$), where N is the number of all SISBEN recent mothers in the estimation sample. The matrix operation $\mathbf{W} \cdot \mathbf{Y}$, where $\mathbf{Y}$ is a vector containing the age at first childbirth for all SISBEN resent mothers in the estimation sample, is a nonparametric estimator of $E[\mathbf{Y}]$. Therefore, a neighbor mother j can influence the fertility decision of mother i, depending on the assigned weight she has in the computation of

$$\sum_{j \neq i} w_{ij} y_j ,$$

where w_{ij} is an element of $\mathbf{W}$ and y_j is $\mathbf{Y}$'s element corresponding to the jth 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 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 mother i. Furthermore, the distance between mother i and her neighbors is certainly not the unique criterion for excluding some mothers from mother i's peer group. There can be other social distances s^{-k} (with $k = 1, 2, \ldots K$), such that neighbors who are very different from mother i in any of the k characteristics should also be excluded from mother i's peer group. These characteristics can be sociodemographic variables such as education, age, or income. In the empirical work, I defined several matrices using different criteria and estimated models using different matrices. Formally, the construction of these matrices can be represented as follows:

$$\text{(10)} \qquad \mathbf{W} = \begin{bmatrix} 0 & S_{12} \cdot \dfrac{1}{d_{12}} & \cdots & S_{1N} \cdot \dfrac{1}{d_{1N}} \\ S_{21} \cdot \dfrac{1}{d_{21}} & 0 & \cdots & S_{2N} \cdot \dfrac{1}{d_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} \cdot \dfrac{1}{d_{N1}} & S_{N2} \cdot \dfrac{1}{d_{N2}} & \cdots & 0 \end{bmatrix},$$

where

$$d_{ij} = \sqrt[2]{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{11}$$

and

$$s_{ij} = 1\left\{d_{ij} < \bar{d}\right\} \cdot 1\left\{\left|s_i^1 - s_j^1\right| < d^{-1}\right\} \cdot 1\left\{\left|s_i^2 - s_j^2\right| < d^{-2}\right\} \cdot \cdots \cdot 1\left\{\left|s_i^K - s_j^K\right| < d^{-K}\right\}. \tag{12}$$

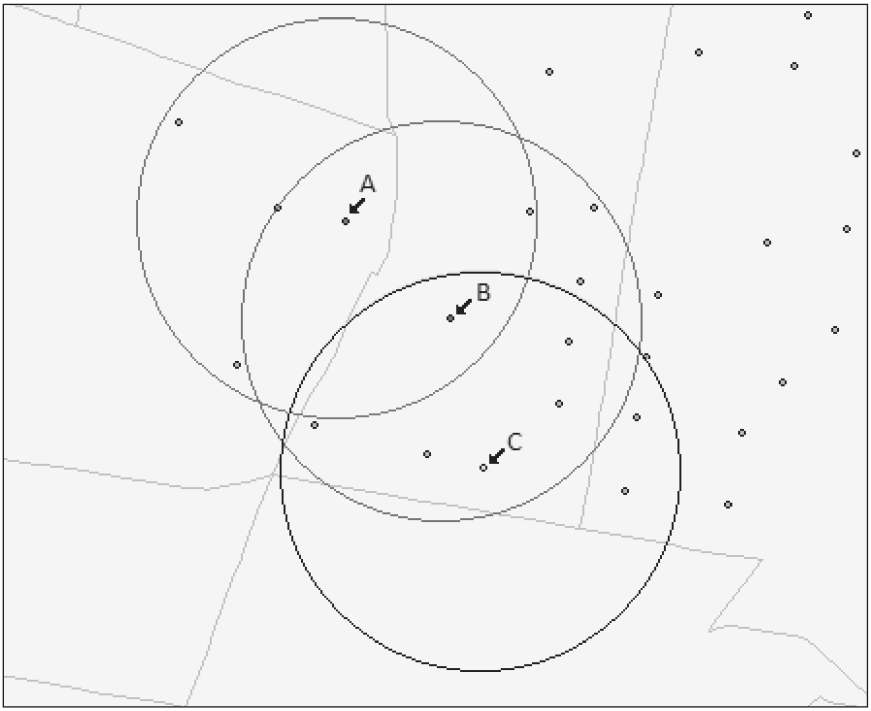
In the empirical work, a standardized version of equation 10 is used.[38] In equation 11, d_{ij} describes the Euclidean distance between mother *i* and mother *j,* where *x, y* stand for the geographical coordinates. In equation 12, S_{ij} describes a multiplication of indicators of functions; these indicators of functions are equal to one when the condition inside of the curly brackets holds, and zero otherwise. The first condition is the distance condition: any neighbor beyond some radius $\bar{d}$ is excluded (that is, a zero weight is assigned for that peer). The other conditions are based on sociodemographic variables: if a neighbor differs from mother *i* in any of these characteristics, $\left|s_i - s_j\right| > \bar{d}$, then the neighbor is excluded. Age and education are the sociodemographic variables used in the construction of this matrix.

In figure 2, each point represents the centroid of a block where a SISBEN mother resides; the polygons represent neighborhoods. Thus, mother B belongs to both mother A's peer group and mother C's peer group, but mother A does not belong to mother C's group and mother C does not belong to mother A's group. When sociodemographic restrictions are operating, the variation is greater. In figure 3, mother A and mother B live in contiguous blocks, but they differ in some of the criteria used to form the social groups. Therefore, despite the fact that they live very close each other, they do not belong to each other's social group, and their peer groups are very different.

FERTILITY DECISION EQUATION. The fertility decision equation describes how family and personal characteristics and the decisions of peers affect the age at which a mother decides to have her first child. The relevant information that explains this decision is the information before or during pregnancy. The SISBEN information covers three different periods. Based on those periods,

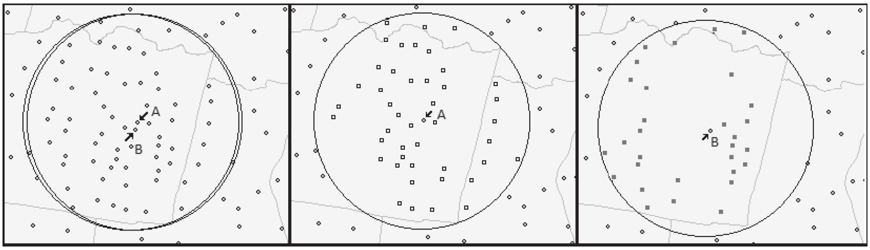
38. The standardized version of **W** is a matrix such that the sum of every row or column is equal to one.

FIGURE 2. Peer Groups[a]



a. Each point represents the centroid of a block, where a SISBEN mother resides; the polygons represent neighborhoods.

FIGURE 3. Peer Group Variation[a]



a. The squares in the far right circle represent peers of individual B; the empty squares represent peers of individual A.

FIGURE 4. Cases[a]

a. The wide line represents the time line for a generic woman in the sample; the thin line represents different possibilities for the pregnancy time.

I defined five different cases that describe how the information is used (see figure 4). In case 1, the woman got pregnant after she was surveyed in the second SISBEN, but before she was surveyed in the third round. In this case, the covariates and the expectation of the dependent variable among her peers are constructed with information from the second round. In case 2, the woman got pregnant after she was surveyed in the first SISBEN, but before she was surveyed in the second. In this case, the covariates and the expectation of the dependent variable among her peers are constructed with information from the first survey. The other three cases (3, 4, and 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 equation 6, which is estimated for recent mothers. To measure the endogenous peer effects, a nonparametric estimation of $E(y_i|G_j)$ is included in the fertility equation (where *j* represents a generic neighbor of mother *i's* neighbor group). The computation of this expectation is

$$E\left(y_i \mid N_j\right) = \sum_{j \neq i} w_{ij} y_j,$$

where y_j is the fertility decision of neighbor *j* and w_{ij} is the weight explained earlier.[39] To control for contextual effects, a nonparametric estimation of $E(\mathbf{z}_i|N_j)$

39. In some of the cases described earlier, 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 the expectation term appear on both sides of equation 2.

is also included in the fertility equation; similar to the previous case, this expectation is computed as

$$E\left(\mathbf{z}_i \middle| G_j\right) = \left[\sum_{j\neq i} w_{ij}\mathbf{z}_{1j} \ldots \sum_{j\neq i} w_{ij}\mathbf{z}_{kj}\right],$$

where $\mathbf{z}_{kj}$ is the neighbor j's exogenous covariate, $\mathbf{z}_k$. The following equation is the one estimated in this paper:

$$(13)\qquad y_i = \alpha + \beta \cdot \sum_{i\neq j} \omega_{ij} y_i + \left[\sum_{j\neq i} w_{ij}\mathbf{z}_{1j} \ldots \sum_{j\neq i} w_{ij}\mathbf{z}_{kj}\right] \cdot \gamma + \sum_{b=1}^{B} \upsilon_b \delta_{bi} + \mathbf{z}_i' \eta + u_i,$$

where y_i is the age of the mother i at first childbirth and $\mathbf{z}_i$ is a vector of individual and household characteristics. The parameter that represents the endogenous peer effects is β. As mentioned above, 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 nonendogenous effects. In the presence of these nonendogenous effects (that is, contextual or correlated effects), the β coefficient could be overestimated. Contextual effects have their root in the fact that mothers in the same peer group have similar socioeconomic composition, and fertility behavior may vary with different socioeconomic characteristics of the group.[40] Correlated effects may be important because individuals with the same characteristics or who face similar institutional environments tend to behave similarly.[41] In many situations, the correlated and contextual 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. To control for these nonendogenous peer effects, I include in the regression the average of the exogenous covariates among the people in the reference group of mother *i*. In addition, I include a set of

40. Exogenous or contextual effects are associated with the fact that the individual's behavior varies with some exogenous characteristic of the group. For example, in a classroom, a student's achievement may be explained by the socioeconomic conditions of the school district; in a comparison of a school district where all the students' parents have a college degree and another district where no parent has a college degree, student achievement would be expected to go in the same direction as average parents' education in the district; but this variation does not obey any endogenous interaction between students. In equation 13, this is precisely what the γ coefficients are capturing: namely, the variation of individual behavior given changes in average exogenous characteristics of the group.

41. Manski (1993).

dummy variables υ_b, where $b = 1 \ldots B$ is an index of the neighborhood, and δ_{bi} is the coefficient measuring the fixed effects of neighborhood b.

The Endogenous Reference Group Problem and IVs

The estimation of social interaction effects is further complicated by the determination of the peer group. The problem arises because the peer group (made up of neighbors or local residents) is often a matter of individual choice.[42] In other words, individuals self-select themselves into the peer group that best fits their possibilities and preferences (unobserved factors). When individuals make their residential location decisions, they are choosing not only the neighborhood in which they want to live, but also their peer group. Therefore, the expectation of the dependent variable, conditional on the social group, is an endogenous variable. Technically, there can be unobservable variables that are correlated with both the location decision and 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 13.

The estimation of equation 13 by ordinary least squares (OLS) will be biased, even after applying the proposed procedure to correct for reflection. The direction of the bias depends on the unobservable factors driving the selection of the peer group and their correlation with the unobservable factors determining the fertility decision. To correct for this bias, I performed an instrumental variables (IV) procedure that uses the peer group structure proposed in this paper to obtain valid instruments by drawing on information on excluded peers—that is, the peers of an individual's peers who are not included in the individual's peer group. Data on excluded peers have been used very recently in the literature for the identification of endogenous social effects.[43]

Expectation of covariates and the fertility decision among the excluded peers should be strongly correlated with the individual's fertility decision. This happens through a series of interconnected social interactions. The exogenous covariates of excluded peers explain their fertility decisions, and then via social interactions, an individual's peers' fertility behaviors are explained by the fertility behaviors of their peers (from the perspective of the individual they are excluded peers). By the same reasoning, the fertility behavior of the individual is explained by the fertility behavior of his or her peers. In addition,

42. Oates, Evans, and Schwab (1992).
43. Bramoullé, Djebbari, and Fortin (2009); De Giorgi, Pellizzari, and Redaelli (2010).

TABLE 1. Networking Criteria

Networking level	*Distance*	*Age*	*Education*
Low	$\lvert d_i - d_j \rvert \leq 500$	$\lvert a_i - a_j \rvert \leq 5$	$\lvert e_i - e_j \rvert \leq 3$
Medium		$\lvert a_i - a_j \rvert \leq 10$	$\lvert e_i - e_j \rvert \leq 5$
High	$\lvert d_i - d_j \rvert \leq 1{,}000$	$\lvert a_i - a_j \rvert \leq 15$	$\lvert e_i - e_j \rvert \leq 7$

it can be formally proved that expectation of covariates and the endogenous outcome among excluded peers are exogenous to the individual's fertility decision.[44] The intuition for this is that these variables have no direct effect on the individual fertility decision; it all happens through the relationship between the individual's peers and the excluded peers. Several instruments are used in the paper, all of which use the principle of the excluded peers. I use the expectation among the excluded peers of the fertility outcome, and the expectation of other covariates among the excluded peers.[45]

Results

The empirical strategy described in the previous section implies that several specifications are possible, 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 for constructing the peer group: physical distance and social distance (age and education). Table 1 defines the conditions inside the brackets of each indicator of function in equation 12 that were used for the regressions presented in this paper.[46]

For each criterion (distance, age, and education), two or three levels of networking are used. Different sets of peers are allowed to affect the fertility decision in each level (the peers receive a nonzero weight). Given the form

44. In a system-of-equations framework, with one equation explaining the endogenous outcome for each individual, exogenous covariates for excluded peers are exogenous to the individual fertility decision; see de Giorgi, Pellizzari, and Redaelli (2010) for more details. In this case, additional conditions are required, including no correlation between the unobservable variables for excluded peers and the individual's unobservable variables.

45. The expectation of exogenous covariates among the excluded peers results in a large number of instruments; to gain efficiency and make the specifications more parsimonuous, I selected the best in terms of the correlation with the endogenous variable and used overidentifying restriction tests.

46. Other criteria were used, as well; the results are comparable to the results presented here.

TABLE 2. Summary Statistics

Variable description	*Obs.*	*Mean*	*Std. dev.*	*Min.*	*Max.*
Age at first childbirth	11,537	21.30	5.04	4	45
Education attainment: incomplete elementary	11,537	0.11	0.31	0	1
Education attainment: complete elementary	11,537	0.15	0.36	0	1
Education attainment: incomplete high school	11,537	0.45	0.50	0	1
Education attainment: complete high school	11,537	0.24	0.43	0	1
Education attainment: some college or completed college	11,537	0.03	0.17	0	1
Currently in school	11,537	0.34	0.47	0	1
Currently unemployed	11,537	0.08	0.27	0	1
Currently cohabitating with sentimental partner	11,537	0.11	0.31	0	1
Currently widow	11,537	0.00	0.04	0	1
Currently divorced	11,537	0.01	0.11	0	1
Currently single	11,537	0.84	0.36	0	1
Women any sort physical or mental disability	11,537	0.01	0.07	0	1
Monthly income of the household (in pesos)	11,537	459,964	408,176	0	6,588,608
Dummy: living in house or apartment=1, other=0	11,537	0.97	0.17	0	1
Dummy: good or standard quality in walls material=1, other=0	11,537	0.78	0.41	0	1
Dummy: good or standard quality in floor material=1, other=0	11,537	0.36	0.48	0	1
The house is owned by the household	11,537	0.48	0.50	0	1
Number of teenager mothers in the household	11,537	0.58	0.77	0	6
Dummy variable: stratum equal to 1 = 1	11,537	0.33	0.47	0	1
Dummy variable: stratum equal to 2 = 1	11,537	0.57	0.50	0	1
Dummy variable: stratum equal to 3 = 1	11,537	0.10	0.29	0	1
Dummy variable: stratum equal to 4 = 1	11,537	0.00	0.01	0	1

a. Stratum is a socioeconomic classification of the neighborhoods in Colombia, used by the government to target subsidy programs, where the first stratum is the poorest.

of the restrictions, these sets are nested. The restriction in the criterion (distance, age, or education) is strongest in the lowest level, resulting in a smaller number of included neighbors than in any other level and, therefore, a larger number of excluded peers. In the medium or high levels, the restriction in the criterion is weaker, so the number of included neighbors is greater than in the lower levels. 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. In the case of age, for example, in the low networking level, a neighbor is included in mother *i*'s peer group if the age difference between the two women is less than or equal to five years, whereas a difference of less than or equal to ten years is allowed in the medium networking level. Combinations of these restrictions are used to determine the different specifications presented in this section.

Table 2 describes the covariates used in the regressions. For explaining the fertility variable, the covariates are variables that intuitively may have an explanatory power in the decision about the age for the onset of childbearing. Based on the panel constructed with the SISBEN information, these covariates are known before pregnancy or at the time of pregnancy (in the case of mothers that were pregnant at the time they were surveyed). The covariates include personal characteristics such as educational attainment, marital status, and employment status, as well as some household characteristics such as income or house features.

OLS and IV Regressions

After testing different restrictions for the maximum distance allowed between peers, I concluded that the specifications that best fit the data are the ones using matrices with nonzero weights for peers inside a circle with a radius of 1,000 meters, centered on mother *i*'s residence. These are the results presented in this section. All regressions include the set of neighborhood fixed effects, $\left[\sum_b \upsilon_b \delta_{bi}\right]$, and the set of contextual effects, $\left[\sum_{j\neq i} w_{ij}\mathbf{z}_{1j} \ldots \sum_{j\neq i} w_{ij}\mathbf{z}_{kj}\right]$.

Since the estimation sample used in the regressions is the sample of SISBEN recent mothers that can be followed throughout the different survey collections (cases 1 to 5 in figure 4), the expectation of the endogenous variable among peers is also defined using that sample. Nevertheless, to reduce potential endogeneity of the contextual effects, all the expectations of the covariates among peers are computed using the sample of SISBEN mothers who fulfill the conditions for being included as the individual's peers, regardless of whether they are included or in the estimation sample. The same strategy is used to compute the instruments. That strategy substantially improves the quality of the estimations in the paper.

Tables 3 and 4 present the results of the estimation for different specifications of the **W** matrix and different sets of instruments for a given **W**. Six different configurations for the peer group are presented in the table; each configuration is a combination of the distance, age, and education restrictions that 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 to the criteria presented in table 1. In the first specification in table 3, for example, the peer group includes peers who live within a one-kilometer radius of the individual, who are within an age range of no more than five years from the individual, and whose education level is within three years of the individual's education level.

TABLE 3. Estimation Results: Age Disparity of Five Years[a]

	Specification and estimation method								
	(1) $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 5$; $\lvert e_i - e_j \rvert \leq 3$			*(2)* $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 5$; $\lvert e_i - e_j \rvert \leq 5$			*(3)* $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 5$; $\lvert e_i - e_j \rvert \leq 7$		
Variable	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[c]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]
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)
Completed 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)
Completed 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)
Some college or completed 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)
Attending 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 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)
Widowed	−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)

(*continued*)

TABLE 3. Estimation Results: Age Disparity of Five Years[a] (*Continued*)

	Specification and estimation method								
	(1) $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 5$; $\lvert e_i - e_j \rvert \leq 3$			*(2)* $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 5$; $\lvert e_i - e_j \rvert \leq 5$			*(3)* $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 5$; $\lvert e_i - e_j \rvert \leq 7$		
Variable	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[c]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]
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)
Monthly household income (in pesos)	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 wall materials	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 floor materials	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)
No. of teenage mothers in 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[e] 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[e] 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
Summary statistic									
No. observations		11,465	11,465		11,465	11,465		11,465	11,465
R squared		0.840	0.830		0.861	0.850		0.866	0.862
F statistic		14.17	11.36		28.44	17.03		30.70	16.49
Hausman test (*p* value)[f]		0.55	0.66		0.98	0.99		0.98	0.77
First simulation[g]		0.28	0.37		0.14	0.35		0.20	0.31
Second simulation[h]		1.57	2.06		0.75	1.83		1.01	1.54

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

a. All instruments use peer group definition $|d_i - d_j| < 1$ km; $|a_i - a_j| \leq 5$; $|e_i - e_j| \leq 3$, which generates the most excluded peers. All instruments based on excluded peers data impose a minimum distance of 250 meters between the individual and the excluded peer. Robust standard errors are in parentheses.

b. Two instruments are used: the expectation of the outcome among the excluded peers, and the expectation of the outcome among the excluded peers of the excluded peers.

c. Two instruments are used: the expectation of the variables incomplete and complete college and standard quality wall materials among the excluded peers.

d. Two instruments are used: the expectation of the variables widowed and standard quality wall materials among the excluded peers.

e. Stratum is a socioeconomic classification of the neighborhoods in Colombia, used by the government as a strategy to target subsidy programs, where the first stratum is the poorest. The omitted alternative is stratum 1.

f. Hausman test of overidentifying restrictions (*p* values).

g. 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.

h. Change in individual fertility behavior caused by a reduction of one standard deviation in the onset of childbearing for each member of the peer group.

TABLE 4. Estimation Results: Age Disparity of Ten Years[a]

	Specification and estimation method								
	(4) $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 10$; $\lvert e_i - e_j \rvert \leq 3$			(5) $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 10$; $\lvert e_i - e_j \rvert \leq 5$			(6) $\lvert d_i - d_j \rvert < 1$ km; $\lvert a_i - a_j \rvert \leq 10$; $\lvert e_i - e_j \rvert \leq 7$		
Variable	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[c]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]
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)
Completed 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)
Completed 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)
Some college or completed 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)
Attending 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 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)
Widowed	−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)
Monthly household income (in pesos)	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 wall materials	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 floor materials	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)
No. of teenage mothers in 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[e] 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)

(*continued*)

TABLE 4. Estimation Results: Age Disparity of Ten Years[a] (*Continued*)

	Specification and estimation method								
	(4) $\lvert d_i - d_j\rvert < 1$ km; $\lvert a_i - a_j\rvert \leq 10$; $\lvert e_i - e_j\rvert \leq 3$			*(5)* $\lvert d_i - d_j\rvert < 1$ km; $\lvert a_i - a_j\rvert \leq 10$; $\lvert e_i - e_j\rvert \leq 5$			*(6)* $\lvert d_i - d_j\rvert < 1$ km; $\lvert a_i - a_j\rvert \leq 10$; $\lvert e_i - e_j\rvert \leq 7$		
Variable	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[c]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]	*OLS*	*2SLS:* **WY***pp*[b]	*2SLS:* **WX***pp*[d]
Stratum[e] 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
Summary statistic									
No. observations		11,465	11,465		11,465	11,465		11,465	11,465
R squared		0.794	0.800		0.831	0.819		0.847	0.839
F statistic		50.35	20.85		85.42	18.46		117.93	24.80
Hausman test (p value)[f]		1.00	0.98		0.30	1.00		0.27	0.98
First simulation[g]		0.53	0.44		0.20	0.45		0.24	0.49
Second simulation[h]		3.12	2.60		1.19	2.71		1.33	2.70

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

a. All instruments use peer group definition $\lvert d_i - d_j\rvert < 1$ km; $\lvert a_i - a_j\rvert \leq 5$; $\lvert e_i - e_j\rvert \leq 3$, which generates the most excluded peers. All instruments based on excluded peers data impose a minimum distance of 250 meters between the individual and the excluded peer. Robust standard errors are in parentheses.

b. Two instruments are used: the expectation of the outcome among the excluded peers, and the expectation of the outcome among the excluded peers of the excluded peers.

c. Two instruments are used: the expectation of the variables incomplete and complete college and standard quality wall materials among the excluded peers.

d. Two instruments are used: the expectation of the variables widowed and standard quality wall materials among the excluded peers.

e. Stratum is a socioeconomic classification of the neighborhoods in Colombia, used by the government as a strategy to target subsidy programs, where the first stratum is the poorest. The omitted alternative is stratum 1.

f. Hausman test of overidentifying restrictions (p values).

g. 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.

h. Change in individual fertility behavior caused by a reduction of one standard deviation in the onset of childbearing for each member of the peer group.

For each configuration, results are presented for the OLS regression and the two-stage least squares (2SLS) regressions. I use two instruments in the first 2SLS regression. The first instrument is the expectation of the fertility outcome among the excluded peers.[47] 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 the individual's peers or to the individual's peer group (that is, the second level of excluded peers). The instruments used in the second 2SLS are the expectation of some exogenous covariates among the excluded peers. I selected two of the best instruments for each regression.[48] The first stages for each of the IV regressions are presented in appendix B. The coefficients of the neighborhood fixed effects and the contextual effects have been omitted for ease of presentation.[49]

In all the regressions in table 3 (with an age disparity of five years) and table 4 (with an age disparity of ten years), the coefficient measuring the endogenous peer effects is positive and less than one; it is highly significant in all 2SLS regressions and in most of the OLS regressions. In all specifications, the OLS coefficients for endogenous peer effects are smaller than the 2SLS coefficients. In this setting, it is theoretically impossible to predict the direction of the OLS bias, which would depend on the correlation of individual factors driving the group selection with unobservable factors driving the fertility decision. For some specific unobserved factors, the bias would certainly be negative, such as an individual's restriction, unobserved to the econometrician, that has a positive effect on the fertility outcome (that is, the restriction delays the individual's onset of childbearing). Consider, for example, the presence of credit restrictions. If a credit restriction is an unobserved component that influences the fertility outcome, it will probably have a positive effect. The greater the credit restriction, the less prone individuals will be to starting a family. Furthermore, a greater credit restriction would also correlate with living in one of the bad-quality neighborhoods, which have the smallest average fertility outcome. Therefore, in that case the correlation of unobservable factors with the endogenous variable would be negative, as would the OLS bias.[50]

47. The process of identifying the excluded peers is as follows. For each of an individual's peers, five of her closest peers are selected (in terms of the criteria discussed before) conditional on their not belonging to the peer group of the original individual. The same process applies for the second level of excluded peers.

48. More details about these instruments can be found in the notes to the tables.

49. This information is available on request.

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

The endogenous peer effect coefficient is interpreted as the effect of 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 her 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 for different instruments. As mentioned, each set of instruments has its advantages; the ones based on excluded peer covariates are theoretically exogenous, and the ones based on excluded peer outcomes are expected to be stronger. The preferred specification in this paper is specification 4 (table 4), in which a mother is considered an individual's peer if she lives within a radius of 1,000 meters of the individual's residential location, the age disparity with the individual is not greater than ten years, and the educational disparity is not greater than three years of schooling. This specification is preferred for several reasons; the F test is remarkably high for the two sets of instruments used, the exogeneity test is strongly not rejected in both 2SLS regressions, and the difference between the 2SLS estimations is not very high. Nevertheless, the endogenous peer effects coefficients are specific for each definition of the peer group, so other coefficients based on alternative specifications are also valid.

The quality of the instruments included in the regressions is tested using an F test of the instruments in the first-stage regression.[51] In every regression presented in the table, the F statistic 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, overidentifying restriction tests are always informative.[52] The null hypothesis of the exogeneity of the instruments in the overidentifying restriction tests cannot be rejected in any of the estimations.

The endogenous peer effects coefficient in the preferred specification and with the best set of instruments is 0.70 (specification 4). This means that an increase of one year in the peer group's average age at first childbirth implies

51. 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.

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

an increase of the mother's age at first childbirth of 0.7 year. The estimated coefficients are similar in several of the alternative specifications, but in others, the magnitude is substantially lower; in all cases, the coefficients are strongly significant. 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 the tables. In the first simulation, the incidence of teenage motherhood among the individual's peer group is increased by 10 percentage points. To do this, I randomly assign an age at first childbirth of sixteen to members of the individual's peer group, until the percentage of teenage mothers increases 10 percentage points. In the second simulation, the age at the onset of childbearing is reduced by one standard deviation for each member of the peer group. In the preferred specification, an increase of 10 percentage points in the incidence of teenage motherhood would reduce the age at the onset of childbearing by 0.5 years. A reduction of one standard deviation in the onset of childbearing for each member of the peer group will reduce the individual's onset of childbearing by more than three years.

In almost all the 2SLS specifications, higher educational achievements, especially completed 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 (specifications 1 and 4). When this restriction is relaxed, the estimated coefficients for educational achievement are such that having some college or beyond implies at least two years of delay in an individual's onset of childbearing. The positive effect of unemployment is significant in all specifications, which captures the fact that unemployed women are less prone to starting a family, holding all else constant. 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 variable for cohabitation with their partner is always negative and significant; therefore, personal relationships outside of marriage have a negative effect in the fertility outcome. The variable for the number of mothers (peers) who had their first child as teenagers and who currently live 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.

To put the endogenous peer effects in context, the magnitude of the endogenous peer effect coefficient can be compared with the magnitude of other coefficients. Based on the results of specification 4, a reduction of one year in the age at the first childbirth for all of an individual's peers results in a reduction of almost 0.69 year in the endogenous fertility outcome (using the **WY***pp* set of instruments in the second column). This reduction is similar in magnitude to the drop associated with the dummy variable for cohabitation (–0.63). In other words, the peer effects have almost the same impact on the endogenous fertility outcome as the important individual decision of cohabitating with a partner. In terms of education, women who have any college education (complete or partial) delay the time at which they decide to have their first child by almost three years. The individual's response to an increase of one standard deviation (five years) in the age at first childbirth of all the members of her peer group would similarly delay the age of the onset of childbearing by almost three years. In other words, a big change in the peers' fertility outcome (one standard deviation) would have the same impact as having some college (completed or partial) on an individual's decision regarding the onset of childbearing. These comparisons give an idea of the importance of the endogenous peer effects.

With regard to the robustness of these results, one potential problem is that the covariates and the expectation of the dependent variable used in the estimations are constructed based on information that may be from several years before a woman's pregnancy. As a robustness check, the same specifications presented above were estimated with a restricted sample of mothers who became pregnant not long before they were interviewed for the SISBEN.[53] In the estimations with the reduced sample, the peer effects coefficient remains positive and significant and, in most cases, of a similar or larger magnitude and always less than one.

Conclusions

This paper has presented 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 would cause a reduction of 0.7 year in the individual's onset of childbearing. The coefficient varies with different definitions of the peer group, but in all 2SLS specifications, the coefficient for β in equation 13 is positive and strongly

53. Three sample restrictions were imposed. The first sample only includes mothers who got pregnant at most two years before they were interviewed for the SISBEN, thereby excluding mothers who got pregnant before this two-year threshold. The second sample only includes mothers who got pregnant at most one year before they were interviewed for the SISBEN. The third sample only includes mothers who were pregnant at the time they were interviewed.

significant. In the case of the preferred model specification, an increase of one standard deviation in the onset of childbearing for each member of the peer group will cause a significant increment of 3.1 years in the individual's onset of childbearing. This increase of 3.1 years in the individual's onset of childbearing for the sample of SISBEN recent mothers will correspond to a reduction in the prevalence of teenage motherhood of 20 percentage points.

In many aspects, Medellín is a good representation of a standard Latin American city, with high levels of teenage motherhood in poor neighborhoods. Peer effects are probably one of the factors explaining the generalized reduction in the average onset of childbearing observed in almost all Latin American and Caribbean countries. In the case of Colombia, this paper presents evidence that social interactions play a crucial role in the determination of fertility outcomes and are one of the explanations for the increase in the teenage fertility rates observed in the last thirty years.

From a theoretical point of view, while individuals are best able to decide the right time to start a family, there are factors that interfere with this process, causing individuals to make inefficient decisions, such as teenage pregnancy. This paper provides evidence that peer effects are one of those factors. In the presence of peer effects, a high incidence of teenage pregnancies among the individual's peer group will cause the individual to have her first child at an earlier age. Simulations based on the estimated coefficients of the model found that an increase of 10 percentage points in the adolescent fertility rate among the individual's peers will cause a reduction of 0.5 years in the individual's age at first childbirth.

It is not easy to think of a policy that can control a social force like peer effects. Nevertheless, spatial socioeconomic segregation exacerbates their negative effects. Many Latin American cities are socioeconomically segregated in the sense that there are big clusters of high-quality neighborhoods geographically separated from big clusters of low-quality neighborhoods. In segregated cities, individuals in poor neighborhoods have as their peers mothers who had their onset of childbearing earlier than the population average. In a situation like this, peer effects can be seen as contributing to the formation of poverty traps. Any social policy that promotes 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.

Education plays an important role in delaying the onset of childbearing and thereby reducing the prevalence of teenage pregnancy. Some of the best estimated models reveal that women with completed college or some college delay the onset of their childbearing by almost three years, on average, in comparison with women with incomplete elementary or no education at all. In addition, these positive effects of education may be exacerbated through social interactions as well. There may be additional strategies that yield good results in terms of preventing adolescent fertility and that operate by changing the accepted reference model of behavior.[54] The effect of nonstandard interventions could be an excellent field of exploration for future literature.

Appendix A: Comparison of the SISBEN Recent Mothers Population with the Estimation Sample

Table A1 presents the summary statistics and a difference-in-means *t* test for comparing the population of SISBEN recent mothers and the estimation sample. The main reason for excluding an observation from the estimation sample is that it cannot be linked to a previous SISBEN collection, in which case information before the pregnancy is not available.[55] While the missing values generated in the estimation construction process should not necessarily follow a special endogenous pattern, some covariates show important differences between the SISBEN population and estimation sample. To verify that the determination of the estimation sample is not driving the results of 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 significant variation after controlling for the selection.[56]

Appendix B: First-Stage Regressions

Table B1 presents the first-stage regression of the 2SLS procedures presented in tables 3 and 4. The instrumental variables change for each specification. A description of the instruments is provided in the main text, as well as in the notes to tables 3 and 4.

54. For example, La Ferrara, Chong, and Duryea (2012) evaluate the effect of soap operas on fertility in Brazil.

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

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

TABLE A1. Comparison of SISBEN and Estimation Samples

	SISBEN recent mothers			*Estimation sample*			
Variable	*No. observations*	*Mean*	*Std. deviation*	*No. observations*	*Mean*	*Std. deviation*	T *test*
Onset of childbearing age	75,768	22.18	5.43	11,461	21.43	5.02	14.8
Some elementary	75,768	0.07	0.25	11,461	0.07	0.25	−1.0
Completed elementary	75,768	0.11	0.31	11,461	0.11	0.32	−1.6
Some high school	75,768	0.31	0.46	11,461	0.34	0.47	−5.7
Completed high school	75,768	0.41	0.49	11,461	0.42	0.49	−0.6
Completed college, some college	75,768	0.09	0.29	11,461	0.05	0.22	17.4
Attending school	75,768	0.12	0.33	11,461	0.14	0.35	−5.9
Unemployed	75,768	0.04	0.20	11,461	0.06	0.24	−7.6
Cohabitating with partner	75,768	0.34	0.48	11,461	0.19	0.40	36.9
Widowed	75,768	0.00	0.06	11,461	0.00	0.06	0.0
Divorced	75,768	0.03	0.17	11,461	0.03	0.17	−0.4
Single	75,768	0.50	0.50	11,461	0.71	0.46	−45.3
Physical or mental disability	75,768	0.01	0.07	11,461	0.01	0.08	−2.4
Monthly household income (in pesos)	75,768	664,585	793,465	11,461	694,033	727,496	−4.0
Living in apartment	75,768	0.96	0.20	11,461	0.97	0.17	−6.6
Good or standard quality wall materials	75,768	0.80	0.40	11,461	0.79	0.41	3.2
Good or standard quality floor materials	75,768	0.49	0.50	11,461	0.43	0.49	11.8
House is owned by the household	75,768	0.36	0.48	11,461	0.53	0.50	−32.2

TABLE B1. First-Stage Regressions

	(1)		(2)		(3)		(4)		(5)		(6)	
Variable	*2SLS:* **WY***pp*[a]	*2SLS:* **WX***pp*[b]	*2SLS:* **WY***pp*[a]	*2SLS:* **WX***pp*[c]	*2SLS:* **WY***pp*[a]	*2SLS:* **WX***pp*[c]	*2SLS:* **WY***pp*[a]	*2SLS:* **WX***pp*[b]	*2SLS:* **WY***pp*[a]	*2SLS:* **WX***pp*[c]	*2SLS:* **WY***pp*[a]	*2SLS:* **WX***pp*[c]
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)
Completed 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)
Completed 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)
Completed 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)
Attending 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)
Widowed	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 (in pesos)	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 wall materials	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 floor materials	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 household	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)
No. of teenage mothers in household	−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)
Stratum 2	−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)
Stratum 3	−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)
Summary statistic												
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
No. 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

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

a. Two instruments are used: the expectation of the outcome among the excluded peers, and the expectation of the outcome among the excluded peers of the excluded peers.

b. Two instruments are used: the expectation of the variables incomplete and complete college and standard quality wall materials among the excluded peers.

c. Two instruments are used: the expectation of the variables widowed and standard quality wall materials among the excluded peers.

References

Barrera, Felipe, and Lucas Higuera. 2004. “Embarazo y fecundudad adolescente.” Working Paper 24. Bogotá: Foundation for Higher Education and Development (Fedesarrollo).

Becker, Gary S. 1981. *A Treatise on the Family.* Harvard University Press.

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

Buvinic, Mayra. 1998. “Costos de la maternidad adolescente en Barbados, Chile, Guatemala y México.” *Studies in Family Planning* 29(2): 201–09.

Case, Anne C., and Lawrence Katz. 1991. “The Company You Keep: The Effects of Family and Neighborhoods on Disadvantaged Youth.” Working Paper 3705. Cambridge, Mass.: National Bureau of Economic Research.

Castañeda, Tarcisio. 2005. “Targeting Social Spending to the Poor with Proxy-Means Testing: Colombia’s SISBEN System.” Social Protection Discussion Paper 32759. Washington: World Bank.

Coleman, James S., and others. 1966. “Equality of Educational Opportunity.” Washington: Government Printing Office.

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.

Flórez, Carmen E., and Jairo Núñez. 2001. “Teenage Childbearing in Latin American Countries.” Research Network Working Paper R-434. Washington: Inter-American Development Bank.

Flórez, Carmen E., and Eugenia Soto. 2007a. “Fecundidad adolescente y desigualdad en Colombia.” Population Notes 83. Santiago: United Nations Economic Commission for Latin America and the Caribbean.

———. 2007b. “Fecundidad adolescente y pobreza: Diagnóstico y lineamientos de política.” Bogotá: National Planning Department.

Gage, Anastasia J. 1999. “The Social Implications of Adolescent Fertility.” *Population and Poverty in Developing Countries,* edited by Massimo Livi-Bacci and Gustavo De Santis, chap. 8, pp. 120–43. Oxford University Press.

Gaviria, Alejandro. 2000. “Decisiones: sexo y embarazo entre las jóvenes colombianas.” *Coyuntura Social* 23. Bogotá: Foundation for Higher Education and Development (Fedesarrollo).

Gaviria, Alejandro, and others. 2010. “The Cost of Avoiding Crime: The Case of Bogotá.” In *The Economics of Crime: Lessons for and from Latin America,* edited by Rafael di Tella, Sebastian Edwards, and Ernesto Schargrodsky, chap. 3, pp. 101–32. University of Chicago Press.

Grogger, Jeff, and Stephen Bronars. 1993. “The Socioeconomic Consequences of Teenage Childbearing: Findings from a Natural Experiment.” *Family Planning Perspectives* 2(4): 156–61.

Hayes, Cheryl D., ed. 1987. *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing.* Washington: National Academies Press.

Hotz, V. Joseph, Susan W. McElroy, and Seth G. Sanders. 1999. "Teenage Childbearing and Its Life Cycle Consequences: Exploiting a Natural Experiment." Working Paper 7397. Cambridge, Mass.: National Bureau of Economic Research.

Hoxby, Caroline M. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90(5): 1209–38.

Iyer, Sriya, and Melvin Weeks. 2009. "Social Interactions, Ethnicity, and Fertility in Kenya." Cambridge Working Paper in Economics 903. University of Cambridge.

Joshi, Shareen, and T. Paul Schultz. 2007. "Family Planning as an Investment in Development and Female Human Capital: Evaluating the Long-Run Consequences in Matlab, Bangladesh." Yale University.

Kohler, Hans-Peter, Jere R. Behrman, and Susan C. Watkins. 2001. "The Density of Social Networks and Fertility Decisions: Evidence from South Nyanza District, Kenya." *Demography* 38(1): 43–58.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4(4): 1–31.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60(3): 531–42.

Montgomery, Mark R., and John B. Casterline. 1993. "The Diffusion of Fertility Control in Taiwan: Evidence from Pooled Cross-Section Time-Series Models." *Population Studies* 47(3): 457–79.

Montgomery, Mark R., and others. 2001. "Social Networks and Contraceptive Dynamics in Southern Ghana." Working Paper 153. New York: Population Council, Research Division.

Oates, Wallace E., William N. Evans, and Robert M. Schwab. 1992. "Measuring Peer Group Effects: A Study of Teenage Behavior." *Journal of Political Economy* 100(5): 966–91.

Rosero-Bixby, Luis, and John B. Casterline. 1994. "Interaction Diffusion and Fertility Transition in Costa Rica." *Social Forces* 73(2): 435–62.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross-Section and Panel Data Models.* MIT Press.