An Analysis of Fertility Behaviour in Mexico

by

Alfonso Miranda Caso Luengo

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Economics

University of Warwick, Department of Economics
September 2004

Table of Contents

List of tables

List of illustrations

Acknowledgements

I am grateful to my supervisors Wiji Arulampalam and Mark Stewart for their patient and continuous guidance during the elaboration of the present thesis. I am also grateful to Joao Santos Silva for his comments to Chapter 6 of the present thesis and his TSP code for estimating quantiles for counts.

I would like to express my gratitude to the people that gave me their invaluable personal support during my studies. To Adriana for all her love, support and patience. Many thanks to my family and to Pavel Gomez, Lud Marquez, Jesus Gonzalez, Juntip Boonprakaikawe, and Monica Correa who shared with me many unforgettable moments during my years at Warwick.

I would like, finally, express my gratitude to the Consejo Nacional de Ciencia y Tecnologia (CONACYT) for its financial support.

Declaration

The present Thesis contains material that is the product of the author's own work. Chapter 2 is a modified and extended version of a literature review contained in my MSc dissertation. This Thesis has not been submitted for a degree to another University.

Abstract

In the last few decades female permanent sterilization became the most used contraception method in Mexico. During this time the demand for short-term contraceptives fell consistently. The shift in the demand for contraceptives raises concerns among demographers that the timing of children may remain unchanged regardless of the observed reductions in period fertility rates. After presenting a brief discussion of the economic theory on fertility behaviour (Chapter 2) and introducing the reader to the main demographic issues of modern Mexico (Chapter 3), Chapter 4 assesses these ideas in the context of modelling the timing of a first child, using duration models as main analysis tool. Findings suggest that young cohorts of women are effectively delaying first birth relative to the experience of older generations.

Chapter 5 reports a study of the determinants of completed fertility. Special attention is given to studying how characteristics such as religion and ethnic group affect the likelihood of transition from low to high order parities. An innovative Double-Hurdle count model is developed for the analysis. Findings indicate that education and Catholicism are associated with reductions in the likelihood of transition from parities lower than four to high order parities. Being an indigenous language speaker increases the odds of a large family.

Chapter 6 enquires how fertility plans of young individuals who live in intact families (i.e., those where both biological parents are present) differ from fertility plans of young individuals who live in non-intact families. The role of family background in the formation of fertility plans is studied. Count data models are used in the analysis, including an innovative technique for estimating quantile regression for count data. Findings suggest that an absent father reduces planned fertility, especially when women have weak preferences towards children. Education decreases planned fertility if strong preferences towards children are felt.

Abbreviations

AIC	Akaike Information Criterion
AML	Asymmetric Maximum Lillihood
APE	Average Partial Effect
CAIC	Consistent Akaike Information Criterion
CONAPO	Consejo Nacional de Poblacion (National Population Council)
DHM	Double-Hurdle Model
EV	Extreme Value
GP	Generalized Poisson
GPM	Generalized Poisson Model
INEGI	Instituto Nacional de Estadistica, Geografia e Informatica
IUD	Intrauterine Dispositive (coil)
LRT	Likelihood Ratio Test
LS	Lamer-Schwarz Metric
NPMLE	Non Parametric Maximum Likelihood Estimator
PFS	Permanent Female Sterilization
TFR	Total Fertility Rate
TIGPM	Two-inflated Generalized Poisson Model

To Adriana

Chapter 1

Introduction

Population ageing is without doubt one of the most significant challenges of modern society. Two main factors contribute to this phenomenon. On one hand, unprecedented scientific and technological advances have made it possible to extend average life span in most latitudes of the globe and across most social groups. On the other hand, fertility in most countries of the world maintains a persistent downward trend. In fact, in Europe total fertility rates fell below replacement level since the second half of the 1970s decade (i.e., below 2.1 children per woman) and it is estimated that by year 2050 nearly 75% of all developing countries will register total fertility rates below two children per woman (UN Population Division 2003c, UN Population Division 2003b).[1] More

---

[1] There are two main descriptive statistics of fertility: the total fertility rate (TFR) and the age-specific fertility rates. The total fertility rate (TFR) is a measure of the number of children that a woman would have at the end of their fertile life if she follows the current year 'typical' fertility behaviour at all the stages of her life (it can be expressed in terms of children per 1,000 women). Age-specific fertility rates indicate the number of births per 1,000 women in different age-specific groups. TRF and age-specific fertility rates are known as well as 'period

significantly, the number of countries with total fertility rates below 2.1 children per woman went from 5 in year 1960 to 64 in year 2000. In other words, there was nearly a ten-fold increment in the number of countries with fertility rates below replacement in the last four decades (UN Population Division 2003a).

Living for longer and having fewer children is transforming the population profile of most countries in the world. While in 1950 there were 13 elder persons per hundred individuals aged between 15 and 64 in Europe, in year 2000 such a dependency ratio was registered to be of the order of 22 elder persons per hundred individuals of working age. During the same period of time the number of children per hundred persons aged between 15 and 64 went from 40 in year 1950 to 26 in year 2000 (UN Population Division 2003c). These trends are obviously creating unprecedented pressure on the pension and health systems, as there are continuously fewer economically active persons to support a larger amount of pensioners who live for longer. Although migration is likely to mitigate the effects on population ageing in the industrialised world, there are no doubts that a better understanding of fertility behaviour and its determinants are needed to design effective public policies so that fertility is set back to replacement levels.

In the developing world total fertility rates are still well over replacement level but follow a steep downward trend as many countries are entering into the final phase of their demographic transitions. In Latino America and the Caribbean, for instance, there were 71 children per hundred persons of working

---

fertility rates' because it are calculated on the basis of times series aggregate data.

age in 1950. For year 2000 it is reported that such a figure was only 51 children per hundred persons aged between 15 and 64 (UN Population Division 2003c). Though total fertility rates are unlikely to fall below replacement level in the short and medium term, there are some important fertility related issues that require active research in the developing world.

There is for instance the fact that fertility decline has been substantially faster in Latino America and Asia than in the Middle East and North Africa. Similarly, fertility decline in all countries of the developing world in the last four decades has been much more rapid than the fertility decline experienced in Europe during its main demographic transition. Further, and in contrast to what common sense would suggest, fertility rates vary widely across countries with similar human development indicators (Bongaarts and Watkins 1996). Although some potential explanations have been offered, there is still little understanding of why such differences are being observed.

Other topics need attention. For instance, it is not clear whether, along with the fall of family size, there are changes in the timing of children of the sort observed in industrialised countries – i.e., postponement of first birth and increase on inter-birth intervals. Clearly, women may enter motherhood early in life and schedule children in a traditional fashion (i.e., allowing only for short inter-birth intervals) and yet reduce lifetime fertility by adopting drastic natal control such as permanent female sterilisation (PFS) once desired family size has been achieved. This 'stopping' strategy leaves untouched the timing of children and, potentially, has important adverse consequences on the accumulation of women's human capital. Hence, research is needed to establish whether or not fertility decline has led to changes in the timing of children in the developing

world. Similarly, there is little understanding of how fertility plans (ideal fertility) are determined in developing countries and what are their main determinants. Finally, research is needed to explain why the composition of the demand for modern contraceptives and the speed of their adoption seem to vary so widely across countries of the developing world (Kohler 1997).

The present Thesis contributes a series of essays on fertility behaviour in Mexico. Three main topics are addressed: (a) timing of first birth, (b) total number of children born to women by the end of fertile life (completed fertility), and (c) fertility plans (ideal fertility). Data from the National Survey of Demographic Dynamics is used for all these studies (ENADID from its acronym in Spanish). The ENADID is a micro data set containing detailed information on Mexico's demographic dynamics that was created by the National Institute of Statistics, Geography and Informatics of Mexico (INEGI).

To establish a general theoretical background, Chapter 2 presents an overview of the economic theory of fertility behaviour. The chapter does not pretend to be a comprehensive review of the literature, but rather presents the main theoretical developments in the field in the last few decades.

Chapter 3 introduces the reader to the basic demographic stylised facts of Mexico in the last four decades. A summary of the main actions taken by the Mexican authorities on population issues during the same period of time is also presented. There aggregate data showing that period fertility rates in Mexico have followed a steep downward trend in the last forty years will be presented. Chapter 3 will also point out that in recent years female permanent sterilization has become the most popular contraceptive method in the country while the

demand for contraceptive pills, condoms and other short-term contraceptives has fallen consistently.

The popularity of female permanent sterilization and the reduction in the demand for short-term contraceptives have created concern among Mexican demographers that the timing of children in Mexico may remain unchanged despite the reduction on period fertility registered in the last four decades.[2] Chapter 4 assesses these ideas in the context of the timing of a first child. The main objective is to test whether or not young cohorts of Mexican women are effectively delaying first birth with respect to older generations. Discrete time (grouped data) duration models are used and unobserved individual heterogeneity is properly controlled for via the estimation of a semi-parametric approximation of the distribution of unobservables (discrete mass points). To avoid potential misspecification of the hazard function, the econometric model explicitly accounts for the existence of individuals that are never at risk of having a first child either because of choice or because they are unable to conceive. Observable characteristics such as education, religion and ethnic background are controlled for.

Chapter five studies the socio-economic determinants of completed fertility. The analysis recognises the fact that in Mexico, like in other countries of the

---

[2] The main concern is that Mexican women may still be entering motherhood at early ages and allowing only for short intervals between consecutive pregnancies. Clearly, this 'traditional' behaviour can have adverse affects on women's accumulation of human capital and their participation and performance in the labour market. For more detail on this topic see, for instance, Welti (1997) and Zavala de Cosio (1989).

developing world, many women have large families and take no active actions to limit their fertility. This chapter argues that women with large families may display such behaviour because they find themselves 'trapped' into a regime in which the opportunity cost of extra children becomes particularly low. In such a context, fertility at low and high order parities may be thought of as determined by two different data generating mechanisms (i.e., determined within two different regimes). To allow for this kind of behaviour chapter five introduces an innovative Double-Hurdle count model. In contrast to previous Double-Hurdle modelling that set two hurdles at zero, the econometric model of chapter 5 considers the case where there is a hurdle at zero and a hurdle at a strictly positive value (interval) of the dependent count variable.[3] The framework is easily extended to control for unobserved heterogeneity and to allow for endogenous switching between regimes. The study intends to establish how socio-economic factors such as religion and ethnic group affect the likelihood of transition from low to high order parities.

Chapter 6 analyses how young Mexican women form their fertility plans (ideal fertility). The main objective of the chapter is to study the implications that family structure may have on fertility plans. To achieve this goal, fertility plans of individuals who live in families with an absent biological parent are compared to fertility plans of individuals who live in intact families (i.e., families where

---

[3] Previous work on Double-Hurdle modelling introduces two barriers at zero so that a strictly positive outcome is observed only if these two "hurdles" are crossed. Double-Hurdle models have been widely used in the study of the demand for goods, tobacco and alcohol. See for instance, Cragg (1971) and Yen and Jensen (1996).

both biological parents are present). The study also enquires about the role that socio-economic characteristics of the family head such as education and income play on the formation of fertility plans of their young female dependants. In order to avoid confusion between actual and planed fertility, the chapter concentrates attention in analysing fertility plans of women aged between 15 and 17 years that live at least with one biological parent and have neither initiated economic independent life nor entered motherhood. A sample selection assessment of the potential bias induced by this data selection process is performed and no empirical evidence of such a bias is found. Count data models are used as the main analysis tool, including an innovative technique for estimating quantile regression for count data.

Chapter 7 presents a brief summary of the main finding of the Thesis and concludes.

Chapter 2

Economic Theory on Fertility Behaviour

2.1 Introduction

Why do people have children? What are the main determinants of the quantity and timing of children? What are the mechanisms of a demographic transition? In the last few decades economic theory has attempted to give a satisfactory answer to these questions. [4] Though up to now no unified theory exists, various influential models that offer a set of alternative explanations have been suggested. This chapter presents a brief overview of such literature. The chapter does not pretend to do a comprehensive review of the literature but rather presents the main theoretical developments in the field in the last forty years.

---

[4] A demographic transition is defined as the transition from a traditional demographic regime in which fertility and mortality are high to a modern regime in which fertility and mortality are much lower (INED 2004).

2.2 Completed Fertility

2.2.1 Quality-quantity trade off of children

Schultz (1973) and Becker and Lewis (1973) pioneered the research on quality-quantity trade-off of children. In their approach number and 'quality' of children are valuable goods that enter, as any other consumption good, in the utility function of parents.[5] In this context, a demand function for children is obtained by means of solving a standard utility maximization problem subject to a modified budget constraint. The framework introduces a non-linearity into the budget constraint because total cost of children is calculated as quality times quantity. Hence, the shadow price of quantity (quality) depends on quality (quantity), and any variation of income leads to changes in the shadow prices of both quantity and quality. More importantly, a substitution effect of quality against quantity is induced if income elasticity for quality is larger than income elasticity for quantity. According to the authors this mechanism may explain the reduction of family size in Occidental societies. However, the quality-quantity trade-off mechanism is not always guaranteed and under certain conditions the model predicts that lump-sum increments of income may lead to a larger demand for children of low quality, other things held constant.

Cigno (1986) uses this basic framework to study family taxation and its effects on fertility behaviour. The author uses an expenditure function to define the `net cost' per child given an arbitrary level of quality. Similarly, quality is

[5] Quality of children is commonly defined as any characteristic that increases the present and future welfare of children, say, education, bequests, and health.

defined as a concave function of monetary expenditure, bequests, and time spent on each child (which is equivalent to forgone labour income).[6] Cigno points out that the effects of fiscal policy depend on the relative degree of substitution of quantity and quality of children. If quantity and quality are complements, an increase in income tax is likely to lead to a higher demand for both quantity and quality as individuals have fewer incentives to work. If quantity and quality are substitutes, however, an increase in income tax may lead to a higher demand for quantity and a lower demand of quality. Further, the author shows that if quantity and quality are substitutes, child benefits encourage high fertility rates and reduces child quality because this sort of subsidy tends to reduce the price of quantity in terms of quality.

2.2.2 Children's time intensity and female forgone income

Willis (1973) suggests that childrearing is a relatively time-intensive activity that is mainly performed by women. He suggests that the opportunity cost of children increases as the female wage becomes higher, resulting in a substitution effect against children. In this context better opportunities of female education and work can be thought as the trigger of the demographic transition.

Willis illustrates this mechanism in a general equilibrium model where the utility of parents depends on the consumption of a 'domestic good' that is produced inside the household using market inputs and a flow of children

[6] The 'net cost' per child is defined as the minimum cost at which a child of a given quality can be produced. Such a costs depends on prices, wage rate, taxes and subsidies.

services. Women are specialised in home production and household production is performed with a constant returns to scale technology. Further, childrearing is supposed to be a time-intensive activity and women decide, by setting their fertility, the proportion of their time that is allocated to market and childrearing activities. Within this framework an increase in female market wage is shown to induce reductions in the demand for child services because the time women spend with children becomes more valuable in terms of forgone female labour income. Pollak and Watchter (1975), however, cast doubts about the generality of this result due to the constant returns to scale assumption.

2.2.3 Easterlin's hypothesis

Easterlin (1975, 1987) argues that individuals compare their current standard of living with an 'internalised' norm, and that such a relative income determines behaviour. According to the author individuals form their preferences towards consumption and children during adolescence. Once in adulthood, individuals compare their current standards of living with those experienced during childhood in the parental household. Easterlin's hypothesis suggests that if a reduction (increase) in relative income is perceived, then a downward (upward) adjustment in lifetime fertility is induced. Under this perspective a shift of preferences (aspirations) explains the baby boom and bust experienced in the US during the second half of the last century.

2.2.4 Fertility and Altruism

Becker and Barro (1986, 1988) and Becker, Murphy and Tamura (1990) introduced the concept of dynastic family into fertility theory. The authors argue that altruistic parents consider the effects of their fertility decisions on the welfare of future generations. As a consequence, present day individuals play the role of a central planner that maximizes a dynastic utility function – which is a weighted sum of the utility of present and future generations – subject to a dynamic budget constraint. Physical and Human capital accumulation is allowed and, as it is commonly assumed in most economic growth models, human capital is produced with a technology that exhibits increasing returns to scale.

The main findings from this sort of model indicate that an increase of income at any point of time is used by the central planner to acquire consumption goods, education, and children in such a way that next generations' resources remain unchanged. In this context, the demand for children is shown to increase (decrease) if the cost of children in terms of goods and time spend on childcare activities decrease (increase). Becker, Murphy and Tamura (1990) show that an economy requires the accumulation of a minimum level of human capital to provide incentives to individuals to make investments in education so that the economy may converge to a steady state growth path with low fertility and positive human capital accumulation. If this minimum level human capital is not reached the economy is permanently 'trapped' in a steady state with high fertility and no human capital accumulation. These results are particularly attractive because they offer a potential explanation for the demographic transition.

However, the overall idea that individuals are altruistic to the extent of maximizing a dynastic utility function has been seriously criticised (Bergstrom 1989).[7]

2.2.5 The Old-age Pension Motive

An alternative approach to study fertility behaviour has been built as an extension of Allais (1947), Samuelson (1958) and Diamond's (1965) overlapping generations model. In this literature the notion that parents derive utility from their offspring does not play a relevant role. Instead, it is stressed that children commonly make monetary transferences to their old parents once they reach adulthood. This old-age pension motive is stated then as the main incentive for parents to have children. In other words, having a child is regarded as an investment rather than an activity that produces joy and satisfaction to parents (for further details see Nugent 1985, Srinivasan 1988). This mechanism is important for individuals' welfare if there are no alternative ways to smooth

---

[7] Altruism does not, in general, abolish incentive problems in inter-generational relationships. This fact cast doubts about the generality of the 'dynastic approach'. For instance, if the head of the dynastic family cannot observe the actions of her children and the 'dynastic outcome' depends on the level of effort performed by each generation, nothing prevents children from deviating from a dynastic optimal path because they do not receive the marginal product of their effort. In this case there is no sense in performing an infinite horizon optimisation. Similar results arise when individuals of different generations weight 'other people's utility' in a different fashion.

consumption over time. However, if alternative assets exist (such as physical capital) children are likely to be dominated as storage technology and, in equilibrium, parents set fertility to zero (Neher 1971). This result has been used to suggest that the development of financial markets may be a trigger for the demographic transition. The framework, nonetheless, fails to explain why most individuals choose to have a positive quantity of children.

Though children may provide parents with a technology to smooth consumption over time, children should not be regarded as passive agents as they possess in general their own utility function. This is a very significant feature that the analyst cannot ignore. Zhang and Nishimura (1993) address the issue in a model where the main sustained hypothesis is that children are altruistic with respect to their parents but parents are not altruistic with respect to their children – i.e., they have children exclusively to ensure an old-age pension. Results suggest that if alternative assets are present (such as physical capital) the old-age pension hypothesis can support a society with a strictly positive fertility rate only if the parameters of utility and costs functions meet a set of rather restrictive conditions. From this perspective the old-age hypothesis seems to be of little use in applied social science.

Morand (1999) introduces human capital into the framework by presenting a model where parents invest in children and their human capital. The main sustained hypothesis suggests that children accumulate human capital during the time that parents spend in child-care activities. Further, it is postulated that the quality of a child's training is a non-decreasing function of the human capital owned by her parents. However, each child obtains at least the average level of human capital in the society. Under these circumstances, Morand finds that

parents have no incentives to invest in children's human capital if their own education is below a given threshold.

The intuition behind this result is that, given a low level of own education, parents find it relatively 'expensive' to make investments in the education of their offspring. Consequently, unskilled parents demand in equilibrium a large quantity of unskilled children rather than a small quantity of skilled children. In this framework, the author shows that if the average human capital in the society is low the economy is stuck in a 'poverty trap'. On the contrary, if the average human capital in the society is high the economy grows permanently. Random events that modify the average level of human capital can push in (or out) an economy from the 'poverty trap'. According to the author's point of view, this mechanism may explain the nature and determinants of a demographic transition.

2.2.6 Children as insurance

Cain (1983), Nugent (1985) and Pörtner (2001) suggest that children provide a general source of insurance when there is uncertainty about future income and alternative insurance is not reliable. For instance, children can provide parents with economic support in emergency conditions such as health disability or uncommon financial difficulties. This mechanism is argued to explain why individuals demand children even if their economic cost of children is larger than the discounted value of future expected transferences. Further, the theory suggests that if uncertainty about the future is large individuals will tend to spread their risk by means of demanding a large quantity of children. More importantly, it is

shown that if education reduces the variance of expected future income, fertility decreases as individuals accumulate human capital.

2.2.7 Norms, innovation and coordination

Most existing economic theory supposes that contraception is perfectly available at low or null cost for all individuals and that contraception use is not subject to any sort of uncertainty. Sociologists and demographers, supported mainly by evidence from developing countries, have for a long time rejected these ideas as a valid description of the data (Bongaarts and Menken 1983, Van de Walle 1992). From their point of view the transition from a regime of high mortality and high fertility to a regime of low mortality and low fertility is a long-term process where contraception and health services become available in an imperfect and gradual way to different groups of people. For these reasons, it is argued, the cost and benefits of contraception differ across individuals and its adoption is subject to a good degree of uncertainly about effectiveness and potential side effects (for more on these hypotheses see for instance Bongaarts and Watkins 1996).

Recent research suggests that adoption of contraceptives and reduction of family size may be thought of as an innovation process (See for example Montgomery and Casterline 1993, Rosero-Bixby and Csterline 1993, Bongaarts and Watkins 1996, Kohler 1997, Kohler 2000). These ideas have contributed to bring closer the points of view of economists, sociologists and demographers. Under the new perspective the transition from a regime of high mortality and high fertility to a regime of low mortality and low fertility is a dynamic process of diffusion of knowledge and adoption of new techniques of contraception and

fertility 'norms'. Following a microeconomic principle, heterogeneous individuals are recognised to be decision units that choose either to keep traditional fertility patterns or to adopt innovative behaviour on the basis of solving a utility maximization problem. Social-network effects are present so that an individual's costs and/or benefits from innovation are a function of the number and identity of other innovators in his/her social network. By this means 'contagion' or diffusion of the new fertility standard is generated.

The hypothesis has various advantages. First, it integrates in a natural way the notion that individuals are heterogeneous, this is, that they have individual-specific characteristics that affect the cost and benefits of innovation. This feature is capable of integrating the views of demographers and sociologist suggesting that, during a demographic transition, contraception technology becomes available in an imperfect and gradual way so that the cost of adoption differs across individuals and social groups. Second, it recognises the existence of 'fertility standards' and the importance of social influence and social-network effects in their diffusion. Finally, it accommodates a micro-founded mechanism of fertility decision-making into the analysis.

Various mechanisms justify the existence of social-network effects in the diffusion of new contraception technology and fertility 'norms'. One argument suggests that women are often uncertain about individual-specific suitability, effectiveness, and health implications of modern contraceptives (low fertility norm). In the presence of such uncertainty, women's contraceptive choice commonly relies on information obtained from early adopters in casual word-of-the-mouth communications with friends and neighbours (Kohler 1997, Montgomery and Casterline 1993). Under these circumstances, the probability

that an individual adopts modern contraception (low fertility norm) depends on the experience on contraceptive use (advantages of low fertility) accumulated in her social-network. Social learning creates then a network effect, and with it, the possibility of information cascades or herding (Ellison and Fudenberg 1995, Kapur 1995, Kirman 1993).

Direct social influence is other source of social-network effects that may affect the diffusion of innovative fertility behaviour. The intuition, which comes from sociology, is that individuals 'conform' to rather than 'revolt' against, established social norms in order to avoid potential punishment for misbehaviour – punishment may have different levels of credibility. Given these incentives, network partners (neighbours, friends) communicate with each other to generate a common local knowledge that is then used to make individual participation decisions – i.e., whether to conform or to revolt. In this context, low fertility (revolt) is a best response strategy for a given individual depending on the number and identity of other individuals in her/his social network that are themselves willing to choose low fertility (contraception adoption) as their own strategy (Kohler 2000). That is, each individual 'participates' only if other individuals participate as well – the "I'll go if you go" of Chwe (2000). The existence of such strategic complementarities creates the possibility of multiple and Pareto-rankable equilibria, inducing potential problems of coordination failure (Kohler 2000).

2.3 The Timing of Children

The basic model is due to Happel et al. (1984). These authors propose a dynamic setting in which parents have one child during an exogenously given (and finite) planning horizon and decide the 'arriving time' of their only heir. After a birth women leave the labour market for an exogenously given period of time. Parent's lifetime utility is the sum of per-period consumption utility and a fixed amount of child services. By assumption, these services do not depend on children's date of birth.[8] For simplicity, it is also assumed that male income increases exogenously with age and reaches eventually a maximum. Women accumulate human capital at the working place in a learning-by-doing fashion.

In this context, Happel et al. (1984) show that a highly qualified woman entering marriage has incentives to postpone first birth until the very end of her fertile life. This behaviour is found to ensure that human capital depreciation due to absences from work will be minimized so that lifetime income reaches its maximum. A low qualified woman, in contrast, finds it optimal not to postpone first birth. Another finding suggests that households delay first birth in order to smooth consumption. This result follows from the fact that, in the absence of well-developed capital markets, households are liquidity constrained and women's absences from the labour market necessarily imply reductions in current income and consumption. Thus, optimal behaviour induces women to postpone first birth until their husband's income has reached its maximum. Clearly, at that

---

[8] This assumption implies that child services and consumption are perfect substitutes from a parent's point of view.

point the marginal utility of female forgone income is minimum. According to the authors access to capital markets can reduce this consumption smoothing incentive. However, it is pointed out that even in developed societies households can only borrow a proportion of their tangible assets. Hence, such an incentive is present even in the most developed societies.

Cigno and Ermish (1989) and Cigno (1994) extend this basic framework to allow lifetime utility of parents to be a function of period fertility rates and the quality of children – i.e., parents care not only about the number of children but also about their 'arriving time' Further, it is postulated that there is a cost (in utility terms) of postponing parenthood so that, other things being equal, early births are preferred to late births. In this context, Cigno and Ermish show that the subjective price (i.e., utility adjusted price) of children varies over the different stages of an individual lifecycle because the opportunity cost of a labour market leave increases with the level of human capital – which in turn is accumulated in a learning-by-doing fashion at the workplace.[9] Therefore, in equilibrium, the rate of change of this subjective price of children should be equalized to the interest rate. This is an arbitrage condition. If children are 'expensive' (rate of growth of opportunity cost of children > interest rate) a household can improve its lifetime welfare by postponing current births, investing in the capital market, and 'catching up' fertility in the future. On the contrary, if children are 'cheap' (rate

[9] The subjective price of children is defined as a utility-adjusted cost of children. The 'raw' price of children is composed by three elements: (a) current expenses of child-bearing, (b) loss of present earnings due to woman's absence from work, and (c) loss of future potential earnings due to woman's forgone human capital investment (for further detail see Cigno and Ermisch 1989).

of growth of opportunity cost of children < interest rate) a household can do better by borrowing from the capital market and having early births. Cigno (1994) finds that women who marry late (or with a large non work-related human capital) have their children earlier than women who marry early (have low non work-related human capital). This result follows from the fact that, at any time, women who marry late pay a larger cost for having children than women who marry early. Therefore, they do better scheduling early births and avoiding paying additional costs for postponing motherhood. Cigno and Ermish (1989) and Cigno (1994) find the usual result that a steep career profile that pays a large rate of return on accumulated work experience is associated with significant postponements in the timing of children.

Hotz, Klerman and Willis (1997) present a similar model. In line with Happel et. al. (1984), Cigno and Ermish (1989), and Cigno (1994) accumulation of female human capital is allowed in a learning-by-doing fashion. Here, however, human capital depreciates when not used. Hotz, Klerman and Willis findings suggest that life-cycle reallocation of child services (number of births and its timing) result from changes in prices, particularly in the case of changes in the female wage rate. If a woman, for instance, has low wage and low human capital at the beginning of her planning period, early fertility is optimal because such behaviour minimises childbearing income losses. In contrast, if female wage and human capital are high at the beginning of her planning period, the best strategy will be delaying first birth until the marginal utility of income becomes low enough. Once the first child has arrived, minimizing inter-birth intervals is optimal. Hortz, Klerman and Willis also suggest that in the absence of perfect

capital markets an incentive to space births appears because households find themselves liquidity constrained and may not be able to afford close births.

2.4 Fertility Plans and Family Background

Since the seminal work of Becker (1960), Willis (1973) and Easterlin (1975), economic demography has been largely influenced by the idea that individuals decide their demand for children in a way such that, given a set of constraints, their utility is maximised. Though early models were mostly static, nowadays it is widely recognised that individuals solve a dynamic optimisation problem that produces a detailed plan of the total number of desired children (planned or ideal fertility) and their timing (Moffit 1984, Happel, Hill, and Low 1984, Ward and Butz 1980). Obviously, plans are formulated conditional on an individual's current information about their present and future wealth, their ability to avoid unwanted pregnancies, and the likely path of all relevant prices. Therefore, having made an initial plan, individuals update their fertility intentions any time a new piece of information is learned. Economic theory explicitly recognises that, since parents cannot dispose of live children, fertility plans are likely to exhibit path dependence (Ward and Butz 1980). The most significant prediction of this sort of literature is that lifetime increments in women's wages (or education as an indicator of future permanent labour income) induce reductions in planned and realised completed fertility (Willis 1973, Moffit 1984, Ward and Butz 1980).

Intuition suggests that family background influences lifetime planned fertility (ideal fertility) because it directly shapes preferences towards children and/or because it intervenes in the formation of expectations about the future. Various

arguments support the relevance of the latter mechanism. Firstly, there is the notion that knowledge about the characteristics of a family – especially those of the main economic support or 'head' – reveals information about the financial ability of the household to maintain dependent children at school and out of the labour market. Secondly, family background offers clues about potential future bequests, 'type' of genetic endowment, and the pool of skills and knowledge an individual could acquire by imitation of other members of her kin. All this information is likely to be exploited when expectations about future employment, wages and wealth are formed. Consequently, clearly it may affect fertility plans (for an empirical assessment of the importance of family background on individuals' education and labour market outcomes see, for instance, San-Segundo and Valiente 2003, Agnarsson and Carlin 2002, Ogawa and Ermisch 1996).

Recent literature has stressed the importance of family structure over all other background characteristics as a key determinant of children's future economic status. The argument has various avenues. First, there is the suggestion that adverse events such as divorce or the death of a parent commonly trigger a wide reorganization of social roles within the family. This change in roles creates considerable stress on children, their nurturing, and their future economic status. Second, adverse events may also affect the capacity of a parent to perform the adequate level of supervision and monitoring to avoid antisocial behaviour – that jeopardizes future economic status – among her children. Third, adverse family events commonly imply long-term economic hardship (for a survey on these ideas see Hill, Yeung, and Duncan 2001). And finally, divorce or death of a spouse commonly changes the relative empowerment of women within a family (Klawon

and Tiefenthaler 2001). Clearly, all the negative shocks induced by an adverse family event are likely to affect, to a certain degree, the future economic status of young individuals. A number of applied studies find empirical evidence that support these ideas (see, for instance, Hill, Yeung, and Duncan 2001, Ermisch and Francesconi 2001, Manski et al. 1992).

Given that family structure affects children's future economic status, intuition suggests that it may also have a relevant role in the formation of their fertility plans and, consequently, in their future demographic outcomes. This line of enquiry has not been systematically pursued in applied work. Among the most relevant available studies, Miles (2001) and Cherlin et. al. (1995) find that parental divorce in childhood increases the likelihood that young adults will cohabit instead of entering a legal marriage. In addition, Miles (2001) reports that, in the case of marriage, adolescents who suffer a parental loss through divorce are significantly more likely to experience marital instability. Similarly, Cherlin et. al. (1995) find that a parental loss during childhood increases the likelihood that an individual will have a child outside a legal marriage.

Chapter 3

An overview of Population Issues in Mexico

3.1 Introduction

Mexico was basically a low-populated country until the start of the 1930s decade. Since pre-colonial times most of its population has been concentrated at the centre and south of the country. Within each sub-region (north, centre, and south) population is mostly settled in a few cities of strong industrial and economic activity. Between 1930 and 1970 the population of Mexico expanded at increasing rates, and for the period 1970-1980 its annual rate growth reached a historic maximum of 3.32% per annum (see Figures 3.1 and 3.2 in the Appendix, page 54). Since then population has consistently grown at declining rates and, according to available data, in year 2000 it was as low as 1.85% (CONAPO 2001b). Today there are nearly 98 million people in the country and it is expected that the population will continue expanding for at least another fifty years (CONAPO 2002b). Reductions in the rate of population growth have been

fundamentally driven by changes in mortality and individual fertility behaviour, and associated with a general improvement in the living standards of Mexico. This chapter presents a brief description of the major demographic transformations of modern Mexico and the main policy actions implemented by the Mexican government during the last few decades.

3.2 Stylised Facts

3.2.1 Changes in the standards of life

The last forty years witnessed the transformation of Mexico from a rural society to an essentially urban nation. While in 1950 nearly 60% of all Mexican citizens were dispersed in settlements of less than 2,500 inhabitants (rural communities), by year 2000 75% of all Mexicans were living in towns of 2,500 or more inhabitants. In fact, available statistics from the 2000 Census indicate that as many as 61% of Mexicans live nowadays in cities of more than 15,000 inhabitants (INEGI 2001a). The accelerated urbanization of Mexico has been primarily determined by the emergence of significant migration waves from rural communities to large and middle-sized cities as well as by the rapid population growth observed in the main cities since the second half of the last century.

Alongside urbanisation other development indicators improved in recent years. In fact, between 1970 and 2000 average education in Mexico increased from 3.4 to 7.6 schooling years and life expectancy increased from 61 to 75 years (INEGI 2001c). Similarly, between 1970 and 1999, real GDP per capita grew from 2,295 to 3,613 US dollars. In other issues, female participation in the labour

force (workers/total women in working age) increased from 11 to 26.6 per cent (World Bank 2001). Finally, the number of people insured by the social security system went from representing 12% of total population in year 1960 to 61% in year 2001 (INEGI 2000, SSA Various years).

Despite the general improvement in the standards of life, income distribution is still highly concentrated. In fact, for year 2000 the earnings of the households at the bottom decile of the income distribution represented less than one percent of the aggregate income (see Table 3.1, page 53). Similarly, the income earned by households at the top decile of the income distribution represented as much as 54% of the aggregate income. In the same reference year the Gini coefficient registered a value of 0.62 points, well above the 0.30-0.40 range observed in most developed countries.

3.2.2 Population Structure

As suggested by its population pyramids in Figures 3.3a and 3.3b, Mexico is still a 'young' country. Most Mexican citizens are aged less than 20 years old and only a relatively small proportion of the elderly reach ages over 75 years. The reduction in the rate of population growth in the last forty years is starting to affect Mexico's population structure. In fact, comparing pyramids for year 1960 and year 2000 in Figures 3.3a and 3.3b the reader can conclude that young children are significantly reducing their participation in the total population. This is a clear signal that the 'ageing' of the country has already started. Therefore, as generations progress in their lifecycle in the future fewer and fewer people will

have to support a larger quantity of retired individuals. The pension and health system will be then subject to considerably higher pressures.

However, the country will enjoy a 'demographic bonus' in the next few decades, as a continuously increasing proportion of Mexicans will enter into their working life in the near future. Clearly, if the labour market expands at the required pace, fewer and fewer people will be economically dependant on other individuals. In fact, available information indicates that between 1990 and 2000 the rate of dependency (dependants per 100 people of working age) came down from 95 to 83 in rural communities and from 62 to 54 in large cities (INEGI 2001c). This tendency will continue for a number of decades.

Obviously this is a huge challenge for the Mexican economy as every year more and more positions in the labour market are (and will be) demanded and, in order to take advantage of its demographic bonus, the country must create the required work opportunities. Though migration is likely to reduce, to some extent, the demand for economic resources, it is clear that sustainable economic growth will require the allocation of enough resources to the education and training of the present and future labour force. Similarly, the country will be in need of a significant improvement in its overall infrastructure.

Another significant demographic challenge for the future seems to be the ever-increasing concentration of the population in large-scale cities. To stop and revert this tendency seems to be a prerequisite for a balanced and sustainable social and economic development in the future as well as a priority to avoid an imminent irrational deterioration of the environment.

3.2.3 Fertility

As is suggested in the introduction to the present chapter, the major demographic changes of Mexico in the last forty years have been associated with important modifications in individual fertility behaviour and a generalised reduction in mortality rates.

Figure 3.4 presents the evolution of the total fertility rate (TFR) in Mexico over the last few decades.[10] From Figure 3.4 the reader may conclude that fertility behaviour in Mexico can be characterised in two different stages. From the start of the 1930s and until (approximately) the middle of the 1960s the TFR in Mexico increased in a steady ascendant way from 6 children per woman to its historical maximum of 7.3 (CONAPO 2001b). Then, since the second half of the 1960s decade, the total fertility rate started its downward trend. Between 1965 and 1970 its registered fall was rather moderate. However, a dramatic downward tendency has been registered from the onset of the 1970s and for year 2000 the TFR was registered to be as low as 2.4 children per woman. Much of this accelerated reduction is likely to be associated with the improvements in the standard of living in Mexico during the last four decades. However, as is discussed later in the text, policy innovations implemented by the Mexican government from the start of the 1970s played, without doubt, a significant role.

The fall in the total fertility rate has been associated with reductions in all age-specific period fertility rates (see Figure 3.5). It is important to underline that fertility rates of young women appear to have started their reduction well before

[10] For a definition of TFR see Chapter 1, footnote 1.

the 1960s. The total fertility rate, however, changed its upward trend only after the fertility of women aged over 35 five years started to fall. Most of the reduction in period fertility rates for women aged over 35 years is explained by the 'cut off' of fertility at high order parities (see Welti 1997).

Though the drop of all age-specific period fertility rates suggest changes in the timing of children, from these demographic descriptive statistics it is difficult to establish to what extend the fall in TFR and age-specific fertility rates are explained by reductions on completed fertility (total number of children at the end of fertile life) and to what extend such reductions are due to modifications in the fertility calendar. In fact, under certain conditions it may be the case that the timing of children could be practically unchanged as some authors have indeed argued is the case (see Welti 1997).[11]

---

[11] To see how this is possible consider the case of a woman that has ten children during her whole fertile life – numbers are arbitrary and not important for the exercise. Suppose further that she has two children every five years starting at her 15th birthday. In other words, she has two children when aged between 15 and 19, two children when aged between 20 and 24, and so on. Now suppose that this particular woman changes her behavior and has only one child every five years but does only marginal changes in the time elapsed between consecutive births. Clearly, if there were many women who behave in this way in the population both age-specific fertility rates and total fertility rates would be reduced regardless no significant changes in the calendar of children have occurred. This example shows the difficulties that the analysts have to discern between changes in timing and number of children when only demographic descriptive statistics are used.

Total fertility rates vary across different groups of people depending on their socio-economic characteristics (see Table 3.2). For instance, women who had no formal education in 1974 were expected to end their fertile life with almost 8 children whereas women who completed secondary school or any other higher education were expected to have only 3.5 children. This is a difference of around 4 children. The decline of fertility triggered since the beginning of the 1970s decade has reduced significantly the fertility gap among women with different education achievements. However, such a gap was still present for year 1996 as the reader may conclude from Table 3.2. Fertility differences between women living in rural and urban zones are also quite remarkable as figures in Table 3.2 show.

3.2.4 Infant mortality

The fertility decline in Mexico has been closely associated to significant reductions on infant mortality.[12] Figure 3.6 presents the evolution of such an indicator in the last fifty years. As Figure 3.6 indicates, infant mortality rates in Mexico started their downward trend well before fertility rates began to decline. This phenomenon is a common feature of all demographic transitions as individuals do not intend to reduce their fertility intensity while the risks of losing children during their early stages of life remain high. Nowadays the Mexican infant mortality rate is estimated to be of around 25 deaths per thousand live births (CONAPO 2001b). Though this figure is low in relation to the

[12] The infant mortality rate is the number of children dying under a year of age divided by the number of live births registered during the same reference year.

previous history of the indicator in Mexico, it is still high if compared to the levels registered in industrialised countries (For year 2000 the indicators are Mexico 25, USA 6.9, UK 5.6, France 4.4, and Spain 3.9. See World Bank 2003).

Infant mortality in rural zones is far larger than infant mortality in urban zones. For instance, between 1991 and 1995 the average infant mortality rate in rural zones was registered to be as high as 48 deaths per thousand live births. In the same reference period, the average infant mortality rate in urban zones was of the order of 26 deaths per thousand births. Such differences reflect to a good extent relative differences in the supply of and access to health services in urban and rural zones.

3.2.5 Demand for contraceptives

The demand for contraceptives has changed significantly in the last twenty years. At the end of the 1970s nearly 35% of all users of contraceptives in Mexico adopted the contraceptive pill, 19% intrauterine dispositive (IUD), and 9% permanent female sterilization (PFS). In contrast, in year 1998 PFS was the most popular method (51%), followed by IUD (24%) and traditional methods (10%). At this last date, the contraceptive pill was selected by less than six per cent of all active users (INEGI 2001b, see Figure 3.7). Various factors are behind the change in the demand for contraceptives. First, there is a deliberate attempt by the public health system to promote the adoption of definitive natal control among women that have three or more children. In fact, most of the 'delivery effort' of contraceptives in the country has so far been concentrated on reaching women looking to initiate natal control (i.e., to take permanent sterilization

measures) after they reach their desired lifetime number of children (Zavala de Cosio 1990). Second, empirical evidence suggests that Mexican women are shifting to PFS because they feel that alternative methods may either lead to serious illness -such as cancer- or result in unwanted pregnancy due to ineffectiveness (Lindstrom 1998). Finally, young individuals do not appear to be adopting contraception at the beginning of their sexual life. Instead, there is a general tendency to wait until the arrival of the first or second child to initiate the use of contraceptives (Gomez 1996a).

3.2.6 Mexico-US migration

The US is by far the most important receptor of Mexican emigration. Available data indicates that during the 1970s decade the net demographic loss of Mexico due to migration to the US was between 1.20 and 1.55 million people. Such a large volume of migration has been increasing over the years. In fact, in the five-year period of 1990-1995 the net demographic loss of Mexico became as large as 1.5 million people. In other words, during this period 315 thousand Mexican citizens emigrated to the US each year in average. Further, in year 1996 it was estimated that between 7 and 7.3 million Mexican were residing in America, mainly in California and Texas (CONAPO 2000).

There are various factors that explain such a high migration intensity. There is first the fact that Mexico and the US share a large border with a long-term tradition of demographic movements. Second, populations on both sides of the border share long-term historical cultural links. Finally, there are important economic factors stimulating migration. On one hand the US economy demands

large quantities of unskilled Mexican workers for the agriculture, industrial and service sectors. In fact, it is estimated that the US will demand nearly 5 million migrants to fill a similar number of job positions that the economy will generate between 1998 and 2008 and for which the growth of the local labor supply will be insufficient. Moreover, 58% of such position will be created in traditional sectors where low skilled labour is required (CONAPO 2000). On the other hand there are, and there will be in the foreseeable future, large wage rate differentials that stimulate migration from the Mexican side of the border towards the American side of the border. To these economic factors one should add the 'demographic pressure' that has been accumulated over the years by the increase of the Mexican population of working age and the rather poor economic performance of Mexico in the last few decades. Clearly, the lack of jobs for young generations in Mexico has expanded significantly emigration volumes towards the US and it is likely to continue doing so in the future.

3.2.7 Current status of the indigenous population

Indigenous and non-indigenous individuals face different realities. For instance, indigenous infant mortality rate in year 2000 was estimated to be as high as 38.5 deaths per each 1,000 births, well above the 25 deaths per each 1,000 births corresponding to Mexico as a whole. Similarly, indigenous life expectancy is around three years lower than life expectancy for a typical Mexican citizen (CONAPO 2002a). Finally, average education among indigenous people is approximately three years lower than average education among non-indigenous people (INEGI 1999). Moreover, CONAPO (2001a) finds that from the 346

indigenous municipios (counties) of Mexico, a total of 209 municipios have a 'very high degree' of marginality, and that, from the remaining 137 indigenous municipios, 133 have a high degree of marginality.[13] Further, from the 186 predominantly indigenous counties, CONAPO classified 163 counties as having either a very high or high degree of marginality.[14] In other words, indigenous

---

[13] According to CONAPO (2001b) a county is said to be indigenous when 70% or more of its population speaks an indigenous language. Similarly, a county is said to be predominantly indigenous when at least 40% and at most 69% of its population speaks an indigenous language.

[14] Marginality is defined as the 'structural...difficulty for diffusing technological progress in the productive structure and in all regions of the country, and...the exclusion of social groups from development and its benefits' CONAPO 2001:11 <<author's translation>>. CONAPO (2001a) builds a County Marginality Index (CMI) for Mexico. The CMI is a summary index that seeks to measure marginality in each Mexican County taking into account diverse aspects of its socio-economic development. For the construction of the CMI, CONAPO uses a number of indicators associated with the marginality phenomenon including percentage of illiterate inhabitants, percentage of population aged 15 or more without primary education, and percentage of population without access to pipe water, sewage or electricity. The range of CMI is then divided into five sections in order to create five degrees of marginality: very high, high, medium, low, and very low. This measure of degree of marginality is thus a relative, country-specific, measure of marginality.

individuals in Mexico live typically in extreme poverty and have no access to the benefits of economic development.

3.3 Public policy

3.3.1 Family planning policy

In 1973 the Mexican government initiated for the first time a public programme to offer free contraceptives and to promote family reduction as a rational and responsible behaviour among Mexican citizens. Simultaneously, all previous legal restrictions for the sale of contraceptives were lifted. Between 1973 and 1979 these 'family planning' campaigns targeted potential users of contraception in urban and sub-urban zones. But at the onset of the 1980s rural zones were also integrated (Cabrera 1994). Since 1977 population growth targets have been established and revised regularly by the Mexican authorities as a tool for the evaluation and formulation of the population and family planning policy. For instance, between 1977 and 1982 a target population growth rate of 2.5% was pursued. However, with the progression of the Mexican fertility transition achieving further reductions on fertility and mortality rates turned difficult. Hence, with the passing of time, such population growth targets tended to lose their utility as a valuable tool for public policy and in 2000 were finally abandoned.

During the last 20 years geographical coverage of such campaigns increased significantly. However, universal access to modern contraceptives is still far from reality. Despite the failure of providing universal access to contraception,

population policy in Mexico is widely considered a success, as the diffusion and adoption of modern contraceptives increased dramatically in the past few decades. In fact, while in 1976 thirty percent of all married women – or living in consensual union – were active users, in 1998 the figure was estimated to be seventy per cent (INEGI 2001c). Today, and since the late 1970s, the public sector constitutes the main supplier of contraceptives in the country, though private supply remains important (INEGI 2001c).

Recently, in year 2003, the Mexican authorities made available for the first time emergency contraceptives ('pill of the day after'). As expected, a strong initial reaction against the new policy was generated – particularly from the Catholic Church. However, despite the initial opposition the introduction of emergency contraception in Mexico seems to be successful and it is expected to play an important role in enhancing the options for fertility control of young generations of Mexicans.

3.3.2 Sexual education and gender orientated policies

In order to promote a new demographic culture among Mexican children, the 1973 policy shift in Mexico led to a significant reform of the academic contents taught at the primary education system (Aguilar 1993, Rodriguez 1991, Camarena 1991). The review concluded that all official textbooks for primary education should include topics on sexuality and family planning, though specific information about contraception practice was, carefully, avoided (Camarena 1991). Moreover, no actions were taken to promote the right of women to play an active role in family and society.

Beyond this initial effort, sexual education was largely neglected as a valuable tool for population policy until the first part of the 1980s. In 1984 a National Population Plan was elaborated for the first time. Such a plan introduced a series of new actions on sexual and family planning education, including the launch of conferences and seminars on demographic issues for students at the various levels of the national education system. Activities for teachers and workers of the social security system were also organized for the first time. Most programs, however, were implemented at a federal level and local authorities did not participate in their elaboration and implementation. It was only in 1987 that a properly coordinated, nationwide, programme for demographic education was created (Sanchez and Monterrubio 2000). A new gender orientated perspective was adopted. Hence, besides the usual family planning contents, information campaigns began to include topics on sexual health and to promote equality among men and women in the family and society. Around the same time the Gente Joven (Young people) project was introduced to meet the specific informational needs of young Mexicans (Pizzonia 1996). Radio and Tv spots informing and promoting family planning were produced and intensively transmitted at top audience hours across all Mexico. Finally, a nationwide telephonic help-line for the young on sexual education and contraception issues was introduced in 1994 (de Joven a Joven service line). All these policy actions have been extended and diversified in the last few years.

Appendix

Table 3.1 Income distribution in Mexico 2000

Decile	Households	Income (Pesos)	% share
Total	22639808.00	34906242757.00	100
I	2273503.00	3032133.00	0.00
II	2255738.00	277381635.00	0.80
III	2263222.00	705881819.00	2.00
IV	2350393.00	1100773032.00	3.20
V	2266584.00	1433303193.00	4.10
VI	2173383.00	1789287096.00	5.10
VII	2284673.00	2471537530.00	7.10
VIII	2236343.00	3297537118.00	9.40
IX	2272004.00	5081898228.00	14.60
X	2263965.00	18745610975.00	53.70
Gini Coefficient			0.62952

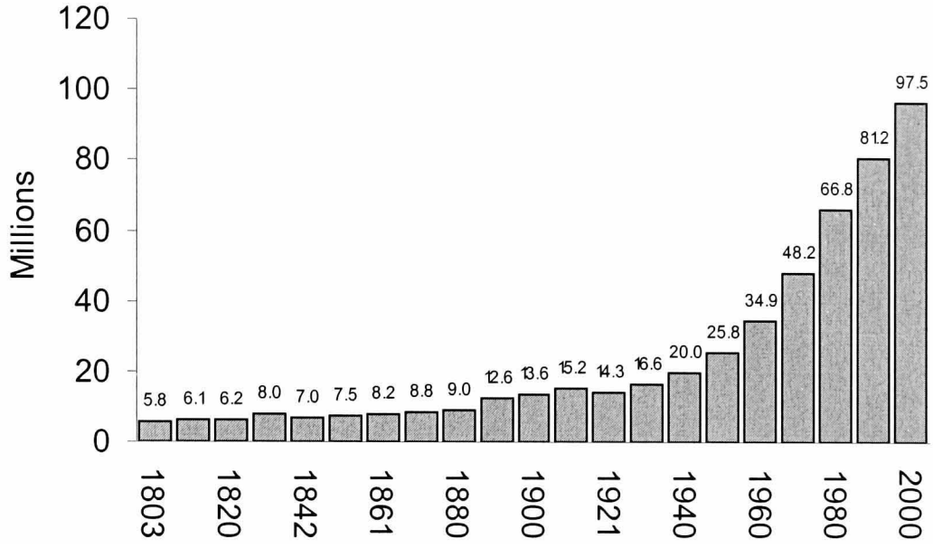
Source: CONAPO (2001a)

Table 3.2 Total fertility rate by some socio-economic characteristics

	1974	1996
No formal education	7.8	4.7
Incomplete primary school	7.0	3.7
Complete primary school	4.9	3.1
Secondary school and higher education	3.5	2.2
rural	7.4	3.5
urban	5.0	2.3

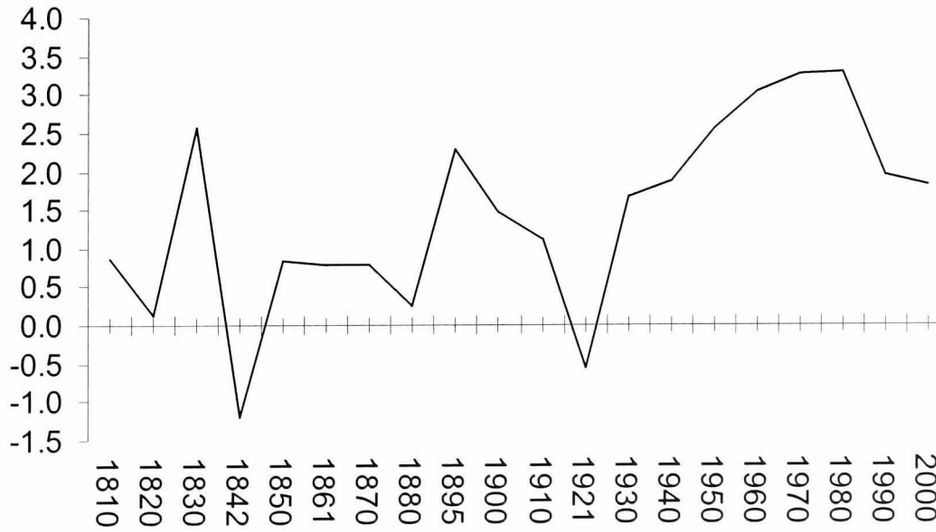
Source: CONAPO (2001a)

Figure 3.1 Population of Mexico



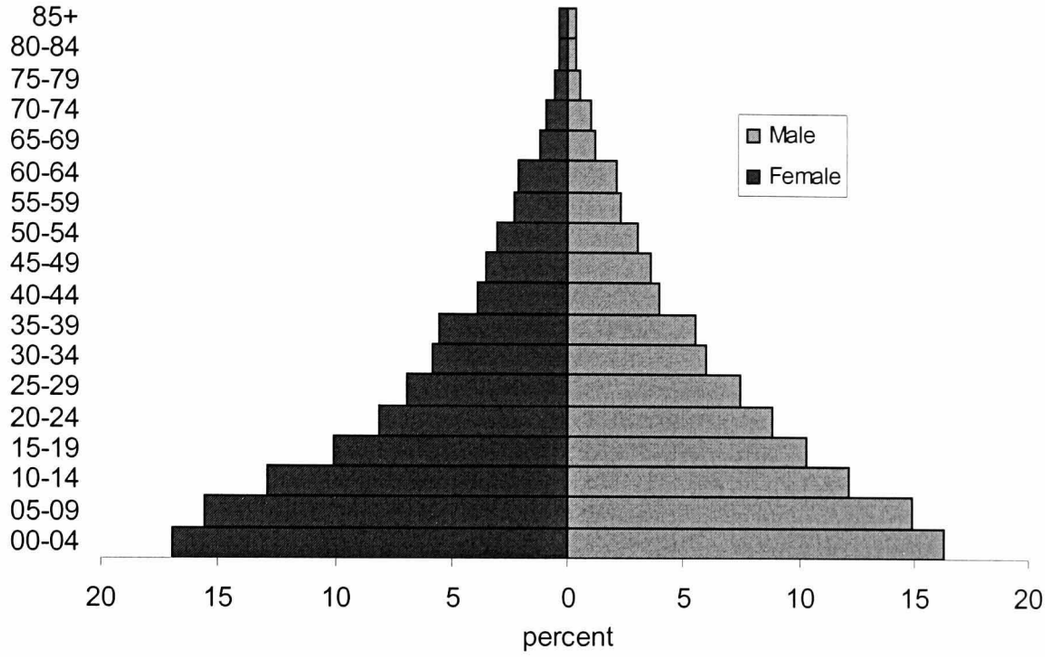
Source: INEGI (2000)

Figure 3.2 Annual Rate of Population Growth



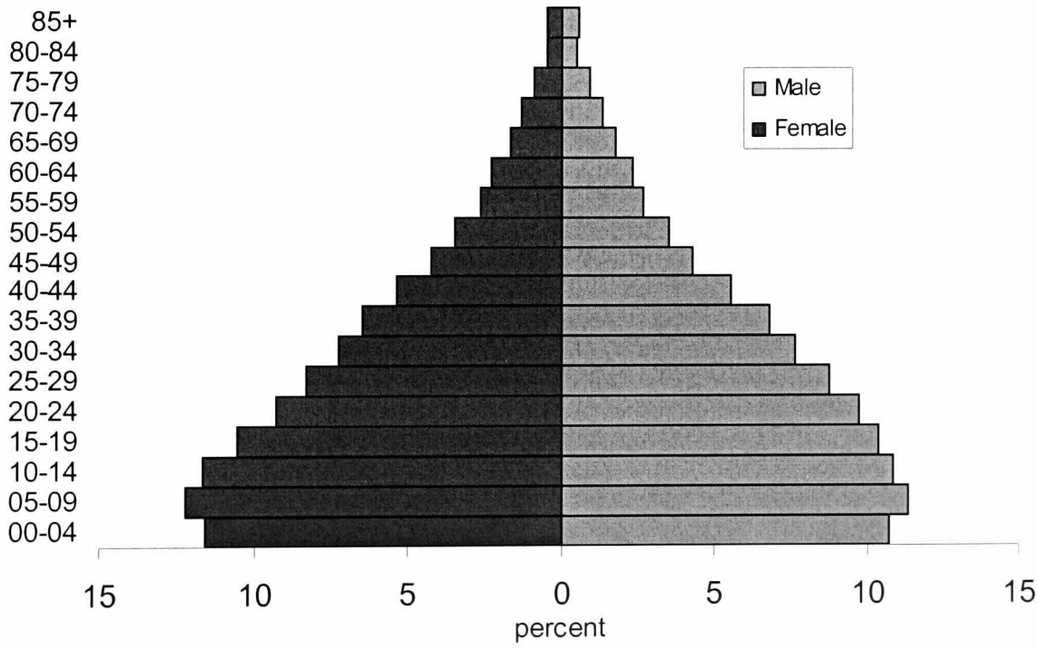
Source: INEGI (2000)

Figure 3.3a Population pyramid 1960



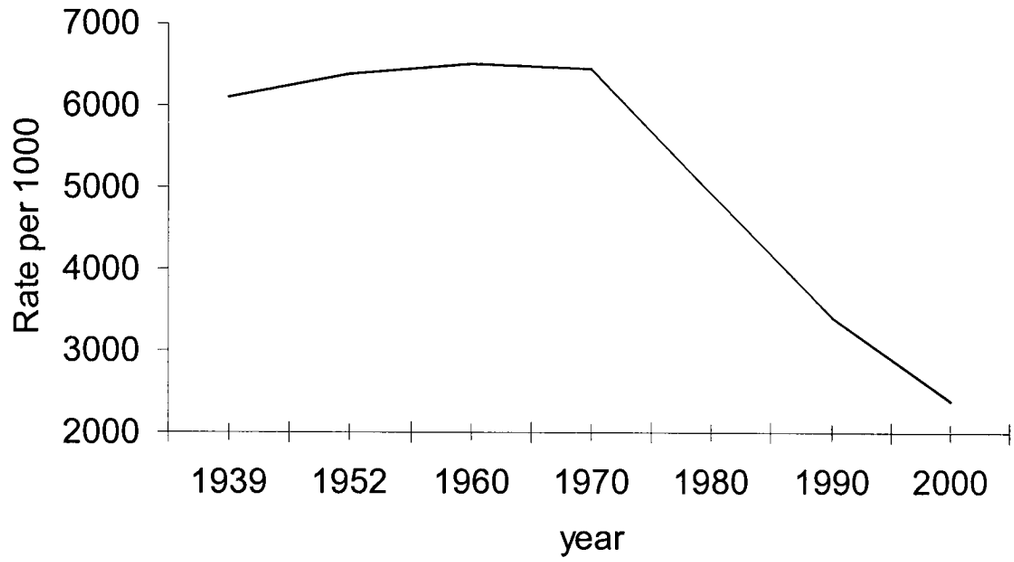
Source: INEGI (2000), CONAPO (2004)

Figure 3.3b Population pyramid 2000



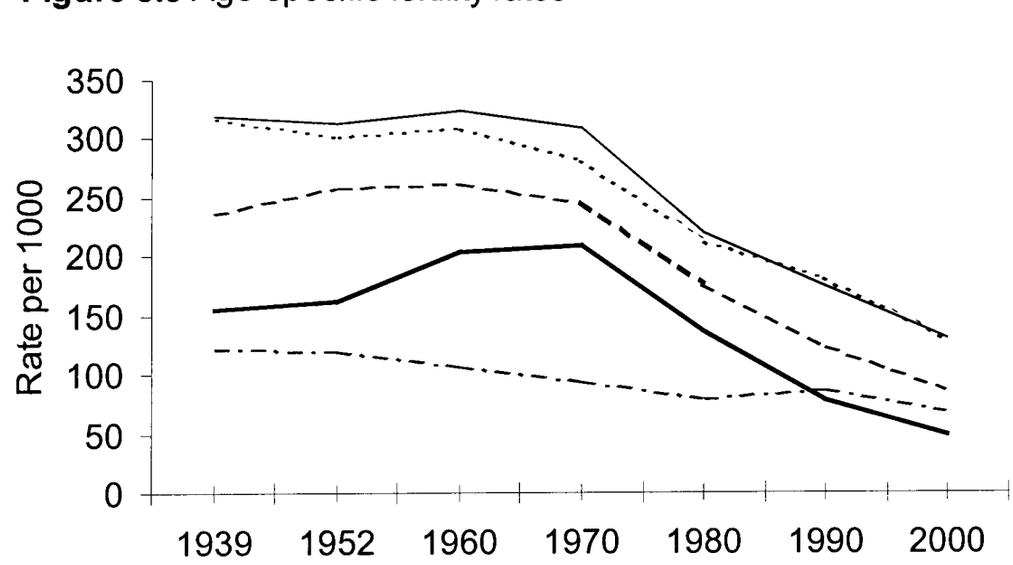
Source: INEGI (2000), CONAPO (2004)

Figure 3.4 Total Fertility Rate



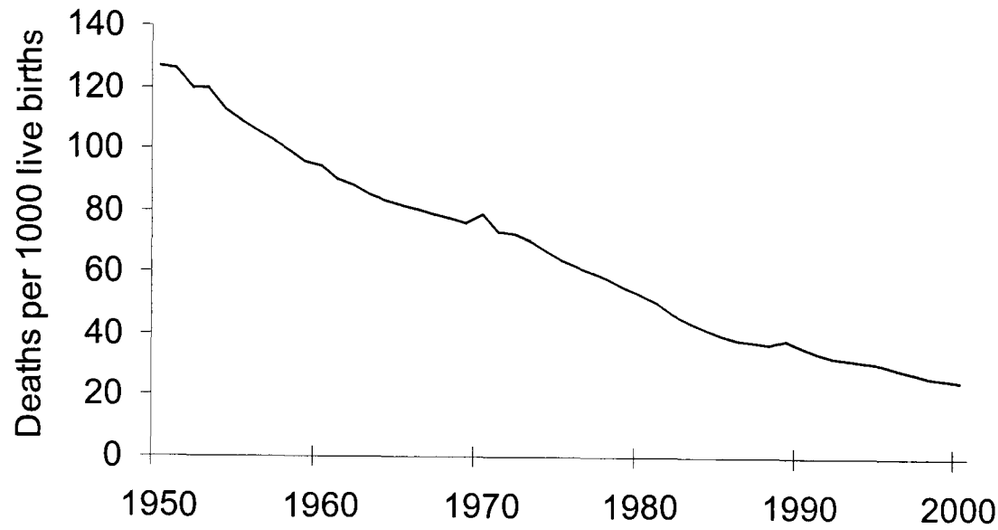
Source: INEGI (2000), CONAPO (2004)

Figure 3.5 Age-specific fertility rates



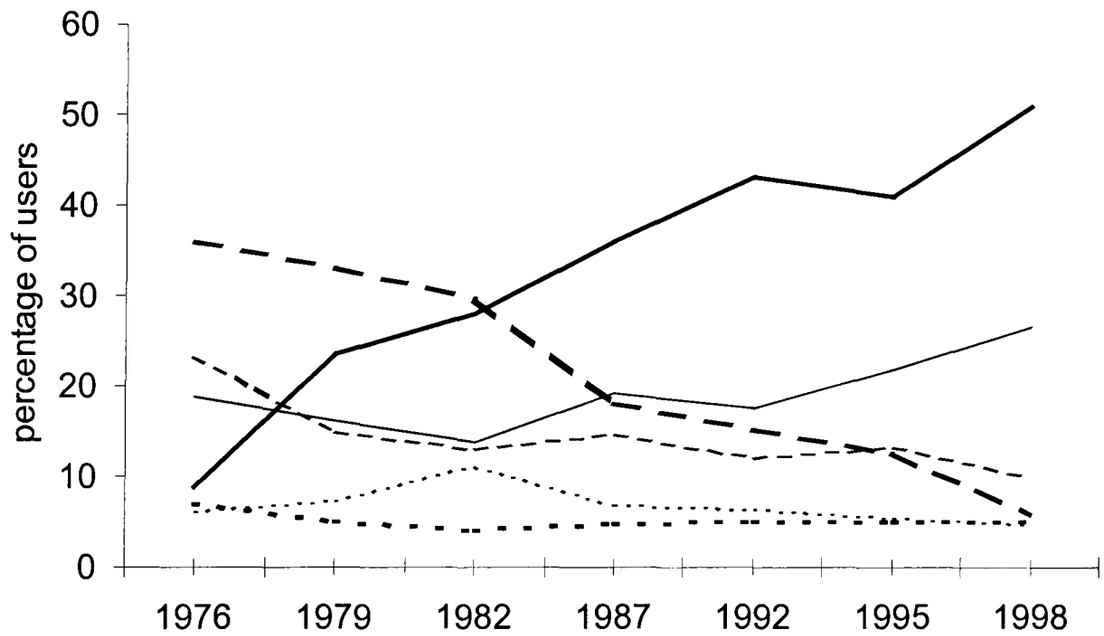
Source: INEGI (2000), CONAPO (2004)

Figure 3.6 Infant Mortality Rate



Source: CONAPO (2004)

Figure 3.7 Composition of the demand for contraceptives



Source: INEGI (2001b)

Chapter 4

Are young cohorts of women delaying first birth in Mexico?

4.1 Introduction

In the last twenty years female permanent sterilization (FPS) has become the most popular contraception method in Mexico. While in 1978 only eight per cent of users of contraceptives demanded FPS, in 1998 that figure was estimated to be as high as fifty per cent. During the same period of time the demand for contraceptive pills, condoms and other short-term contraceptives dropped in a sharp and consistent fashion (INEGI 2001c). These tendencies in the demand for contraceptives have raised concerns that the timing of children in Mexico might remain unchanged despite the reduction of fertility rates in the last few decades.[15] The impression is that women might be cutting off fertility at high parities

[15] See the discussion on page 17 (Chapter 1) and footnote 2 (page 19).

without modifying significantly their reproductive calendar – adopting no contraception until they reach their desired number of children and then using definite natal control to limit lifetime fertility (see Welti 1997, Gomez 1996a, Mier y Teran and Rabell 1990, Zavala de Cosio 1989)

The present work contributes to this discussion by reporting a study on the timing of first birth in Mexico. The main objective is to test whether or not young cohorts of Mexican women are effectively delaying first birth with respect to older generations. To accomplish this objective duration models are estimated using individual-level data from the National Survey of Demographic Dynamics 1997 (INEGI 1999).

The interest in studying the timing of first birth emerges mainly from the notion that the timing of childbirth, especially that of the first, might have significant implications for the total number of children a woman has over her whole life. Women who enter motherhood at early ages, it is argued, have commonly more children than women who enter motherhood at late stages of their fertile life (for more on this topic see, for instance, Gustafsson 2001, Melkersson and Rooth 2000, Chen and Morgan 1991, Heckman, Hotz, and Walker 1985, Bloom and Trussell 1984). First birth timing may also be a key factor determining the pace of subsequent fertility as suggested by Trussell and Menken (1978), Bumpass, Rindfuss and Janosik (1978), and discussed by Heckman, Hotz and Walker (1985). It is thus important, from the point of view of policy evaluation and design, to determine if the Mexican demographic transition has lead to first birth postponements.

Various aspects of the present work are worth noting. First, no previous study has discussed systematically – and with the use of advanced econometric

techniques – the issue of first birth postponement in Mexico. This is an initial contribution to a topic that, from the point of view of the author, deserves a close examination in the future. Second, and unlike the common practice in the analysis of transition data, the hazard function is estimated in a semi-parametrical fashion so that no a-priori restrictions on the form of the baseline hazard are imposed. Hence, the possibility of inducing bias due to misspecification of the functional form of the hazard is avoided. This is a critical point giving that misspecification of the baseline hazard generally produces inconsistent estimators and, as a consequence, might lead to wrong conclusions about sign and significance of generation (cohort) effects. Third, as suggested by Heckman and Singer (1975), unobserved individual heterogeneity is controlled for by estimating a discrete approximation of the distribution of unobservables. This non-parametric technique reduces potential sensitivity of results to prior distributional assumptions about unobservables that are otherwise required. In any case, an alternative hazard model with normal unobserved heterogeneity is considered and results contrasted. Both procedures control for the effect of potential differences across individuals in fecundity, tastes, skills, and other excluded variables that, if left unaccounted, may result in severe bias (for more details on this issue see Heckman and Singer 1984). Finally, the existence of individuals who remain childless until the end of their fertile life is explicitly allowed for by means of the inclusion of a mass point at minus infinity (left-end mass) in the distribution of unobservables. Taking account of this 'stayer-mover' problem avoids again bias due to misspecification of the hazard.

The remainder of this chapter is organised as follows. Section two presents the data and deals with the definition of all variables used in the analysis. Section

three discusses econometric issues and section four presents empirical results. Finally, section five concludes.

4.2 Data and variable definition

Data from the National Survey of Demographic Dynamics 1997 (ENADID from its acronym in Spanish) is used. The ENADID is a micro-data-set containing detailed information on Mexico's demographic dynamics that was created by the National Institute of Statistics, Geography and Informatics of Mexico (INEGI) between September and December of 1997. The INEGI is the main source of economic, demographic and geographic information in Mexico. The data set contains economic and demographic information on 88,022 Mexican women aged between 15 and 55 years. Detailed dates of birth of all women and their children were collected retrospectively. Excluding observations with missing information for education or/and with unreasonable dates of birth of either the reference woman or her first child, a total of 78,467 valid cases constitute the sample for the analysis.[16]

A strict definition of a waiting time to first birth would necessarily use age at menarche as the event by which women become at risk of entering motherhood. Most fertility surveys, however, do not collect such information. That is the case

[16] Cases where the given dates of birth of mothers and their children implied a negative duration interval were excluded. The analysis is done conditional on this selection.

of the ENADID. Following the approach of Newman and McCulloch (1984) and Heckman et. al. (1985), a common event to all women is used here as opening event, or starting point, for calculating duration intervals. In particular, it is supposed that women become at risk at age 12. Previous studies have used instead age at marriage as the `opening event' (see Newman and McCulloch 1984, Heckman, Hotz, and Walker 1985, Heckman and Walker 1990). Here we prefer not to follow that practise because age at marriage is potentially endogenous and its use as starting point might lead to sample selection bias (for more on this topic see Heckman 1979). The dependent variable, **duration**, is defined as number of years from age 12 to first birth if first birth occurred by the time of the survey, or number of years from age 12 to 1997 otherwise. A dummy variable, **fail**, indicates when a duration interval represents a completed spell (**fail** = 1) and when it is a censored observation (**fail** = 0). Mean duration is around 10 years and nearly 60% of all cases report a completed spell – i.e., a duration interval ended by a first birth (see Table 4.1, page 96). The dependent variable is built in terms of whole years in order to facilitate the estimation of discrete time duration models and avoid, by those means, imposing a-priori restrictions on the form of the baseline hazard.

Five explanatory variables are considered: cohort of age, education at age 12, religion, a proxy for ethnic group, and place of birth. Descriptive statistics are included in Table 4.1.

Using information on women's date of birth, three cohorts of age are defined: 1942-1962 (base group), 1963-1972 and 1973-1982. Three dummy variables indicating the cohort of age of the index woman are then generated (=1 if born in the corresponding calendar period and zero otherwise): **c4262**, **c6372**, **c7382**.

This generation split creates three age groups that contribute approximately the same number of individuals to the sample (about thirty per cent). Notice that the base group contains cases of women born over a twenty-year period while the other two cohorts contain cases of women born over two consecutive ten-year periods. Choosing a base age group that spans a relatively long period of time is largely innocuous as the definition of 'old' and 'young' generations is somehow arbitrary. In addition, given that the Mexican government initiated its new 'family planning' campaigns in 1973, it is intuitive (and interesting) to compare groups of women that were already at risk at the time of the policy innovation (i.e., generation 1942-1962) with groups of women that become at risk after the start of the new policy (generations 1963-1972 and 1973-1982).

Religion is controlled by the inclusion of a zero/one variable splitting Catholics (**catholic** = 1) from non-Catholics (**catholic** = 0). This two-cell classification of religious groups in Mexico seems sensible given that nearly 90% of Mexicans are Catholic and a further 7% admit to be Protestant. Place of birth is taken into account by defining a set of 31 geographic dummies - 32 federal entities compose Mexico, Mexico City (D.F) is left to be the base group.

To proxy ethnic group an indicator variable, **indspker**, that takes the value of one if the interviewed woman speaks an indigenous language and zero otherwise is included. Women who speak an indigenous language contribute in total six per cent of the sample. Notice that **Indspker** proxies broad ethnic group (indigenous/mixed) rather than specific socio-cultural community. Though indigenous and non-indigenous groups are widely heterogeneous socio-cultural entities in Mexico, this two-group ethnic splitting is justified in the sense that attitudes towards contraception, family size, and female labour market

participation are predominantly traditional across the different indigenous groups of Mexico (that is, against female job-market participation and modern contraception), while being predominantly liberal among mixed individuals. The use of **indspker** as a proxy variable for broad ethnic group supposes implicitly that indigenous individuals keep the ability of speaking their own language and truthfully declared so during the ENADID interview. There is, obviously, the possibility that some individuals who consider themselves indigenous may have lost their indigenous language. Further, some bilingual individuals could intentionally hide their language skills. As a consequence, **indspker** is potentially recorded with measurement error. However, if present, such an error is likely to be small and non-correlated with observed and unobserved variables that may affect fertility, including **indspker** itself.

Finally, variable **Edu12** controls for completed years of education at age 12. **Edu12** is a proxy variable and is included in the model as a general indicator of skills and human capital accumulated before the start of reproductive life. Given that Mexican law enforces six years of compulsory education stating at age six, **Edu12** is bounded between zero and six and, in general, is not subject to individual choice. Nearly 50% of individuals in the sample have less than six years of successfully completed years of education at age 12. Variation is induced because in rural and marginal urban zones there is a limited supply of education services and in some cases schools do not offer the six compulsory primary education grades. Long-term financial difficulties of the parental household may also result in a permanent dropout of their dependent children from primary education, especially in marginal zones where education law is not properly enforced. Hence, **Edu12** measures to some extent family background.

Another source of variation are course failure and temporal drop outs – though course repetition is rarely extended beyond age 12. All these childhood 'contextual' factors induce variation on education at age 12 in Mexico and reflect in some extent family traits. Clearly, though children have little influence on their early education there is still the possibility that **Edu12** may be endogenous. However, as it is usual in most data sets, no valid instruments for education are available in the ENADID. Thus, **Edu12** is treated as an exogenous variable and the reader should interpret results with the due care.

Due to the lack of detailed information **Edu12** is built under a set of assumptions. First, as enforced by the federal law, it is supposed that all children initiate their primary education at age 6. Second, it is supposed that all children attend continuously school until the date of their definite dropout. Finally, it is assumed that none fails an attended course. These assumptions guarantee that completed years of education at age 12 may be calculated on the basis of information on women's date of birth and their current completed years of education – data indeed available in the ENADID. In practice, obviously, children may start education after age 6, dropout temporally, and/or repeat some courses. **Edu12** contains thus some potential measurement error. This error, however, is likely to be small and, if present, it is supposed to be random and uncorrelated with all observed and unobserved explanatory variables entering the hazard function (including **Edu12** itself). This is, once again, a strong assumption and results should be properly qualified.

Women's income, work and marriage status are not considered here as explanatory variables. These variables are likely to be endogenous in a fertility model and their use might lead to simultaneous equation bias and invalid

inference. No information on women's family background or other valid instruments for income, work and marriage status is available. Hence, estimation of a system of simultaneous equations is infeasible and the researcher should focus instead in obtaining reduced form duration models. Clearly, excluding variables such as income, work, and marriage status could, by itself, lead to biased estimators. However, consistency can be achieved if explicit account for the presence of unobserved heterogeneity is considered.

4.3 Econometric Issues

Discrete time (grouped) duration models are used. This semi-parametric econometric strategy does not impose the a-priori restrictions on the form of the baseline hazard that other fully parametric methods do. Consequently, a common source of misspecification bias in the analysis of transition data is avoided. Duration is defined in terms of whole years. That is, number of years between age 12 and first birth if first birth is observed by the date of the ENADID interview, and number of years between age 12 and 1997 if the spell remained censored.

4.3.1 The Model

As a departing point we suppose that the underlying continuous-time hazard for individual i, $\theta_i(t)$, belongs to the Proportional Hazards family

$$\theta(t) = \lambda(t)\exp[\mathbf{x}_i'\boldsymbol{\beta}] \tag{1}$$

where $\mathbf{x_i}$ represents a kx1 vector of characteristics for individual i (including the constant term), $\boldsymbol{\beta}$ represents a kx1 vector of coefficients to be estimated, and $\lambda(.)$ represents the baseline hazard. Notice that any negative (positive) term in the vector $\boldsymbol{\beta}$ implies a reduction (increase) of the hazard of observing a failure and, consequently, induces a longer (shorter) mean duration. Individual characteristics might be observed at various points during the duration spell so that $\mathbf{x_i}$ might contain time-varying variables. Since duration is coded in terms of completed years, the probability of observing a failure at time t given that at time t-1 first birth had not yet occurred may be written as

$$\begin{aligned} h_{it} &= h_i(t \mid \mathbf{x_i}) = \Pr[T_i < t \mid T_i \geq t-1] \\ &= 1 - \exp\left[- \int_{t-1}^{t} \theta_i(s)\, ds \right] \\ &= 1 - \exp\left[\int_{t-1}^{t} \lambda(s) \exp(\mathbf{x_i}'\boldsymbol{\beta}) ds \right]. \end{aligned} \tag{2}$$

If the vector of individual characteristics remains unchanged between time t-1 and time t, equation (2) reduces to

$$h_{it} = 1 - \exp\{-\exp[\mathbf{x_i}'\boldsymbol{\beta} + \gamma(t)]\} \tag{3}$$

where,

$$\gamma(t) = \ln\left[\int_{t-1}^{t} \lambda(s) ds \right].$$

From equation (3) the reader may conclude that the discrete time hazard takes an extreme value form. Notice that no restrictions on the form of the function $\gamma(t)$

have been imposed. This feature creates the opportunity of estimating $\gamma(t)$ using semi-parametric techniques.

Model (3) can be thought of as a sequence of non-identical Bernoulli trials where each individual contributes one observation per survived period (see Kiefer 1988). To see this, let d_i be recorded duration and c_i be a dummy indicating whether d_i is a completed ($c_i = 1$) or censored ($c_i = 0$) spell. Next, define a new variable

$$w_{it} = \begin{cases} 1 & if \quad d_i = t \quad and \quad c_i = 1 \\ 0 & \qquad\qquad otherwise, \end{cases}$$

so that if, for instance, we observe $d_i = 10$ and $c_i = 1$ individual i would contribute 10 observation to the sample. Namely, $w_{i1} = w_{i2} = \ldots = w_{i9} = 0$ and $w_{i10} = 1$. Using these definitions a binary choice model for w_{it} may be written,

$$\begin{aligned} \Pr[w_{it} = 0 \mid \mathbf{x_i}] &= \prod_{k=1}^{t} [1 - h_{ik}] \\ \Pr[w_{it} = 1 \mid \mathbf{x_i}] &= h_{it} \prod_{k=1}^{t-1} [1 - h_{ik}], \end{aligned} \tag{4}$$

and write the overall contribution of the i-th individual to the likelihood as

$$L_i = \prod_{t=1}^{d_i} \left\{ [h_{it}]^{w_{it}} [1 - h_{it}]^{1-w_{it}} \right\}. \tag{5}$$

Interpreting the discrete time duration model as a sequence of binary choice regressions creates the opportunity of relaxing the proportional hazards

assumption and use alternative binary choice models such as Logit and Probit. Logistic and Normal hazards exhibit a symmetric distribution and constitute a valuable contrast to the skewed Extreme Value distribution. This sort of model has been estimated in the field of labour economics by various authors including Meyer (1990), Narendranathan and Stewart (1993), Sueyoshi (1995), and Arulampalam and Stewart (1995).

4.3.2 Unobserved Heterogeneity

To accommodate unobserved heterogeneity, an unobserved random term, v_i, is introduced into equation (1). This random term captures differences in the hazard induced by heterogeneity in omitted characteristics such as fecundity and/or skills. Unobserved heterogeneity is supposed to be orthogonal to the vector of explanatory variables $\mathbf{x_i}$. Identification is secured if at least one continuous variable is included into the design matrix $\mathbf{x_i}$ (for more details on this issue see Heckman and Singer 1984). Conditional on unobservables the hazard in equation (3) becomes

$$h_{it}(v_i) = h_i(t \mid \mathbf{x_i}, v_i) = 1 - \exp\{-\exp[\mathbf{x_i}'\boldsymbol{\beta} + \gamma(t) + v_i]\}. \tag{6}$$

On the basis of equation (6) the likelihood function might be written in terms of the conditional hazard $h_{it}(v_i)$ and the probability density function of v_i, $f(v_i)$,

$$L_i = \int \left\{ \prod_{t=1}^{d_i} [h_{it}(v_i)]^{w_{it}} [1 - h_{it}(v_i)]^{1-w_{it}} \right\} f(v_i)\, dv_i. \tag{7}$$

The model is closed once a functional form for the mixing density $f(v_i)$ is chosen. Normal and Gamma densities are a common choice. Alternatively, as suggested by Heckman and Singer (1984), a non-parametric maximum likelihood estimator (NPMLE) can be used to estimate an empirical approximation to $f(v_i)$. This method is based on results establishing that, given a functional form of the hazard, a NPMLE can be approximated by a finite discrete mixture. The result implies that equation (7) might be re-expressed as,

$$L_i = \prod_{t=1}^{d_i} \left\{ \sum_{k=1}^{m} [h_{it}(v_k)]^{w_{it}} [1 - h_{it}(v_k)]^{1-w_{it}} p_k \right\}, \tag{8}$$

where m, v_k, and p_k are to be estimated along with the parameters of the hazard. The log-likelihood is maximised sequentially adding mass points, v_k, until the probability associated to the last point, p_m, is close to zero and no significant improvements in the log-likelihood are detected. This methodology is used in the present work. For comparison proposes, models with Normal unobserved heterogeneity are also considered.

4.3.3 The mover-stayer problem

It remains the issue that some individuals may be sterile or dislike children to such extend that they reach the end of fertile life without offspring. Those individuals might never be at risk of entering parenthood and, consequently, report extremely long duration spells. In order to account explicitly for the presence of lifetime-childless individuals the model is extended by means of

introducing a mass point at minus infinity (referred simply as 'final mass point' in future discussion). This feature allows for a 'mover-stayer' framework.

$$L^*_i = \frac{\delta}{1+\delta}\left\{ \prod_{t=1}^{d_i}(1-w_{it}) \right\} + \frac{L_i}{1+\delta}, \qquad (9)$$

with L_i as in equation (7). Notice that L_i^* is built without reference to the method used for evaluating L_i. One can, for instance, specify $f(v_i)$ as Normal and approximate the integral in (7) with Gaussian-Hermite quadrature. Alternatively, the non-parametric method of Heckman and Singer might be used. The basic idea behind model (9) is that a large negative value of the unobservable v_i induces a extremely long duration interval and, consequently, increases the likelihood of observing a censored observation.

Model (9) is equivalent to the Split-Population model of Schmidt and Witte (1989). The unique difference is that Schmidt and Witte divide the population into two groups, permanent childless and eventually mothers, and L_i represents the hazard of the group of people that eventually enter parenthood. $\delta/(1+\delta)$ is then the probability of being in the permanent childless (sterile) group. In model (9) population is not divided into fertile and sterile groups but permanent childlessness results from the realisation of a very large negative v_i. In addition, $f(v_i)$ is thought to have a thick left tail that is represented as a final attraction point at minus infinity – that in practice implies the existence of a group of lifetime childless individuals. In such a case, $\delta/(1+\delta)$ represents the probability mass attached to the final mass point (i.e., the probability mass at $v_i = -\infty$). In

practice Split-Population and final mass point are two equivalent forms of modelling the same 'mover-stayer' phenomenon (see appendix A).

4.3.4 Econometric Specification

Moving to issues on econometric specification it is clear that potential confusion between duration dependence and generation (cohort) effects should be properly avoided. While the former effect describes fluctuations of the hazard over duration time from the onset of a spell, the later effect describes fluctuations of the hazard due to changes in the calendar time of entry. It is then important to allow enough flexibility in the baseline hazard so that all fluctuations that the failure rate may have over duration time are correctly described as duration dependence. Otherwise, misspecification of the hazard function might lead to misleading inference. A non-parametric approach is therefore implemented, consisting of estimating an empirical approximation to,

$$\gamma(t) = \ln\left[\int_{-1}^{t} \lambda(s)ds\right]$$

at each possible – discrete – duration length. The technique involves generating a set of binary indicator variables, dum_t, for each duration length, t, where at least a single failure is observed. If the i-th individual is reported to be at risk at time $t=10$, then the corresponding dummy variable dum_{10} will be set to one in the 10th entry contributed by her. All other dum_t with $t \neq 10$ are correspondingly set to zero in the 10th entry contributed by the i-th individual. Clearly, one category should be kept as the reference group (here the initial 0-3 segment) so that a constant

might be included in the hazard function. Once the interval-specific dummies are generated, it can be added into the model. Coefficients on these dummies provide a non-parametric estimator of $\gamma(t)$. Notice that the inclusion of the continuous variable Edu12 guarantees that the model is properly identified (see Heckman and Singer 1984).

Various model specifications are considered. A benchmark discrete Extreme Value (EV) hazard with no unobserved heterogeneity is first estimated. Since neglecting unobserved heterogeneity may lead to serious bias, unobserved heterogeneity is then explicitly allowed and a non-parametric maximum likelihood estimator (NPMLE) implemented. Finally, the model is extended to account for the presence of lifetime childless individuals by means of including a mass point at minus infinity (final mass point) in the empirical distribution of the unobservables.

Once an EV hazard model has been estimated, a Probit hazard specification is implemented in the various versions considered for the EV case. As it is discussed in section 3, the Probit specification relaxes the assumption of proportional hazards imposed by the benchmark EV case. Therefore, by comparing partial effects from EV and Probit hazards the econometrician may assess the sensibility of the results to the proportional hazards assumption.

A final set of estimates will be obtained from models that impose prior assumptions on the distribution of the unobservables. Namely, that the random term v_i is normally distributed. Imposing prior assumptions on $f(v_i)$ is a wide-spread practice in the analysis of duration data, and a Central Limit Theorem argument would suggest that a Normal specification is a natural choice (see Narendranathan and Stewart 1993, Arulampalam and Stewart 1995). Producing

estimates from models with normal unobserved heterogeneity seems to be an interesting exercise. On one hand, a set of alternative results might be obtained so that findings from models with parametric and non-parametric unobserved heterogeneity could be contrasted and qualified. On the other hand, a general assessment of the advantages of the NPMLE method over other standard econometric techniques could be performed in the context of the ENADID data.

4.3.5 Model Selection

Given that the alternative model specifications considered in the present study are non-nested, various information criteria are used to compare them. Namely, the Akaike information criterion (AIC) and its consistent version (CAIC) are used to compare alternative models. Model selection is also performed on the basis of a Lamer-Schwarz metric (LS). If AIC is employed for model selection, a best fitting model will be the one that achieves the minimum AIC against all other considered alternatives. CAIC and LS work in a similar fashion. All these information criteria penalize for the loss of degrees of freedom when an additional parameter is included and are widely used for model selection in many fields of applied work. Formally the statistics are defined as,

$$
\begin{aligned}
AIC &= -2\ln(L) + 2k \\
CAIC &= -2\ln(L) + k\{(\ln(n) + 1\} \\
LS &= \ln(L) - k\frac{\ln(n)}{2}.
\end{aligned}
$$

4.4 Empirical Results

The analysis begins performing a preliminary examination of the data using non-parametric methods. Figure 4.1 (page 100) presents smoothed kernel estimators of the hazard for the three cohorts of age considered here: **c4262**, **c6372** and **c7382**. An Epanechnikov kernel with varying bandwidth is used in such a way that the bandwidth is allowed to increase as the number of cases at risk decreases. Figure 4.1 suggests that the hazard function exhibits an inversed-U form. This functional form for the hazard conforms the intuition that the cost of first pregnancy is high at the beginning of fertile life, decreases with the passing of time, and then, at a certain point, starts to rise again as the end of fertile life is approximated. Comparing curves in Figure 4.1 the reader may conclude that the hazard of having the first child at the start of the spell is slightly higher for women in cohort **c4262** than for women in cohort **c6372** and **c7382**. Moreover, excluding the peak point, Figure 4.1 suggests that the hazard for young cohorts remains well below the hazard for the eldest cohort. It is just at the end that women in the young cohorts have a relatively high risk of failure in relation to the experience of women in the eldest cohort – here evidence from cohort **c4262** and **c6372** is stressed, as long durations for the youngest cohort are not observed. All these observations suggest then that women in young cohorts do tend to delay first birth for a longer period of time than women in the eldest cohort. Conclusive inference, however, can only be done after observed and unobserved heterogeneity have been properly controlled for.

4.4.2 Model Selection

Tables 4.2 and 4.3 (pages 97) present regression results for EV and Probit hazards respectively. Three columns are presented. Column (1) reports estimated coefficients and standard errors from a benchmark model that neglects unobserved heterogeneity. Then column (2) reports coefficients obtained once unobserved heterogeneity has been accounted for by the inclusion, and empirical estimation, of two mass points for approximating the distribution of unobservables. Finally, column (3) reports results once a final mass point (at minus infinity) is added into the model. Including an additional mass point in models (2) and (3) of Tables 4.2 and 4.3 did not produce significant improvements in the log-likelihood and the vector of coefficients on explanatory variables remained unchanged.

Notice that, except for the probability mass attached to the final mass, coefficients in Tables 4.2 and 4.3 are not directly comparable. It is remarkable though that all coefficients on explanatory variables have the same sign across the different versions of either the EV or Probit hazard. Similarly, signs on coefficients of all covariates are similar across Table 4.2 and 4.3. Moreover, excluding the case of **c6372** in column (1), all coefficients are different from zero at all conventional significance levels.

Contrasting estimates in columns (2) and (3) within Table 4.2 (or alternatively, within Table 4.3) the reader can conclude that controlling for unobserved heterogeneity improves significantly the fit of the model. All information criteria indicate that a model with two masses performs better than the benchmark model – i.e., the model that neglects unobserved heterogeneity.

These conclusions are supported by the fact that the estimated location one point (mass1) and the probability attached to it, Pr(mass1), are highly significant in columns (2) and (3) of Table 4.2 (Table 4.3). Hence, unobserved heterogeneity is present and significant.

Discrimination on the basis of AIC, CAIC and LS suggest that models with two plus final masses are preferred over models with only two masses (see bottom of columns 2 and 3 in Tables 4.2 and 4.3, page 97). In either case, EV or Probit hazard, the probability attached to the final mass is around 0.12. This finding suggest then that a 12% of Mexican women are expected to remain childless for her entire lifetime.

A Probit hazard seems to fit better the data. In fact, comparing column results in Tables 4.2 and 4.3 the reader can find that a Probit hazard always attains lower log-likelihood values than the EV hazard specification – in the case of models in column (3), for instance, the actual figures are -158,208.5 and -158,294.1 respectively. Model selection on the basis of AIC, CAIC and LS statistics would conclude as well that a Probit hazard should be preferred over the Extreme Value. Relaxing the proportional hazard assumption seems thus to be the best thing to do in the present context.

Table 4.4 presents results from a random effects Probit hazard with Normal heterogeneity. Sixteen Gauss-Hermite quadrature points were used to approximate the integral in equation (7) of section 3. Including additional quadrature points did not result in either a significant improvement of the log-likelihood or significant changes in the parameters of the estimated models. Contrasting figures from Table 4.3 and 4.4 it is possible to conclude that imposing a Normal density for the distribution of unobservables is of little

consequence for the sign, size, and significance of the coefficients on explanatory variables. Moreover, and in line with previous findings, unobserved heterogeneity is detected to be present as the estimate for the proportion of the total variance contributed by the random effect v_i, rho, is found to be significant in either model (2) or (3). In fact, a boundary-value likelihood ratio test for rho=0 easily rejects the null hypothesis at any conventional significance level in both cases (with a chi-square[01] statistic of 553 and 1666, respectively). As before, selection on the basis of AIC, CAIC and LS in Table 4.4 favours the model that includes a final mass point in the distribution of the unobservables.

Results from Table 4.3 and 4.4 suggest that as far as likelihood, AIC, CAIC and LS concerns a Probit hazard with non-parametric heterogeneity performs better than a Probit hazard with Normal heterogeneity. This observation is valid for all alternative specifications being considered. Therefore, evidence indicates that a Probit hazard with non-parametric heterogeneity and a final mass point is the best fitting model in the present context – that is to say, the model reported in column 3 of Table 4.3.

4.4.3 Effect of Explanatory variables

Concentrating attention on the best fitting model – Probit hazard with two plus final unobserved non-parametric heterogeneity masses – results suggest that Catholic individuals have a significant lower risk of failure than non-Catholic individuals (and consequently a longer mean duration). In fact, the probability of failure at mean duration for a Catholic individual is estimated to be on average 1.57 points lower that the corresponding probability for a non-Catholic

individual. That is, the average partial effect (APE) on Catholic at mean duration is approximately -0.0157 (see Table 4.5, APEs are obtained calculating marginal effects at each empirical location of v_i and then averaging over the distribution of the random term v_i).[17]

At first sight this result might seem counter-intuitive given that the Catholic Church traditionally opposes the adoption of contraceptives as a way to regulate fertility, an attitude that is generally thought to increase the odds of unwanted pregnancy among catholic individuals. However, the result is better understood if the reader considers that, besides being reluctant to adopt modern birth control, the Catholic Church also discourages sexual activity out-of-wedlock, an attitude that in many cases leads to the delay of first sexual intercourse. Hence, the net effect of Catholicism on the hazard may well be negative rather than positive. In the case of Mexico, where Catholic individuals represent nearly 90% of the population and a further 9% are either Protestant or Atheist (information from INEGI 1999), it is intuitive that non-Catholic individuals may initiate sexual activity earlier than Catholic individuals as Protestants and Atheists are traditionally more liberal about such issues. Therefore, though Catholic youngsters may be highly reluctant to adopt modern contraception, they still may have lower sexual activity at early ages than their non-Catholic peers.

---

[17] Average Partial effects of continuous variables are calculated at the mean value, mean duration, and setting all dummies to zero. Average Partial effects for discrete variables are calculated by the difference measure at the mean of all continuous variables, mean duration, and setting all other dummy variables to zero.

In contrast, being an Indian language speaker implies a significant increased hazard of failure at mean duration of almost three percent, for the average partial effect on **indspker** at mean duration is 0.0288. This effect is the synthesis of a number of factors. Most Mexican Indians live in rural zones on settlements of less than 1,500 inhabitants that are far from the main regional cities and where the main economic activity is agriculture. In their majority, indigenous individuals have a small extension of land and produce mainly for self-consumption. As a general rule, these indigenous zones have very limited supply of education and health services. Hence, commonly, Indian individuals do not study beyond the basic instruction available in their localities, which in many cases is below six grades. Post basic education is even more limited in rural and indigenous zones despite the fact that from 1996 it became compulsory for all Mexican children. This is one of the main reasons that keep average education among indigenous individuals at 5.38 years, well below the 8.13 years corresponding to Mexico as a whole.

Limited health services are associated with a relative high rate of infant mortality among Indians, which in 2000 reached 38.5 deaths per 1,000 births, above the national rate of 17 deaths per 1,000 births (CONAPO 2002a). Culture might also play some role in the relatively high hazard of failure among Mexican Indians, especially if tradition keep women outside the labour market (that is, away from paid jobs) and reduce their bargaining power inside the household. From the present analysis, however, there is no way of inferring the relative importance of these cultural factors as determinants of early motherhood.

As economic intuition would suggest, education at age 12 is found to reduce the hazard of a first pregnancy (or equivalently, to increase mean duration). In

fact, the average partial effect at mean duration on **Edu12** calculated on the basis of results from the Probit hazard with two plus final masses attains a value of – 0.0245. In other words, an extra year of schooling at age 12 reduces the likelihood of failure at mean duration by approximately 2.5 points. These results are consistent with previous findings reported by Newman and McCulloch (1984) for the case of Costa Rica and by Ermisch and Owaga (1994) for the case of Japan. Further, the empirical results agree with the predictions of the economic theory of fertility behaviour discussed in Chapter 2 (see for instance, Happel, Hill, and Low 1984, Cigno and Ermisch 1989, Hotz, Klerman, and Willis 1997). It is important to note here that variation in **Edu12** is mainly induced by course failure, economic difficulties in the parental household during childhood and a lack of non-compliance of federal laws regarding compulsory education in Mexico. Hence, **Edu12** measures in good extent family background. The reader should keep these facts in mind and interpret the effects on **Edu12** with the due qualifications – as it is a coefficient pooling the effects of family background and human capital at the onset of the family planning period.

An inspection of column (3) of Table 4.3 reveals that women in cohort 1963-1972 have a higher risk of failure than women in the base age group (1942-1962). This is, the positive coefficient on **c6372** is significant at all standard levels. The implied average partial effect at mean duration is though just about 0.0076, indicating a slight reduction on mean duration. Hence, according to this evidence women born between year 1963 and year 1972 seem to have accelerated the pace of their entrance to motherhood in relation to women in the base group. The effect, however, is rather small.

Results indicate that women in the last group of age (1963-1982) effectively delayed first birth relative to the experience of women in the base age group. This finding is clearly witnessed by the negative and significant coefficient on **c7382** in column 3. The implied average partial effect at mean duration attains a value of -0.0131, indicating a reduction on the risk of failure of around 1.3 points (and thus and increase on mean duration) in reference to the 1942-1962 base group. Though the delay effect is rather small, it is highly significant. Hence, evidence seems to support the idea that the fall of total fertility rate of the last four decades in Mexico has been associated not only to reductions on lifetime fertility but also to delays in first birth timing. In the light of these findings the conjecture that Mexican women might be cutting off fertility at high parities without modifying the reproductive calendar of their first child seems not to be supported by the data (contradicting then the ideas of Welti 1997, Gomez 1996b, Mier y Teran and Rabell 1990, Zavala de Cosio 1989).

Table 4.5 (page 98) presents a comparison of the average partial effects (APE) obtained from EV and Probit hazard regressions with non-parametric heterogeneity and final mass point. For completeness, APE calculated from a Probit hazard with normal heterogeneity and final mass point are also reported. From the figures in Table 4.5 it is possible to conclude that the APE figures from the three different models are largely consistent. Thus, there is confidence that the findings are robust to changes in prior assumptions about the functional form of the hazard and about prior assumptions about the distribution of unobservables.

4.4.4 Estimated average hazard

Figure 4.2 (page 100) presents estimates for the average hazard based on results from EV and Probit with non-parametric heterogeneity and final mass point. Results from a Probit hazard with Normal heterogeneity and final mass point are also displayed. Likewise APE, for any potential realisation of the heterogeneity term v the conditional hazard of failure at each duration time, $h_t(v)$, was calculated on the basis of the coefficients on the interval-specific dummies d_t. Next, the average hazard was obtained taking the expected value of $h_t(v)$ over the distribution of the random term v, $f(v)$. This procedure delivers a series of points – one for each discrete duration interval where at least a single failure was reported – that are then plotted in Figure 4.2. Graphs in Figure 4.2 represent a non-parametric estimator of the average hazard. All these calculations were performed for a typical individual. The typical individual was found to be Catholic and had 5.35 years of education at age 12.

Duration dependence in all considered specifications follows the same pattern. First, at the beginning of the spell, the hazard rate exhibits positive duration dependence. Then, with the passing of time, the hazard becomes flat before it starts to exhibit negative duration dependence. Hence, the hazard has an inverted-U form, just as suggested by the kernel estimates in Figure 4.1.

From Figure 4.2 the reader may conclude that the average hazards from the EV and Probit models with non-parametric and final mass point heterogeneity look fairly similar (being almost a straight line) for duration intervals shorter than 20 years after age 12 (age 32). For duration spells longer than 20 years important differences are however detected. While the EV hazard peaks around duration

20, a Probit hazard peaks three years later – at duration 23. Moreover, the Probit hazard for durations longer than 20 is found to be systematically higher than the EV hazard. In other words, the Probit hazard predicts a larger proportion of failures at the end of fertile life than an EV hazard would predict. It is likely that this feature is the reason of why a Probit hazard was found to fit better the data than the EV hazard. This evidence suggests then that the proportional hazards assumption is restrictive in the present context. Both EV and Probit hazard detect an acute peak in the hazard function at duration 27; that is, at age 39. This final peak is probably associated to reductions on women's contraception effort at the very end of fertile life, maybe because the risk of a pregnancy is wrongly sub-estimated with the detection of temporary losses of fecundity. Alternatively, the final peak in the hazard might indicate the existence of a group of women who wait until the last periods of their fertile life to do decisive efforts to enter motherhood.

Figure 4.2 presents as well estimates of the average hazard calculated on the basis of a Probit model with Normal unobserved heterogeneity and final mass point. It is clear from the curves in Figure 4.2 that the assumption of normal heterogeneity results in a hazard function that is smoother in relation to the non-parametric heterogeneity version of the model. This result is rather expected as the Normal heterogeneity version of the model supposes that the support of the distribution of v is a set with infinite points. In this context the advantages of using the non-parametric maximum likelihood method of estimation suggested by Heckman and Singer (1984) become apparent, as the Normal heterogeneity version of the model seems to over-estimate the hazard at short durations and under-estimate it at long durations.

4.4.5 Results by Cohort of Birth

Up to now the analysis has maintained the assumption that entry at different calendar times lead to parallel shifts in the hazard function. Further, explanatory variables have been constrained to affect equally the hazard across different age groups. Finally, a unique distribution for the unobservables has been estimated for individuals belonging to different age cohorts. These assumptions are stringent and unlikely to be met in practice. Hence, important insights could be gained by estimating a separate hazard function for each cohort. Clearly, splitting the data into cohort sub-samples removes all aforementioned constraints. The present section discusses results from such an exercise.

Following a strategy similar to that used in previous sections, various specifications for hazard and unobserved heterogeneity were used for each cohort. Model selection on the basis of AIC, CAIC and LS indicate that a Probit hazard with non-parametric heterogeneity is the best fitting model for cohort 1942-62 and 1963-72.[18] In both cases a model with two plus final mass points performed better than any other alternative. For cohort 1973-1982 a Probit hazard with exclusively one mass point, the final mass, was best supported by the data.

Table 4.6 (page 99) reports average partial effects (APE) calculated on the basis of the best fitting hazard model for each age group. It appears first that,

[18] Extreme Value and Probit hazards with non-parametric unobserved individual heterogeneity were estimated for each cohort. EV and Probit hazard models with normal heterogeneity were also obtained.

though limited, variation in the APE across the different cohorts is non-negligible given the size of the calculated average partial effects. **Catholic**, for instance, is associated to a 1.63 point reduction in the likelihood of failure for women born between 1942 and 1962. In comparison, **Catholic** in the 1963-1972 group leads to a decreased risk of failure of around 1.80 points. This implies that the Catholic/Non-Catholic relative risk of failure in the later group is 1.43 lower than in the former group.[19] If cohorts 1942-62 and 1973-82 are contrasted instead it is the younger and no the older generation who bear an increased Catholic/Non-Catholic relative risk of failure of 1.35. A similar story describes variation of the APE of **indspker** and **Edu12** across age groups. In the case of **indspker**, however, there is a clear pattern indicating that the passing of calendar time has led to increments in the risk of failure of indigenous language speakers in relation to non-indigenous language speakers.

An important point to note form Table 4.6 is the fact that the coefficients on **Edu12** have a rather limited variation over the different cohorts, though there is a greater variation in the years of education for older cohorts. It is difficult to offer an intuitive explanation of this result. Notice, however, that it is also for old cohorts that duration is likely to have a larger variation. Clearly, the relative stability in the effect of **Edu12** may be the net outcome of these two interacting factors.

---

[19] The group A/B relative risk of failure is obtained as the ratio of the hazard of failure for individuals in group A and B. Hazard is calculated at the mean of continuous variables, mean duration, and setting all dummy variables to zero.

Estimated average hazards for each cohort are reported in Figure 4.3 (page 101). As in the case of APE, calculations are based in the best fitting model for each age-specific group. Various pieces of new information are obtained. First, there is clear empirical evidence that younger cohorts of women are delaying first birth in reference to older cohorts. This is witnessed by an almost everywhere inwards shift of the **C6372** hazard line with respect to **C4262** line, and by the inwards shift of the **C7382** with respect to the **C6372** line.

Inspection of Figure 4.3 reveals that the assumption that entry at different calendar times leads exclusively to parallel shifts in the hazard is hardly supported by the data. Further, as discussed some lines above, explanatory variables are reported to have different effects across the three considered age groups. Thus, since misspecification of the hazard may lead to serious bias, there is evidence that an analysis performed under such underlying assumptions may result in wrong inference. This sort of misspecification may be behind previous findings suggesting that women born in the 1963-1972 generation accelerated first birth in relation to women in the 1942-1962 generation. Clearly new evidence does not support such findings and the analyst should conclude that women in the 1963-1972 cohort delayed first birth in relation to women in the 1942-1962 generation.

4.4.6 Consistency Analysis: Using different age groups

The preceding discussion has enquired about the socio-economic determinants of the timing of first birth in Mexico. In particular, the study has intended to establish if young generations of women are delaying first birth in reference to

the experience of older generations. To accomplish such an objective a set of age groups were defined and a hazard model for the timing of first birth was estimated for each group. In particular, three groups were considered: women born between 1942 and 1962, women born between 1963 and 1972, and women born between 1973 and 1982. This generation split creates a partition of the sample such that each age group contributes approximately the same number of observations to the sample (about thirty per cent).

In order to test for the robustness of previous findings a new set of generation dummies are used here. Two cases are considered: (a) change the limits of the generation groups by minus five years, and (b) change the limits of the generation groups were by plus five years. The new age groups in case (a) are then 1942-1957, 1958-1967, and 1968-1982. A set of corresponding binary indicator variables were generated (i.e, **c4257**, **c5867** and **c6882**). In a similar fashion in case (b) the new set of generation dummies **c4267**, **c6877** and **c7882** were defined. Once these new generation groups were defined a hazard model for each group was estimated. Figures 4.4 and 4.5 present, correspondingly, the estimated average hazard for each generation in case (a) and (b). Only best fitting models are reported.[20]

Comparing graphs in Figures 4.3 and 4.4 (page 101) the reader can conclude that moving the limits of generations groups by minus five years causes the hazard function of the two younger age groups – **c5867** and **c6882** – to overlap at

[20] Best fitting model for **c4257**, **c5867** and **c6882** is a Probit Hazard with two plus final mass points. Best fitting model for **c4267** and **c6877** is a Probit Hazard with two plus final mass points. Finally, best fitting model for **c7882** is a Probit Hazard with final mass point.

almost all duration times. Further, the gap between the hazard of the oldest cohort, **c4257**, and the hazard functions of the two younger cohorts is now wider than before (compare Figures 4.3 and 4.4). These findings are intuitive because taking out five years leads to age groups that, on average, are composed by older women. Hence, women who entered at risk of first pregnancy shortly after and well after the 1973 change on population policy are now included in the youngest cohort (for further reference see Chapter 3, section 3.3). Similarly the base group, **c4257**, now contains women that were born well before the 1973 policy change. Therefore, the researcher should expect the new partition to mitigate the differences between the hazards of the two youngest cohorts, and to stress the differences between the hazards of the oldest and the two younger cohorts. In any case it is clear from Figure 4.4 that women born after 1958 have delayed first birth in reference to the base age group 1942-1957.

Figure 4.5 (page 102) repeats the exercise but now with generation groups that are moved by plus five years. The hazard function of the three age groups **c4267**, **c6877** and **c7882** are now very close to each other. Once again the findings are intuitive because moving the calendar limits of the cohort dummies by plus five years leads to age groups that, on average, contain younger women. As a consequence, the base group **c4267** has now women that entered at risk of first pregnancy close to the 1973 policy change and were therefore exposed for a longer time to the new policy regime. Having more women that were born close to the 1973 policy innovation causes, intuitively, a downward shift of the hazard function of the base group **c4267** in reference to the hazard function of the original base group **c4262** – compare Figures 4.3 and 4.5. Moreover, as it is shown by figure 4.5, having a younger base group should be expected to mitigate

the differences between the hazard of the oldest age group and the hazard of the remaining two age groups. Even though differences between old and younger cohorts are expected to be mitigated, Figure 4.5 shows that younger generations have tended to delay first birth in reference to the base **c4267** group – here represented by a slight downward shift of the **c6877** and **c7882** hazard functions in reference to the **c4267** hazard.

In conclusion one could say that differences of the hazard function across the compared generations depend largely on how women who were born around the first part of the 1970s are distributed into the three constructed age groups. Clearly, this is an indication that the 1973 innovation on population policy has played a significant role in stimulating first birth postponement in Mexico. In all cases considered empirical evidence suggest that young generations of Mexican women are indeed delaying first birth in relation to the experience of older generations.

To close the discussion the top of Table 4.7 (page 99) presents the average partial effects for **catholic**, **indspker** and **edu12** that were derived from hazard models estimated for generation **c4257**, **c5867** and **c6882**. Similarly, the bottom of Table 4.7 contains average partial effects calculated on the basis of hazard models estimated for generation **c4267**, **c6877** and **c7882**. From Table 4.7 the reader can conclude that although there is some variation, average partial effects for **catholic**, **indspker** and **edu12** are broadly consistent (in both sign and magnitude) across the whole table – and in relation to the figures contained in Table 4.6. Hence, findings indicate that changing the definition of age cohorts does not result in unexpected large variations of estimated average partial effects.

4.5 Conclusions

In the last twenty years female permanent sterilization (FPS) has become the most popular contraceptive method in Mexico. During the same period of time the demand for contraceptive pills, condoms and other short-term contraceptives dropped sharply and consistently. The changes in the demand for contraceptives have raised concerns that the timing of children in Mexico might remain unchanged despite the reduction of fertility rates in the last few decades. In particular, it is thought that women might be cutting off fertility at high parities without modifying their reproductive calendar. Such behaviour has been argued to be potentially harmful to women's welfare because it may provide disincentives to the accumulation of human capital and to adversely affect women's performance (and participation) into the labour market (see Welti 1997, Gomez 1996a, Mier y Teran and Rabell 1990, Zavala de Cosio 1989).

In the present chapter it is shown that despite the popularity of PFS and the drop in the demand for short-term contraceptives, young generations of Mexican women have tended to delay first birth.

The study finds that Catholic individuals have lower hazard of entering motherhood than non-Catholic individuals. A result that supports the idea that Catholicism has not been a relevant factor preventing women from modifying their fertility behaviour in Mexico. Instead, Catholicism seems to play a relevant role for the delay of first birth. From the analysis is not possible to infer the reasons behind this result. But a postponement of marriage and sexual abstinence previous to it seems to be a plausible explanation. As predicted by economic theory, education at age 12 was found to induce postponements of motherhood

(see Chapter 2, section 2.3 and, among others, Happel, Hill, and Low 1984, Cigno and Ermisch 1989, Hotz, Klerman, and Willis 1997). Finally it is found that indigenous language speakers enter motherhood faster than non-indigenous language speakers, a result that is intuitive given the limited access to education and health services that, as a general rule, individuals belonging to the indigenous ethnic groups have in Mexico.

Apendix A

In the Split-Population model of Schmidt and Witte (1989) there are two different types of individuals. One type of individuals (group A) remain childless for their whole lifespan either because they are biologically unable to procreate or because they freely choose to do so. A second type of individuals (group B) will eventually enter parenthood during their fertile period of life. Following Schmidt and Witte, suppose that a dichotomous random variable F determines whether an individual belongs to group A or B, with $F_i=1$ if the i-th individual belongs to group B. Further, suppose that F is independent of all observed and unobserved variables affecting fertility decisions and that $\Pr[F=0] = \kappa$. Then, if the i-th individual enters motherhood at duration t, the likelihood of observing such an event will be clearly

$$\Pr[Failure\, at\, t \mid Suvival\, up\, to\, t-1] = \Pr[F=1]\Pr[Failure\, at\, t \mid Suvival\, up\, to\, t-1,\, F=1]. \qquad (A.1)$$

Here the assumption that F is independent of all observed and unobserved variables affecting fertility behaviour is being exploited. Now, if instead of a failure a censored observation is registered at duration t, the likelihood of such an event will be

$$\Pr[Survival\, up\, to\, t] = \Pr[F=0] + \Pr[F=1]\Pr[Survival\, up\, to\, t \mid F=1]. \qquad (A.2)$$

Notice that if a hazard function for the group of 'eventually mothers' is specified, a log-likelihood function can be written on the basis of equations (A.1) and (A.2), and a whole set parameters may then be estimated by usual maximum likelihood techniques. Nothing prevents the conditional hazard of eventually mothers to depend on unobservables. Therefore, the model might be easily extended to account for the presence of unobserved heterogeneity. Using the notation from section 3, the contribution of the i-th individual to the likelihood in such a generalized model can be written as:

$$L_i^* = \int \left\{ \prod_{t=1}^{d_i} \left[(1-\kappa)h_{ti}(v_i, F_i = 1)\right]^{w_{it}} \left[\kappa + (1-\kappa)\{1 - h_{it}(v_i, F_i = 1)\}\right]^{1-w_{it}} \right\} f(v_i)\, dv_i \qquad (A.3)$$

where d_i, v_i, h_{it}, w_{it} and $f(v_i)$ remain as in section 3 with the only difference that h_{it} is now defined exclusively for individuals belonging group B. From (A.3) it is clear that if $w_{it}=0$ for all t, the contribution of the i-th individual to the likelihood becomes,

$$L_i^* = \kappa + (1-\kappa) \int \left\{ \prod_{t=1}^{d_i} \{1 - h_{it}(v_i, F_i = 1)\} \right\} f(v_i)\, dv_i. \qquad (A.4)$$

Moreover, if $w_{it}=1$ for $t=d_i$ the i-th individual belongs to group B and her contribution to the likelihood reduces to

$$L_i^* = (1-\kappa) \int \left\{ \prod_{t=1}^{d_i} \left[h_{ti}(v_i, F_i = 1)\right]^{w_{it}} \left[1 - h_{it}(v_i, F_i = 1)\right]^{1-w_{it}} \right\} f(v_i)\, dv_i. \qquad (A.5)$$

Hence, from equations (A.4) and (A.5) it follows that,

$$L_i^* = \kappa\left\{\prod_{t=1}^{d_i}(1-w_{it})\right\} + (1-\kappa)\int\left\{\prod_{t=1}^{d_i}\left[h_{ti}(v_i,F_i=1)\right]^{w_{it}}\left[1-h_{it}(v_i,F_i=1)\right]^{1-w_{it}}\right\}f(v_i)\,dv_i$$
$$= \kappa\left\{\prod_{t=1}^{d_i}(1-w_{it})\right\} + (1-\kappa)L_i(F_i=1). \qquad (A.6)$$

where $L_i(F_i=1)$ is similar to equation (7) with the exception that in (A.6) $L_i(F_i=1)$ is defined only for individuals in group B. Notice, however, that if $v_i=-\infty$ in equation (9) the contribution of the i-th individual to the likelihood reduces to $\Pr[v_i=-\infty] = \delta/(1+\delta)$ because in such a case $L_i(v_i=-\infty) = 0$. Therefore, the probability attached to the mass point at minus infinity (called simply final mass point throughout the text) in (9) is equivalent to the probability that $F_i=0$ in equation (A.6). This is, $\kappa= \delta/(1+\delta)$. As a consequence, models (A.6) and (9) are equivalent. They are simply two equivalent ways of allowing a 'mover-stayer' feature into the model. The intuition behind this result is that if in model (9) there exist a significant proportion of individuals with unobserved characteristic $v_i=-\infty$, the existence of a group of people that never become at risk of entering motherhood is therefore implicitly recognized.

Apendix B

Table 4.1 Descriptive Statistics

Variable	Description	Mean	Std. Dev.	Min	Max
Age	age in years	29.66	10.47	15	55
duration	See footnote 1	9.94	6.41	3	43
fail	=1 if first birth observed; 0 otherwise	0.60	-	-	-
Education, Religion and Ethnic group			-	-	-
Catholic	=1 if Catholic; 0 otherwise	0.89	-	-	-
indspker	=1 if indian language speaker; 0 otherwise	0.06	-	-	-
Edu12	education at age 12 in years	5.36	1.38	0	6
Cohort (base 1942-1962)			-	-	-
c6372	=1 if born within 1963-1972; 0therwise	0.30	-	-	-
c7382	=1 if born within 1973-1982; 0therwise	0.38	-	-	-
+ 32 birth place dummies (base Mex. City)					
Number of observations					78,467

1. Duration is defined as years between age 12 and first birth if first a completed spell was observed, and years between age 12 and 1997 otherwise.

Table 4.1a Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Cohort 1942-1962 (base)					
Age	25,454	42.42	5.54	34	54
duration	25,454	12.92	8.78	3	43
fail	25,454	0.85	0.35	-	-
Education, Religion and Ethnic group					
Catholic	25,454	0.90	-	-	-
indspker	25,454	0.06	-	-	-
Edu12	25,454	4.84	1.75	-	-
Cohort 1963-1972					
Age	23,214	29.06	2.92	24	35
duration	23,214	10.86	4.64	3	22
fail	23,214	0.75	0.43	-	-
Education, Religion and Ethnic group					
Catholic	23,214	0.89	-	-	-
indspker	23,214	0.06	-	-	-
Edu12	23,214	5.50	1.23	-	-
Cohort 1973-1982					
Age	29,799	19.24	2.83	15	25
duration	29,799	6.67	2.56	3	12
fail	29,799	0.26	-	-	-
Education, Religion and Ethnic group					
Catholic	29,799	0.89	-	-	-
indspker	29,799	0.06	-	-	-
Edu12	29,799	5.71	0.95	-	-

Note: Duration is defined as years between age 12 and first birth if first a completed spell was observed, and years between age 12 and 1997 otherwise.

Table 4.2 Extreme Value Hazard--Empirical Mass Points
Coefficient [Std. Err.]

	(1) No heterogeneity	(2) Two mass points	(3) Two + final mass points
Education, Religion and Ethnic group			
Catholic	-0.0585 [0.0151]**	-0.1339 [0.0189]**	-0.1484 [0.0201]**
indspker	0.1555 [0.0211]**	0.2684 [0.0267]**	0.2625 [0.0281]**
Edu12	-0.0920 [0.0031]**	-0.2016 [0.0044]**	-0.2203 [0.0050]**
Cohort (base 1942-1962)			
c6372	-0.0099 [0.0105]	0.0620 [0.0136]**	0.0732 [0.0146]**
c7382	-0.2252 [0.0141]**	-0.1495 [0.0161]**	-0.1398 [0.0168]**
Constant	-4.7142 [0.0397]**	-4.0527 [0.0456]**	-3.8830 [0.0500]**
Birthplace dummies	Yes	Yes	Yes
mass1	-	-3.2308 [0.0814]**	-1.6649 [0.0575]**
Pr(mass1)	-	0.8110 [0.0043]**	0.1708 [0.0130]**
Pr(massend)	-	-	0.1220 [0.0040]**
Log-likelihood	-159,196.73	-158,344.82	-158,294.17
chi2	22,621.52	19,277.19	16,209.13
Prob > chi2	0.0000	0.0000	0.0000
AIC	318,521.46	316,823.64	316,724.34
CAIC	319,178.77	317,511.76	317,422.73
LS	-159,557.38	-158,722.38	-158,677.36
Number of observations	78,467	78,467	78,467

Note: ML estimates. * significant at 5%; ** significant at 1%

Table 4.3 Probit Hazard--Empirical Mass Points
Coefficient [Std. Err.]

	(1) No heterogeneity	(2) Two mass points	(3) Two + final mass points
Education, Religion and Ethnic group			
Catholic	-0.0373 [0.0080]**	-0.0730 [0.0102]**	-0.0802 [0.0109]**
indspker	0.0891 [0.0113]**	0.1469 [0.0147]**	0.1471 [0.0153]**
Edu12	-0.0526 [0.0017]**	-0.1163 [0.0043]**	-0.1233 [0.0028]**
Cohort (base 1942-1962)			
c6372	-0.0050 [0.0056]	0.0338 [0.0075]**	0.0387 [0.0079]**
c7382	-0.1104 [0.0071]**	-0.0703 [0.0084]**	-0.0668 [0.0088]**
Constant	-2.2878 [0.0180]**	-3.8154 [0.0485]**	-2.9860 [0.0547]**
Birthplace dummies	Yes	Yes	Yes
mass1	-	1.8540 [0.0422]**	1.0744 [0.0464]**
Pr(mass1)	-	0.1670 [0.0031]**	0.7706 [0.0105]**
Pr(massend)	-	-	0.1256 [0.0042]**
Log-likelihood	-159,119.74	-158,236.57	-158,208.50
chi2	12,775.5	22,826.3	17,008.8
Prob > chi2	0.0000	0.0000	0.0000
AIC	318,367.48	316,603.14	316,547.0
CAIC	319,024.79	317,270.72	317,214.58
LS	-159,480.39	-158,602.86	-158,574.79
Number of observations	78,467	78,467	78,467

Note: ML estimates. * significant at 5%; ** significant at 1%

Table 4.4 Probit Hazard--Normal Heterogeneity
Coefficient [Std. Err.]

	(1) No heterogeneity	(2) Normal Hetererogeneity	(3) NH + final mass point
Education, Religion and Ethnic group			
Catholic	-0.0373 [0.0080]**	-0.0737 [0.0132]**	-0.0884 [0.0123]**
indspker	0.0891 [0.0113]**	0.1611 [0.0189]**	0.1551 [0.0172]**
Edu12	-0.0526 [0.0017]**	-0.1104 [0.0043]**	-0.1317 [0.0042]**
Cohort (base 1942-1962)			
c6372	-0.0050 [0.0056]	0.0090 [0.0094]	0.0394 [0.0090]**
c7382	-0.1104 [0.0071]**	-0.1321 [0.0110]**	-0.0810 [0.0099]**
Constant	-2.2878 [0.0180]**	-2.6537 [0.0422]**	-2.1529 [0.0308]**
Birthplace dummies	Yes	Yes	Yes
rho	-	0.3484 [0.0189]**	0.1632 [0.0187]**
chi2(01) for rho	-	533.28	1,666.80
Prob > chi2(01)	-	0.0000	0.0000
Pr(massend)	-	-	0.1287 [0.0033]**
Log-likelihood	-159,119.74	-158,852.10	-158,286.34
chi2	12,775.5	8,960.84	5,555.64
Prob > $\chi 2$	0.0000	0.0000	0.0000
AIC	318,367.48	317,836.20	316,704.68
CAIC	319,024.79	318,503.78	317,382.53
LS	-159,480.39	-159,219.39	-158,658.26
Number of observations	78,467	78,467	78,467

Note: ML estimates. * significant at 5%; ** significant at 1%
Note 2. rho represents the proportion of the total variance that is explained by the random effect, v_i.

Table 4.5 Average Partial Effects

	Extreme Value Hazard Two + final mass points	Probit Hazard Two + final mass points	Probit Hazard Normal het. + final mass point
Education, Religion and Ethnic group			
Catholic	-0.0167**	-0.0157**	-0.0179**
indspker	0.0350**	0.0288**	0.0347**
Edu12	-0.0264**	-0.0245**	-0.0277**
Cohort (base 1942-1962)			
c6372	0.0090**	0.0076**	0.0084**
c7382	-0.0158**	-0.0131**	-0.0165**

Note 1. ** (*) indicates significance at 1% (5%) of the estimated coefficient, β_i, in Tables 2, 3 and 4.
Note 2. If dj(xi,vi) represents the partial efect of xj on the conditional hazard, given the vector of observed variables xi and the unobserved random effect vi, the average partial effect of xj is calculated as the expected value of dj(xi,vi) over the density function of the random effect f(vi).
Note 3. Average Partial effects of continuous variables are calculated at the mean value, mean duration,and setting all dummies to zero. Average Partial effects for discrete variables are calculated by the difference measure at the mean of all continuous variables, mean duration, and setting all other dummy variables to zero.

Table 4.6 Average Partial Effects – Cohort of Birth Models
(based on the best fitting model for each cohort)

	1942-1962	1963-1972	1973-1982
Education, Religion and Ethnic group			
Catholic	-0.0163**	-0.0180**	-0.0102**
indspker	0.0161**	0.0291**	0.0370**
Edu12	-0.0244**	-0.0308**	-0.0243**

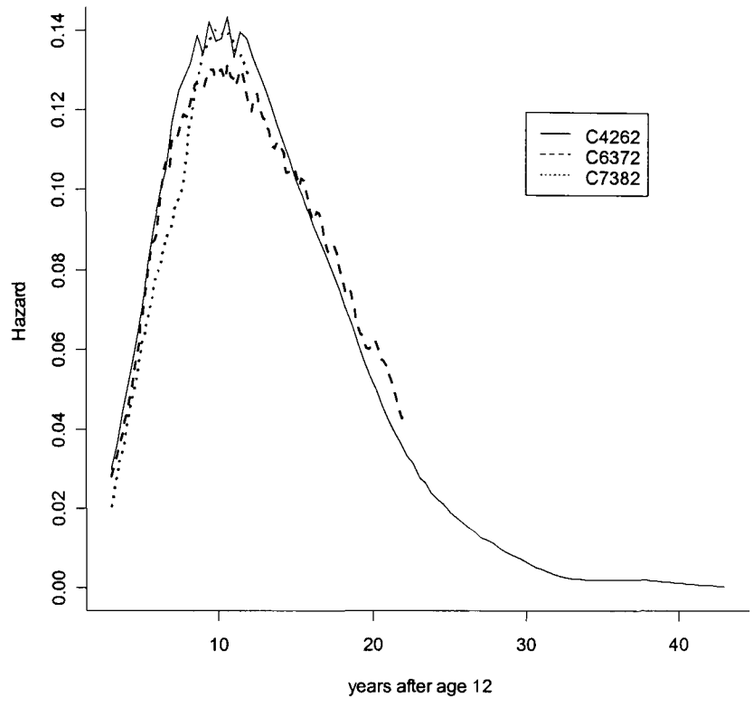
Note 1. ** (*) indicates significance at 1% (5%) of the estimated coefficient, β_i, in Tables 2, 3 and 4.
Note 2. Best fitting model for **C4262** and **C6372** is a Probit Hazard with 2 + final mass points. Best fitting model for **C7382** is a Probit hazard with final mass point
Note 3. If dj(xi,vi) represents the partial efect of xj on the conditional hazard, given the vector of observed variables xi and the unobserved random effect vi, the average partial effect of xj is calculated as the expected value of dj(xi,vi) over the density function of the random effect f(vi).
Note 4. Average Partial effects of continuous variables are calculated at the mean value, mean duration,and setting all dummies to zero. Average Partial effects for discrete variables are calculated by the difference measure at the mean of all continuous variables, mean duration, and setting all other dummy variables to zero.

Table 4.7 Average Partial Effects -- Cohort of Birth Models (plus and minus 5 years) based on the best fitting model for each cohort

	Catholic	**indspker**	**Edu12**
Minus five years			
1942-1957	-0.0124**	0.0143*	-0.0244**
1958-1967	-0.0213**	0.0302**	-0.0275**
1968-1982	-0.0127**	0.0374**	-0.0297**
Plus five years			
1942-1967	-0.0181**	0.0224**	-0.0221**
1968-1977	-0.0113**	0.0327**	-0.0279**
1978-1982	-0.0138**	0.0316**	-0.0187**

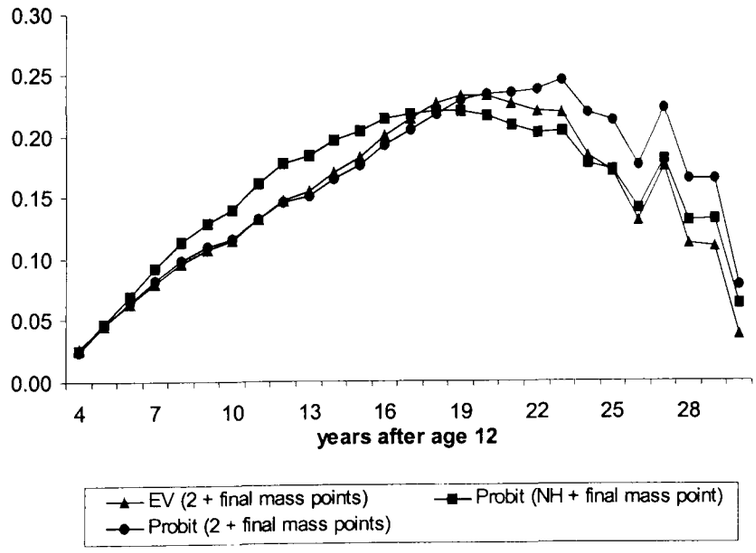
Note 1. ** (*) indicates significance at 1% (5%) of the corresponding coefficient in the estimated hazard function.
Note 2. Best fitting model for **C4257** and **C5867** and **C6882** is a Probit Hazard with 2 + final mass points.
Note 3. Best fitting model for **C4267** and **C6877** is a Probit Hazard with 2 + final mass points. Best fitting model for **C7882** is a Probit hazard with final mass point.
Note 4. If dj(xi,vi) represents the partial efect of xj on the conditional hazard, given the vector of observed variables xi and the unobserved random effect vi, the average partial effect o f xj is calculated as the expected value of dj(xi,vi) over the density function of the random effect f(vi).
Note 5. Average Partial effects of continuous variables are calculated at the mean value, mean duration, and setting all dummies to zero. Average Partial effects for discrete variables are calculated by the difference measure at the mean of all continuous variables, mean duration, and setting all other dummy variables to zero.

Figure 4.1 Kernel estimates of the hazard function



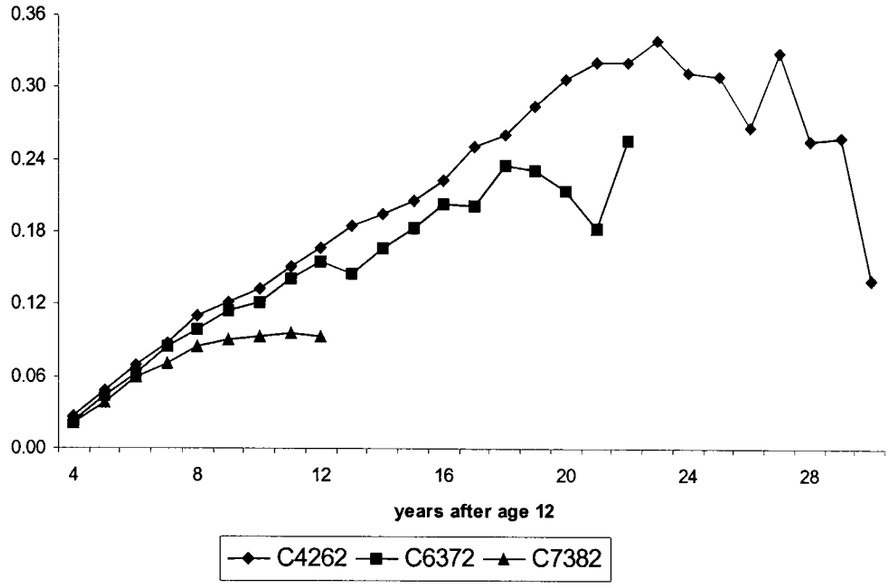
Note: An Epanechnikov kernel with varying bandwidth is used for generating Kernel estimates of the hazard function. Bandwidth increases as the number of cases at risk decrease. Estimates calculated with the muhaz package of the R statistical software.

Figure 4.2 Estimated Average Hazard



Note: Average taken over the density function of the unobserved random effect. Hazard calculated for a typical individual. The typical individual is catholic and had 5.35 years of education at age 12.

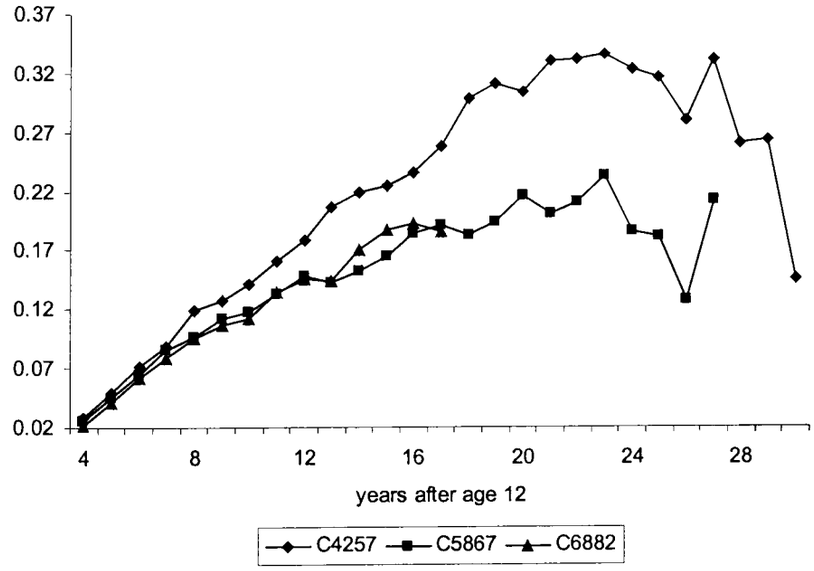
Figure 4.3 Estimated Average Hazard -- Cohort of Birth Models (based on the best fitting model for each cohort)



Note 1. Best fitting model for **C4262** and **C6372** is a Probit Hazard with 2 + final mass points. Best fitting model for **C7382** is a Probit hazard with final mass point

Note 2. Average taken over the density function of the unobserved random effect. Hazard calculated for a typical individual. The typical individual is catholic and had 5.35 years of education at age 12.

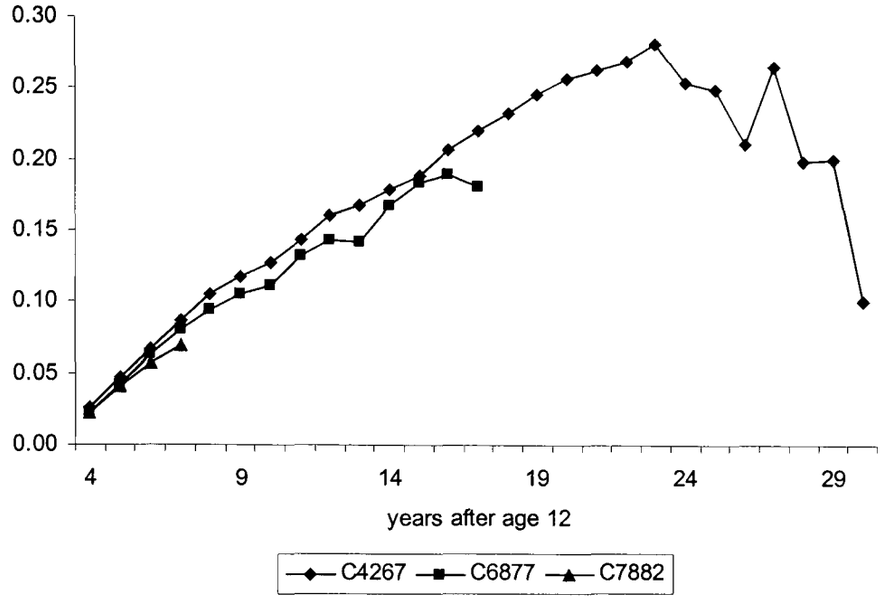
Figure 4.4 Estimated Average Hazard – Cohort of Birth Models (-5 years) Based on the best fitting model for each cohort



Note 1. Best fitting model for **C4257**, **C5867** and **C6882** is a Probit Hazard with 2 + final mass points.

Note 2. Average taken over the density function of the unobserved random effect. Hazard calculated for a typical individual. The typical individual is catholic and had 5.35 years of education at age 12.

Figure 4.5 Estimated Average Hazard -- Cohort of Birth Models (+5 years)
Based on the best fitting model for each cohort



Note 1. Best fitting model for **C4267**, **C6877** is a Probit Hazard with 2 + final mass points. Best fitting model for **C7882** is a Probit Hazard with final mass point.
Note 2. Average taken over the density function of the unobserved random effect. Hazard calculated for a typical individual. The typical individual is catholic and had 5.35 years of education at age 12.

Chapter 5

Completed fertility and the transition from low to high order parities: A double-hurdle approach

5.1 Introduction

The present chapter presents a study on completed fertility in Mexico. As is well known, fertility data have special features that need explicit econometric modelling. In the case of the developed world, for instance, data often exhibit under-dispersion and a relative excess of zero and two counts. Data from developing countries like Mexico, in contrast, are commonly over-dispersed and do not contain a particularly large excess of two outcomes. This sort of data, however, poses other important challenges to the analyst. Namely, that a non-negligible proportion of cases are contributed by women who have a large number of children and who tend to move to high order parities without taking any action to limit their fertility. In fact, in the case of Mexico nearly 21% of women end their fertile life with more than six children (INEGI 1999) and use contraceptives much less intensively than women with fewer children (Gomez 1996a).

Among other potential explanations, this sort of behaviour may be displayed because women with large families find themselves 'locked' in a regime in which the opportunity cost of extra children becomes particularly low. A large family, for example, may imply a permanent exit from the labour market and lead to further increases in family size. Clearly, some explicit account of this sort of behaviour is required when, as reported in Mexico, a good proportion of women give birth to a large number of children. Otherwise results will be difficult to interpret and most likely subject to serious bias.

Two main econometric avenues may be taken. One alternative would be to specify a Generalized Poisson Process, or *pure birth process*, as the main analysis technique and allow transition intensities to depend on women's accumulated stock of children – i.e., to introduce occurrence dependence in the stochastic process that generates completed fertility data (for further details see Winkelmann 2000). This possibility is exploited in Faddy (1997), Faddy and Bosch (2001), and Podlich, Faddy and Smyth (2004) in an extended count data framework. Applications, however, require the solution of a set of differential equations for which an exact analytical solution is not available. Numerical methods are needed and thus considerable computing power demanded.

An alternative approach would consider the assumption that low and high order fertility counts are drawn from different data generating mechanisms which do not exhibit occurrence dependence on their own. In such a context women move from one to another regime when their fertility crosses certain pre-established thresholds – say, zero and three children. Such an avenue, which is in line with the literature on hurdle count models (Mullahy 1986), is taken in the present work to develop a Double-Hurdle count model. The Double-Hurdle

model is estimated by standard maximum likelihood techniques and can be easily extended to account for unobserved individual heterogeneity and endogenous switching across regimes. No special demands on computing power are involved. The Double-Hurdle model is used to study in detail how socio-economic characteristics such as religion and ethnic group affect the probability of transition from low to high order parities in Mexico.

The rest of the chapter is organised as follows. Section two discusses all relevant data issues. Section three presents the econometric model and section four discusses empirical results. Finally, section five concludes.

5.2 Data and Variable definition

Data from the National Survey of Demographic Dynamics 1997 (ENADID from its acronym in Spanish) is used. Since completed fertility is the main concern of this study, a total of 19,477 cases of women aged 40 or over at the time of the ENADID interview (December 1997) are selected.

From a theoretical point of view it is not clear whether fertility decisions are taken in terms of lifetime number of pregnancies, lifetime number of live births, or lifetime number of surviving children. Obviously, lifetime number of pregnancies is the broadest concept as it is the cumulative sum of every conception a woman has during her fertile life. Number of live births excludes voluntary and involuntary miscarriages as well as stillbirths. Finally, number of surviving children removes infant deaths up to a certain age, say, age five. Most economic models of fertility choice consider that individuals decide in relation to the number of surviving children rather than over number of pregnancies or live

births (see Chapter 2, section 2.2, and among others Becker and Lewis 1973, Becker and Barro 1988, Willis 1973). That is, individuals choose the number of children they would like to have at the end of their fertile life, without regard to the number of pregnancies required to reach such a number of decedents (see for instance Bergstrom 1989, Willis 1973). Hence, the death of a child is thought to induce a new pregnancy (or a series of failed pregnancies) such that final family size remains constant. In the same line of thought, unwanted children would be abandoned to die in the absence of better means of birth control.

In the child mortality literature these ideas have been subject to criticism. First, (1991) suggest that if parents decide in terms of lifetime surviving children rather than in terms of lifetime pregnancies, a decline in child mortality may either be positively or negatively related with fertility. This clearly contradicts experience and the basic intuition that in a world under uncertainly parents 'ensure' themselves from the death of a child by increasing the number of pregnancies. This 'insurance', however, does not imply that a child death will induce a 'perfect' replacement – as the likelihood of survival is not under the control of parents. Wolpin (1984) suggest as well that if child mortality is exogenous rational behaviour is consistent with either none or perfect replacement. Cigno (1998) shows, however, that if child mortality is exogenous, fertility and child mortality are negatively related. It is only when child mortality is endogenous that Cigno (1998) finds the usual positive relationship between child mortality and fertility.

In applied work, the common practice is to define lifetime fertility as the number of children ever born live to a woman by the end of her childbearing period (see for instance Santos Silva and Covas 2000, Melkersson and Rooth

2000). In most pieces of applied work child mortality is considered explicitly because child mortality is very likely to an endogenous variable and most fertility surveys do not contain valid instrumental variables for mortality. Clearly, this strategy has the disadvantage of ignoring that some pregnancies are induced by a replacement behaviour.

Like in most fertility surveys, the ENADID does not have valid instruments for child mortality. Therefore, child mortality will not be explicitly addressed in the present work. Since replacement behaviour is, nonetheless, a significant feature of fertility decisions, the present study will assume that there is perfect replacement and completed fertility will be defined as total number of at least 5-years-old surviving children ever born to a woman during her lifetime. Following the work of Santos Silva and Covas [#covas2000 /d] and Melkerson and Rooth [#melkersson2000 /d] age 40 will be taken as the end of the childbearing period. Hence, the dependent variable, **Children**, is defined as the number of at least 5-years-old surviving children ever born to a woman aged 40 or older. According to the descriptive statistics (see Table 5.1 and Table 5.2, page 140) children has mean 4.43 and variance 7.56. The data is therefore over-dispersed.

A note of caution should be done at this point. Clearly, by age 40 some women may still fertile and have children. Further, live children with less of five years – but that will survive such threshold – are excluded from **Children**. Thus the dependent variable may not represent surviving completed fertility in some cases. This should not have a large impact in the study as most women end their childbearing before age 40.

Figure 5.1 (page 142) and Table 5.3 (page 141) present details on the empirical distribution of **children**. For comparison purposes a theoretical

Poisson distribution with mean 4.4 is also depicted. Notice first that, like data generated in developed countries, Mexican data exhibits an excess of zeroes relative to a Theoretical Poisson. This feature is found in most fertility data and various strategies for dealing with it have been introduced in the literature, including hurdle and zero-inflated count models (see the very informative surveys of Cameron and Trivedi 1986, Winkelmann 1995, Winkelmann 2000). Second, unlike data collected in developed countries, Mexican data do not contain a relative excess of one and/or two counts in reference to a Poisson distribution. Thus, there is no need here to inflate the probability of one and/or two counts. Finally, and more importantly, the Poisson distribution over-predicts the probability of observing counts 4, 5 and 6.

Looking closely at Figure 5.1 one may conclude that women who have more than three children seem to behave differently with respect to women who have a completed fertility of up to three. While women with less than four children, excluding zero outcomes, are well described by a standard Poisson, women with more than three children tend to transit to high parities more frequently than predicted. In fact, according to the data in Table 5.4, 53% of women who have more than three children transit to parities higher than five. And among those with more than five, 69% end fertile life with seven children or more. Intuitively, women who have four or more children may find themselves in a regime where the cost of an extra child is lower than the cost they would pay if their current fertility were lower than four. A fourth child could imply, for instance, a permanent exit from the labour market and a corresponding reduction in the opportunity cost of extra children. Although observed and unobserved

heterogeneity are yet to be accounted for, these are relevant features of the data that the analyst should not neglect.

Controls for women's religion, ethnic group, education at age 12, cohort of age, and place of birth are included as explanatory variables (see Table 5.1). The definition of these variables is as follows:

Catholic. Binary indicator that takes value one if the woman is catholic and zero otherwise. Defining two broad religious groups seems to be the finest sensible classification for Mexico given that nearly 90% of Mexicans are Catholics and a further 7% are Protestants.

Indspker. Dummy variable indicating whether an individual is able (**indspker** = 1) or unable (**indspker** = 0) to speak an indigenous language. **Indspker** proxies broad ethnic group (indigenous/mixed) rather than specific socio-cultural community. This variable is constructed in a parallel fashion to the ethnic binary indicator used in chapter 4. Therefore, no further comment will be added here.

Edu12. Like in the study on the timing of first birth reported in chapter 4, Edu12 represents education at age 12. Edu12 was constructed in a similar way to the education variable used in chapter 4. Hence, the reader should see section 4.2 of the previous chapter for a detailed discussion on Edu12.

Cohort of age. Using information on women's date of birth five cohorts can be defined, from 1940-1944 to 1955-1957. Four binary dummy variables indicating cohort of age are then generated (=1 if born in the corresponding 5-year period): **c4044**, **c4549**, **c5054** and **c5557**. The first cohort is taken as reference group. Notice that for the youngest cohort some women may have not finished their childbearing – see discussion in page 107.

Place of birth. Four regional geographic dummies for place of birth are defined: **MexCity** (base group), **North**, **Centre** and **South**.[21] There are important differences in the features of the data across the four geographical zones. Mean value and standard deviation of the dependent variable vary significantly from one region to the other, the South being the zone where the highest mean count is registered. Moreover, Mexican Indians are clearly concentrated in the South and Centre of the county. Important variations of education at age 12 are also detected across the different geographic zones (see Table 5.2, page 140).

5.3 Econometric issues

As was discussed earlier in the text, Mexican completed fertility data exhibit some characteristic features: an excess of zeros, a recognizable proportion of women choosing a completed fertility between one and three children, and a characteristic excess of large counts contributed by women that seem to move from low to high order parities without taking measures for limiting their fertility. Clearly, successful modelling should therefore consider that the various values of

[21] **North** is integrated by Baja California, Baja California Sur, Coahuila, Chihuahua, Durango, Nuevo León, Sinaloa, Sonora and Tamaulipas. **Centre** is integrated by Aguascalientes, Colima, Guanajuato, Guerrero, Hidalgo, Jalisco, Estado de México, Michoacán, Morelos, Nayarit, Puebla, Querétaro, San Luis Potosí, Tlaxcala, Veracruz, and Zacatecas. Finally, Campeche, Chiapas, Oaxaca, Quintana Roo, Tabasco, and Yucatan integrate the **South**.

the dependent variable might be generated by different mechanisms. Otherwise results are difficult to interpret and important bias might be present.

5.3.1 A double-hurdle model

Let individual's i-th completed fertility be y_i. The objective is to estimate a model for the probability that a fertility count j would be observed for the i-th individual from a random sample $Y=\{y_1,\ldots,y_n\}$. The model is formulated as follows. First a standard Poisson Hurdle model (Mullahy 1986) is considered,

$$\Pr(y_i = j) = \begin{cases} \exp(-\mu_{0,i}), & j = 0 \\ \left[1-\exp(-\mu_{0,i})\right]\Pr(y_i \mid y_i > 0), & j = 1,2,3\ldots \end{cases} \tag{1}$$

where the parameter $\mu_{0,i}$ maintains a deterministic log-linear relationship with a kx1 vector $\mathbf{x}_{i,0}$ of explanatory variables (including the constant term),

$$\mu_{0,i} = \exp\left(\mathbf{x}_{0,i}'\boldsymbol{\beta}_0\right), \tag{2}$$

$\boldsymbol{\beta}_0$ is its kx1 vector of associated coefficients, and $\Pr(y_i \mid y_i > 0)$ represents the probability distribution function of y_i given that a positive count has been observed. Notice that, unlike most Hurdle models reported in the literature, equation (1) uses an Extreme Value (EV) distribution for modelling the probability of observing a zero count. Specifying EV rather than the commonly selected Normal or Logistic distributions has two advantages in the present context. First, in contrast to Normal and Logistic, Extreme Value delivers a non-

symmetric distribution for the binary outcome model in equation 1 (see Arulampalam and Booth 2001). Second, since EV and Poisson predict the same Pr(y_i = 0), for practical purposes the hurdle in equation (1) can be seen as governed by a standard Poisson model.

Equation (1) represents a standard Hurdle Model. The model stresses the fact that the decision of entering parenthood is qualitatively different from the decision on the actual number of children, given that a strictly positive count is desired. To put it in other words, the Hurdle stresses the fact that zero and strictly positive counts may be generated by two different mechanisms. In order to allow for a second hurdle modifications are introduced in Pr($y_i \mid y_i > 0$),

$$\Pr(y_i = j \mid y_i > 0) = \begin{cases} \left[1-\exp(-\mu_{1,i})\right]^{-1} \dfrac{\exp(-\mu_{1,i})\mu_{1,i}^{\ j}}{j!}, & j = 1,2,3 \\ \left[1-\sum_{k=1}^{3}\left[1-\exp(-\mu_{1,i})\right]^{-1} \dfrac{\exp(-\mu_{1,i})\mu_{1,i}^{\ k}}{k!}\right] \Pr(y_i \mid y_i \geq 4), & j = 4,5,6... \end{cases} \quad (3)$$

with,

$$\mu_{1,i} = \exp(\mathbf{x}_{1,i}'\boldsymbol{\beta}_1). \quad (4)$$

A standard Hurdle specifies Pr($y_i \mid y_i > 0$) as a zero-truncated Poisson distribution. In contrast, equation (3) considers the case where counts in the [1,3] and [4,∞) intervals are drawn from two different data generating processes. For the [1,3] interval a zero-truncated Poisson distribution is written as usual. However, for counts larger than three, a new distribution Pr($y_i \mid y_i \geq 4$) is introduced. Clearly Pr($y_i \mid y_i \geq 4$) will be truncated at three and, to guarantee a

well behaved probabilistic model, it should be re-scaled so that Pr($y_i \mid y_i > 0$) sums up to one. Since equation (3) is similar to equation (1) in its philosophy, one could interpret the count process for the [1,3] interval as a second hurdle. From this perspective the probability of crossing such a barrier is given by

$$\Pr(y_i > 3 \mid y_i > 0) = \left[1 - \sum_{k=1}^{3} \left[1 - \exp(-\mu_{1,i}) \right]^{-1} \frac{\exp(-\mu_{1,i}) \mu_{1,i}^{\ k}}{k!} \right].$$

To close the model a functional form for Pr($y_i \mid y_i \geq 4$) must be specified. For convenience a Poisson distribution is, once again, selected:

$$\Pr(y_i = j \mid y_i \geq 4) = \left[1 - \sum_{h=0}^{3} \frac{\exp(-\mu_{2,i}) \mu_{2,i}^{\ h}}{h!} \right]^{-1} \frac{\exp(-\mu_{2,i}) \mu_{2,i}^{\ j}}{j!}, \qquad j = 4,5,6... \quad (5)$$

As usual,

$$\mu_{2,i} = \exp(\mathbf{x}_{2,i}' \boldsymbol{\beta}_2). \quad (6)$$

In principle $\mathbf{x}_{0,i}$, $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ may contain some (or all) common elements and no exclusion restrictions are required to achieve identification. Similarly, the vector of parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are estimated without constraints. Notice that if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ the Double-Hurdle model (DHM) collapses to a standard Poisson Hurdle model. Moreover, if $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ a simple Poisson model is obtained. Hence, the advantages of DHM over standard Poisson Hurdle and Poisson models may be assessed by testing for the equality of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Parameters are estimated by

maximum likelihood. The contribution of the i-th individual to the overall likelihood is simply

$$L_i = \prod_{y_i=0} \exp(-\mu_{0,i}) \prod_{y_i>0} \left[1-\exp(-\mu_{0,i})\right] \prod_{1\le y_i \le 3} \left[1-\exp(-\mu_{1,i})\right]^{-1} \frac{\exp(-\mu_{1,i})\mu_{1,i}^{\ y_i}}{y_i!}$$
$$\prod_{y_i \ge 4} \left[1-\sum_{k=1}^{3}\left[1-\exp(-\mu_{1,i})\right]^{-1} \frac{\exp(-\mu_{1,i})\mu_{1,i}^{\ k}}{k!}\right] \tag{7}$$
$$\prod_{y_i \ge 4} \left[1-\sum_{h=0}^{3} \frac{\exp(-\mu_{2,i})\mu_{2,i}^{\ h}}{h!}\right]^{-1} \frac{\exp(-\mu_{2,i})\mu_{2,i}^{\ y_i}}{y_i!}.$$

At convergence minus the inverse of the Hessian matrix $-H^{-1}$ estimates the covariance matrix. Usual asymptotic hypothesis testing is valid. The likelihood function is separable. Therefore, estimates can be obtained by maximizing separately three different likelihood functions. First, a binary outcome model (the first two terms of equation 7) can report consistent and efficient estimators for β_0. Then, a model for a left truncated and right censored Poisson variable can properly estimate β_1 (third and fourth terms of equation 7: for further details see Terza 1985). Finally, a model for a left truncated Poisson (the fifth term of equation 7) can estimate β_2. Separating the likelihood function into three independent elements is possible because selection into zero, one-to-three, and larger-than-three fertility groups is exogenous.

To summarize, notice that Double-Hurdle models are composed of three parts: (i) an Extreme Value distribution governing the likelihood that a woman will remain childless for her entire lifetime, (ii) conditional on having a strictly positive outcome, a Poisson distribution governing the likelihood of observing any particular count in the [1,3] interval, and finally (iii) conditional on having

more than three children, a Poisson distribution governing the likelihood of observing any count larger than or equal to four. The model has a Double Hurdle interpretation because in order to observe an outcome equal or larger than four it is necessary first to register a strictly positive count (i.e., to cross the first hurdle) and then to move to parities higher than three (i.e., to cross the second hurdle). The structure of the model is graphically represented in Figure 5.2 (page 142).

Selection among different specifications will be based on an Akaike information criterion (AIC) statistic. For completeness, selection on the basis of a consistent Akaike information criterion (CIAC) statistic will be also performed,

$$\begin{aligned} AIC &= -2\ln(L) + 2k \\ CIAC &= -2\ln(L) + k\{\ln(n) + 1\}, \end{aligned} \tag{8}$$

where k represents the number of parameters to be estimated. A best fitting model achieves the minimum AIC and CIAC among all its potential competitors.

In the count data literature competing models are also assessed by means of a goodness-of-fit χ^2 statistic. To calculate such a statistic the analyst must first predict, for each individual, the probability of observing $r = 0,1,2,\ldots$ children on the basis of the estimated model. The resulting probabilities are thus summed over individuals to obtain the predicted number of women with r children, $\hat{n}_r$. Finally the statistic is calculated as,

$$\chi^2 = \sum_{r=0}^{R} \frac{(n_r - \hat{n}_r)^2}{\hat{n}_r}, \tag{9}$$

where n_r represents the actual number of women with r children in the sample. The statistic has a χ^2 distribution with R-1 degrees of freedom (Melkersson and Rooth 2000, Heckman and Walker 1990). A low value χ^2 is evidence of good fit and a best preferred model should have minimum χ^2 among all potential alternatives.

5.3.2 Unobserved heterogeneity

The model is easily extended to allow for unobserved individual heterogeneity. A general strategy would consider the inclusion of a random term in each section of the Double Hurdle,

$$\mu_{k,i} = \exp\left[\mathbf{x}_{\mathbf{k},\mathbf{i}}'\boldsymbol{\beta}_{\mathbf{k}} + v_{k,i}\right], \quad k = 0,1,2 \tag{10}$$

Next, some assumptions about the distribution of $v_{0,i}$, $v_{1,i}$, and $v_{2,i}$ will be required to fully specify the model. Joint Normality is a natural choice.

This general approach has, however, two important drawbacks. First, various levels of numerical integration are needed so that estimation will be computing-intensive

- particularly in the most interesting case where $v_{0,i}$, $v_{1,i}$, and $v_{2,i}$ are not orthogonal. Clearly, in many applications the computing cost may become large or even prohibitive. Second, and more substantially, there are no theoretical reasons to believe that selection into each fertility group is dependent on different unobservables. Tastes towards children, for instance, are likely to enter every

single part of the Double-Hurdle model. To avoid the aforementioned problems one could rewrite equation (8) as

$$\mu_{k,i} = \exp\left[\mathbf{x}_{k,i}'\boldsymbol{\beta}_k + \theta_k v_i\right]; \quad \theta_2 = 1,\ k = 0,1,2. \tag{11}$$

Under the new specification there is conceptually only one unobserved random factor but its impact varies in each part of the Double-Hurdle via the inclusion of three factor loadings θ_0, θ_1, and θ_2. Since only two factor loads are identified θ_2 will be normalised to one. If σ^2 represents the variance of the random effect v, one could show that

$$\begin{aligned} &\mathrm{var}\left[\log(\mu_2)\right] = \sigma^2 \\ &\mathrm{var}[\log(\mu_k)] = \theta_k^2\sigma^2, \quad k = 0,1 \end{aligned}$$

and,

$$\begin{aligned} &\mathrm{cov}[\log(\mu_0), \log(\mu_1)] = \theta_0\theta_1\sigma^2 \\ &\mathrm{cov}\left[\log(\mu_2), \log(\mu_k)\right] = \theta_k\sigma^2, \quad k = 0,1. \end{aligned}$$

Hence, over-dispersion is allowed in any component of the Double-Hurdle and correlation of any sign between the μ's may be accommodated. In a few words, the simplification does not impose serious loss of flexibility.

Once unobserved heterogeneity is included the likelihood function is no longer separable. Therefore, from this perspective selection into zero, one-to-three, and larger-than-three fertility groups is now endogenous and all parameters $\{\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \theta_0, \theta_1, \sigma^2\}$ must be estimated in a simultaneous fashion (other models with

endogenous selectivity have been suggested by Greene 1997, Terza 1998, Winkelmann 1998). Notice, however, that given v_i all sections of the conditional likelihood function remain independent. Consequently, the unconditional likelihood function is simply written as

$$L_i = \int_{v_i} L_i(v_i) g(v_i) dv_i, \tag{12}$$

where $L_i(v_i)$ represents the conditional likelihood function. The model is closed once a distribution for the unobserved heterogeneity term, $g(v_i)$, is specified. Here a Normal distribution will be used. Since the integral in equation (12) does not accept a closed solution Gauss-Hermite quadrature may be used to approximate it. As usual, the model is estimated by maximum likelihood and at convergence – H^{-1} estimates the covariance matrix.

Tests for the significance of θ_0, θ_1, and σ^2 may be used to assess the adequacy of the specification for the unobservables in the Double-Hurdle model. If the null θ_0=0 cannot be rejected, then unobserved heterogeneity does not enter the first hurdle (i.e., the count process that determines the probability of remaining childless for a entire lifetime). Similarly, if θ_1=0 then there is no unobserved heterogeneity in the second hurdle. Finally, if σ^2=0 unobserved heterogeneity will be absent in the overall model. Clearly, testing σ^2=0 requires a boundary-value likelihood ratio test. Given that the admissible range of θ_0 and θ_1 is the whole real line, testing for θ_0=0 and θ_1=0 may be performed on the basis of standard likelihood tests.

5.3.3 Relation to the literature

To the knowledge of the author no previous study has used a Double-Hurdle count data model similar to the one introduced in the present work. There are, however, two main previous efforts to control explicitly for the special characteristics that completed fertility data exhibit. On one hand, Melkersson and Rooth (2000) point out that, due to social norms, completed fertility data from developed countries commonly exhibit an excess of zero and two counts. In such a context Melkersson and Rooth suggest the use of a zero and two inflated count model. On the other hand, Santos Silva and Covas (2000) argue that social norms discourage individuals in developed societies from having an only child. Thus, if for instance a woman enters motherhood, the chances of observing an only child at the end of her fertile life are lower than predicted by standard count models. To control for this tendency to avoid an only child, Santos Silva and Covas develop a modified hurdle model that deflates the probability of observing such an outcome.

Double Hurdle models are widely used in the econometrics literature in various application fields. Existing models, however, are based on the modified Tobit-like model of Cragg (1971) and have a different philosophy from the Double-Hurdle model presented here. In particular, previous work has considered the case where the variable of interest must cross two different hurdles to achieve a strictly positive value. In the case of tobacco (alcohol) consumption, for instance, it is argued that a zero outcome might be equally reported for individuals who never smoke (drink) during their life - or up to the date of data collection - and for individuals who have smoke (have drunk) once but have quit the habit in the past (Yen and Jensen 1996, Blaylock and Blisard 1993, Jones

1989, Labeaga 1999). Clearly, at-least-once and current participation in the smoking (drinking) activity are potentially two different decisions. Thus, observing a strictly positive level of consumption implies that two hurdles have been crossed. Yen, Tan and Su (2001) offer a count data model with similar characteristics to the Tobit-like Double-Hurdle of Cragg (1971). Unlike previous work, the Double-Hurdle presented in this chapter considers the case where the second hurdle occurs in a strictly positive value (interval) of the variable of interest. Hence, the approach is essentially different.

5.4 Empirical results

In this section the empirical results of a study on the socio-economic determinants of completed fertility in Mexico are presented. Special emphasis is given to enquiring how socio-economic factors such as religion and ethnic group affect the likelihood of transition from low to high parities.

5.4.1 Insights from standard hurdle models

Table 5.5 (page 143) contains empirical results from standard Poisson hurdle models. For comparison purposes the hurdle at zero is modelled with an EV binary variable model in place of the usual Probit or Logit specification. Two cases are considered. Column (1) reports estimates from a hurdle model with no added unobserved heterogeneity, while column (2) reports estimates from a model where Normal unobserved heterogeneity is allowed in the post hurdle count process – i.e., for counts larger than zero. Model (2) is an important

extension of model (1) as it relaxes the restrictive equi-dispersion assumption of the Poisson distribution.

To start with, notice that, though v_i is detected to have small variance, the presence of unobserved heterogeneity is strongly supported by the data via a significant positive estimate for σ^2 (see column 2 of Table 5.5). In fact, a boundary-value likelihood ratio test for H_0: $\sigma^2 = 0$ rejects the null at any conventional significance level with a $\chi^2(01)$ of 296. These results are consistent with the previously discussed observation that unconditional variance (7.5) is larger than unconditional mean (4.43).

According to Table 5.5 the likelihood of remaining permanently childless is significantly affected only by the education of the index woman – see the top panel of Table 5.5. In fact, a likelihood ratio test for the exclusion of **catholic**, **indspker**, **c4549** through **c5557**, and **north** through **south** is not rejected with a χ^2 (8) = 14.6 and p-value = 0.067. The coefficient on **edu12** is reported to be negative, implying that women with a higher level of education at age 12 are more likely to remain permanently childless than women with a lower level of education at age 12. These findings confirm economic theory in the sense that individuals with a higher level of education are expected to have a large opportunity cost of bearing children in relation to the cost paid by individuals with a lower level of education (Willis 1973).

Regarding strictly positive outcomes, a negative and significant coefficient on **Catholic** in models (1) and (2) indicates that Catholic individuals have fewer children than individuals with other religious backgrounds – see the bottom panel of Table 5.5. This is an interesting finding given the widespread opposition of the Catholic Church to the use of contraceptives as a way of limiting family size, an

attitude that is traditionally thought to be a barrier to fertility reduction. The result is better understood if one considers that despite its formal opposition, the Catholic Church in Mexico has in practice been tolerant towards the adoption of contraceptives as a way of limiting family size. In fact, beyond some insignificant negative campaigns implemented by radical catholic associations – not directly related to the Catholic Church – no efforts to fight against the use of contraceptives have been undertaken in Mexico (Cabrera 1994). Under these circumstances other group-specific characteristics of the Catholic community may induce a negative coefficient on **Catholic**, say, its opposition towards out-of-wedlock sex. Other factors may also be at work. For instance, the existence of a large base of contraception users within the Catholic community may imply that a Catholic individual receives better information about the advantages of family planning relative to a non-Catholic individual (for more detail on these ideas see Chapter 2, section 2.2.7, and Kohler 1997, Kohler 2000).

The proxy for broad ethnic group **Indspker** has a positive coefficient attached, though it is significant only at a 5% significance level. Besides differences in culture, it is likely that the coefficient on **Indspker** may reflect differences in standards of living between indigenous and non-indigenous individuals in Mexico. As is well known, most indigenous individuals in Mexico live in small rural communities (particularly in the south) that are far from the main industrial centres. In such localities health and education services are very limited and most individuals live with a high degree of marginality (CONAPO 2001a).

According to the results in Table 5.5, education at age 12 has a negative and significant effect on completed fertility. This finding clearly supports theory

suggesting that investment in human capital increases the opportunity cost of children (Willis 1973). A negative coefficient on **Edu12** is also consistent with recent literature stressing the idea that education might increase the bargaining power of women within the household (see for instance Klawon and Tiefenthaler 2001, Eswaran 2002, Hindin 2000).

All coefficients on cohort-of-age dummies are negative and significant (base group 1940-1944.) These results are clearly in line with the general trend that Mexican period fertility rates, including the total fertility rate TFR, have showed in the last forty years. Pair-wise tests for the equality of the coefficients on **c4549**, **c5054** and **c5557** reject the null at any conventional confidence level. More importantly, results indicate that younger cohorts of women have larger coefficients attached to their age-specific dummy. Hence, there is strong evidence that younger cohorts of Mexican women are reducing their lifetime fertility in comparison to the experience of older cohorts.

5.4.2 Results from double-hurdle models

5.4.2.1 Model selection

Table 5.6 (page 144) presents the empirical results. For comparison proposes various specifications are reported. Column (1) contains estimates for a Double Hurdle model that does not control for the presence of unobserved individual heterogeneity. Similarly, Column (2) through (4) contain estimates for Double Hurdle models with Normal unobserved heterogeneity and three different assumptions about factor loadings. Namely, these are (a) $\theta_0 = \theta_1 = 0$, (b) $\theta_0 = \theta_1 = 1$, and (c) θ_0 and θ_1 free. Notice that θ_2 has been normalised to one in all cases.

Case (a) corresponds to a model where unobserved heterogeneity enters exclusively in the count process (iii) – conditional on having more than three children. In addition, selection among regimes is exogenous in the sense that the log-likelihood function can be factored into three independent components. Case (b) removes the assumption of exogenous selection but constrains unobserved heterogeneity to have a symmetric effect in all (i), (ii) and (iii). Finally, case (c) removes all restrictions on the unobservables so that for each regime a different random effect is estimated. Correlation (of either sign) among random effects is explicitly allowed. Hence, the log-likelihood cannot be factored into three independent components. In other words, there is endogenous regime selection.

A significant positive estimate for σ^2 is detected in all the alternative models with heterogeneity (column 2 through 4). In fact, a boundary-value likelihood ratio test for $\sigma^2 = 0$ rejects the null at any conventional significance level with a $\chi^2(01)$ of 78.53 for model (2), 48.62 for model (3), and 78.52 for model (4). Further, pair-wise selection performed on the basis of Akaike and Consistent Akaike information criteria strongly favours (2), (3) or (4) over (1). In a few words, unobserved heterogeneity is present and significant.

Table 5.7 (page 145) presents a series of likelihood ratio tests that help discrimination among the different models. The first row of the top panel considers a test on the overall significance of θ_0 taking $\sigma^2 \neq 0$ as a premise and imposing no constraints on θ_1. Clearly, this is a test for H_0: $\text{var}(\log(\mu_0)) = \theta_0\sigma^2 = 0$ against H_1: $\text{var}(\log(\mu_0)) \neq 0$. Table 5.7 reports a $\chi^2(1)$ statistic of 0.016 for this test. Hence, the null hypothesis cannot be rejected at any conventional significance level. A similar LRT (see second row of Table 5.7) fails to reject H_0: $\text{var}(\log(\mu_1)) = 0$ against H_1: $\text{var}(\log(\mu_1)) \neq 0$. But if H_0: $\sigma^2 = 0$ is tested against

H_1: $\sigma^2 \neq 0$ a $\chi^2(01) = 78.53$ [p-val = 0.000] is obtained, indicating that unobserved heterogeneity cannot be ignored overall. These results support, then, a model where unobserved heterogeneity enters exclusively in the process that governs the realisation of large outcomes. That is, in the truncated-at-three Poisson distribution (iii). The bottom panel of Table 5.7 reports further evidence that $\theta_0 = \theta_1 = 0$ and $\sigma^2 \neq 0$ is the correct specification. Selection on the basis of Akaike and Consistent Akaike information criteria supports the same conclusion (see bottom of Table 5.6). Notice that these results imply that selection into zero, one-to-three, and larger-than-three fertility groups is then exogenous and that the Double-Hurdle likelihood function is separable into three independent elements.

Before moving to discuss how explanatory variables affect fertility behaviour, it is worth pointing out that alternative assumptions about the distribution of unobservables have a limited, almost negligible, impact on the estimates. Thus results seem to be robust to various assumptions about unobservables.

5.4.2.2 Test for the joint equality of the coefficients

The following discussion reports findings from a model where unobserved heterogeneity enters exclusively in the Poisson process that governs the realisation of large outcomes (i.e., θ_0 and θ_1 are set to zero). As discussed in the previous section, this is the specification that fits best the ENADID data. The results are reported in Table 5.6. From now on the vector of parameters that enter count process (i) of the Double Hurdle model will be referred to as β_0. Similarly, parameters that enter count process (ii) and (iii) are referred to as β_1 and β_2.

Table 5.8 (page 145) contains a formal likelihood ratio test for the joint equality of the coefficients β_1 and β_2. The reported $\chi^2(10)$ statistic takes a value of 164.27, which is enough evidence to reject the null at a 1% significance level. Similar tests strongly reject $\beta_0 = \beta_1$ with a $\chi^2(10) = 1610.30$ [p-val=0.000], and $\beta_0 = \beta_1 = \beta_2$ with a $\chi^2(20) = 2339.49$ [p-val=0.000]. In a few words, neither Poisson nor hurdle at zero Poisson are supported by the data (notice that in either case unobserved individual heterogeneity is being controlled for). The Double-Hurdle model is therefore preferred.

Comparing the elements of vector β_1 and β_2 various interesting observations can be made. Education at age 12, religion and ethnic group have a larger effect in the transition from low to high parities – i.e., the likelihood of crossing the 1-3 hurdle – than in determining fertility once the second hurdle has been crossed. This observation is supported by the fact that the coefficients on **Catholic**, **Indspker** and **Edu12** are larger in absolute value in vector β_1 than in vector β_2. However, pair-wise tests for (Coefficient on variable j in β_1) = (Coefficient on variable j in β_2) reject the null hypothesis exclusively in the case of **Edu12** with a t-stat = -2.27 [p-val=0.0115]. A similar exercise reveals that there are significant pair-wise differences in the coefficients on **c4549** (t-stat = 1.61, pval = 0.053), **c5054** (t-stat = 2.55, pval = 0.0054), **c5557** (t-stat = 4.89, pval = 0.0000), **centre** (t-stat = -1.70, pval = 0.0444) and **south** (t-stat = -3512, pval = 0.0000). Hence, differences in the likelihood of crossing the one-to-three children and the likelihood of observing any particular count larger than three are mainly driven by education, cohort of age and place of birth. It is important to underline here that cohort of age and birthplace dummies have larger coefficients in β_2 than in

β_1, implying that the impact of these socio-economic characteristics on family size is stronger once the second hurdle has been crossed.

5.4.2.3 Advantages of the Double-Hurdle model

Table 5.9 (page 146) contains a detailed comparison of predicted sample distributions generated on the basis of standard Hurdle and Double-Hurdle models. Only predicted probabilities from a best fitting Double-Hurdle are reported (i.e, a model with $\theta_0 = \theta_1 = 0$). To obtain the figures presented in Table 5.9 the likelihood of observing any particular count, from zero to eighteen, must be estimated for each individual using the relevant model and conditioning on their observed characteristics. Individual-specific predicted probabilities should then be averaged over all individuals (cell by cell) and the results collected for tabulation. In the bottom section of Table 5.9 a goodness-of-fit chi-square statistic is reported for each competing model along with Akaike and Consistent Akaike information criterion statistics.

If models that do not control for unobserved heterogeneity are compared, goodness-of-fit chi-square statistics for standard Hurdle and Double-Hurdle are, respectively, 371 and 150. Even controlling for unobserved heterogeneity Double-Hurdle (chi-square = 150) does better than standard Hurdle (chi-square = 213). Therefore, empirical evidence suggests that Double-Hurdle models fit noticeably better the data than the standard Hurdle – similar conclusions may be obtained on the basis of Akaike and Consistent Akaike information criteria. It must be stressed here that even the best fitting Double-hurdle with Normal

unobserved individual heterogeneity does not offer a complete description of the data, as is witnessed by its relative large goodness of fit chi-square.

Inspecting in detail Table 5.9, the reader can conclude that a standard hurdle with no heterogeneity under-predicts 2 and 3 counts, and over-predicts 4,5,6 counts. Clearly, a Double-hurdle model with no heterogeneity fits better 2,3,5, and 6 counts but does marginally worse predicting 1 and 4 outcomes. Accounting for unobserved heterogeneity improves the fit of both models. In particular, standard Hurdle reduces its degree of under-prediction of 2 and 3 counts. Counts 4,5 and 6 are still over-predicted but not to the same degree as in the case where unobserved individual heterogeneity is completely neglected. Similarly, controlling for unobserved heterogeneity causes the Double-Hurdle model to improve its prediction power of 4, 5, and 6 counts and to do better in predicting 2 outcomes. It seems that the relative ability to predict well 4,5, and 6 counts is what causes the Double-Hurdle model to perform better than a standard Hurdle model.

5.4.2.4 Effect of explanatory variables

Estimates from various specifications of a Double Hurdle Poisson model are reported in Table 5.6 (page 144). The present section discusses results for a model in which $\theta_0 = \theta_1 = 0$. This is the best fitting specification (see column 2 of Table 5.6). Additionally, Table 5.11 contains predicted probabilities for various representative individuals. Since most Mexicans are Catholic and non-indigenous language speakers, let a Catholic and non-indigenous language speaker who was born in Mexico City between 1940 and 1944 be the benchmark case (see row 2).

Set as well **Edu12** to its mean value of four years of schooling. This individual, referred as individual II for the rest of the discussion, has a likelihood of remaining childless for her whole lifetime of approximately seven per cent. Moreover, if a non-negative count has been observed individual II is expected to have a family of one, two or three children 47 out of a hundred times. To put it in other words, conditional on observing a positive count, individual II will move to parities higher than three with probability $\{1\text{-}Pr[1 < j \leq 3 \mid j > 0\ ]\} = 0.5316$. Finally, once a fourth child is observed Individual II will have a family larger than six with $Pr[j > 6 \mid j > 3] = 0.3947$.

5.4.2.4.1 Probability of a zero count

In section 5.4.2.1 unobserved heterogeneity was found to enter only in count process (iii) of the Double-Hurdle model. This result implies that selection into zero, one-to-three, and larger-than three fertility groups is exogenous and that the Double-Hurdle likelihood function can be separated into three independent elements. For these reasons results from Standard Hurdle and Double-Hurdle models will be identical as long as count the hurdle at zero concerns – compare column (1) in Table 5.5 and columns (1) and (2) in Table 5.6. Hence, results discussed in section 5.4.2.1 regarding the likelihood of observing a zero count remain valid under the Double-Hurdle and no further comment will be done here. Just notice that, as in the case of a standard Hurdle model, the Double-Hurdle model suggests that except for constant and **Edu12** all the elements of β_0 are insignificant (see Table 5.6 column 2). Moreover, from Table 5.11 the reader may learn that, ceteris paribus, a woman who had no formal education at age 12 is

3.33% less likely of remaining childless for her entire life than a woman who had six years of education at age 12 (a result that supports economic theory of fertility behaviour. For more detail see Chapter 2, section 2.2). Hence, though statistically significant, the effect of **Edu12** on Pr[j=0] seems to be rather small.

5.4.2.4.2 Transition from low to high parities given a positive count

Conditional on having at least one child, the probability of observing any particular count in the interval [1,3] is determined by a truncated-at-zero Poisson distribution that depends on the vector of parameters β_1. Notice then that, since $Pr(j > 3 \mid j > 0)$ is a function of β_1, the probability of crossing the second hurdle - or say, getting out of the [1,3] interval - is also a function of β_1.

Using this interpretation for the elements of vector β_1 the reader can conclude from the estimates in Table 5.6 (page 144) that Catholic individuals are less likely to cross the second hurdle than non-Catholic individuals. In order to assess the relevance of such an effect Table 5.11 contains predicted probabilities for a non-Catholic woman (individual I) who is otherwise identical to the benchmark woman II. There the reader can learn that individual I scores a $Pr[1 < j \leq 3 \mid j > 0] = 0.4302$ while individual II scores a $Pr[1 < j \leq 3 \mid j > 0] = 0.4684$. That is, Catholicism reduces the chances of transition from low to high order parities by as many as 3.8 percentage points.

Various factors may be behind the negative and significant coefficient on **Catholic** in the middle panel of Table 5.6. Among the most significant reasons there is a rather weak opposition of the Catholic Church towards the diffusion and adoption of contraceptives among the Catholic community in Mexico.

Coming back to Table 5.6, it seems that being an indigenous language speaker increases the chances of crossing the second hurdle, as the coefficient on **Indspker** is estimated to be positive – though the coefficient is different from zero only at 5%. The finding is intuitive because, as was discussed earlier in the text, indigenous individuals in Mexico have in general a lower economic status than non-indigenous individuals. Row 3 of Table 5.11 reports predicted probabilities for an indigenous language speaker individual who is otherwise identical to the benchmark individual II. Comparing figures in row 2 and 3 of Table 5.11 it is easy to conclude that the marginal effect of **Indspker** on $Pr[1 < j \leq 3 \mid j > 0]$ is around -.0306. In other words, holding other things constant, an indigenous language speaker has a 3% higher chance of having a family larger than three than a non-indigenous language speaker.

A negative coefficient on **Edu12** in vector β_1 of Table 5.6 suggests that an extra year of education at age 12 increases the likelihood that a woman will remain with less than four children during her entire lifespan. The finding confirms general economic intuition. More importantly, the effect of **Edu12** on the probability of observing such an event is estimated to be rather large. For instance, according to Table 5.11 increasing **Edu12** from five to six years will lead to an increment in $Pr[1 < j \leq 3 \mid j > 0]$ of 5.93 points, other things being constant. Further, a rise of schooling at age 12 from zero to six years implies that the odds of crossing the second hurdle would shrink by as much as 36.48 percentage points.

Vector β_1 in Table 5.6 contains sequentially more negative coefficients on **c4549** through **c5557**. Hence, the evidence is that young generations have lower chances of crossing the second hurdle. In fact, a woman born between 1945 and

1949 who is in other aspects similar to the benchmark woman II is estimated to bear 4% lower chances of ending her fertile life with more than three children in relation to the reference individual. Such a reduced risk becomes 10% and 13% for women in cohort 1950-54 and 1955-1957 respectively (see row 4 through 6 of Table 5.11, page 148-49).

As expected, being born in a region other than Mexico City implies increments in the odds of crossing the one-to-three hurdle. For instance, an individual who was born in the North of the Country will cross the second hurdle 18.6 out of a hundred more times than individual II, other things being equal. Similarly, marginal effects of **Centre** and **South** on $\{1\text{-}Pr[1 < j \leq 3 \mid j > 0\,]\}$ are respectively 0.1931 and 0.1197. Thus, being born in different geographical areas of the country leads to wide variations in the likelihood of a large family.

5.4.2.4.3 Probability of Counts Larger than Six given that the second Hurdle has been crossed

Conditional on having more than three children, a truncated-at-three Poisson distribution governs the likelihood of observing any particular count equal or higher than four. This last distribution depends on a vector of coefficients β_2.

Notice first from Table 5.6 that conditional on observing a count larger than three the coefficient on **Indspker** is insignificant at all conventional levels. In other words, ethnic group seems to have no influence on completed fertility once the second hurdle has been crossed. In other issues, the negative coefficient on **Catholic** is different from zero at 5% but not 1% significance level. Such a negative coefficient on **Catholic** implies that, conditional on crossing the second

hurdle, the Catholic reference individual II of Table 5.11 (pages 148-49) will end her fertile life with more than six children with probability 0.3947 while her non-Catholic equivalent individual I will register the same event with probability 0.4165. That is, Catholicism is associated with a reduction of 0.02181 units in Pr[$j > 6 \mid j > 3$]. Since the previous discussion has already offered some intuition for explaining this result no further comment on the issue will be made here.

Cohort of age affects significantly Pr[$j > 6 \mid j > 3$] as well. Namely, a woman born in the 1945-1949 cohort – i.e., individual IV of Table 5.11 – that has crossed the second hurdle is estimated to end fertile life with a family size larger than six with probability 0.3416. In comparison, woman II scores a Pr[$j > 6 \mid j > 3$] of 0.3947. Hence, ceteris paribus, a woman in the cohort 1945-1949 bears a reduced risk of 5.31 per cent of registering a large count in relation to a woman in the control group. Younger generations have even lower odds of a large completed fertility. In fact, marginal effects of **c5054** and **c5557** on Pr[$j > 6 \mid j > 3$] are -0.1105 and -0.1547 respectively.

Marginal effects for **North**, **Centre** and **South** on Pr[$j > 6 \mid j > 3$] might be obtained on the basis of row 2, and 7 through 9 of Table 5.11. Marginal effects are positive and large: 0.1873, 0.2414 and 0.1822 respectively.

5.4.2.5 Regional Results

Given that the different geographical regions of Mexico present important heterogeneity regarding economic development, availability of educational and health services and culture it is interesting to extend the analysis and allow coefficients in the Double-Hurdle model to vary according to the place of birth of

the women studied. Table 5.10 presents regression results for a Double-Hurdle model fitted to various sub-samples of the data constructed according to women's birthplace. Four Regions are considered: Mexico City, North, Centre, and South. In each region various specifications were estimated and Table 5.10 (page 147) reports exclusively the resulting best fitting model. Model selection was performed on the basis of the strategy followed at the National level. With the exception of the Centre, unobserved individual heterogeneity was detected exclusively in the post second hurdle count process (that is, evidence suggested $\theta_0 = \theta_1 = 0$). In the case of the Centre, θ_1 is reported to be significantly different from zero. Except for the North, likelihood ratio tests for the joint equality of the coefficients β_1 and β_2 easily reject the null (see Table 5.8). In the case of the North a standard Hurdle model is supported by the data. In all cases $\beta_0 = \beta_1$ and $\beta_0 = \beta_1 = \beta_2$ are rejected at least at 5% of significance. Interpretation of the coefficients remains the same and marginal effects might be calculated on the basis of Table 5.11 (pages 148-49).

Some differences in the coefficients on explanatory variables across the various regions are detected. In the first place, the evidence suggests that the likelihood of observing a zero count is independent of all the explanatory variables for women who were born in Mexico City and the South. And education at age 12 affects significantly Pr[j=0] only for women who were born in the North and Centre of the country.

Regarding the probability of crossing the one-to-three hurdle, Pr[j >3 | j>0], empirical evidence indicates that religious background is irrelevant for women who were born in Mexico City and the Centre, while relevant for women who were born the in North and South of the Country. Similarly, with the exception of

women who were born in the South, ethnic group seems not to affect the probability of crossing the second hurdle. Finally, education at age 12 is found to reduce the likelihood of having a large family in all cases. There are, however, some differences in the size of its effect. In particular, **Edu12** seems to have a far larger effect for women who were born in Mexico City than for women who were born in any other geographical region of the country.

Conditional on observing a count larger than three, Catholic individuals are expected to have a significantly lower fertility than non-Catholics only if they were born in the South. A similar observation is valid for ethnic group. That is, being an indigenous language speaker is associated significantly with increases in $Pr[\, j \mid j > 3]$ exclusively for women who were born in the South of the country. Education at age 12 reduces significantly $Pr[\, j \mid j > 3]$ in all the geographic regions of the country.

In conclusion, the evidence suggests that the effect of explanatory variables on completed fertility varies across the different regional birthplace areas of the country. For some birthplace areas religion and ethnic background have significant impact on fertility behaviour while in other birthplace regions such characteristics are largely irrelevant. Education at age 12 is a relevant factor across the whole country.

5.4.2.6 Results by Selected Groups of Age

To close the discussion Table 5.12 (page 150) presents results for two selected groups of age. This exercise relaxes the assumption previously sustained that, with exception of the constant, coefficients on the explanatory variables for

women belonging to different generations remain the same. In order to stress differences across generations only two cohorts of age were defined, namely, age group 1940-1949 and age group 1950-1957. As before, various versions of the Double-Hurdle model were estimated for each cohort and selection was performed in the same fashion as outlined in section 5.4.2.1. Table 5.12 reports only the best fitting models.

One of the most interesting results obtained from this exercise is the fact that unobserved heterogeneity appears to be present at all the three count processes that compose the Double-Hurdle estimated for women born between 1940 and 1949. In contrast, for women born between 1950 and 1959, unobserved heterogeneity only enters in the last count process (i.e., the Poisson governing outcomes equal or longer than four). Among other potential interpretations, the negative and significant estimate for θ_0 for generation 1940-1949 implies that women who dislike the most children have a higher probability of remaining childless for their whole life. Similarly, a positive estimate for θ_1 suggest that women who like the most children are more likely to cross the one-to-three hurdle. Clearly these results conform economic intuition. Table 5.8 presents a series of likelihood ratio test that showing that β_0, β_1, and β_2 are statistically different of each other so that the Double-Hurdle model is supported.

Another relevant characteristic of the results in Table 5.12 is the fact that a very large factor loading, $\theta_0 = -13.7$, is estimated. This substantial factor loading probably is associated with the fact that zero cases reported in the 1940-1949 cohort represent only 2.2% and the fact that previous 1973 the sale of contraceptives in Mexico was banned. In such a context it is then likely that the probability of observing a zero count was largely driven for unobservable factors,

which is represented by the large factor loading. In fact, according to Table 5.11 the probability of observing a zero outcome, at the mean of the random effect v, for women in the 1940-1949 cohort is, in practical terms, zero regardless the value of the observable characteristics. Hence, empirical evidence suggest that for generations of women that started their fertility planning period well before the 1973 innovation in population policy, childlessness was the outcome a very large realisation of unobservable factors – most probably, a lack of biological ability to conceive.

Like in previous analysis, findings suggest that education at age 12 significantly affects the likelihood of observing a zero count in both age groups. However, as shown in Table 5.11 results suggest that the probability of observing a zero outcome for the older generation (1940-1949) is, in practical terms zero (with four digit precision) regardless the value taken by the explanatory variables. In fact, it is only for the generations of women born after 1950 that the predicted probability of lifetime childlessness is different from zero. The fact that contraceptives became widely available in Mexico only after 1974 may explain why women in the older cohort of age had practically no chances of a zero outcome.

Another interesting result obtained from Table 5.8 is the fact that religion became a factor reducing the likelihood of crossing the second hurdle only for the generation of women born in the group 1950-1957. That is, religion did not have a significant statistical effect for the older generation born between 1940 and 1949. This is a remarkable result that deserves further attention in future research. As is suggested in the previous discussion, a potential explanation for this finding is that, given the relative size of the Catholic community in Mexico, Catholic

individuals belong to a social network that is broad and heterogeneous in reference to a non-Catholic social network. This feature of the Catholic community may imply that Catholic individuals adopted in a faster way the modern contraception technology that became available since the 1970s in Mexico. A similar argument may be used to explain why **indspker** and **catholic** have significant coefficients in the bottom panel of Table 5.8 (post second hurdle count process) only for the 1950-1957 generation.

Marginal effects of explanatory variables on $Pr[j=0]$, $Pr[1 < j \leq 3 \mid j > 0]$ and $Pr[j > 6 \mid j > 3]$ are in general larger for generation 1940-1949 than for generation 1950-1957. Hence, findings indicate that with the passing of time the effect of socio-economic factors on lifetime fertility has been reduced. Clearly, this result may be simply a reflection of the fact that, with the passing of the time, more and more women adopt a low fertility strategy and new generations, independently of their socio-economic characteristics, become more likely to imitate such behaviour.

5.5 Conclusions

The present work reports a study on the socio-economic determinants of completed fertility in Mexico. Special attention is given to how socio-economic factors such as religion and ethnic group affect the likelihood of transition from low to high parities. An innovative Poisson Double-Hurdle count model is developed for the analysis. This methodological approach allows low and high order parities to be determined by two different data generating mechanisms, and

explicitly accounts for potential endogenous switching between both regimes. Unobserved heterogeneity is properly controlled.

Catholicism is found to be associated with reductions in the likelihood of transition from low to high parities. This result may be associated with the relatively weak opposition of the Catholic Church to the diffusion of contraceptives in Mexico, and its much stronger opposition to the initiation of sexual life before marriage (Cabrera 1994). Other factors may be at work. For instance, the existence of a large base of contraception users within the Catholic community may imply that a Catholic individual receives better information about the advantages of family planning relative to a non-Catholic individual (Kohler 1997, Kohler 2000).

Empirical evidence suggests that being an indigenous language speaker increases the likelihood of transition from low to high parities, especially for women who were born in the South and Centre of the country. Further, as suggested by economic intuition, education at age 12 is found to reduce women's odds of having a large family (for further detail on the economic theory of fertility behaviour see Chapter 2 and, among others, Willis 1973, Becker and Lewis 1973, Becker and Barro 1988, Easterlin 1975).

Conditional on observing a count larger than three, Catholic individuals are expected to have a significantly lower fertility than non-Catholics only for women who were born in the south of the country. A similar observation is valid for ethnic group. That is, being an indigenous language speaker is associated significantly with increases in completed fertility exclusively for women who were born in the South.

Appendix

Table 5.1 Descriptive Statistics

Variable	Description	Mean	Std. Dev.	Min	Max
Age	age in years	45.93	4.21	40	54
Children	number of children ever born alive	4.43	2.75	0	18
Edu12	Completed years of schooling at age 12	4.01	2.33	0	6
Religion and Ethnic group					
Catholic	=1 if Catholic; 0 otherwise	0.90	-	-	-
indspker	=1 if indian language speaker; 0 otherwise	0.09	-	-	-
Cohort			-	-	-
c4044 (base group)	=1 if born within 1940-1944; 0 otherwise	0.10	-	-	-
c4549	=1 if born within 1945-1949; 0 otherwise	0.29	-	-	-
c5054	=1 if born within 1950-1954; 0 otherwise	0.36	-	-	-
c5557	=1 if born within 1955-1957; 0 otherwise	0.25	-	-	-
Birth Place					
MexCity (base group)	=1 if born in Mex City; 0 otherwise	0.05	-	-	-
North	=1 if born in North; 0 otherwise	0.23	-	-	-
Centre	=1 if born in Cebtre; 0 otherwise	0.54	-	-	-
South	=1 if born in South; 0 otherwise	0.18	-	-	-
Number of observations					19,477

Table 5.2 Descriptive Statistics -- Region (split according to birthplace dummies)

	Mean	Std. Dev.	Min	Max		Mean	Std. Dev.	Min	Max
Mexico City					**Centre**				
Age	45.40	4.10	40	54	Age	46.01	4.24	40	54
Children	2.91	1.78	0	12	Children	4.68	2.87	0	18
Edu12	5.69	1.12	0	6	Edu12	3.75	2.39	0	6
Catholic	0.90	-	-	-	Catholic	0.93	-	-	-
indspker	0.01	-	-	-	indspker	0.07	-	-	-
c4044	0.07	-	-	-	c4044	0.10	-	-	-
c4549	0.26	-	-	-	c4549	0.30	-	-	-
c5054	0.39	-	-	-	c5054	0.36	-	-	-
c5557	0.29	-	-	-	c5557	0.25	-	-	-
N. obs				967	N. obs				10537
North					**South**				
Age	45.90	4.20	40	54	Age	45.91	4.16	40	54
Children	4.11	2.48	0	16	Children	4.51	2.78	0	16
Edu12	4.80	1.85	0	6	Edu12	3.29	2.45	0	6
Catholic	0.89				Catholic	0.81	-	-	-
indspker	0.02	-	-	-	indspker	0.29	-	-	-
c4044	0.09	-	-	-	c4044	0.09	-	-	-
c4549	0.29	-	-	-	c4549	0.28	-	-	-
c5054	0.37	-	-	-	c5054	0.38	-	-	-
c5557	0.25	-	-	-	c5557	0.25	-	-	-
N. obs				4532	N. obs				3441

Table 5.3 Empirical distribution of Children and a Poisson distribution with mean of 4.4

Count	Obs.	Share	Poisson
0	1,211	0.0622	0.012
1	1,134	0.0582	0.054
2	2,504	0.1286	0.119
3	3,383	0.1737	0.174
4	2,905	0.1492	0.192
5	2,349	0.1206	0.169
6	1,818	0.0933	0.124
7	1,390	0.0714	0.078
8	1,036	0.0532	0.043
9	746	0.0383	0.021
10	474	0.0243	0.009
11	241	0.0124	0.004
12-18	286	0.0147	0.002
Total	19,477	1.000	1.000

Table 5.4 Likelihood of high parities given y > 3

Count	4	5	6	7-18	Total
No. obs.	2,905	2,349	1,818	4,173	11,245
Pr(count | y>3)	0.26	0.21	0.16	0.37	1.00

Figure 5.1 Empirical distribution of Children and a theoretical Poisson with mean 4.4

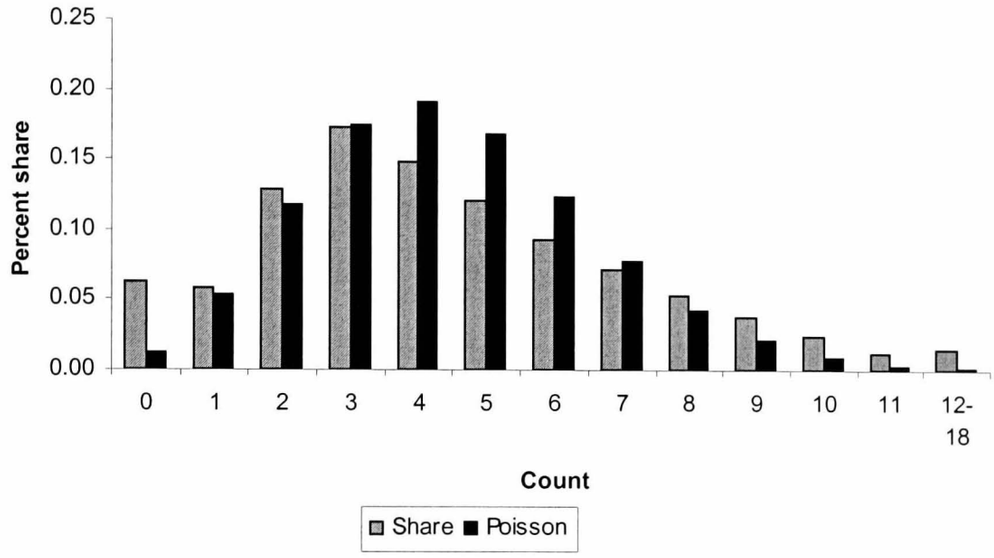


Figure 5.2 Double-Hurdle Model Structure

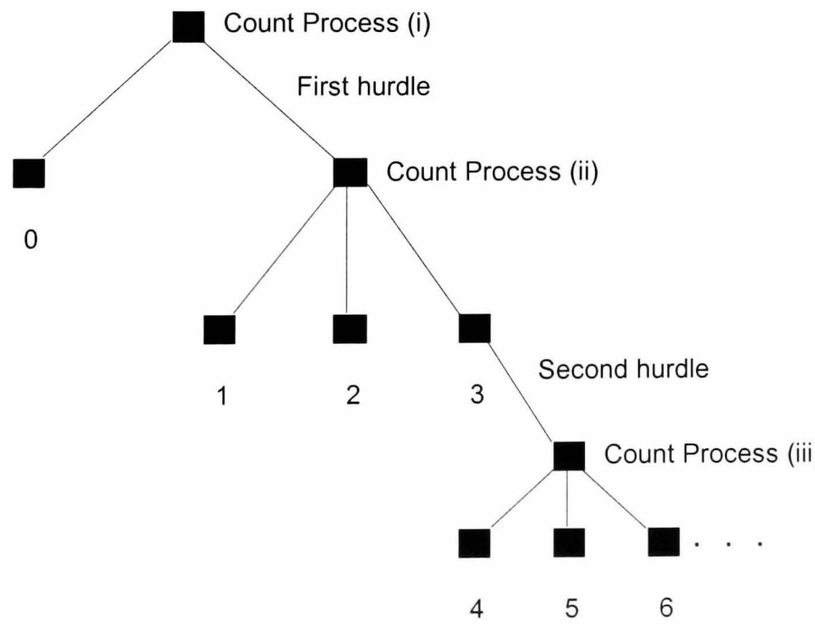


Table 5.5 Standard Hurdle Model -- National Data
Coefficient [Std. Err.]

Count Process	(1) No Het.	(2) Normal Het.
At Zero		
Constant	1.1547 [0.0675]**	1.1547 [0.0675]**
Education, Religion and Ethnic group		
Catholic	-0.0525 [0.0342]	-0.0525 [0.0342]
Indspker	-0.0728 [0.0381]	-0.0728 [0.0381]
Edu12	-0.0314 [0.0047]**	-0.0314 [0.0047]**
Cohort (base 1940-1944)		
c4549	0.0230 [0.0382]	0.0230 [0.0382]
c5054	0.0494 [0.0374]	0.0494 [0.0374]
c5557	0.0225 [0.0390]	0.0225 [0.0390]
Birthplace (base Mexico City)		
North	0.0558 [0.0487]	0.0558 [0.0487]
Centre	0.0001 [0.0465]	0.0001 [0.0465]
South	0.0460 [0.0519]	0.0460 [0.0519]
Larger than zero		
Constant	1.7903 [0.0260]**	1.7740 [0.0280]**
Education, Religion and Ethnic group		
Catholic	-0.0475 [0.0112]**	-0.0482 [0.0124]**
Indspker	0.0289 [0.0120]*	0.0321 [0.0133]*
Edu12	-0.0878 [0.0015]**	-0.0891 [0.0017]**
Cohort (base 1940-1944)		
c4549	-0.0836 [0.0120]**	-0.0848 [0.0134]**
c5054	-0.1868 [0.0120]**	-0.1895 [0.0133]**
c5557	-0.2563 [0.0129]**	-0.2588 [0.0143]**
Birthplace (base Mexico City)		
North	0.2669 [0.0220]**	0.2676 [0.0233]**
Centre	0.3053 [0.0214]**	0.3060 [0.0227]**
South	0.2057 [0.0228]**	0.2036 [0.0243]**
σ^2	-	0.0411 [0.0027]**
Log-likelihood	-44144.42	-43996.48
AIC	88,328.84	88,034.95
CIAC	88,506.38	88,221.37
Number of observations	19,477	19,477

Note: ** significant at 1% ; * significant at 5%. An Extreme Value function is used for the hurdle at zero.

Table 5.6 Poisson Double Hurdle Model -- National Data
Coefficient [Std. Err.]

	No Het.	Normal Het.		
Count Process	(1)	(2) $\theta_0=\theta_1=0$	(3) $\theta_0=\theta_1=1$	(4) θ_0 , θ_1 free
At Zero [vector β_0] -- Process (i)				
Constant	1.1547 [0.0675]**	1.1547 [0.0675]**	1.1800 [0.0698]**	1.1567 [0.0755]
Education, Religion and Ethnic group				
Catholic	-0.0525 [0.0342]	-0.0525 [0.0342]	-0.0543 [0.0353]	-0.0527 [0.0344]
Indspker	-0.0728 [0.0381]	-0.0728 [0.0381]	-0.0753 [0.0393]	-0.0731 [0.0383]
Edu12	-0.0314 [0.0047]**	-0.0314 [0.0047]**	-0.0324 [0.0049]**	-0.0314 [0.0049]**
Cohort (base 1940-1944)				
c4549	0.0230 [0.0382]	0.0230 [0.0382]	0.0237 [0.0394]	0.0230 [0.0383]
c5054	0.0494 [0.0374]	0.0494 [0.0374]	0.0513 [0.0386]	0.0496 [0.0376]
c5557	0.0225 [0.0390]	0.0225 [0.0390]	0.0235 [0.0402]	0.0226 [0.0391]
Birthplace (base Mexico City)				
North	0.0558 [0.0487]	0.0558 [0.0487]	0.0575 [0.0502]	0.0559 [0.0489]
Centre	0.0001 [0.0465]	0.0001 [0.0465]	0.0001 [0.0480]	0.0001 [0.0467]
South	0.0460 [0.0519]	0.0460 [0.0519]	0.0475 [0.0535]	0.0462 [0.0521]
At one-to-three [vector β_1] -- Process (ii)				
(vector β_1)				
Constant	1.7142 [0.0328]**	1.7142 [0.0328]**	1.7370 [0.0344]**	1.7142 [0.0328]**
Education, Religion and Ethnic group				
Catholic	-0.0509 [0.0157]**	-0.0509 [0.0157]**	-0.0535 [0.0165]**	-0.0509 [0.0157]**
Indspker	0.0408 [0.0181]*	0.0408 [0.0181]*	0.0430 [0.0191]*	0.0408 [0.0181]*
Edu12	-0.0842 [0.0022]**	-0.0842 [0.0022]**	-0.0888 [0.0024]**	-0.0842 [0.0022]**
Cohort (base 1940-1944)				
c4549	-0.0535 [0.0184]**	-0.0535 [0.0184]**	-0.0564 [0.0194]**	-0.0535 [0.0184]**
c5054	-0.1326 [0.0179]**	-0.1326 [0.0179]**	-0.1391 [0.0190]**	-0.1326 [0.0179]**
c5557	-0.1770 [0.0187]**	-0.1770 [0.0187]**	-0.1853 [0.0198]**	-0.1770 [0.0187]**
Birthplace (base Mexico City)				
North	0.2523 [0.0248]**	0.2523 [0.0248]**	0.2605 [0.0256]**	0.2523 [0.0248]**
Centre	0.2616 [0.0239]**	0.2616 [0.0239]**	0.2702 [0.0248]**	0.2616 [0.0239]**
South	0.1597 [0.0262]**	0.1597 [0.0262]**	0.1638 [0.0271]**	0.1597 [0.0262]**
Larger than three [vector β_2] -- Process (iii)				
Constant	1.7752 [0.0522]**	1.7564 [0.0542]**	1.7429 [0.0537]**	1.7554 [0.0550]**
Education, Religion and Ethnic group				
Catholic	-0.0348 [0.0156]*	-0.0359 [0.0168]*	-0.0379 [0.0164]*	-0.0361 [0.0168]*
Indspker	0.0129 [0.0156]	0.0163 [0.0169]	0.0161 [0.0165]	0.0160 [0.0169]
Edu12	-0.0753 [0.0023]**	-0.0768 [0.0024]**	-0.0798 [0.0025]**	-0.0769 [0.0027]**
Cohort (base 1940-1944)				
c4549	-0.0911 [0.0153]**	-0.0934 [0.0166]**	-0.0944 [0.0162]**	-0.0933 [0.0167]**
c5054	-0.2025 [0.0156]**	-0.2075 [0.0170]**	-0.2103 [0.0166]**	-0.2073 [0.0171]**
c5557	-0.3030 [0.0180]**	-0.3086 [0.0193]**	-0.3130 [0.0190]**	-0.3084 [0.0195]**
Birthplace (base Mexico City)				
North	0.2831 [0.0494]**	0.2810 [0.0509]**	0.2913 [0.0504]**	0.2811 [0.0510]**
Centre	0.3570 [0.0486]**	0.3559 [0.0500]**	0.3657 [0.0496]**	0.3557 [0.0501]**
South	0.2787 [0.0499]**	0.2740 [0.0515]**	0.2816 [0.0510]**	0.2740 [0.0516]**
σ^2	-	0.0340 [0.0042]**	0.0239 [0.0038]**	0.0341 [0.0042]**
θ_0	-	set to zero	set to one	-0.0110 [0.2300]
θ_1	-	set to zero	set to one	0.2745 [1.6784]
Log-likelihood	-43,980.42	-43,941.15	-43,956.11	-43,941.16
AIC	88,020.84	87,944.30	87,974.22	87,946.32
CIAC	88,287.15	88,219.49	88,249.41	88,230.38
Number of observations	19,477	19,477	19,477	19,477

Note: ** significant at 1% ; * significant at 5%.

Table 5.7 Model Selection
Poisson Double-Hurdle with Normal Heterogeneity -- National Data

Case	H_0	H_1	Test type	$\chi 2$ [p-val]	Inference
1	$\theta_0 = 0, \sigma^2 \neq 0$	$\theta_0 \neq 0, \sigma^2 \neq 0$	LRT	0.016 [0.8993]	Do not reject H_0
2	$\theta_1 = 0, \sigma^2 \neq 0$	$\theta_1 \neq 0, \sigma^2 \neq 0$	LRT	0.018 [0.8933]	Do not reject H_0
3	$\sigma^2 = 0$	$\sigma^2 \neq 0$	BVLRT	78.53 [0.0000]	Reject H_0
4	$\theta_0 = \theta_1 = 0, \sigma^2 \neq 0$	$\theta_0 \neq 0, \theta_1 = 0, \sigma^2 \neq 0$	LRT	0.032 [0.858]	Do not reject H_0
5	$\theta_0 = \theta_1 = 0, \sigma^2 \neq 0$	$\theta_0 = 0, \theta_1 \neq 0, \sigma^2 \neq 0$	LRT	0.002 [0.9643]	Do not reject H_0
6	$\theta_0 = \theta_1 = 1, \sigma^2 \neq 0$	$\theta_0 \neq \theta_1 \neq 1, \sigma^2 \neq 0$	LRT	29.90 [0.0000]	Reject H0

Note: Boundary-value likelihood ratio test is abbreviated as BVLRT. Likelihood ratio test is abbreviated as LRT.

Table 5.8 Likelihood Ratio Tests

	LR	P-val	Inference
H_0: $\beta_0=\beta_1$ vs. H_1: $\beta_0\neq\beta_1$			
National	1610.30	0.0000	Reject H_0
Mex City	22.64	0.0122	Reject H_0
North	269.44	0.0000	Reject H_0
Centre	1295.47	0.0000	Reject H_0
South	251.18	0.0000	Reject H_0
C4049	26.08	0.000496	Reject H_0
C5059	788.16	0.0000	Reject H_0
H_0: $\beta_1=\beta_2$ vs. H_1: $\beta_1\neq\beta_2$			
National	164.27	0.0000	Reject H_0
Mex City	20.36	0.0260	Reject H_0
North	12.58	0.2483	Do no reject H_0
Centre	255.82	0.0000	Reject H_0
South	35.92	0.0001	Reject H_0
C4049	56.05	0.0000	Reject H_0
C5059	22.58	0.0020	Reject H_0
H_0: $\beta_0=\beta_1=\beta_2$ vs. H_1: $\beta_0\neq\beta_1\neq\beta_2$			
National	2339.49	0.0000	Reject H_0
Mex City	35.41	0.0180	Reject H_0
North	308.92	0.0000	Reject H_0
Centre	1584.50	0.0000	Reject H_0
South	345.58	0.0000	Reject H_0
C4049	105.19	0.0000	Reject H_0
C5059	906.22	0.0000	Reject H_0

Note: Tests based on best fitting Double-Hurdle Models.

Table 5.9 Observed and predicted sample distribution -- National data

Count	Obs.	Standard Hurdle		Double-Hurdle (best fit)	
		No Het.	Normal Het.	No Het.	Normal Het. ($\theta_0=\theta_1=0$)
0	0.062	0.062	0.062	0.062	0.062
1	0.058	0.058	0.070	0.066	0.066
2	0.129	0.113	0.122	0.125	0.125
3	0.174	0.152	0.152	0.163	0.163
4	0.149	0.160	0.153	0.136	0.145
5	0.121	0.142	0.132	0.128	0.128
6	0.093	0.111	0.103	0.106	0.102
7	0.071	0.079	0.074	0.079	0.074
8	0.053	0.052	0.050	0.054	0.051
9	0.038	0.032	0.033	0.035	0.033
10	0.024	0.018	0.020	0.021	0.021
11	0.012	0.010	0.012	0.012	0.013
12-18	0.015	0.010	0.016	0.013	0.016
chi-square		371	213	150	116
Pr > chi-square		0.0000	0.0000	0.0000	0.0000
logL		-44144	-43996	-43980	-43941
AIC		88,329	88,035	88,021	87,944
CIAC		88,506	88,221	88,287	88,219

Note: Sample size is 19,477.

Table 5.10 Poisson Double Hurdle Model -- Regional Results (Best fitting model)
Coefficient [Std. Err.]

	(1) Mex City	(2) North	(3) Centre	(4) South
Count Process				
At Zero [vector β_0] -- Process (i)				
Constant	1.4155 [0.3385]**	1.2455 [0.1087]**	1.1121 [0.0692]**	1.2510 [0.1080]**
Education, Religion and Ethnic group				
Catholic	-0.1366 [0.1575]	0.0046 [0.0682]	-0.0489 [0.0532]	-0.0968 [0.0646]
Indspker	-0.2638 [0.4871]	-0.2505 [0.1596]	-0.0482 [0.0562]	-0.0555 [0.0571]
Edu12	0.0024 [0.0398]	-0.0434 [0.0121]**	-0.0331 [0.0060]**	-0.0193 [0.0105]
Cohort (base 1940-1944)				
c4549	-0.2458 [0.2161]	-0.0443 [0.0819]	0.0733 [0.0499]	-0.0201 [0.0969]
c5054	-0.3898 [0.2084]	0.0480 [0.0809]	0.0839 [0.0489]	0.0250 [0.0944]
c5557	-0.3265 [0.2129]	0.0127 [0.0840]	0.0914 [0.0513]	-0.1216 [0.0973]
At one-to-three [vector β_1] -- Process (ii)				
Constant	2.2446 [0.1622]**	2.0979 [0.0495]**	1.9182 [0.0370]**	1.8954 [0.0483]**
Education, Religion and Ethnic group				
Catholic	0.0681 [0.0790]	-0.0958 [0.0313]**	-0.0019 [0.0257]	-0.1010 [0.0287]**
Indspker	0.0454 [0.3040]	-0.0379 [0.0806]	0.0012 [0.0281]	0.1015 [0.0260]**
Edu12	-0.1770 [0.0220]**	-0.0940 [0.0054]**	-0.0859 [0.0036]**	-0.0770 [0.0049]**
Cohort (base 1940-1944)				
c4549	-0.1037 [0.0896]	-0.0748 [0.0385]	-0.0354 [0,0255]	-0.0801 [0.0447]
c5054	-0.2653 [0.0870]**	-0.1887 [0.0376]**	-0.0889 [0.0251]**	-0.1748 [0.0432]**
c5557	-0.3413 [0.0904]**	-0.2384 [0.0392]**	-0.1469 [0.0262]**	-0.1676 [0.0454]**
Larger than three [vector β_2] -- Process (iii)				
Constant	1.8688 [0.2595]**	2.1311 [0.0516]**	2.0352 [0.0304]**	2.0693 [0.0413]**
Education, Religion and Ethnic group				
Catholic	-0.0919 [0.1845]	-0.0493 [0.0396]	0.0120 [0.0250]	-0.1071 [0.0282]**
Indspker	-0.3206 [0.6569]	-0.0621 [0.0971]	-0.0361 [0.0234]	0.0938 [0.0257]**
Edu12	-0.0733 [0.0279]**	-0.0867 [0.0060]**	-0.0761 [0.0032]**	-0.0843 [0.0061]**
Cohort (base 1940-1944)				
c4549	-0.2200 [0.1592]	-0.1274 [0.0380]**	-0.0671 [0.0212]**	-0.1314 [0.0390]**
c5054	-0.4044 [0.1644]**	-0.3055 [0.0455]**	-0.1804 [0.0216]**	-0.1810 [0.0390]**
c5557	-0.8182 [0.2170]**	-0.4044 [0.0455]**	-0.2785 [0.0245]**	-0.2823 [0.0446]**
σ^2	0.1520 [0.0507]**	0.0507 [0.0111]**	0.0277 [0.2688]**	0.0304 [0.0096]**
θ_0	set to zero	set to zero	set to zero	set to zero
θ_1	set to zero	set to zero	0.7686 [0.2688]**	set to zero
Log-likelihood	-1,793.98	-9,839.05	-24,332.7	-7,799.3
AIC	3,649.96	19,740.10	48,729.40	15,660.60
CIAC	3,832.06	19,970.09	48,993.80	15,882.05
Number of observations	967	4,532	10,537	3,441

Note: ** significant at 1% ; * significant at 5%.

Table 5.11 Predicted Probabilities -- Double Hurdle Poisson Model

	Caracteristics	Pr(j = 0)	Pr(1 < j ≤ 3 | j > 0)	Pr(j > 6 | j > 3)
National				
(1)	edu12=mean, all dummies set to zero	0.0609	0.4302	0.4165
(2)	edu12=mean,catholic=1, other dummies set to zero	0.0703	0.4684**	0.3947*
(3)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.0847	0.4378*	0.4044
(4)	edu12=mean,catholic=1, c4549=1, other dummies set to zero	0.0661	0.5081**	0.3416**
(5)	edu12=mean,catholic=1, c5054=1, other dummies set to zero	0.0615	0.5648**	0.2842**
(6)	edu12=mean,catholic=1, c5559=1, other dummies set to zero	0.0662	0.5955**	0.24**
(7)	edu12=mean,catholic=1, north=1, other dummies set to zero	0.0604	0.2818**	0.582**
(8)	edu12=mean,catholic=1,centre=1, other dummies set to zero	0.0703	0.2753**	0.6361**
(9)	edu12=mean,catholic=1, south=1, other dummies set to zero	0.0620	0.3487**	0.5769**
(10)	edu12=0,catholic=1, other dummies set to zero	0.0493**	0.2243**	0.6015**
(11)	edu12=5,catholic=1, other dummies set to zero	0.0762**	0.5298**	0.3511**
(12)	edu12=6,catholic=1, other dummies set to zero	0.0826**	0.5891**	0.3107**
Mex City				
(1)	edu12=mean, all dummies set to zero	0.0169	0.3126	0.4987
(2)	edu12=mean,catholic=1, other dummies set to zero	0.0285	0.2646	0.4381
(3)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.0650	0.2342	0.2648
(4)	edu12=mean,catholic=1, c4549=1, other dummies set to zero	0.0619	0.3386	0.3124**
(5)	edu12=mean,catholic=1, c5054=1, other dummies set to zero	0.0899	0.4596**	0.2299**
(6)	edu12=mean,catholic=1, c5559=1, other dummies set to zero	0.0768	0.5159**	0.11**
(10)	edu12=0,catholic=1, other dummies set to zero	0.0275	0.0096**	0.6423**
(11)	edu12=5,catholic=1, other dummies set to zero	0.0287	0.3919**	0.3934**
(12)	edu12=6,catholic=1, other dummies set to zero	0.0290	0.5239**	0.3514**
North				
(1)	edu12=mean, all dummies set to zero	0.0539	0.1886	0.6471
(2)	edu12=mean,catholic=1, other dummies set to zero	0.0532	0.2493**	0.6113
(3)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.1020	0.2752	0.5666
(4)	edu12=mean,catholic=1, c4549=1, other dummies set to zero	0.0604	0.3012	0.5207**
(5)	edu12=mean,catholic=1, c5054=1, other dummies set to zero	0.0461	0.3851**	0.4044**
(6)	edu12=mean,catholic=1, c5559=1, other dummies set to zero	0.0513	0.4224**	0.3474**
(10)	edu12=0,catholic=1, other dummies set to zero	0.0305**	0.0624**	0.8459**
(11)	edu12=5,catholic=1, other dummies set to zero	0.0602**	0.3143**	0.5498**
(12)	edu12=6,catholic=1, other dummies set to zero	0.0678**	0.3838**	0.4897**
Centre				
(1)	edu12=mean, all dummies set to zero	0.0698	0.2849	0.6084
(2)	edu12=mean,catholic=1, other dummies set to zero	0.0792	0.2862	0.6171
(3)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.0892	0.2853	0.5911
(4)	edu12=mean,catholic=1, c4549=1, other dummies set to zero	0.0653	0.3115	0.5687**
(5)	edu12=mean,catholic=1, c5054=1, other dummies set to zero	0.0634	0.3506**	0.4898**
(6)	edu12=mean,catholic=1, c5559=1, other dummies set to zero	0.0621	0.3939**	0.4259**
(10)	edu12=0,catholic=1, other dummies set to zero	0.0553**	0.092**	0.8259**
(11)	edu12=5,catholic=1, other dummies set to zero	0.0859**	0.3478**	0.5629**
(12)	edu12=6,catholic=1, other dummies set to zero	0.0931**	0.412**	0.5096**
South				
(1)	edu12=mean, all dummies set to zero	0.0394	0.2759	0.6092
(2)	edu12=mean,catholic=1, other dummies set to zero	0.0531	0.3486**	0.5328**
(3)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.0622	0.2755**	0.5996**
(4)	edu12=mean,catholic=1, c4549=1, other dummies set to zero	0.0563	0.4086	0.4444**
(5)	edu12=mean,catholic=1, c5054=1, other dummies set to zero	0.0493	0.4796**	0.4132**
(6)	edu12=mean,catholic=1, c5559=1, other dummies set to zero	0.0743	0.4742**	0.3541**
(10)	edu12=0,catholic=1, other dummies set to zero	0.0419	0.1477**	0.7724**
(11)	edu12=5,catholic=1, other dummies set to zero	0.0561	0.4056**	0.4757**
(12)	edu12=6,catholic=1, other dummies set to zero	0.0593	0.4636**	0.4214**

Note: ** (*) indicates that the relevant coefficient in Table 6 and 10 is significant at 1% (5%) of significance.

Table 5.11 Predicted Probabilities -- Double Hurdle Poisson Model (Cont)

	Caracteristics	Pr(j = 0)	Pr(1 < j ≤ 3 | j > 0)	Pr(j > 6 | j > 3)
Cohort 1940-1949				
(I)	edu12=mean, all dummies set to zero	0.0001	0.3900	0.4055
(II)	edu12=mean,catholic=1, other dummies set to zero	0.0001	0.4320	0.3943
(III)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.0001	0.3937	0.3774
(VII)	edu12=mean,catholic=1, north=1, other dummies set to zero	0.0001	0.2580**	0.5687**
(VIII)	edu12=mean,catholic=1,centre=1, other dummies set to zero	0.0001	0.2766**	0.6059**
(IX)	edu12=mean,catholic=1, south=1, other dummies set to zero	0.0001	0.3453*	0.5493**
(X)	edu12=0,catholic=1, other dummies set to zero	0.0001**	0.1924**	0.5709**
(XI)	edu12=5,catholic=1, other dummies set to zero	0.0001**	0.4947**	0.3565**
(XII)	edu12=6,catholic=1, other dummies set to zero	0.0001**	0.5562**	0.3210**
Cohort 1950-1957				
(I)	edu12=mean, all dummies set to zero	0.0746	0.5639	0.2424
(II)	edu12=mean,catholic=1, other dummies set to zero	0.0765	0.5989**	0.2200*
(III)	edu12=mean,catholic=1,indspker=1, other dummies set to zero	0.0899	0.5727	0.2410*
(VII)	edu12=mean,catholic=1, north=1, other dummies set to zero	0.0528*	0.4008**	0.3769**
(VIII)	edu12=mean,catholic=1,centre=1, other dummies set to zero	0.0625	0.3771**	0.4410**
(IX)	edu12=mean,catholic=1, south=1, other dummies set to zero	0.0562	0.4622**	0.3877**
(X)	edu12=0,catholic=1, other dummies set to zero	0.0620**	0.3369**	0.3976**
(XI)	edu12=5,catholic=1, other dummies set to zero	0.0804**	0.6563**	0.1882**
(XII)	edu12=6,catholic=1, other dummies set to zero	0.0844**	0.7090**	0.1602**

Note: ** (*) indicates that the relevant coefficient in Table 5.6 and 5.10 is significant at 1% (5%) of significance. Marginal Effetcs can be obtained taking differences between two alternative rows. For intance, marginal effects of Catholic can be obtained as row (I) minus row (II).

Table 5.12 Poisson Double Hurdle Model -- Selected Groups of Age
Best fitting model

Coefficient [Std. Err.]

	1940-1949	1950-1957
Count Process		
At Zero [vector β_0] -- Process (i)		
Constant	4.3307 [1.6474]**	1.0323 [0.0734]**
Education, Religion and Ethnic group		
Catholic	-0.3873 [0.1982]	-0.0096 [0.0422]
Indspker	-0.2673 [0.2161]	-0.0649 [0.0482]
Edu12	-0.1552 [0.0550]**	-0.0196 [0.0061]**
Birthplace (base Mexico City)		
North	-0.3431 [0.2846]	0.1345 [0.0590]*
Centre	-0.5101 [0.2945]	0.0756 [0.0560]
South	-0.3231 [0.2976]	'0.1134 [0.0628]
At one-to-three [vector β_1] -- Process (ii)		
Constant	1.7704 [0.0553]**	1.5522 [0.0374]**
Education, Religion and Ethnic group		
Catholic	-0.0557 [0.0291]	-0.0509 [0.0198]**
Indspker	0.0508 [0.0333]	0.0382 [0.0229]
Edu12	-0.0849 [0.0045]**	-0.0892 [0.0029]**
Birthplace (base Mexico City)		
North	0.2382 [0.0446]**	0.2721 [0.0313]**
Centre	0.2111 [0.0432]**	0.3036 [0.0302]**
South	0.1156 [0.0473]*	0.1904 [0.0329]**
Larger than three [vector β_2] -- Process (iii)		
Constant	1.6966 [0.0684]**	1.4742 [0.0830]**
Education, Religion and Ethnic group		
Catholic	-0.0186 [0.0230]	-0.0568 [0.0245]*
Indspker	-0.0289 [0.0239]	0.0535 [0.0240]*
Edu12	-0.0663 [0.0033]**	-0.0909 [0.0036]**
Birthplace (base Mexico City)		
North	0.2631 [0.0650]**	0.3293 [0.0806]**
Centre	0.3147 [0.0640]**	0.4347 [0.0792]**
South	0.2357 [0.0663]**	0.3479 [0.0811]**
σ^2	0.0207 [0.0073]**	0.0570 [0.0068]**
θ_0	-13.7088 [6.2986]**	set to zero
θ_1	1.6505 [0.3898]**	set to zero
Log-likelihood	-17657.25	-26,267.20
AIC	35,362.5	52,578.4
CIAC	35,552.40	52,741.13
Number of observations	7427	12050

Note: ** significant at 1% ; * significant at 5%.

Chapter 6

Planned fertility and family background: A quantile regression for counts analysis*

6.1 Introduction

Sociologists and demographers have long argued that one of the driving forces of a demographic transition is the change in individual's attitudes towards family size. In pre-transitional societies, it is said, people aim to form large families. In post-transitional societies, in contrast, large families are rather avoided. The diffusion of new social norms and economic development are commonly thought to be the causes of the change in people's attitudes towards fertility (see for instance Gustavus and Nam 1970).

---

* I am grateful to Joao Santos Silva for useful comments to this Chapter. I am also grateful to Joao Santos Silva for his TSP code to estimate Quantiles for Counts. His code was a valuable basis for writing the Stata code that is used in the present Chapter.

The idea that cultural changes may be behind important demographic innovations has led population analysts to measure, or attempt to measure, fertility preferences. Today, most fertility surveys contain questions enquiring about a woman's preferred lifetime number of children (ideal or planned fertility). Though various ways of formulating the question exist, the common premise in all surveys is that respondents should imagine themselves to be at the beginning of their fertility-planning horizon, as surveys aim to collect a time-independent preference index – which is an individual-specific characteristic. Ideal fertility data has proved to be highly predictive of women's future fertility behaviour (Knodel and Prachuabmoh 1973, Pritchett 1994).

There are, however, some criticisms. Firstly, there is the suspicion that a woman with children would never state that some of her offspring were unwanted, even when that is the case (Knodel and Prachuabmoh 1973). Consequently, current and planned fertility are difficult to set apart once motherhood has been initiated. Secondly, there is the possibility that fertility plans might change over time in response to experience and the arrival of relevant information (Lee 1980) – say, for instance, that individuals learn the cost and benefits of children only after they have entered parenthood. This argument implies then that ideal (planned) fertility should be treated as a variable reflecting the outcome of a dynamic optimisation exercise, rather than one reflecting an exogenous time-independent individual characteristic. In fact, it is clear from this last argument that a distinction between planned fertility and preferences towards children should be made. While preferences are most likely unobserved individual characteristics, planned fertility is the result of an optimisation process in which preferences, income and prices play relevant roles. Therefore, although

ideal fertility surveys aim to measure preferences towards children, an economic interpretation of the data would suggest that it is information on planned fertility, and not preferences towards children, which is actually recorded.

Criticism of ideal fertility data has tended to hold back the study of how individuals form their fertility plans. Important questions remain. Does economic theory of fertility behaviour apply within this context? Does family background matter? The present chapter reports a study of these issues for the case of Mexico. Attention is concentrated on analysing the fertility plans of young Mexican women aged between 15 and 17 years who live with at least one biological parent and who have neither initiated independent economic life nor entered motherhood. Clearly, in such a context Knodel and Prachuabmoh's and Lee's criticisms of ideal fertility data become weak as no confusion between current and planned fertility may exist and individuals only have information about their initial endowments: i.e., women are, effectively, at the beginning of their fertility and labour planning horizon. In this chapter count data models are used as the main tool of analysis, including an innovative technique for estimating quantile regressions for counts.

Two topics are primarily discussed. Firstly, the study enquires about the potential impact that family structure may have on fertility plans. In particular, this chapter analyses how the behaviour of individuals who live in families with an absent biological parent differ from the behaviour of individuals who live in intact families (i.e., families where both biological parents are present). Secondly, the study intends to establish whether or not socio-economic characteristics of the head of the family are relevant in the formation of young women's plans. To avoid ambiguity, family structure is defined as an individual-specific

characteristic rather than a common feature of all youngsters living in the same household. This approach avoids the confusion that otherwise arises in the cases of families composed of step-parents and step-children.

The remainder of the present chapter is organized as follows. Sections two and three describe relevant data issues. Sections four presents the econometric methodology and section five discusses empirical findings. Section six concludes.

6.2 Data issues

Most fertility surveys collect data on desired fertility as the direct response of women to questions enquiring about their preferred lifetime number of children. Various ways of presenting the question exist. Nonetheless in all cases there is the common feature of instructing people to hypothetically set themselves at the onset of their fertility-planning horizon. The idea is to generate a time-independent preference index that might be treated as an individual-specific characteristic. Despite this effort, ideal fertility data has traditionally been subject to scepticism about its quality and meaning. For this reason most studies of desired fertility concentrate on validating the data as a meaningful tool for scientific analysis.

Two early studies established that ideal fertility data achieve a minimum of quality. Knodel and Prachuabmoh (1973) found in Thailand that a large majority of women were able to give a numerical response to questions on desired family size. In addition, the authors documented that in most cases women's ideal fertility – a non-negative integer number – was significantly different from their current (realised) fertility, even though some degree of ex-post rationalization of achieved family size appeared to occur. More importantly, the study showed that

women consistently state that no additional children are wanted when the ideal target has been reached, or that more pregnancies are planned otherwise. According to Knodel and Prachuabmoh, in either case appropriate contraception actions are taken. Freedman et. al. (1975) report similar findings for a longitudinal study on Taiwanese data. In addition, Freedman et. al. (1975) conclude that statements about ideal fertility are highly predictive of future fertility behaviour and contraception use.

More recently Lee (1980) suggested that ideal fertility should not be interpreted as a preference index but as a moving target. From Lee's perspective, individuals form their fertility plans at the beginning of their fertile life on the basis of all available information at that time. Then, as new pieces of information arrive, individuals update expectations and plans. In this context, desired fertility is thought to be the outcome of a dynamic optimisation problem – a point that is not explicitly noted by Lee but is largely recognized by economic theory (see for instance Moffit 1984, Happel, Hill, and Low 1984, Ward and Butz 1980). Therefore, parenthood and labour market experience, being major sources of information, are likely to induce significant changes in desired fertility. Lee's point suggests then that ideal fertility should be treated as a time-varying election variable rather than as a time-fixed individual-specific characteristic. Hence, if fertility plans are to be studied, the analyst should take due care in selecting individuals who are at the same point in their lifecycle and have, broadly, similar informational sets. Otherwise, significant bias may be induced.[22] Since

[22] If fertility plans of women at different points of their life cycle are studied, ex-post rationalization of realised fertility may bias the analysis. This bias is likely to exist even though fertility surveys try to mitigate the problem by providing

controlling for all relevant experience is difficult, restricting attention to individuals who are effectively at the beginning of their fertility-planning horizon and have not initiated independent life seems to be the best strategy. This avenue is taken in the present work.

Another important point raised by Lee's criticism is that planned (ideal) fertility and preferences towards children should be clearly distinguished by the researcher. While preferences are most likely unobservable characteristics, planned fertility is the outcome of an optimisation process. Regarding this point economic intuition suggests that, despite the efforts of the interviewer, women are likely to give information about their planned fertility whenever questions about their most preferred lifetime (ideal) fertility are asked. Ideal fertility data should, then, be interpreted as information on fertility plans rather than information on preferences towards children.

Exploiting ideal fertility data poses other important challenges to the researcher. Jensen (1985) notices that non-numerical responses to ideal fertility questions – including the 'Up-to-God' and 'Don't know' response – account for about 2 to 13.8 percent of all recorded observations in fertility surveys collected in a number of Asian and Latin American countries during the 1970s. In the face of this problem some analysts treat non-numerical responses as missing values that are simply dropped from the studied sample. Other analysts, alternatively, treat non-numerical answers as an additional outcome that reflects the largest possible planned fertility. In the latter option, missing observations are re-coded

---

hypothetical assumptions – such as the 'going back to the time' statement- to women who are enquired about their desired family size.

as a somewhat arbitrary large number (for further reference on this topic see Riley, Hermalin, and Rosero-Bixby 1993). Jensen criticises both procedures because in neither case is there explicit recognition that 'fatalistic' responses might be generated as a result of a non-random selection process. If non-random selectiveness is present, the author argues, important problems of sample selection bias may appear (see Heckman 1979). In an exercise to assess the potential size of such a bias in Guatemala and India, Jensen finds no evidence to support his argument of non-random selectiveness – even though in both cases non-numeric responses represented at least 10 percent of the sample.

The present study is based on the analysis of data from the National Survey of Demographic Dynamics 1997 (ENADID). The study is based on data for women aged between 15 and 17 who at the time of the ENADID interview were living with at least one biological parent and had neither started independent economic life nor entered motherhood. Restricting attention to this set of women avoids any possible confusion between current and planned fertility and ensures that all individuals are broadly at the same point of their lifecycle and possess similar informational sets. The study is done conditional on this selection.

The ENADID contains 7,376 cases of women aged between 15 and 17. Information for family background is only recorded if the index women had not left the parental household by the time of the ENADID interview. From the original 7,376 observations available it is reported that 5% of women had left the parental household by the time of the ENADID interview, 5% had entered motherhood, and 18% were already active in the labour market (see Table 6.1, page 193). Excluding all these observations a sample of 5,768 women is

obtained.[23] The dependent variable is ideal fertility (planned completed fertility), **ideal**. The ENADID collected this variable as the direct response to the following question:

If you could choose the number of children for a lifetime, how many would you have?

From the 5,768 data entries that meet the selection criterion only 2.4% report a non-numerical value for **ideal**. Except for non-numerical responses, **ideal** is a non-negative variable that takes integer values. Non-numerical responses include the 'up to God' response and the 'will decide when I get married' response. A total of 5,628 valid cases remain once such observations are excluded. These 5,628 data entries constitute the sample used in the present work. Clearly, conditioning on the observation of a numerical response might lead to sample selection bias. However, as witnessed by the study of Jensen (1985), it is likely that, if present, such a bias will be small given that nearly 98% of all responses are numerical.

Ideal has an unconditional mean of 2.49 and an unconditional variance of 1.37. Hence, the data is under-dispersed (further details of this issue will be given later in the chapter). Table 6.2 (page 194) presents the main descriptive statistics. Additionally, Table 6.3 gives details of the empirical distribution of the **ideal**. For comparison proposes, Table 6.3 contains a theoretical Poisson distribution with a

[23]An assessment of the bias that this selection criteria potentially induces into the analysis is presented in section 6.6.4. Empirical evidence in section 6.6.4 suggest that, if present, such a bias is rather small.

mean of 2.5. Note that, unlike most data on realised completed fertility, **ideal** contains no evident excess of zeros relative to the Poisson distribution. There is, however, a considerable excess of two counts. This is a relevant feature of the data.

6.4 Variable definition

As was said in the introduction, the present work enquires about the effect that adverse events during childhood such as the death of a parent(s) or divorce might have on the fertility plans of young women in Mexico. To perform the study family structure is defined as an individual-specific characteristic. This approach avoids ambiguity that otherwise arises in cases of families composed of step-parents and step-children. To visualise this point, think of a couple with two children, one being the daughter of both spouses and the other being the daughter of just one spouse, say, the mother. In this case, if family structure were to be defined as a common feature of all children living in the same household, an intact/non-intact classification would be the best definition available. Such a definition would neglect the fact that one child is a biological daughter of both spouses while the other is a biological daughter of just one spouse. Clearly, this information is likely to be valuable. Thus, defining family structure as a common feature of all children in a family can neglect important pieces of information.

Three 'types' of family are identified: families where the index woman lives with both her biological parents, families where the biological father of the index woman is absent, and finally, families where the biological mother of the index woman is absent. To indicate to which type of family a young woman is attached

three indicator dummy variables are generated: **BothParents**, **AbsentFather**, and **AbsentMother**. 84% of the studied women live with both their parents, 14% live only with their mother, and the remaining 2% live only with their father (see Table 6.2, page 194).

Other explanatory variables are considered. Controls for characteristics such as **Age** and successfully completed years of education, **Edu**, are included (both variables recorded in years), along with the number of siblings of the reference woman, **Siblings**. Religion is controlled via a dummy variable, **Catholic**, which takes the value of one if the interviewed woman is Catholic and zero otherwise. Given that 90% of all Mexicans are Catholics this split of religious groups seems to be the most sensible. Ethnic background is controlled by a dummy variable, **Indspker**, which indicates if the reference woman can speak an indigenous language (**indspker**=1) or not (**indspker**=0). Notice that **Indspker** does not distinguish between individuals that belong to different indigenous groups. Therefore, the many indigenous ethnicities of Mexico are pooled in one category that is then contrasted with the non-indigenous population. Though there are important cultural differences across indigenous groups, this indigenous/non-indigenous classification is broadly justified because indigenous individuals, regardless of their specific ethnicity, commonly hold conservative attitudes towards contraception and family size (i.e., they prefer large families and oppose contraception) in relation to non-indigenous individuals (for a detailed discussion of **indspker** see chapter 4, section 2.3).

To control for the location of the parental household a set of dummy variables are included. Three categories are considered. **Urban** is defined as a binary indicator that takes the value of one if the parental household of the reference

woman is located in an urban zone and zero otherwise. Similarly, **Suburban (Rural)** identifies women living in a parental household that is located in a suburban (rural) zone. Notice that, by definition, all the studied women live with at least one biological parent and have not initiated independent life. Therefore, the place of residence of the index woman is entirely determined by her parents. As a consequence, **Urban**, **Suburban** and **Rural** may be treated as exogenous variables.

Family background is controlled by two alternative sets of variables. One set of variables includes the average education (years) and income (thousands of pesos) of the members of the household – excluding the index individual. These variables are labelled **Hedu** and **Hincome** respectively. A second set of variables includes the socio-economic characteristics of the head of the family. This last set of variables includes the age of the family head **Hfage** (years), her/his education **HFedu** (years) and income **HFincome** (thousands of pesos).

6.5 Econometric Methodology

6.5.1 Parametric Approach: Two-inflated Generalized Poisson

The starting point is a simple Poisson regression. Let y_i be a non-negative integer representing planned fertility for the i-th woman. The Poisson distribution function is then specified as

$$P(y_i) = \Pr[y_i = j] = \frac{\exp[-\mu_i][\mu_i]^{y_i}}{y_i!}, \; j = 0,1,2,... \tag{1}$$

where μ_i is the mean failure rate underlying the Poisson process followed by the count y_i. The parameter μ_i is then specified as a deterministic exponential function of a kx1 vector of observed individual characteristics, $\mathbf{x_i}$, and a kx1 vector of its associated coefficients β,

$$\mu_i = \exp[\mathbf{x_i}'\boldsymbol{\beta}]. \tag{2}$$

Notice that for this model, conditional on the observed characteristics, mean and variance of the count are equal

$$E(y_i \mid \mathbf{x_i}) = V(y_i \mid \mathbf{x_i}) = \mu_i. \tag{3}$$

In the econometrics literature this property is known as the 'equi-dispersion' assumption of the Poisson model. Two sorts of violations to this assumption are possible. It is said that the data exhibits under-dispersion if $V(y_i \mid \mathbf{x_i}) < E(y_i \mid \mathbf{x_i})$, and that the data exhibits over-dispersion if $V(y_i \mid \mathbf{x_i}) > E(y_i \mid \mathbf{x_i})$. According to Winkelmann (1995), violations of the equi-dispersion assumption result in loss of efficiency and biased inference. However, an important advantage of this model is that if the mean function is correctly specified the Poisson regression estimates in a consistent way the vector of coefficients even if over-dispersion/under-dispersion is present (for more on this see, Gourieroux, Monfort, and Trognon 1984, Cameron and Trivedi 1986). The Poisson model is estimated by standard techniques of Maximum Likelihood.

If the data are over-dispersed the Poisson model can be extended to accommodate the presence of unobserved individual heterogeneity. For this purpose various versions of the Negative Binomial Model and the Poisson model with lognormal unobserved heterogeneity have been suggested. In the case of under-dispersion, however, such extensions are not available. Still, relaxing the equi-dispersion assumption of the Poisson model if under-dispersion is suspected promises efficiency gains and improved inference. This is why Wang and Famoye (1997) suggest the use of a Generalised Poisson Regression model (GP) in this context. Another alternative might be King's Generalised Event Count model (King 1989). Both methodologies accept under-dispersion and, beyond the fact that Generalised Poisson is easier to implement, there are no statistical reasons to prefer one to the other. Here a Generalised Poisson model is estimated. The initial assumption is then that the distribution of the count y_i is given by:

$$P(y_i) = \Pr[y_i = k] = \left(\frac{\mu_i}{1+\varphi\mu_i}\right)^{y_i} \frac{(1+\varphi y_i)^{y_i - 1}}{y_i!} \exp\left[-\frac{\mu_i(1+\varphi y_i)}{1+\varphi\mu_i}\right]; \; k = 0,1,2,\ldots \quad (4)$$

Where μ_i remains as in equation (2) and

$$E(y_i \mid \mathbf{x}_i) = \mu_i$$
$$V(y_i \mid \mathbf{x}_i) = \mu_i(1+\varphi\mu_i)^2. \quad (5)$$

Parameter α determines the 'dispersion' properties. If $\varphi = 0$ there is equi-dispersion and Generalised Poisson reduces to Poisson. Similarly, if $\varphi > 0$ over-

dispersion is obtained. Finally, the model exhibits under-dispersion if $\varphi < 0$.[24] In the case of under-dispersion two constraints should be imposed: namely, that the expressions $1+\varphi\mu_i > 0$ and $1+\varphi y_i > 0$ must hold for all i (for more details on how these constrains are implemented for estimation see Wang and Famoye 1997). Notice that the standard Poisson model is nested within the Generalised Poisson. Hence, estimating the Generalised model and testing the null of $\varphi = 0$ against $\varphi \neq 0$ provides a general test for the validity of the Poisson model. Wang and Famoye suggest the use of a standard likelihood ratio statistic (LR), which is asymptotically distributed as a chi-square variable with one degree of freedom under the null hypothesis.

A relevant feature of the data analysed here is the presence of an excess of two counts relative to what a Poisson distribution would predict (see Table 6.3). This feature may be the source of under-dispersion and it is generally induced by the existence of social norms to which individuals tend to conform. A desire to take advantage of potential economies of scale in childbearing activities may also be a factor contributing to the popularity of a two count. Independently of the reasons that determine the relative excess of two counts, it is clear that this

[24] The use of a Generalized Poisson model in the context of under-dispersed data has been criticized because under $\varphi<0$ probabilities do not sum up to one. Behind this criticism lies the fact that the Generalised Poisson distribution is truncated up to the integer part of $(-1/\varphi)$ when $\varphi<0$. Thus, effectively, if phi is negative the probabilities may not sum up 1. It has been shown, however, that the truncation error is very small, so that GP provides a good statistical description of the data (see Consul and Shoukri 1985).

feature of the data requires explicit econometric modelling. Otherwise the model may be seriously misspecified and lead to invalid inference. To avoid this source of misspecification the Generalized Poisson model is two-inflated in line with the methodology proposed by Mullahy (1986),

$$\begin{aligned} \Pr(y_i) &= \delta + (1-\delta)g(y_i) \quad if \quad y =_i 2 \\ \Pr(y_i) &= (1-\delta)g(y_i) \quad\quad if \quad y_i = 0,1,3,4,... \end{aligned} \tag{6}$$

where $g(y_i)$ represents the non-inflated (or 'parent') distribution and $0 < \delta < 1$ represents the extra probability of observing a two-count. Here the parent distribution is implemented as a Generalized Poisson like in equations (4) and (5). Parameter δ may be either a constant for all women in the sample or may depend on a set of individual characteristics, $\mathbf{w_i}$. To allow for this last alternative a Logit specification of δ will be used,

$$\delta = \frac{\exp(\mathbf{w_i}'\gamma)}{1+\exp(\mathbf{w_i}'\gamma)}. \tag{7}$$

Since identification does not require exclusion restrictions, vectors $\mathbf{x_i}$ and $\mathbf{w_i}$ may contain the same elements. However, securing some exclusions and having a parsimonious model for δ is always a good practice. The two-inflated Generalized Poisson model (TIGPM) is estimated by standard maximum likelihood techniques. At convergence minus the negative of the Hessian matrix, $-H^{-1}$, may be used to estimate the covariance matrix. Usual inference is valid.

Akaike information criterion (AIC) and consistent Akaike information criterion (CAIC) are used to discriminate among different specifications. These statistics are defined as

$$
\begin{aligned}
AIC &= -2\ln(L) + 2k \\
CAIC &= -2\ln(L) + k\{(\ln(n) + 1\},
\end{aligned}
\tag{8}
$$

where k represents the number of parameters estimated. When comparing two alternative models lower values of AIC and CAIC will be preferred.

6.5.2 Semi-parametric Approach: Quantiles for Counts

As discussed in previous paragraphs, a two-inflated Generalized Poisson model (TIGP) constitutes a good alternative to the standard Poisson specification when under-dispersion is suspected and a noticeable excess of two counts detected. Unlike the standard Poisson, TIGP possesses a dispersion parameter α which, if empirically required, induces under-dispersion in the model. More importantly, estimating α along the other parameters of the model guarantees efficiency gains.

A limitation of the GP model, and of most fully parametric count data models, is its underlying assumption that explanatory variables affect exclusively the mean and variance of the conditional distribution of y_i (for further details see Winkelmann 2000). Aspects of the distribution such as shape are supposed to be unaffected by the value that covariates take on. Until recently econometricians had no clear way of evaluating the suitability and relevance of such an assumption in the context of count data. However, the work of Machado and

Santos Silva (2003) on Quantile Regression for count data has created the opportunity of undertaking this sort of analysis.

Quantile regression for count data is fundamentally complicated by the combination of a non-differentiable objective function with a discrete dependent variable. As noted by Huber (1981), under such conditions the asymptotic distribution of the conditional quantiles may not be approximated by a Taylor expansion. There is, then, a lack of known distributional results on the basis of which the researcher may do valid inference. In fact, quantile regression is available for count data only if some smoothness is artificially incorporated into the problem. Three alternatives have been suggested. Lee (1992) suggests interpreting the count as a discrete realisation of a latent continuous process crossing a set of thresholds, as in the ordered Logit model. Conditional quantiles are then estimated on the basis of the smoothed process. Though this procedure is technically appealing it has the disadvantage that all estimated parameters have no interpretation in terms of the original count. A second alternative considers smoothing the objective function so that it can be interpreted as the asymmetric maximum likelihood (AML) estimator described by Efron (1992). Once in the AML context a set of conditional location functions may be obtained. The method, however, does not estimate conditional quantiles but a set of location functions for which asymptotic inference is available. Both Lee's and Efron's methodologies have unappealing features that, though being available in the literature since the beginning of the 1990s, have made them unattractive to the applied econometrician. The third alternative, introduced by Machado and Santos (2003), is based on a procedure that involves smoothing the data. The present study uses this last approach.

The basic idea behind Machado and Santos's method consists of building a continuous random variable whose quantiles have a one-to-one relationship with the quantiles of the count, y_i. As the authors show, the task is achieved by creating the auxiliary variable $z_i = y_i + u_i$, where u_i represents the realisation of a uniform random variable with support [0,1). The procedure delivers a continuous variable but it remains the problem that the distribution of z_i is not smooth over its whole domain. As a consequence, standard results on quantile regression cannot be applied to z_i. To solve this issue it is supposed that at least one continuous explanatory variable exists, so that if a monotone transformation $T(z_i\ ,\ \alpha)$ that meets the restriction

$$Q_{T(z_i|\alpha)}(\alpha \mid \mathbf{x}_i) = \mathbf{x}_i{}'\boldsymbol{\beta}(\alpha) \tag{9}$$

is used for estimation, $\Pr[T^{-1}\ (\mathbf{x}_i{}'\boldsymbol{\beta}(\alpha)) \in \mathbb{N}] = 0$ would be warranted. Here $Q_T(\alpha \mid \mathbf{x}_i)$ represents the α-quantile of the jittered data $T(z_i \mid \alpha)$ and $0 < \alpha < 1$.[25] The procedure ensures that, under some regularity conditions, at almost every realisation of x_i the conditional density of z_i at the quantile of interest will be continuous – notice that the distribution of z_i is discontinuous precisely at the points belonging to $\mathbb{N}$, and that such points are excluded by (9) and the continuous variable in vector $\mathbf{x}_i$.

[25] Here the property that conditional Quantiles are invariant to monotonic transformations of the dependent variable is exploited (see Koenker and Bassett 1978).

To implement the procedure a transformation $T(z_i;\alpha)$ and its associated representation of the conditional quantile of z, $Q_z(\alpha \mid \mathbf{x_i})$, must be specified. Here the following parameterisation is considered:

$$T(z_i,\alpha)=\begin{cases}\log(z_i-\alpha) & if \quad z_i>\alpha \\ \log(\varsigma) & if \quad z_i\le\alpha\end{cases} \tag{10}$$

and

$$Q_z(\alpha \mid x_i)=\alpha+\exp[\mathbf{x_i}'\boldsymbol{\beta}(\alpha)], \tag{11}$$

where ς represents a small positive number (here taken to be 1.0E-10).[26] Once T(.) and $Q_z(.)$ are specified the vector of parameters $\boldsymbol{\beta}(\alpha)$ are estimated by means of a standard quantile regression of $T(z_i,\alpha)$ on the vector of explanatory variables $\mathbf{x_i}$. Machado and Santos show that the derived estimator is asymptotically distributed as a normal variable. Namely,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}(\alpha)-\boldsymbol{\beta}(\alpha)\right)\xrightarrow{D}N(\mathbf{0},\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}) \tag{12}$$

with **D** and **A** as described in the Appendix A. Notice that the quantile regression of T(.) on $\mathbf{x_i}$ depends not only on the data but also on the realisation of the random variable u. Since this "noise" has been artificially created, the econometrician would naturally seek to reduce the sensitivity of the estimates to the realisations of u. Machado and Santos propose averaging out the random noise: that is, to create m draws of u and average the QR estimates for the m "jittered" samples. The average-jittering estimator would then be:

$$\boldsymbol{\beta}_m^A(\alpha) = \frac{1}{m}\sum_{h=1}^{m}\boldsymbol{\beta}^h(\alpha), \tag{13}$$

where $\beta^h(\alpha)$ is the QR estimator based on the h draw of u. As in the case of a single-draw, the average-jittering estimator is distributed asymptotically as a normal with mean vector zero. In this case, however, the covariance matrix **V** should be calculated as

$$\mathbf{V} = \frac{1}{m}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1} + \left(1 - \frac{1}{m}\right)\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1} \tag{14}$$

with **D** and **A** as in equation (13) and **B** as described in Appendix A. Matrix **V** is built as a misspecification robust covariance matrix (Chamberlain 1994) and inference may be based on usual t and Wald tests.

Since there is a one-to-one relationship of the conditional quantiles of z_i and y_i, the interpretation of the coefficients $\beta(\alpha)$ in terms of y_i is similar to the interpretation of $\beta(\alpha)$ in terms of z. In fact from equation (11) it follows that

$$\frac{\partial Q_z(\alpha \mid \mathbf{x}_i)}{\partial x_{ij}} = \exp[\mathbf{x}_i'\boldsymbol{\beta}(\alpha)]\beta_j(\alpha).$$

Hence, the change needed in x_{ij} to induce a change in the conditional quantile of y_i of one unit is inversely proportional to $\beta_j(\alpha)$, other things held constant. This is the usual interpretation of the coefficients in a standard non-linear model.

---

[26] Machado and Santos Silva (2003) use the same value for ς.

6.6 Empirical Results

Table 6.2 (page 194) contains the main descriptive statistics of the data. As noted previously, the dependent variable **Ideal** has a mean of 2.49 and a standard deviation of 1.17. Thus the unconditional data exhibits under-dispersion. Table 6.3 presents the empirical distribution of **Ideal**. From there the reader may conclude that mode and median are both a count of two children. Unlike what is common in realised fertility data, there is no evident excess of zero counts. Instead, the data contains fewer zeros and ones than a standard Poisson distribution with a mean value of 2.5 would predict. That is, remaining childless or having an only child is a relatively unattractive fertility plan. A two-child family, in contrast, is more popular than expected. The existence of a well-established two-child social norm may be behind this relative excess of two counts. Alternatively, the popularity of a two-child family may be a result of deliberate intentions to avoid an only child outcome (Santos Silva and Covas 2000).

Mean education is around eight years but its distribution is hardly normal. There are two main attraction points: six and nine schooling years. More than 50% of the studied cases fall into either of these two categories. Obviously the bimodal distribution of education is highly influenced by institutional rules that made compulsory at least six years of education for all Mexicans. Clearly, this bimodal distribution of **edu** implies that there is a good proportion of women who already left school but have not entered the labour market – i.e., a paid job. It is likely those women are living in rural areas where most female youngsters participate in unpaid agricultural and household activities in the parental house

after leaving school and before entering a marriage or consensual union. Indigenous language speakers represent only four percent of the sample and non-Catholics represent just ten per cent.

Regarding the characteristics of families and households, the descriptive statistics reveal that over 50% of the studied women grew up in cities (urban zones). On the other extreme, 29% of women grew up in settlements with less than 2,500 inhabitants. The average head of family is aged around 47 and has six years of schooling. As in the case of young women, the distribution of education among heads of family has two peaks at six and nine years of schooling. But unlike the young group, there is also a significant attraction point at zero. In fact, 11% of the heads did not complete the first grade of primary education. Finally, average annual income of the heads of family amounts to around two thousand Mexican pesos. Nearly 10% declare not having monetary income at all – probably because they are already retired.

A total of 84% of the studied women live within a family that includes both their biological parents. In the remaining cases the most common family structure is that where the father is absent. This sort of family represents 14% of the sample. Individuals who grew up in urban zones are more likely to live in a non-intact family (i.e., a family where both parents are present). In fact, a 19% of urban women live in a family where at least one biological parent is absent. The corresponding figure for non-urban women amounts to just 13%. These differences probably reflect disparities in the frequency of divorces between urban and non-urban zones.

6.6.1 Prior beliefs

Prior beliefs are as follows. Economic theory suggests that education has a negative effect on ideal fertility because it increases the opportunity cost of childbearing activities (see Willis 1973, Moffit 1984, Ward and Butz 1980 among others). Thus, the coefficient on **edu** is expected to be negative. Similarly, intuition suggests that youngsters follow to some degree fertility norms stabilised in their communities and families, so that **siblings** should have a positive coefficient attached. In the case of **indspker** prior information suggests that a positive coefficient is to be expected given the common disadvantaged economic status of most indigenous groups in Mexico. Regarding location, the prediction is that individuals who grew up in rural areas will prefer on average larger families than individuals from urban or suburban zones. Here the intuition is that information about family planning is better transmitted in large and highly heterogeneous social networks – like those found in a city – than in short and homogeneous networks – like those found in a rural community (Kohler 2000). For similar reasons, adoption costs of innovative fertility norms would be expected to be lower in a city than in a rural community. Hence, coefficients on **urban** and **suburban** should be negative. Moving to the background characteristics, **HFedu** and **Hedu** are expected to have negative coefficients, as human capital may be transferred from one household member to the other via imitation, teaching, or genetic endowment (for further detail on the economic theory of planned fertility see Chapter 2, section 2.4). Finally, since the death or absence of a parent generates economic hardship, coefficients on **AbsentFather** and **AbsentMother** are expected to be negative. **HFincome** is expected to

increase planned fertility if the income elasticity of quantity of children is larger than the income elasticity quantity of children. Otherwise, **HFincome** should induce reduction on planned fertility (see Becker and Lewis 1973). No prior information can be used to predict the signs of the coefficients on **Catholic**, **Hincome**, and **Hfage**.

6.6.2 Poisson and Two-inflated Generalised Poisson

Table 6.4 (page 195) contains results for a standard Poisson regression of **ideal**. Three different specifications are presented:

A. *Basic model:* Only youngsters' personal characteristics and family structure dummies are included as explanatory variables.

B. *Household characteristics*: Besides variables in specification A, household characteristics are integrated as explanatory variables.

C. *Head of family characteristics*: Besides variables in specification A, characteristics of the head of family and the location dummies are included as explanatory variables.

Since the mother heads most families with an absent father, and the father heads most intact families, a dummy indicating the sex of the family head is highly collinear with the family-structure dummies. For this reason, sex of the family head is not included as an explanatory variable. The initial specification (column 1 of Table 6.4) abstracts from the characteristics of households and heads of family. This simplified reduced-form model functions as a departure

point. Then, background characteristics are integrated in two alternative ways. Firstly, family background is approximated by variables registering the average level of education and income of the members of the index household (Household Characteristics model). Obviously, in the construction of these average background variables the information of the index woman was excluded. Secondly, family averaged variables are substituted by the individual characteristics of the family head (Head of Family Characteristics model). These last two econometric specifications are substitutes up to a certain point. Comparing results from specifications B and C then constitutes a sort of robustness exercise. Specifications that include interactions among explanatory variables were also estimated. In all these specifications the interaction terms resulted jointly insignificant and hence they are not discussed here.

Column 1 of Table 6.4 contains Poisson results for model specification A (basic model). All coefficients have their expected signs. However, with the exception of **constant**, **edu** and **siblings** coefficients on explanatory variables appear to be insignificant at all conventional confidence level. Moreover, a likelihood-ratio test for the joint exclusion of **age**, **catholic** and **indspker** fails to reject the null with a chi-square of 3.69 and a p-value of 0.2419. **Absentfather** and **AbsentMother** are jointly insignificant.

Moving to column 2, Poisson results for model specification B are presented. As the reader may conclude, introducing parental-house-location dummies and average family background characteristics into the model (i.e., **Hedu**, **Hincome**) induces some variation in the estimated parameters. Signs, however, remain unchanged and consistent with economic theory (see, for instance, Willis 1973, Becker and Lewis 1973, Easterlin 1975). As before, coefficients on **edu** and

siblings are significant in specification B. Regarding the new controls, location dummies are found to be significant, unlike average family education and income. A likelihood-ratio test for the joint exclusion of **age**, **catholic, indspker, Hedu**, **Absentfather** and **AbsentMother** does not reject the null hypothesis at any conventional confidence level.

Substituting **Hedu** and **Hincome** for the socio-economic characteristics of the family head – e.i., **HFfage**, **HFedu** and **Hfincome** – does not modify the general picture (see column 3 of Table 6.4). That is, besides education and number of siblings all individual characteristics have no influence on fertility plans. As before, a negative coefficient is estimated for the location dummies **surban** and **urban**. Similarly, a negative coefficient on **HFedu** confirms prior expectations that the higher education of the family head imply reductions in the planned fertility of the index woman (see Chapter 2, section 2.4). And income of the family head is found to increase planned fertility significantly. Notice that under model specification C a negative coefficient on **indspker** is estimated. However, such a coefficient is not different from zero at any conventional significance level.

Table 6.5 (page 196) reports regression results from a TIGPM that sets the extra probability of observing a two-count, δ, to a constant that is common to all individuals in the sample. Alternatively, a TIGPM that allows variation on δ according to a set of explanatory variables was also estimated. In the latter case, and after trying various specifications, it was found that only women's education and the number of their siblings have significant coefficients on the Logit form used for δ. Hence, to avoid potential over-parameterisation, all non-significant

variables in the Logit for δ were excluded and the model re-estimated. This final specification is reported on Table 6.6.

Column 1 of Table 6.5 presents TIGPM results for the initial model specification – i.e, the specification that has only personal characteristics and family structure dummies as explanatory variables. The reader should notice that a negative and highly significant estimate for the auxiliary parameter φ is found. There is then empirical evidence to reject the standard Poisson model as a valid description of the data. Since φ is found to be negative, the model exhibits under-dispersion. The constant in the Logit specification of δ is also found to be highly significant, implying that there is an extra probability of 0.35 of planning a two-child family in the sample in relation to what a standard Generalised Poisson model (GPM) would predict. This finding suggests then that if a GPM is fitted in place of the TIGPM estimated here, the econometrician will most likely obtain a misspecified model. With the exception of **indspker**, which has now a negative but insignificant coefficient attached, coefficients from TIGPM and standard Poisson are similar (see column 1 of Tables 6.4 and 5). Further, excluding **Absentfather** that is significant at 5% in TIGPM but insignificant in Poisson, all explanatory variables that have a significant (insignificant) coefficient in Poisson maintain their significance status in TIGPM.

Including the characteristics of the households (i.e., location dummies, **Hedu** and **Hincome**) to estimate the TIGPM version of model specification B does not change the main conclusions drawn from the corresponding specification in the Poisson model (see column 2 of Tables 6.4 and 6.5). That is, that education, number of siblings, and location dummies are the only factors affecting planned fertility.

Moving to analyse model specification C in column 3 of Table 6.5, the reader can find that head-of-family characteristics have significant coefficients, though only at a 5% level. Hence, it seems that family background has indeed an important role in determining desired lifetime number of children of young Mexican women. It is not the characteristics of the average family member which matters. Instead, results suggest that young women's ideal fertility is affected mainly by the socio-economic characteristics of the family head. An intuitive explanation of this finding might be that it is the characteristics of the head which contribute the most information about the current and future wealth of the family. Therefore, averaging characteristics across family members other than the index individual may introduce uninformative noise. As before, **HFedu** is found to reduce planned fertility and **HFincome** to increase it. In terms of the quantity-quality trade off of children literature the latter result is an indication that young women expect the income elasticity of quantity of children to be larger than the income elasticity of quality of children (see, for instance, Becker and Lewis 1973). **Age**, **catholic**, **indspker**, **Absentfather** and **AbsentMother** are individually and jointly insignificant.

A further extension of the TIGPM is reported in Table 6.6 (page 197). In this case, besides a constant, δ depends on the education and number of siblings of the index woman. Results are similar to those previously discussed. However, various new conclusions may be drawn. First, it is found that the extra probability of observing a two-count, δ, increases as the education of the index woman becomes higher and decreases with a larger count of siblings. These results, which are witnessed by a positive (negative) and significant coefficient on **edu** (**siblings**) in the 'Delta' section of Table 6.6, suggest that it is the more educated women who

tend to impose the two-child fertility norm and confirm previous findings obtained from the study of completed fertility in Mexico (Miranda 2004). The fact that δ decreases with the number of siblings is intuitive in economic terms – it implies that women with many siblings desire themselves a large family – and requires no further comment (for further detail see the literature review on economic theory of fertility behaviour presented in Chapter 2).

An interesting result from Table 6.6 is the fact that once δ is allowed to vary with women's education the confident on **edu** in the Generalized Poisson distribution becomes insignificant in model specification B and C (columns 2 and 3 respectively). This finding implies that, once household and family characteristics have been controlled for, women's education only affects the likelihood of observing a two-count. In other words, that the chances of observing any value of **ideal** different from two are not affected by **edu**. Once again, this result indicates that it is educated women who are establishing the two-child fertility norm in Mexico.

Comparing TIGPM and Poisson in Tables 6.4, 6.5 and 6.6 (pages 195, 196 and 197) via an Akaike information criterion statistic (AIC) it is easy to conclude that letting δ be a function of **edu** and **siblings** delivers the best-fitting model. Using a consistent Akaike information criterion (CAIC) for the selection produces the same set of conclusions. Focussing on Table 6.6 and comparing model specification A through C via an AIC statistic, the reader may conclude that the best-fitting model is the one which contains the characteristics of the family head as explanatory variables (i.e., specification C). Similar conclusions can be drawn if a consistent Akaike information criterion (CAIC) is used instead (see the bottom panel of Table 6.6).

6.6.3 Quantile Regression

So far, the results from Poisson and Generalised Poisson regressions have been discussed. The findings suggest that education and number of siblings constitute the only personal characteristics that affect ideal fertility. Among the family background variables, location of the parental household and age of the family head appear to be relevant. Education and income of the family head are also important. Results from Poisson and TIGPM models show that failing to account for the special characteristics of the data (that is, under-dispersion and relative excess of two counts) may lead to wrong inference.

A limitation of the Generalised Poisson model is its underlying assumption that, conditional on explanatory variables, individuals are identical. In practice unobserved heterogeneity is commonly found even after controls for all observed individual characteristics have been included into the model. In general, if detected, this remaining heterogeneity is induced by the presence of unobserved individual features that ultimately affect fertility behaviour. The important point here is that if unobserved heterogeneity interacts in some way with the explanatory variables, then an exogenous shift of planned fertility could affect the conditional distribution of **ideal**. In such a case the most optimistic view would allow at least for shifts in location of the conditional distribution. However, changes in the shape and scale of the conditional distribution should not be ruled out *a priori*. Therefore, the possibility that explanatory variables could affect aspects other than mean and variance of the conditional distribution should be carefully evaluated.

As is discussed earlier in the chapter, a Generalised Poisson model – and its extended two-inflated GP version – supposes that explanatory variables affect exclusively mean and variance of the conditional distribution. Though this assumption is rather unsatisfactory, in the past the econometrician had no practical way of assessing the suitability and relevance of such a hypothesis in the context of count data. Machado and Santos Silva (2003), however, recently developed a quantile regression methodology that opens the door to this type of analysis. In the new setting, covariates may affect the conditional distribution of the count in a general way and, as is customary in quantile regression, no prior assumptions about the distribution of the dependent variable are required. The semi-parametric technique proposed by Machado and Santos is used here to check the consistency of the results obtained from the two-inflated Generalised Poisson model.

To help economic intuition in the present Chapter unobserved individual heterogeneity is interpreted as women's preferences towards children. Clearly, if a woman has strong (weak) preferences towards children she is likely to plan a relatively large (small) family regardless of her individual socio-economic characteristics. Under this interpretation, therefore, a large positive realisation of the unobserved random characteristic implies that a large family is planned and that an observation in the right tail of the conditional distribution of **ideal** is recorded. Obviously, other economic interpretations of the nature of unobserved heterogeneity may be suggested in the present context. Hence, taking unobserved heterogeneity to represent women's preferences toward children is simply a device to help the economic interpretation of the empirical results.

Table 6.7 (page 198) contains quantile regression results for **ideal**. Since model specification C delivered the best-fitting model in the previous analysis, to ease exposition, only quantile results for specification C are discussed here. Attention is concentrated on the first, second and third quartiles. In order to select the number of jittered samples needed for the analysis various preliminary regressions were estimated, starting with 1,000 samples. Then for each regression 100 additional jittered samples were included and the model re-estimated. This procedure was followed iteratively until no significant changes in the parameters were detected. In most cases 1,500 jittered samples were found to be enough.

From table 6.7 the reader may learn that explanatory variables have different impacts in the various regions of the conditional distribution of **ideal**. There are even changes in the signs of some coefficients. Notice first that education is only significant in the third quartile while **siblings** are relevant across all the regions of the conditional distribution. Therefore, education seems to be a relevant determinant of planned fertility only if the index woman has strong preferences towards children. As expected, **siblings** keeps a positive relationship with planned fertility regardless of the value of α (the quantile indicator). Finally, **catholic** and Indian language (**indspker**) appear to have no significant effect in the first, second or third quartile.

In other issues, results from Table 6.7 indicate that location of the parental household affects fertility plans across the whole distribution of **ideal**. Moreover, the coefficients on **surban** and **urban** are higher for α=0.5 and α=0.75 than for α=0.25. In other words, urban/rural background plays a more relevant role in the transition from large to large counts, say from 5 to 6, than in the transition from low to low counts, say from 0 to 1.

In the category of family background characteristics Table 6.7 shows a negative coefficient on the education of the family head (**HFedu**) that is significantly different from zero at every considered value of α. Notice, however, that the impact of **HFedu** is larger at the top of the distribution. Rephrasing this last point one could say that it is among women who like children the most that education of the family head has its stronger effects. Similarly, the effect of income of the family head increases planned fertility. Further, according to Table 6.7 **HFIncome** has a larger effect at the top than at the bottom of the conditional distribution. Notice that, in terms of the quantity-quality trade-off of children of Becker and Lewis (1973), it is intuitive that income elasticity of quantity (quality) at the top of the conditional distribution is larger (smaller) than income elasticity of quantity (quality) at the bottom of the conditional distribution – as a large realisation on the unobservable is an indicator of strong preferences towards children. This observation may explain why the effects of **HFIncome** are larger at the top than at the bottom of the conditional distribution. Finally, Table 6.7 shows that, unlike the absence of a mother, an absent father induces significant reductions on ideal fertility for any value of α.

To close the discussion, Table 6.8 (page 199) presents marginal effects calculated on the basis of both Two-inflated Generalised Poisson and Quantile Count regression.[27] Marginal effects of continuous variables are evaluated at their mean value with all dummies set to zero. Similarly, marginal effects of dummy variables are evaluated as the unit change in the conditional quantile (or conditional mean in the case of GP) induced by the change of the relevant dummy

[27] Marginal effects are calculated on the basis of the formula provided in page 162.

from zero to one, and setting all continuous covariates at their sample means. Note that for TIGPM marginal effects have two components. One partial component is the effect that a variable x has on **ideal** via the Generalized Poisson distribution. The other partial component is the effect that a variable x has on **ideal** via its impact on the likelihood of observing a two count. The total marginal effect of x on **ideal** is simply the sum of these two components. In Table 6.8 total marginal effects for **siblings** and **edu** are reported in the case of TIGPM. From Table 6.8 the reader may learn that marginal effects are rather small. However, the figures generally increase as a higher value of **ideal** is recorded (see columns corresponding to the first, second and third quantile). This result suggests that the covariates have a larger effect on planned fertility when women have stronger preferences towards children.

Among all variables, the urban/suburban location of the parental household is by far the control that induces the largest change in **ideal**. This observation is true either at the mean (TIGPM model) or at the three different quartiles considered (Quantile model). For instance, according to Table 6.8 a change of **urban** from zero to one induces a reduction of about –0.22 units in the conditional mean of **ideal** (see column 1), other things held constant. The marginal effect of **urban**, however, varies across the different regions of the distribution of **ideal**, ranging from –0.10 in the first quartile to –0.39 in the third quartile.

In order to offer more detail, Figure 6.1 (pages 201-202) presents a series of graphs with the estimated marginal effects of various explanatory variables evaluated at a value of α which goes from 0.10 to 0.90. To help the reader, a 95% confidence interval has been included alongside marginal effects. Confidence intervals were obtained using the results provided by Machado and Santos Silva

(2003) which show that, at convergence, the quantile estimators are asymptotically normally distributed. The following conclusions can be drawn. First, the negative and relatively steep slope of the graph for **edu** confirms that the marginal effect of **edu** is larger among women with strong preference towards children than among women with weak preference towards children.[28] Similarly, a steep positive slope in the graph for **siblings** confirms previous suggestions that **siblings** has a larger effect on planned fertility the stronger her preference towards children. Second, the graph for **Hfedu** gives further evidence that the education of the head of family has an important role in the fertility plans of young Mexican women, and that its effects are larger if the reference woman has strong preferences towards children (i.e., at the top of the conditional distribution). Further, a similar analysis reveals that income of the family head has larger effects the stronger the preferences towards children. An absent father, finally, seems to have a larger effect on **ideal** at the tails of the distribution – though in the top the marginal effects of **AbsentFather** are hardly significant. Parallel conclusions can be drawn from graphs for **urban** and **suburban**.

[28] Remember that in the present chapter unobserved heterogeneity is interpreted as women's preferences towards children. In this context, a large positive realisation of unobserved heterogeneity implies that a large value for **ideal** is recorded. Therefore, to state that education has a larger effect in the top higher quantiles than in lower quantiles is equivalent to say that education has a larger effect on **ideal** when women have strong preferences towards children than when women have weak preferences towards children.

6.6.4 Robustness assessment: the sample selection issue

The present work is based on the analysis of data for Mexican women aged between 15 and 17 who at the time of the ENADID interview were living with at least one biological parent and had neither started independent economic life nor entered motherhood. From a total of 7,376 cases of women aged between 15 and 17 a total of 5,628 observations met the selection criterion. The selected sample represents 76% of cases of all women aged within the relevant range contained in the ENADID. As was explained earlier in the text, to obtain the selected sample all women who at the time of the interview had moved from the parental household and/or had entered motherhood and/or had entered the labour market were excluded. Cases that reported a non-numerical entry for **ideal** were similarly dropped (Table 6.1 and 6.2 present details of the descriptive statistics, see pages 193 and 194). The study presented here is done conditional on this selection.

Providing a comprehensive assessment of the potential bias induced by the data selection criteria used in the present analysis is beyond the scope of the present work. First, as has been already discussed, there are good reasons to believe that data on ideal fertility that do not meet the chosen selection criteria are difficult to interpret. Second, family background characteristics are only available in the ENADID if the reference woman remains in the parental household.

An initial assessment of the sample selection issue is nonetheless performed on the basis of the method suggested by Heckman (1979). Obviously, the use of such a methodology is unsatisfactory because critical features of the dependent

variable are ignored – such as the fact that **ideal** is a non-negative and discrete variable – and a standard model for continuous variables is estimated. Further, though the model is technically identified through functional form, good econometric practice would require the existence of an instrument for the selection equation. In other words, the analyst would need to specify a variable that affects selection but not fertility – or if it does, it should be required that it effects fertility only through selection. Such a kind of instrument is not available in the ENADID and, thus, identification must rely on functional form. Despite its limitations, the exercise is a valuable initial assessment of the relevance of the selection issue in the context of the present work.

Table 6.9 (page 200) contains the results. Three different selection criteria were considered. First, it is assessed whether or not excluding the cases of women who report a non-numerical response for **ideal** and/or have left the parental house by the time of the ENADID interview results in serious bias (first column of Table 6.9). A second criterion excludes additionally all women who entered motherhood by the time of the interview (second column). Finally, the more restrictive criterion excludes all previous observations plus those cases of women who entered the labour market by the time of the ENADID interview (third column). In the mean equation all explanatory variables considered in the present work are included. For the selection equation only personal characteristics are considered, as they are the only available data when women have moved from the parental household by the date of the ENADID interview. Estimates for mean and selection equations are reported.

The interest is centered on the likelihood ratio test for the parameter *rho* reported at the bottom of Table 6.9. *Rho* represents the correlation coefficient

between the error term in mean and selection equations. If *rho* = 0 then there is no selection bias, and OLS can be used to obtain a consistent estimator of the parameters in the mean equation using exclusively data from the selected sample. However, if *rho* ≠ 0 selection bias is present and a suitable correction is required. According to Table 6.9, the null of *rho* = 0 cannot be rejected at all conventional significance levels in all the three cases considered. Obviously, this is not definite proof that sample selection a bias does not exist, as a test that accounts for the features of **ideal** is still needed. The evidence however suggests that, if present, such a bias should be rather small.

6.7 Conclusions

The present chapter reports a study of the determinants of planned fertility (ideal fertility) in Mexico. To avoid confusion between actual and planned fertility, attention is concentrated on analysing the fertility plans of young Mexican women aged between 15 and 17 years who live with at least one biological parent and have neither initiated economic independently life nor entered motherhood. Analysing data for young women also ensures that individuals have information exclusively about their initial endowments as they are effectively at the beginning of their fertility and labour-planning horizon. Thus, it is certain that fertility plans of the analysed women have not been updated by the arrival of information that unfolds only with labour market and parenthood experience. A Heckman-like assessment of the potential bias induced by the data selection criterion used in the present work finds no evidence of such a bias. Count data models are used as the

main analysis tool, including an innovative technique for estimating quantile regressions for count data (see Machado and Santos Silva 2003).

Two issues are primarily discussed. Firstly, the study enquires about the potential impact that family structure may have on fertility plans. Secondly, the study attempts to establish whether or not the socio-economic characteristics of the family head have a significant role in determining the fertility plans of young female dependants.

The findings indicate that, as predicted by economic theory, the education of a woman is associated with reductions on her planned fertility (see, for instance, Willis 1973, Becker and Lewis 1973, Easterlin 1975). However, this negative effect is statistically significant only among women that have relatively strong preferences towards children. In other issues, it is found that the number of siblings a woman has increases significantly the number of her planned offspring. Further, results indicate that this effect is larger the stronger the preference for children is felt. Surprisingly, religion (Catholic vs. non-Catholic) and ethnic background are found to have no statistically significant impact on young women's planned fertility in Mexico. Among all considered explanatory variables, the urban/rural location of the parental household seems to be the variable with the largest single marginal effect.

Regarding family structure, empirical results suggest that an absent father impacts negatively on planned fertility, especially when women have weak tastes for children. An absent mother seems to have a relevant role only for women with strong preferences towards children.

Women who live in a household that is headed by a highly educated individual are found to plan fewer children during their lifetime than women who live in a

household headed by a poorly educated individual. More importantly, education of the family head is found to have a larger effect the stronger an inclination for children is felt. In other words, it is among women who the most like children where education of the family head reduces planned fertility the most. Finally, income of the family head is found to maintain a positive relationship with planned fertility.

Appendix A

The three elements needed to estimate the covariance matrix in a Quantile regression for counts are:

$$\hat{\mathbf{A}} = \frac{1}{n}\sum_i \left[\alpha - 1\{T(z_i;\alpha) \le \mathbf{x_i}'\boldsymbol{\beta}(\alpha)\}\right]^2 \mathbf{x_i}\mathbf{x_i}'$$

$$\hat{\mathbf{D}} = \frac{1}{n}\sum_i \exp[\mathbf{x_i}'\boldsymbol{\beta}(\alpha)]1\{F_n(\hat{z}_{\alpha i}) \le z_i \le F_n(\hat{z}_{\alpha i}+1)\}\mathbf{x_i}\mathbf{x_i}'$$

$$\hat{\mathbf{B}} = \frac{1}{n}\sum_i \left\{\alpha^2 + (1-2\alpha)1(y_i \le \hat{z}_{\alpha i} - 1) + [(\hat{z}_{\alpha i} - y_i)1(\hat{z}_{\alpha i} - 1 \le y_i \le \hat{z}_{\alpha i})]\right\}\mathbf{x_i}\mathbf{x_i}'.$$

Function 1(.) represents the usual indicator function and F_n is defined as,

$$F_n(x) = \begin{cases} \lfloor x \rfloor - 0.5 + \dfrac{(x - \lfloor x \rfloor)}{2c_n} & if \quad x - \lfloor x \rfloor < c_n \ and \ x \ge 1 \\ \lfloor x \rfloor & if \quad c_n \le x - \lfloor x \rfloor < 1 - c_n \ or \ x < 1 \\ \lfloor x \rfloor + 0.5 + \dfrac{(x - \lfloor x \rfloor)}{2c_n} & if \quad x - \lfloor x \rfloor \ge 1 - c \end{cases}$$

where $\lfloor x \rfloor$ represents the floor function which returns the largest integer smaller than, or equal to, x. Finally, $\{c_n\}$ represents a sequence of real numbers in (0,0.5) such that c_n=o(1) and

$$\frac{\sup_{1<i\le n} |T^{-1}(\mathbf{x_i}'\hat{\boldsymbol{\beta}}(\alpha)) - T^{-1}(\mathbf{x_i}'\boldsymbol{\beta}(\alpha))|}{c_n} = o_p(1)$$

as $n \to \infty$. As suggested by Machado and Santos Silva, results in the present work are obtained using the following functional form for c_n,

$$c_n = \frac{0.5 \ln(\ln(n))}{\sqrt{n}}.$$

See Machado and Santos Silva (2003) for further reference.

Appendix B

Table 6.1 Descriptive Statistics -- All ENADID cases

Variable	Description	Obs	Mean	Std. Dev	Min	Max
Personal characteristics						
Ideal	desired number of children	7376	2.50	1.16	0	12
nresponse	= 1 if numerical response for ideal	7560	0.98	-	-	-
moveout	= 1 if away of the parental household	7560	0.05	-	-	-
motherhood	= 1 if entered motherhood	7560	0.05	-	-	-
work	= 1 if work	7560	0.18	-	-	-
siblings	number of siblings	7212	3.00	1.91	0	13
age	age in years	7560	15.98	0.82	15	17
edu	education in years	7560	8.12	2.20	0	11
catholic	=1 if catholic	7560	0.89	-	-	-
indspker	=1 if indian speaker	7560	0.05	-	-	-
+ 32 birth state dummies (base Mexico City)						
Household Characteristics						
rural (base)	=1 if < 2,500 inhabitants.	7212	0.30	-	-	-
surban	=1 if > 2,500 & < 99,999 inhabitants.	7212	0.19	-	-	-
urban	=1 if > 100,000 inhabitants.	7212	0.51	-	-	-
Hedu	average household years of education excluding index indv.	7212	6.48	2.75	0	19
Hincome	Household income, thousand of mexican pesos	7212	3.44	5.61	0	306.64
Head of Family Caracteristics						
HFage	Head of family age in years	7212	46.15	7.55	35	75
HFedu	Head of family education in years	7212	5.79	4.67	0	24
HFincome	Head of family income, thousand of mexican pesos	7212	2.00	3.97	0	200
Family Structure						
BothParents (base)	=1 if both parents present	7212	0.83	-	-	-
AbsentFather	=1 if father missing	7212	0.15	-	-	-
AbsentMother	=1 if mother missing	7212	0.03	-	-	-

Table 6.2 Descriptive Statistics -- Selected Sample

Variable	Description	Obs	Mean	Std. Dev	Min	Max
Personal characteristics						
Ideal	desired number of children	5628	2.49	1.17	0	12
siblings	number of siblings	5628	2.86	1.83	0	12
age	age in years	5628	15.89	0.81	15	17
edu	education in years	5628	8.45	2.07	0	11
catholic	=1 if catholic	5628	0.90	-	-	-
indspker	=1 if indian speaker	5628	0.04	-	-	-
+ 32 birth state dummies (base Mexico City)						
Household Characteristics						
rural (base)	=1 if < 2,500 inhabitants.	5628	0.29	-	-	-
surban	=1 if > 2,500 & < 99,999 inhabitants.	5628	0.18	-	-	-
urban	=1 if > 100,000 inhabitants.	5628	0.53	-	-	-
Hedu	average household years of education excluding index indv.	5628	6.79	2.84	0	19
Hincome	Household income, thousand of mexican pesos	5628	3.44	4.68	0	141.8
Head of Family Caracteristics						
HFage	Head of family age in years	5628	46.10	7.50	35	74
HFedu	Head of family education in years	5628	6.30	4.83	0	24
HFincome	Head of family income, thousand of mexican pesos	5628	2.16	3.48	0	81
Family Structure						
BothParents (base)	=1 if both parents present	5628	0.84	-	-	-
AbsentFather	=1 if father missing	5628	0.14	-	-	-
AbsentMother	=1 if mother missing	5628	0.02	-	-	-

Table 6.3 Empirical distribution of Ideal and a Poisson distribution with mean of 2.5 children

Count	0	1	2	3	4	5	6	7+	Tot
No. obs.	162	374	2952	1345	549	124	70	52	5628
% share	2.9	6.65	52.45	23.9	9.75	2.2	1.24	0.93	100
Poisson	8.2	20.52	25.65	21.38	13.36	6.68	2.78	1.42	100

Table 6.4 Poisson Regression for Ideal
Coefficient [Std. Error]

Variable	(1) Basic Model	(2) Household Char.	(3) Head of Fam. Char.
Intercept	0.6239 [0.1730]**	0.7257 [0.1742]**	0.6755[0.1798]
Personal characteristics			
age	0.0139 [0.0107]	0.0103 [0.0107]	0.0072 [0.0108]
edu	-0.0224 [0.0044]**	-0.0131 [0.0051]**	-0.0115 [0.0049]*
siblings	0.0295 [0.0048]**	0.0247 [0.0049]**	0.0253 [0.0049]**
catholic	0.0189 [0.0296]	0.0224 [0.0296]	0.0233 [0.0296]
indspker	0.0313 [0.0441]	0.0014 [0.0444]	-0.0010 [0.0444]
Birth state dummies	Yes	Yes	Yes
Household Characteristics			
surban	-	-0.0671 [0.0259]**	-0.0622 [0.0258]*
urban	-	-0.1249 [0.0228]**	-0.1140 [0.0227]**
Hedu	-	-0.0036 [0.0041]	-
Hincome	-	0.0028 [0.0019]	-
Head of Family Characteristics			
HFage	-	-	0.0018 [0.0012]
HFedu	-	-	-0.0049 [0.0024]*
HFincome	-	-	0.0052 [0.0026]*
Family Structure			
AbsentFather	-0.0404 [0.0255]	-0.0287 [0.0256]	-0.0304 [0.0257]
AbsentMother	-0.0395 [0.0558]	-0.0310 [0.0558]	-0.0375 [0.0559]
LogL	-8997.82	-8980.19	-8976.22
Chi2	238.80	274.07	282.00
Pr > Chi2	0.0000	0.000	0.000
AIC	9073.82	9064.19	9062.22
CIAC	18368.00	18365.06	18366.76
Number of observations	5,628	5,628	5,628

Note: ML estimates. * significant at 5% ; ** significant at 1%.

Table 6.5 Two-inflated Generalized Poisson Regression for Ideal -- Constant inflation factor
Coefficient [Std. Error]

Variable	Basic Model	Household Char.	Head of Fam. Char.
Generalized Poisson			
Intercept	0.6210 [0.1795]**	0.7564[0.1799]**	0.6823 [0.1856]**
Personal characteristics			
age	0.0143 [0.0110]	0.0090 [0.0110]	0.0051 [0.0111]
edu	-0.0241 [0.0044]**	-0.0126 [0.0051]*	-0.0104 [0.0048]*
siblings	0.0319 [0.0047]**	0.0265 [0.0048]**	0.0273 [0.0048]**
catholic	0.0395 [0.0303]	0.0441 [0.0302]	0.0455 [0.0302]
indspker	-0.0085 [0.0427]	-0.0371 [0.0429]	-0.0394 [0.0429]
Birth state dummies	Yes	Yes	Yes
Household Characteristics			
surban	-	-0.0720 [0.0257]**	-0.0664 [0.0256]**
urban	-	-0.1395 [0.0232]**	-0.1235 [0.0230]**
Hedu	-	-0.0054 [0.0043]	-
Hincome	-	0.0029 [0.0017]	-
Head of Family Caracteristics			
HFage	-	-	0.0025 [0.0012]*
HFedu	-	-	-0.0076 [0.0025]*
HFincome	-	-	0.0059 [0.0024]*
Family Structure			
AbsentFather	-0.0570 [0.0269]*	-0.0431 [0.0268]	-0.0450 [0.0269]
AbsentMother	-0.0502 [0.0580]	-0.0425 [0.0577]	-0.0533 [0.0576]
Phi	-0.0539 [0.0026]**	-0.0550 [0.0026]**	-0.0554 [0.0025]**
Inflation factor (δ)			
Intercept	-0.6059 [0.0406]**	-0.6120 [0.0407]**	-0.6108 [0.0406]**
LogL	-7920.75	-7898.59	-7890.38
Chi2	328.94	378.43	481.09
Pr>Chi2	0.0000	0.000	0.000
AIC	8000.75	7986.59	7980.38
CAIC	16226.91	16221.15	16214.36
Number of observations	5,628	5,628	5,628

Note: ML estimates. * significant at 5% ; ** significant at 1%.

Table 6.6 Two-inflated Generalized Poisson Regression for Ideal
Coefficient [Std. Error]

Variable	Basic Model	Household Char.	Head of Fam. Char.
Generalized Poisson			
Intercept	0.6000 [0.1814]**	0.7387 [0.1818]**	0.6679 [0.1875]**
Personal characteristics			
age	0.0134 [0.0111]	0.0081 [0.0111]	0.0046 [0.0111]
edu	-0.0183 [0.0044]**	-0.0070 [0.0051]	-0.0051 [0.0048]
siblings	0.0240 [0.0048]**	0.0187 [0.0049]**	0.0194 [0.0049]**
catholic	0.0387 [0.0304]	0.0437 [0.0303]	0.0452 [0.0303]
indspker	0.0074 [0.0427]	-0.0232 [0.0429]	-0.0256 [0.0429]
Birth state dummies	Yes	Yes	Yes
Household Characteristics			
surban	-	-0.0732 [0.0257]**	-0.0680 [0.0256]**
urban	-	-0.1417 [0.0233]**	-0.1265 [0.0231]**
Hedu	-	-0.0057 [0.0044]	-
Hincome	-	0.0031 [0.0017]	-
Head of Family Characteristics			
HFage	-	-	0.0024 [0.0012]*
HFedu	-	-	-0.0076 [0.0025]*
HFincome	-	-	0.0062 [0.0025]*
Family Structure			
AbsentFather	-0.0595 [0.0273]*	-0.0453 [0.0272]	-0.0475 [0.0273]
AbsentMother	-0.0571 [0.0580]	-0.0500 [0.0577]	-0.0602 [0.0576]
Phi	-0.0523 [0.0027]**	-0.0535 [0.0026]**	-0.0539 [0.0026]**
Inflation factor (δ)			
Intercept	-1.3204 [0.2321]**	-1.3330 [0.2325]**	-1.3348 [0.2320]**
edu	0.1481 [0.0232]**	0.1492 [0.0232]**	0.1489 [0.0232]**
siblings	-0.1948 [0.0280]**	-0.1964 [0.0281]**	-0.1942 [0.0280]**
LogL	-7848.22	-7825.35	-7817.80
Chi2	264.63	315.22	330.23
Pr>Chi2	0.0000	0.000	0.000
AIC	7932.22	7917.35	7911.80
CAIC	16101.13	16093.94	16088.46
Number of observations	5,628	5,628	5,628

Note: ML estimates. * significant at 5% ; ** significant at 1%.

Table 6.7 Quantile Regression for Ideal
Coefficient [Std. Error]

Variable	1st Quartile	2nd Quartile	3th Quartile
Intercept	0.4628 [0.0793]**	0.6367 [0.1005]	0.7993 [0.1315]
Personal characteristics			
age	0.0055 [0.0048]	0.0089 [0.0062]	0.0105 [0.0080]
edu	0.0048 [0.0028]	-0.0059 [0.0033]	-0.0224 [0.0042]**
siblings	0.0145 [0.0027]**	0.0241 [0.0037]**	0.0297 [0.0042]**
catholic	0.0132 [0.0140]	-0.0048 [0.0163]	0.0250 [0.0202]
indspker	-0.0649 [0.0342]	0.0115 [0.0363]	0.0662 [0.0428]
Birth state dummies	Yes	Yes	Yes
Household Characteristics			
surban	-0.0419 [0.0136]**	-0.0682 [0.0176]**	-0.0662 [0.0207]**
urban	-0.0497 [0.0109]**	-0.1034 [0.0140]**	-0.1447 [0.0173]**
Head of Family Caracteristics			
HFage	0.0014 [0.0006]*	0.0014 [0.0007]*	0.0021 [0.0009]*
HFedu	-0.0025 [0.0010]*	-0.0041 [0.0012]**	-0.0050 [0.0017]**
HFincome	0.0037 [0.0016]*	0.0043 [0.0013]**	0.0071 [0.0021]**
Family Structure			
AbsentFather	-0.0294 [0.0110]**	-0.0433 [0.0125]**	-0.0353 [0.0180]*
AbsentMother	-0.0101 [0.0249]	-0.0117 [0.0299]	-0.0594 [0.0349]
Number of observations	5,628	5,628	5,628

Note 1. * significant at 5% ; ** significant at 1%

Note 2. In order to select the number of jittered samples needed for the analysis various preliminary regressions were estimated, starting with 1,000 samples. Then for each regression 100 additional jittered samples were included and the model re-estimated. This procedure was followed iteratively until no significant changes in the parameters were detected. In most cases 1,500 jittered samples were found to be enough.

Note 3. I am grateful to Joao Santos Silva for his TSP code to estimate Quantiles for counts. His code was a valuable basis for writing the Stata code that is used in the present study.

Table 6.8 Marginal Effects for Ideal
[Std. Error]

Variable	TIGPM	Quantile Regression		
		1st Quartile	2nd Quartile	3rd Quartile
Personal characteristics				
age	0.0079 [0.0193]	0.0115 [0.0100]	0.0207 [0.0145]	0.0286 [0.0217]
edu	-0.0312 [0.0085]**	0.0100 [0.0057]	-0.0138 [0.0077]	-0.0609 [0.0116]**
siblings	0.0630 [0.0086]**	0.0302 [0.0056]**	0.0559 [0.0087]**	0.0808 [0.0114]**
catholic	0.0772 [0.0508]	0.0272 [0.0288]	-0.0112 [0.0380]	0.0674 [0.0539]
indspker	-0.0440 [0.0728]	-0.1310 [0.0669]	0.0268 [0.0852]	0.1860 [0.1241]
Birth state dummies		Yes	Yes	Yes
Household Characteristics				
surban	-0.1157 [0.0427]**	-0.0858 [0.0276]**	-0.1551 [0.0392]**	-0.1766 [0.0541]**
urban	-0.2208 [0.0407]**	-0.1033 [0.0227]**	-0.2412 [0.0331]**	-0.3966 [0.0478]**
Head of Family Caracteristics				
HFage	0.0041 [0.0021]*	0.0028 [0.0012]*	0.0033 [0.0016]*	0.0058 [0.0025]*
HFedu	-0.0132 [0.0044]**	-0.0051 [0.0022]*	-0.0095 [0.0028]**	-0.0135 [0.0047]**
HFincome	0.0107 [0.0043]*	0.0077 [0.0033]*	0.0100 [0.0030]**	0.0194 [0.0056]**
Family Structure				
AbsentFather	-0.0812 [0.0458]	-0.0604 [0.0224]**	-0.0990 [0.0282]**	-0.0950 [0.0480]*
AbsentMother	-0.1017 [0.0946]	-0.0208 [0.0513]	-0.0271 [0.0687]	-0.1575 [0.0900]
Number of observations	5,628	5,628	5,628	5,628

Note 1: * significant at 5% ; ** significant at 1%

Note 2. For the two-inflated GPM marginal effects have two partial components. Firstly, a variable x has an effect on ideal via the GP distribution. Secondly, x has an effect on ideal via its impact on the likelihood of observing a two count, delta. The total marginal effect of x on ideal equals the sum of these two partial effects. Here total marginal effects are presented for TIGPM.

Table 6.9 Sample selection regression for Ideal (Heckman model)
Coefficient [Std. Error]

Variable	Selection Criterion		
	(1) nresponse=1 moveout=0	(2) nresponse=1 moveout=0 motherhood=0	(3) nresponse=1 moveout=0 motherhood=0 work=0
Mean Equation			
Intercept	1.8940 [0.2828]*	1.9483 [0.2882]*	1.8904 [0.3332]**
Personal characteristics			
age	0.0173 [0.0172]	0.0131 [0.0178]	0.0234 [0.0224]
edu	-0.0223 [0.0079]*	-0.0207 [0.0082]*	-0.0324 [0.0107]**
siblings	0.0547 [0.0079]*	0.0564 [0.0079]*	0.0670 [0.0090]**
catholic	0.0434 [0.0456]	0.0397 [0.0460]	0.0569 [0.0517]
indspker	0.0919 [0.0715]	0.0516 [0.0725]	-0.0083 [0.0805]
Birth state dummies	Yes	Yes	Yes
Household Characteristics			
surban	-0.1582 [0.0403]*	-0.1744 [0.0408]*	-0.1686 [0.0468]**
urban	-0.3014 [0.0355]*	-0.3119 [0.0358]*	-0.2922 [0.0407]**
Head of Family Characteristics			
HFage	0.0053 [0.0019]*	0.0056 [0.0019]*	0.0047 [0.0021]*
HFedu	-0.0085 [0.0037]*	-0.0082 [0.0037]*	-0.0113 [0.0041]**
HFincome	0.0092 [0.0036]*	0.0082 [0.0036]*	0.0133 [0.0049]**
Family Structure			
AbsentFather	-0.0563 [0.0386]	-0.0522 [0.0390]	-0.0706 [0.0443]
AbsentMother	-0.1583 [0.0854]	-0.1519 [0.0870]	-0.0947 [0.0967]
Number of observations	7,032	6,892	5,628
Selection Equation			
Intercept	4.9271 [0.4877]*	5.6827 [0.4583]*	5.2601 [0.3364]**
Personal characteristics			
age	-0.2931 [0.0292]*	-0.3560 [0.0275]*	-0.3782 [0.0205]**
edu	0.1476 [0.0101]*	0.1599 [0.0095]*	0.1801 [0.0077]**
catholic	0.2397 [0.0682]*	0.1953 [0.0653]*	0.1421 [0.0532]**
indspker	-0.2061 [0.0980]*	-0.1522 [0.0947]	0.0719 [0.0814]
Birth state dummies	Yes	Yes	Yes
Number of observations	528	668	1,932
Sigma	1.1150 [0.0094]*	1.1143 [0.0095]*	1.1189 [0.0106]**
Rho	-0.0265 [0.0571]	-0.0281 [0.0571]	-0.0136 [0.0605]
LR Chi2 for rho=0	0.20	0.22	0.05
Pr > Chi2	0.6546	0.6375	0.8256

Note: * significant at 5% ; ** significant at 1%

Figure 6.1. Quantile Regression
Marginal Effects for ideal

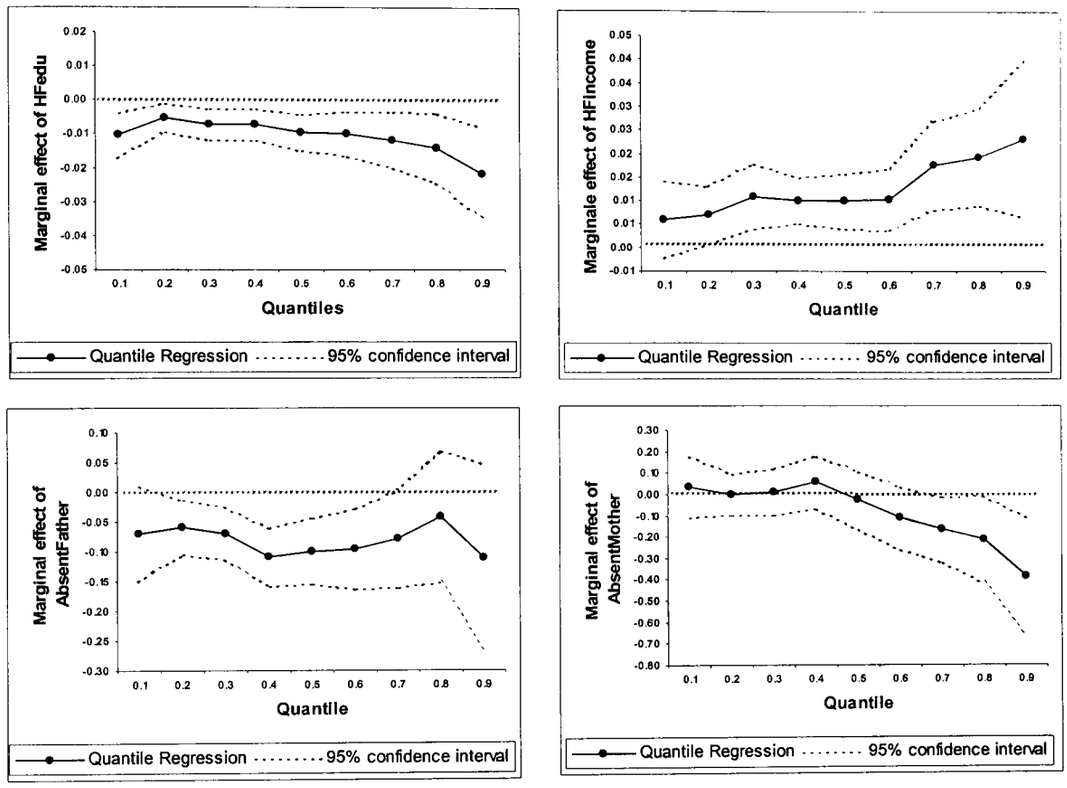
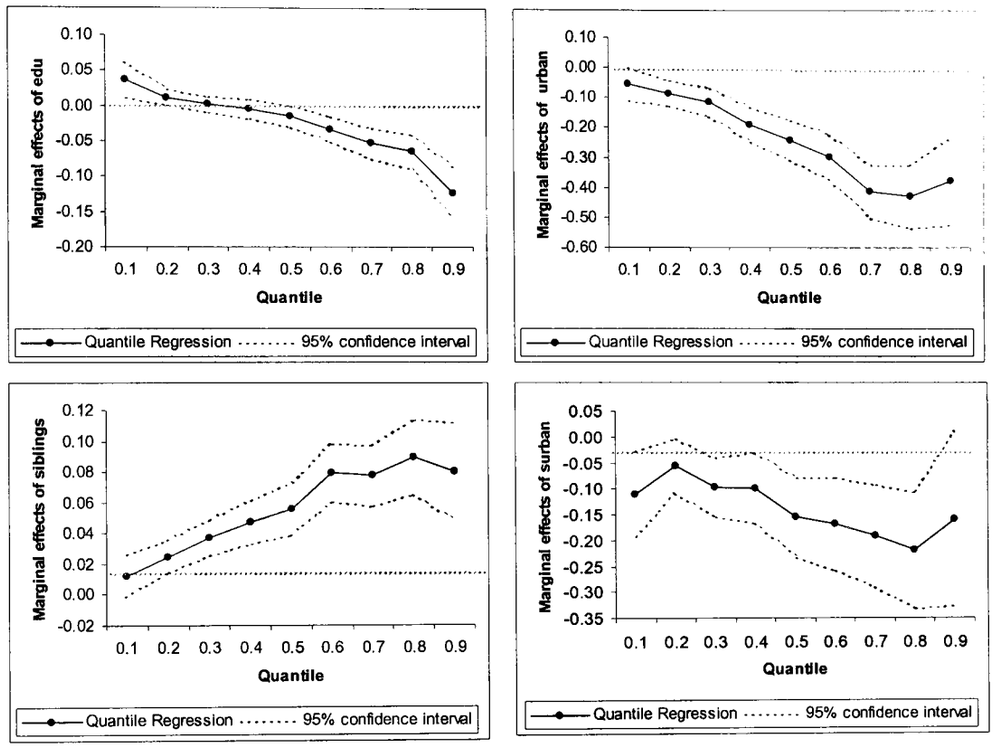


Figure 6.1. Quantile Regression
Marginal Effects for ideal (Cont.)



Chapter 7

Summary and conclusions

In the last forty years period fertility rates in Mexico have followed a steep and consistent downward trend. The fall in period fertility rates indicate that significant changes in individual fertility behaviour are taking place in the country. Without doubt, such modifications in individual behaviour are transforming the present and future demographic profile of Mexico.

Though demographic descriptive statistics suggest that reductions in lifetime fertility are present in the Mexican fertility transition, it is not clear whether, along with the fall of family size, there are changes in the timing of children of the sort observed in industrialised countries – i.e., postponement of first birth and increase on inter-birth intervals. Clearly, women may enter motherhood early in life and schedule children in a traditional fashion (i.e., allowing only for short inter-birth intervals) and yet reduce lifetime fertility by cutting off fertility at high parities using drastic natal control such as permanent female sterilisation. This 'stopping' strategy leaves untouched the timing of children and, potentially, can have adverse affects on women's accumulation of human capital and their

participation and performance in the labour market. An assessment of this issue is particularly needed given that female permanent sterilization has become the most popular contraceptive method in Mexico during the last twenty years while the demand for pills, condoms and other short-term contraceptives has dropped consistently.

After presenting a brief discussion of the economic theory on fertility behaviour (Chapter 2) and introducing the reader to the main demographic issues of modern Mexico (Chapter 3), Chapter 4 assesses these ideas in the context of the timing of a first child. The main objective is to test whether or not young cohorts of Mexican women are effectively delaying first birth with respect to older generations. Duration models are used as the main analysis tool. To the knowledge of the author, no previous study has discussed systematically and with the use of advanced econometric techniques the issue of first birth postponement in Mexico. Various aspects of the study should be underlined. First, unlike the common practice in the analysis of transition data, potential misspecification of the shape of the baseline hazard is avoided by means of estimating the hazard function in a semi-parametrical fashion so that no a-priori restrictions on the form of the baseline hazard are imposed. Second, and critically important to avoid spurious duration dependence, unobserved individual heterogeneity is controlled for using a discrete approximation of the distribution of unobservables (Heckman and Singer's mass points method). Finally, the presence of individuals that are never at risk of entering motherhood and remain childless until the end of fertile life is explicitly addressed to avoid potential misspecification of the hazard function.

Empirical results suggest that despite the popularity of permanent female sterilization as a contraceptive method in Mexico, young generations of Mexican women have tended to delay first birth in the last few decades. The study finds as well that Catholic individuals have a lower hazard of an early entry to motherhood than non-Catholic individuals. A result that supports the idea that Catholicism has not been a relevant factor preventing women from modifying their fertility behaviour in Mexico. Similarly, it is found that indigenous language speakers enter motherhood earlier than non-indigenous language speakers. This last finding is intuitive given the limited access to education and health that, as a general rule, individuals belonging to the indigenous ethnic groups have in Mexico. As suggested by economic theory, education is found to reduce the hazard of an early entry to motherhood.

Chapter five presents an analysis of completed fertility. The analysis stresses the fact that Mexico, being a country in the middle of a demographic transition, a non-negligible proportion of Mexican women have a large number of children and move to high order parities without taking any action to limit their fertility. Among other potential explanations, it is suggested that such behaviour may be displayed by women with large families who find themselves 'locked' in a regime where the opportunity cost of extra children becomes particularly low. Chapter five presents an innovative Double-Hurdle count model that explicitly allows for this sort of behaviour. Unlike previous work on Double-Hurdle modelling that set two different barriers at a zero count, the methodology developed for the study introduces a hurdle at zero and a hurdle at a strictly

positive value (interval) of the dependent count variable.[29] Hence, low and high order parities may be determined by two different data generating mechanisms (in other words, there are two different regimes). The approach accommodates potential endogenous switching between regimes and controls for unobserved heterogeneity. The study pays especial attention to establishing how socio-economic factors such as religion and ethnic group affect the likelihood of transition from low to high order parities.

In line with the results obtained in chapter four, Catholicism is found to be associated with reductions in the likelihood of transition from low to high order parities. Clearly, this result may be related to the relatively weak opposition of the Catholic Church to the diffusion of contraceptives in Mexico and its much stronger opposition to the initiation of sexual life before marriage. However, other factors may be at work. For instance, the existence of a large base of contraception users within the Catholic community may imply that a Catholic individual receives better information about the advantages of family planning relative to what is available for a non-Catholic individual.

Being an indigenous language speaker is found to increase the likelihood of transition from low to high order parities, especially for women born in the South and Centre of the country. Further, as suggested by economic intuition, education is found to reduce women's odds of having a large family. Conditional on observing a count larger than three children, Catholic individuals are found to

---

[29] Previous work on Double-Hurdle modelling introduces two barriers at zero so that a strictly positive outcome is observed only if these two "hurdles" are crossed. For further reference see, for instance.

have a significantly lower fertility than non-Catholic individuals only for women who were born in the south of the country. Similarly, being an indigenous language speaker is associated significantly with increases in completed fertility exclusively for women who were born in the South.

Chapter six moves the study to the determinants of planned fertility (ideal fertility) in Mexico. To avoid confusion between actual and planed fertility, attention is concentrated in analysing fertility plans of young Mexican women aged between 15 and 17 years that live at least with one biological parent and have neither initiated economic independent life nor entered motherhood. A Heckman-like assessment of the potential bias induced by this data selection process is performed. No empirical evidence of such a bias is found. Count data models are used as the main analysis tool, including an innovative technique for estimating quantile regressions for count data. The study enquires about the potential impact that family structure may have on fertility plans and on whether or not the socio-economic characteristics of the family head have a significant role in determining fertility plans of young female dependants.

The main findings support the idea that education reduces women's planned fertility. However, such a negative effect is significant only among women that have relatively strong preferences towards children. This result is consistent with economic intuition in the sense that women who demand many children are expected to substitute away from larger quantities of children when their opportunity cost becomes high.

As economic intuition would suggest, the number of siblings is found to increase women's planned fertility. Moreover, empirical evidence suggests that number of siblings has a higher effect the stronger preference towards children is

felt. Religion and ethnic background seem to have no statistically significant roles on the formation of fertility plans in Mexico. Finally, women who live in a parental household located in an urban zone are found to plan significantly fewer children than women who live in a parental household located in a rural zone. Clearly, besides urban/rural differences in the cost of children, this result may be an indicator of the differences on the supply of health care services in rural and urban zones.

Findings reported in chapter six indicate that an absent father impacts negatively on planned fertility, especially when women have weak tastes for children. An absent mother seems have a relevant role only if planned fertility is set to a large number. More importantly, empirical results indicate that women who live in a household that is headed by a highly educated individual plan fewer children than women who live in a household headed by a non-educated individual. Education of the family head has a stronger effect the stronger are preferences towards children.

A number of interesting topics related with fertility behaviour in Mexico remain to be studied in the future. For instance, research is needed to establish whether of not the demographic transition has led to changes in the timing of children for parities larger than one. It is likewise interesting to determine if socio-economic characteristics such as religion and ethnic background play a relevant role in the spacing of children once women have entered parenthood. In other issues, work is needed to study the links between fertility and labour market outcomes such as unemployment, training and female labour force participation. Finally a detailed analysis of the determinants of the demand for

contraceptives in Mexico is, without doubt, an important task to enhance the understanding of fertility behaviour in the future.

Bibliography

Agnarsson S and Carlin P (2002) Family Background and the Estimated Return to Schooling: Swedish Evidence. *Journal of Human Resources* **37** (1): 680-692.

Aguilar C (1993) Educación en Población: una necesidad inaplazable en el curriculum de la educación básica. *Demos* **6**: 34-35.

Allais M (1947) *Economie et interet.* Paris: Imprimerie National.

Arulampalam W and Booth A (2001) Learning and Earning: Do Multiple Training Events Pay? A Decade of Evidence from a Cohort of Young British Men. *Economics* **68**: 379-400.

Arulampalam W and Stewart M (1995) The Determinants of Individual Unemployment in an Era of High Unemployment. *The Economic Journal* **105**: 321-332.

Becker G (1960) An Economic Analysis Of Fertility And Demographic Change In Developed Countries. Princeton University Press and NBER.

Becker G and Barro R (1986) Altruism and the Economic Theory of Fertility. *Population and Development Review* **12** (supplement): 69-76.

Becker G and Barro R (1988) A Reformulation of the Economic Theory of Fertility. *The Quarterly Journal of Economics* **103** (1): 1-25.

Becker G and Lewis H (1973) On the Interaction between the Quantity and

Quality of Children. *The Journal of Political Economy* **81** (2): S279-S288.

Becker G, Murphy K, and Tamura R (1990) Human Capital, Fertility, and Economic Growth. *Journal of Political Economy* **98** (5): S12-S37.

Bergstrom T (1989) A Fresh Look at the Rotten Kid Theorem--and Other Household Mysteries. *The Journal of Political Economy* **97** (5): 1138-1159.

Blaylock J and Blisard W (1993) Women And The Demand For Alcohol: Estimating Participation And Consumption. *The Journal of Consumer Affairs* **27** (2): 319-334.

Bloom D and Trussell J (1984) What are the Determinants of Delayed Childbearing and Permanent Childlessness in the United States? *Demography* **21** (4): 591-611.

Bongaarts J and Menken (1983) The Supply of Children: A critical essay. In *Determinants of Fertility in Developing Countries: A summary of Knowledge.* Vol. 1. Edited by RA Bulatao and RD Lee. New York: Academic Presss. 27-60.

Bongaarts J and Watkins S (1996) Social Interactions and Contemporary Fertility Transitions. *Population and Development Review* **22** (4): 639-682.

Bumpass L, Rindfuss R, and Janosik R (1978) Age and Marital Status at First Birth and the Pace of Subsequent Fertility. *Demography* **12**: 75-86.

Cabrera G (1994) Demographic Dynamics and Development: The role of Population Policy in Mexico. *Population and Development Review* **20**: 105-120.

Cain M (1983) Fertility as an Adjustment to Risk. *Population and Development Review* **9** (4): 688-702.

Camarena M (1991) La escuela y el conocimiento demográfico. *Demos* **4**: 32-33.

Cameron A and Trivedi P (1986) Econometric Models Based On Count Data: Comparison And Applications Of Some Estimators and Test. *Journal of Applied Econometrics* **1** (1): 29-53.

Chamberlain G (1994) Quantile Regression, Censoring and the Structure of Wages. In *Advances in Econometrics.* Cambridge: Cambridge University Press.

Chen R and Morgan S (1991) Recent Trends in the Timing of First Births in the United States. *Demography* **28** (4): 513-533.

Cherlin A, Kiernan K, and Chase-Lansdale P (1995) Parental Divorce in Childhood and Demographic Outcomes in Young Adulthood. *Demography* **32** (3): 299-318.

Chwe M (2000) Communication and Coordination in Social Networks. *Review of Economic Studies* **67**: 1-16.

Cigno A (1986) Fertility and the Tax-Benefit System: A Reconsideration of the Theory of Family Taxation. *The Economic Journal* **96** (384): 1035-1051.

Cigno A (1994) Economic Considerations in the Timing of Births: Theory and Evidence. In *The Family, the Market and the State in Ageing Societies.* Edited by J Ermisch and N Ogawa. Oxford: Claredon Press.

Cigno A and Ermisch J (1989) A microeconomic analysis of the timing of births. *European Economic Review* **33** (4): 737-760.

Cigno A (98) Fertility Decisions When Infant Survival Is Endogenous. *Journal of Population Economics* **11** (1): 21-28.

CONAPO (2000) Migración México-EU presente y futuro.

----- (2001a) Marginality Index 2000. Document downloadable at www.conapo.gob.mx.

----- (2001b) La Poblacion de Mexico En El Nuevo Siglo. Document downloadable at www.conapo.gob.mx.

----- (2002a) National Population Development Plan 2001-2006.

----- (2002b) Proyecciones de la población de México, 2000-2050.

Consul P and Shoukri M (1985) The Generalized Poisson Distribution When The Sample Mean is Larger Than The Sample Variance. *Communications in Statistics-Simulation and Computation* **14** (3): 667-681.

Cragg J (1971) Some Statistical Models for Limited dependent Variables with Application to the Demand for Durable Goods. *Econometrica* **39** (5): 829-844.

Diamond P (1965) National Debt in a Neoclassical Growth Model. *The American Economic Review* **55** (5): 1126-1150.

Easterlin R (1975) An Economic Framework For Fertility Analysis. *Studies in Family Planning* **6** (3): 54-63.

Easterlin R (1987) *Birth and Fortune: The Impact of Numbers on Personal Welfare.* Chicago: University of Chicago Press.

Efron B (1992) Poisson Overdispersion Estimates Based on the Method of Asymetric Maximum Likelihood. *Journal of the American Statistical Association* **87** (417): 98-107.

Ellison G and Fudenberg D (1995) Word-of-mouth Communication and Social Learning. *The Quarterly Journal of Economics* **110** (1): 93-125.

Ermisch J and Ogawa N (1994) Age at motherhood in Japan. *Journal of Population Economics* **7** (4): 393-420.

Ermisch J and Francesconi M (2001) Family Structure and Children's Achievements. *Journal of Population Economics* **14** (2): 249-270.

Eswaran M (2002) The Employment Of Women, Fertility, And Child Mortality: Towards A Theoretical Analysis. *Journal of Population Economics* **15** : 433-454.

Faddy M (1997) Extended Poisson Process Modelling And Analysis Of Count Data. *Biometric Journal* **39**: 431-440.

Faddy M and Bosch R (2001) Likelihood-based Modelling And Analysis Of Data Underdispersed Relative To The Poisson Distribution. *Biometrics* **57**: 620-624.

Freedman R, Hermalin A, and Chang M (1975) Do Statements About Desired Family Size Predict Fertility? The Case of Taiwan, 1967-1970. *Demography* **12** (3): 407-416.

Gomez J (1996a) Fecundidad y Anticoncepcion . *Demos* **9**: 8-10.

Gomez L (1996b) La fecundidad y el crecimiento de la descendencia. *Demos* **9**: 12-13.

Gourieroux C, Monfort A, and Trognon A (1984) Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica* **52** (3): 701-721.

Greene W (1997) FIML Estimation of Sample Selection Models for Count Data. Manuscript.

Gustafsson S (2001) Optimal Age at Motherhood. Theoretical and Empirical Considerations On Postponement Of Maternity in Europe. *Journal of Population Economics* **14** (2): 225-47.

Gustavus S and Nam C (1970) The formation and Stability of Ideal Family Size Among Young People. *Demography* **7** (1): 43-51.

Happel S, Hill J, and Low S (1984) An Economic Analysis of the Timing of Childbirth. *Population Studies* **38** (2): 299-311.

Heckman J (1979) Sample Selection Bias as an Specification Error. *Econometrica* **47** (1): 153-162.

Heckman J, Hotz V, and Walker J (1985) New Evidence On The Timing And Spacing Of Births . *American Economic Review* **75** (2): 179-184.

Heckman J and Singer B (1984) The Identifiability of the Proportional Hazard Model. *The Review of Economic Studies* **51** (2): 231-241 .

Heckman J and Walker J (1990) Estimating Fecundability From Dara On Waiting Times To First Conception. *Journal of The American Statistical Association* **85** (410): 283-294.

Heckman J and Singer B (1984) A Method for Minimising the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* **52** (2): 271-320.

Hill M, Yeung W, and Duncan G (2001) Childhood Family Structure and Young Adult Behaviors. *Journal of Population Economics* **14** (2): 271-299.

Hindin M (2000) Women's Autonomy, Women's Status and Fertility-Related Behavior in Zimbabwe. *Population Research and Policy Review* **19**: 255-282.

Hotz V, Klerman, and Willis (1997) The Economics of Fertility in Developed Countries. In *Handbook of population and family economics.* Vol. 1A . Edited by MR Rosenzweig and Oded Stark. Amsterdam, New York and Oxford: Elsevier Science, North-Holland. 275-347.

Huber P (1981) *Robust Statistics.* New York: John Wiley & Sons.

INED (2004) Glosary of Demographic Terms. Document available at www.ined.fr.

INEGI (1999) National Survey of Demographic Dynamics 1997.

----- (2000) Estadisticas Historicas de Mexico.

----- (2001a) Censo de Población 2000.

----- (2001b) Sistema De Indicadores Para El S eguimiento De La Situacion De La Mujer En Mexico.

----- (2001c) Sistema De Indicadores Para El S eguimiento De La Situacion De La Mujer En Mexico.

Jensen E (1985) Desired Fertility, The "Up to God" Response, and Sample Selection Bias. *Demography* **22** (3): 445-454.

Jones A (1989) A Double-Hurdle Model for Cigarette Consumption. *Journal of Applied Econometrics* **4** (1): 23-39.

Kapur S (1995) Technological Diffusion With Social Learning. *The Journal of Industrial Economics* **43** (2): 173-195.

Kiefer N (1988) Economic Duration Data and Hazard Functions. *Journal of Economic Literature* **26** (2): 646-679.

King C (1989) Variance Specification in event count models: from restrictive assumptions to a generalized estimator. *American Journal of Political Science* **33** (3): 762-784.

Kirman A (1993) Ants, Rationality, and Recruitment. *The Quarterly Journal of Economics* **108** (1): 137-156.

Klawon E and Tiefenthaler J (2001) Bargaining over family size: The determinants of fertility in Brazil. *Population Research and Policy Review* **20**: 423-440.

Knodel J and Prachuabmoh V (1973) Desired Family Size in Thailand: Are the Responses Meaningful? *Demography* **10** (4): 619-637.

Koenker R and Bassett G (1978) Regression Quantiles. *Econometrica* **46** (1): 33-50.

Kohler H (1997) Learning in Social Networks and Contraceptive Choice. *Demography* **34** (3): 369-383.

----- (2000) Fertility decline as a coordination problem. *Journal of Development Economics* **63**: 231-263.

Labeaga J (1999) A Double-Hurdle Rational Addiction Model with Heterogeneity: Estimating the Demand for Tobacco. *Journal of Econometrics* **93**: 49-72.

Lee M (1992) Median Regression for Ordered Discrete Response. *Journal of Econometrics* **51** (1-2): 59-77.

Lee R (1980) Aiming at a Moving Target: Period Fertility and Changing Reproductive Goals. *Population Studies* **34** (2): 205-226.

Lindstrom D (1998) The Role of Contraceptive Supply and Demand in Mexican Fertility Decline: Evidence From a Microdemographic Study. *Population Studies* **52**: 255-274.

Machado J and Santos Silva J (2003) Quantiles for Counts. *Centre for Microdata Methods and Practice working paper* **CWP22/02**.

Manski C, Sandefur G, McLanahan S, and Powers D (1992) Alternative Estimates of the Effects of Family Structure During Adolescence on High School Graduation. *Journal of the American Statistical Association* **87** (417): 25-37.

Melkersson M and Rooth D (2000) Modeling Female Fertility Using Inflated Count Data Models. *Journal of Population Economics* **13** (2): 189--203.

Meyer B (1990) Unemployment Insurance And Unemployment Spells. *Econometrica* **58** (4): 757-782.

Mier y Teran M and Rabell C (1990) Introduccion: la transicion demografica en

la decada de los ochenta . *Revista Mexicana de Sociologia* **52** (1): 3-13.

Miles C (2001) Death and Divorce: The Long-Term Consequences of Parental Loss on Adolescents. *Journal of Labor Economics* **19** (3): 682-715.

Miranda A (2004) Completed fertility and the transition from low to high order parities in Mexico: A double-hurdle approach. Manuscript.

Moffit R (1984) Profiles of Fertility, Labour Supply and Wages of Married Women: A complete life-cycle model. *Review of Economic Studies* **51** (2): 263-278.

Montgomery M and Casterline J (1993) The Diffusion of Fertility Control in Taiwan: Evidence from Pooled cross-section time-series models. *Population Studies* **47** (3): 457-479.

Morand O (1999) Endogenous Fertility, Income Distribution, and Growth. *Journal of Economic Growth* **4** (3): 331-349.

Mullahy J (1986) Specification And Testing Of Some Modified Count Data Models. *Journal of Econometrics* **33**: 341-365.

Narendranathan W and Stewart M (1993) How Does The Benefit Effect Vary As Unemployment Spells Lengthen? *Journal of Applied Econometrics* **8**: 361-381.

Neher P (1971) Peasants, Procreation, and Pensions. *The American Economic Review* **61** (3): 380-389.

Newman J and McCulloch C (1984) A Hazard Rate Approach To The Timing Of Births. *Econometrica* **52** (4): 939-961.

Nugent J (1985) The Old-Age Security Motive for Fertility. *Population and Development Review* **11** (1): 75-97.

Ogawa N and Ermisch JF (1996) Family Structure, Home Time Demands, and

the Employment Patterns of Japanese Married Women. *Journal of Labor Economics* **14** (4): 677-702.

Pizzonia C (1996) El Programa Gente Joven. *Demos* **9**: 38-40.

Podlich H, Faddy M, and Smyth G (2004) Likelihood Computations For Extended Poisson Process Models. Manuscript.

Pollak R and Wachter M (1975) The Relevance of the Household Production Function and Its Implications for the Allocation of Time . *The Journal of Political Economy* **83** (2): 255-278.

Pörtner C (2001) Children as insurance. *Journal of Population Economics* **14** (1): 119-136.

Pritchett L (1994) Desired Fertility and the Impact of Population Policies. *Population and Development Review* **20** (1): 1-55.

Riley A, Hermalin A, and Rosero-Bixby L (1993) A New Look at the Determinants of Nonnumeric Response to Desired Family Size: The Case of Costa Rica. *Demography* **30** (2): 159-174.

Rodriguez G (1991) Educación sexual y políticas de Población. *Demos* **4**: 34-35.

Rosero-Bixby L and Csterline J (1993) Modelling Diffusion Effects in Fertility Transition. *Population Studies* **47** (1): 147-167.

Sah RK (91) The Effects of Child Mortality Changes on Fertility Choice and Parental Welfare. *Journal of Political Economy* **99** (3): 582-606.

Samuelson P (1958) An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money. *The Journal of Political Economy* **66** (6): 467-482.

San-Segundo M and Valiente A (2003) Family Background and Returns to Schooling in Spain. *Education Economics* **11** (1): 39-52.

Sanchez M and Monterrubio (2000) La experiencia de 25 años de educacion en poblacion en Mexico. In *La educación en población en México: hacia un nuevo modelo.* Edited by CONAPO.

Santos Silva J and Covas F (2000) A Modified Hurdle Model for Completed Fertility. *Journal of Population Economics* **13** (2): 173--88.

Schmidt P and Witte A (1989) Predicting Criminal Recidivism Using Split Population Survival Time Models. *Journal of Econometrics* **40**: 141-159.

Schultz T (1973) The Value of Children: An Economic Perspective. *The Journal of Political Economy* **81** (2): S2-S13.

Srinivasan T (1988) Fertility and old-age security in an overlapping generations model. *Journal of Quantitative Economics* **4**: 11-17.

SSA (Various years) Boletín de Información Estadística.

Sueyoshi G (1995) A Class of Binary Response Models for Grouped Duration Data . *Journal of Applied Econometrics* **10**: 411-431.

Terza J (1998) Estimating Count Data Models with Endogenous Switching: Sample selection and endogenous treatment effects. *Journal of Econometrics* **84**: 129--154.

Terza J (1985) A Tobit-Type Estimator For The Censored Poisson Regression Model. *Economics Letters* **18**: 361-365.

Trussell J and Menken J (1978) Early Childbearing and Subsequent Fertility. *Family Planning Perspectives* **10** (4): 209-218.

UN Population Division (2003a) Parnership and reproductive behaviour in low-fertility countries.

----- (2003b) Press Realease POP/850.

----- (2003c) World Population Prospects: The 2002 revision.

Van de Walle E (1992) Fertility Transition, Conscious Choice, and Numeracy. *Demography* **29** (4): 487-502.

Wang W and Famoye F (1997) Modeling Household Fertility Decisions with Generalized Poisson Regression. *Journal of Population Economics* **10** (3): 273-83.

Ward M and Butz W (1980) Completed Fertility And Its Timing. *Journal of Political Economy* **88** (5): 917-40.

Welti C (1997) Cambios En La Fecundidad. *Demos* **10**: 16-18.

Willis R (1973) A New Approach to the Economic Theory of Fertility Behavior. *Journal of Political Economy* **81** (2): s14-s63.

Winkelmann R (1995) Recent Developments in Count Data Modelling: Theory and Application. *Journal of Economic Surveys* **9** (1): 1-24.

Winkelmann R (1998) Count Data Models with Selectivity. *Econometric Reviews* **17**: 339-359.

Winkelmann R (2000) *Econometric Analysis of Count Data.* Berlin Heidelberg, New York: Springer.

Wolpin K (84) An Estimable Dynamic Stochastic Model of Fertility and Child Mortality. *The Journal of Political Economy* **92** (5): 852-874.

World Bank (2001) World Development Indicators.

----- (2003) World Development Indicators.

Yen S and Jensen H (1996) Determinants of Household Expenditures on Alcohol. *Journal of Consumer Affairs* **30** (1): 48-67.

Yen S, Tang C, and Su S (2001) Demand for Traditional Medicine in Taiwan: A Mixed Gaussian-Poisson Model Approach. *Health Economics* **10**: 221-232.

Zavala de Cosio E (1989) Dos momentos en la transicion demografica. *Demos* **2**:

6-7.

Zavala de Cosio E (1990) *Cambios de la fecundidad en Mexico y Politicas de Poblacion.* Mexico: Fondo de Cultura Economica/Colegio de Mexico.

Zhang J and Nishimura K (1993) The old-age security hypothesis revisited. *Journal of Development Economics* **41** (1): 191-202.