I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13357

Econometric Models of Fertility

Alfonso Miranda
CIDE and IZA

Pravin K. Trivedi
University of Queensland

JUNE 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Econometric Models of Fertility*

This paper reviews some key contributions to econometric analysis of human fertility in the last 20 years, with special focus on discussion of prevailing econometric modeling strategies. We focus on the literature that highlights the role of the key drivers of the birth outcomes, including age at entry into motherhood, the number of children, and the time between births. Our overall approach is to highlight the use of single equation reduced form modelling, which has important advantages but has the limitation of typically being unable to shed light on detailed causal mechanisms through which exogenous factors such as birth control and infant mortality, and policy variables such as child allowances and tax incentives, impact fertility. Structural models that embed causal mechanisms explicitly are better suited for this objective. We start with a description of the subject matter, including a brief review of existing theories of fertility behaviour and a detailed discussion of the sources of data that are available to the analyst. At this point we stress the intrinsic dynamic nature of fertility decisions and how such dynamics create data with empirical features that pose important challenges for modelling. Once the nature of the problem and the characteristics of the data are spelled out, we proceed to review the different econometric approaches that have been used for modelling fertility outcomes with cross-section and panel data.

JEL Classification: C21, C23, C41, J13

Keywords: econometrics of fertility, cross-section and panel data models, count data models, hazard (survival) models

Corresponding author:
Alfonso Miranda
División de Economía
Centro de Investigación y Docencia Económicas (CIDE)
A.C, sede Región Centro
Circuito Tecnopolo Norte 117
Aguascalientes 20313
México

E-mail: alfonso.miranda@cide.edu

* A revised version of this discussion paper is forthcoming as a chapter of Nigar Hashimzade and Michael Thorton (eds.), *Handbook of Research Methods And Applications In Empirical Microeconomics*, Edward Elgar Publishing. The usual disclaimers apply. We thank Rainer Winkelmann, whose insightful comments helped to improve this chapter.

1 Introduction

This paper reviews some key contributions to econometric analysis of human fertility in the last 20 years, with special focus on discussion of prevailing econometric modeling strategies. To keep the task manageable and the result useful, we restrict our focus on the strand in the literature that seeks to highlight the role of the key drivers of the birth outcomes, i.e. where investigations model outcomes such as the age at entry into motherhood, the number of children, and the time between births. Hence, we are mainly concerned with work that views fertility and related variables as the outcome variables of interest and socioeconomic characteristics are key explanatory covariates. Thus we exclude works where the main interest lies in interdependence between and interactions with fertility and other response variables such as: infant mortality, labour force participation, labour supply, and/or educational attainment, in which fertility enters as an important predetermined or jointly determined control variable.

Our overall approach is to use single equation reduced form type modelling of fertility. Such an approach suffers from limitation that typically it does not shed much light on detailed causal mechanisms through which exogenous factors such as birth control and infant mortality, and policy variables such as child allowances and tax incentives, impact fertility. Structural models that embed causal mechanisms explicitly are better suited for this objective. Nevertheless the reduced form approach is widely used, especially in demographic literature. It is useful for studying associations, projecting fertility patterns and trends, and related demographic features.

We start with a description of the subject matter, including a brief review of existing theories of fertility behaviour and a detailed discussion of the sources of data that are available to the analyst. At this point we stress the intrinsic dynamic nature of fertility decisions and how such dynamics create data with empirical features that pose important challenges for modelling.

Once the nature of the problem and the characteristics of the data are spelled out, we pro-

ceed to review the different econometric approaches that have been used for modelling fertility outcomes with cross-section and panel data. We shall discuss the properties of a standard OLS estimator before reviewing more popular count and hazard models. Notwithstanding the popularity of this class of models, there are important aspects of the fertility process that can be better handled econometrically using a discrete-time dynamic multi-spell duration model which we then go on to analyze. A detailed discussion of the main advantages and limitations of each approach is provided. The paper ends with an illustrative example of an econometric modeling strategy based on panel data.

2 Nature of the subject matter

2.1 Theories of fertility behaviour

How do couples decide on the number of children and the timing of arrival of each child? Do they set a plan at the outset for the full fertility cycle that is strictly followed to completion or are the outcomes sequentially determined? Does economics play a role or are traditional and established socioeconomic norms dominant? Why does fertility decline (or transition to a lower level) as economies transition from low average income to high average income? Generations of demographers, sociologists and economists have wrestled with these questions and yet a consensus view has not yet emerged, even though it is widely acknowledged that the *homo economicus* and *homo sociologicus* positions are not necessarily mutually exclusive. Both components are present in a typical reduced form model.

The remainder of this section will briefly review the main ideas in the field. To keep the scope of the paper manageable, here we focus on reviewing 'micro theories', which are concerned with explaining individual behavior at the level of family, and for which econometric models are typically based on cross-section or longitudinal data collected over relatively short periods of time; including studies on desired fertility fertility. These studies have a microeconometric flavor. At the other end of the spectrum are 'macro theories' concerned with

explaining population or society-wide phenomena—such as the epidemiological and fertility transition triggered by the industrial revolution (see, for instance, Landry 1909, Thompson 1929, Notestein 1945). These studies attempt to model long-term movements in population fertility rates—such as the total fertility rate[1]—using country or region level time series or panel data that span many decades and generations. Studies of the so-called "great transition" from low-income-high-fertility state to high-income-low-fertility state, especially in Europe, are a leading example of this type of research with macroeconometric flavor (see, for instance, Boldrin et al. 2015, Sánchez-Barricarte 2017).

Another topic with a similar macro emphasis concerns short run variations in age- or group-specific fertility rates that are observed in wars, famines, and humanitarian crises stemming from social and political upheavals. There are numerous examples in the literature. For example, Vandenbroucke (2014) finds that, the birth rate decreased dramatically during WWI, and then later recovered. Caldwell (2006) shows that European countries, the United States and Japan had their fertility levels reduced due to local armed conflicts. There are also other studies of fertility changes using data from Angola, Agadjanian and Prata (2002), Rwanda (Jayaraman et al. 2009), Eritrea (Woldemicael 2008) and so forth, documenting similar significant short-term fluctuations. A different strand in the literature uses, for example, data from Columbia (Torres and Urdinola 2019) and Mexico (Torche and Villarreal 2014), and explores the connection between within-country violence and its differential impact on adolescent and older women's pregnancies, and, more generally, on changes in age at first pregnancy. A third strand in the literature concerns impact of humanitarian crises arising from famines, tsunamis, and earthquakes on birth-rates. An example is the Dutch famine of 1944-45 (Stein and Susser 1975). We group these three strands in the literature under the heading of short term variations in fertility. From a microeconometric viewpoint these studies raise several unresolved issues such as: How do these shocks fit into the established micro theories? Is the observed behaviour

[1] A total fertility rate (TFR) is defined as the number of children a woman is expected to have over her lifetime if she behaves according to a current schedule of age-specific fertility rates. Sometimes the TFR is written in terms of expected children per 1,000 women. In this definition, an age-specific fertility rate is taken to be the annual number of live births to women of a specified age or age group per 1,000 women in that age group.

optimal in some sense? Are the observed short-term changes driven by behavioural responses to transitory shocks or are there biological mechanisms at work? Are observed changes purely transitory and mean-reverting or do they interact with underlying long-term trends? These issues are important but beyond the scope of this paper.

2.1.1 Main ideas from economics

We may begin by asking: Why do people have children? There are two main, not necessarily contradictory, views. One theory is based on the idea that people derive utility from having and raising children, so that children are just like any other consumption good c that gives utility to the consumer. We call this the child-in-the-utility-function approach. A major complication is that parents may care not only about the number n but also about the 'quality' q of their children. Parents face the following problem:

$$\begin{aligned} \max_{\{c,qn\}} U(c,qn) &= \alpha log(c) + (1-\alpha) log(qn) \\ st.\ c + qn &= m. \end{aligned} \tag{1}$$

where, without loss of generality, we have set the price of the quantity and quality of children to one and use the price of the consumption good as a numeraire — so that q plays the role of relative price for n and n plays the role of relative price for q. The solution is

$$\begin{aligned} qn &= (1-\alpha)m \\ c &= \alpha m \end{aligned} \tag{2}$$

Notice that the number and quality of children remain indeterminate but the relationship between the quantity and quality of children takes the form of

$$n = \frac{(1-\alpha)m}{q}, \tag{3}$$

so that for every level of quality there is an optimal number of children. Hence, a quantity-quality trade-off of children arises because the shadow price of quantity depends on quality—and vice versa. In their seminal work Becker (1960) and Becker and Lewis (1973) introduce these ideas in a general form utility function and show that, if the income elasticity for quality is larger than the income elasticity of quantity, parents will substitute quality for quantity when income increases. This mechanism, they conclude, may explain the fertility transition that was triggered by the industrial revolution: fertility declines when households become wealthier.

To consider the role of technological change we may introduce a quality production function (see, for instance Morand 1999, Rosenzweig and Wolpin 1980)

$$q = b(t)n^{-\delta} \tag{4}$$

that keeps the inverse relationship between quantity and quality of children but introduces a new multiplicative term $b(t)$ that is a function of time. Then we solve for n and q

$$n = \left[\frac{(1-\alpha)m}{b(t)}\right]^{\frac{1}{1-\delta}} \tag{5}$$

$$q = b(t)^{\frac{1}{1-\delta}}\left[(1-\alpha)m\right]^{-\frac{\delta}{1-\delta}} \tag{6}$$

and conclude that technological change plays the role of increasing the quality of children over time, which leads to a reduction in the quantity of children. Think of $b(t)$, for instance, as the effect of the discovery of antibiotics and the improvement of health care on child mortality and life expectancy. Now the fertility transition can be seen not just as the result of an income effect but, more generally, as the result of technological change.

A different theory is based on the idea that people do not enjoy children—and thus, do not directly derive utility from them—but use children to transfer consumption over time. This is an overlapping generations model where individuals live for two periods, receive income m_1 only in period one, and cannot save for old-age (see, for instance, Allais 1947, Samuelson 1958, Diamond 1965). The only way of securing an old-age pension is having children in

period one. Hence parents solve the following problem:

$$\max_{\{c_1,c_2\}} U(c_1,c_2) = \alpha log(c_1) + (1-\alpha) log(c_2) \tag{7}$$

$$st.\ c_1 + \frac{c_2}{1+\pi qn} = m_1, \tag{8}$$

where for an investment of $(m-c_1)$ monetary units in child services qn in period one, parents are paid an old-age pension of $(1+\pi qn)(m-c_1)$ monetary units from their children in period two. The solution is:

$$\begin{aligned} c_1 &= \alpha m_1, \\ c_2 &= (1-\alpha)(1+\pi qn)m_1. \end{aligned} \tag{9}$$

Here, again, the number and quality of children remain indeterminate. However, if we fix the rate of return of child services to $\bar{\kappa} = \pi qn$, appealing to the existence of a known and binding contract between overlapping generations, the relationship between the quantity and quality of children takes the form of

$$n = \frac{\bar{\kappa}}{\pi q}, \tag{10}$$

which describes a quantity-quality trade-off of children that is very similar to the one we had before. Again, for every level of quality there is an optimal quantity of children. We call this the child-in-the-budget-constraint approach, which describes what is known in the literature as the old-age pension motive for fertility (Nugent 1985, Srinivasan 1988).

An alternative description of the old-age pension motive argues that children are an insurance device that allows parents to reduce uncertainty about unforeseeable shocks to their health and/or income (Nugent 1985, Pörtner 2001). In modern times, however, where well developed security markets and pension systems exist, the old-age pension motive for fertility is weaker and couples may set fertility near zero (Neher 1971).

In this model we may also introduce a quality production function as the one in (4) to investigate the role of technology. Consider the following technology:

$$q = b(t) + n^{-\delta} \tag{11}$$

so that parents solve the problem of

$$max_{\{n\}} Q(n) = m_1 (1-\alpha) \left(1 + \pi \left(b(t)n - n^{1-\delta}\right)\right) \tag{12}$$

which has solution

$$n = \left(\frac{(1-\delta)}{\pi b(t)}\right)^{\frac{1}{\delta}} \tag{13}$$

$$q = b(t) + \left(\frac{\pi b(t)}{(1-\delta)}\right). \tag{14}$$

Here, once again, the quality increases with technological change whereas the quantity decreases with it.

There are many other complementary ideas. For instance, Willis (1973) explores how women's participation in the labor market could affect their fertility decisions. As women enter the labor market, the argument goes, the opportunity cost of children increases because the time women spend in child-rearing activities is time they cannot spend at work. Therefore, it is predicted, couples demand fewer children when female education and wage increase. A similar argument is put forward by Becker et al. (1990) and Galor and Weil (1996) in the context of a growth model with endogenous fertility and by Rosenzweig and Wolpin (1980) and Heckman and Walker (1990b) in a joint model of labor supply and fertility.

Becker and Barro (1986; 1988) and Becker et al. (1990) investigate how introducing altruism into the child-in-the-utility-function framework affects fertility decisions. Parents, they argue, do not care only about their own welfare, but also about the welfare of their children, their grand children, and their great grand children. So, when deciding about their own fertility,

parents act like a central planner who takes into account how current fertility decisions affect the welfare of their whole 'dynasty'. Physical and human capital accumulation are allowed. While physical capital is subject to diminishing or constant returns to scale, human capital accumulation exhibits increasing returns to scale. All these features come together in what is known today as a growth model with endogenous fertility in the literature, which in turn, is a type of endogenous growth model; see for instance Uzawa (1965), Nelson and Phelps (1966), Arrow (1972), Romer (1990). Besides the classic quantity-quality trade-off that is present in all child-in-the-utility-function specification, these models show that the demand of children is a function of all goods and time that is spent on child-care activities.

Moreover, Becker et al. (1990) show that an economy needs a minimum stock of human capital to create enough incentives for individuals to invest in education and be able to reach a steady state with low fertility and high human capital. When this minimum human capital is not present, the economy converges to a steady state with high fertility and no human capital accumulation. Becker et al. (1990) think of this mechanism as an explanation of the fertility transition and the way large economic disparities were created between modern developed and developing countries.

Regarding the timing of children, the main blocks of theory are due to Happel et al. (1984). The authors start with a child-in-the-utility framework and force parents to have only one child within their lifetime, which span is known and limited to, say, three "days".[2] Women are young during the first two days and may work and have children in that period of life. On day three women are old, retired, and may not have children. So, once the quantity of children issue is gone: Which day should parents have their child 'delivered'?

If parents like children and no cost is paid for rearing them, then it is dynamically optimal to have their child delivered on day one. However, when parenthood involves some costs things are no longer clear cut. For instance, as Happel et al. (1984) put forward, if women leave work for some time after a birth and human capital 'depreciates' while away of the workplace, then

[2]Limiting the life span to three days and considering that time is measured in discrete units is not the way Happel et al. (1984)'s dynamic model is set-up, however, here we use here such a simplification to discuss the main results of their model without going in much mathematical detail.

those highly qualified will have incentives to postpone motherhood towards the end of their fertile period — i.e. the end of day two — in order avoid as much as possible the depreciation of their human capital that inevitably motherhood will bring about. Those with low human capital, on the contrary, may find postponement not quite attractive and choose to have their child delivered on day one.

Summarising, consistent with Becker (1960)'s human capital theory, Happel et al. (1984) predict that women postpone motherhood as they accumulate human capital.

2.1.2 Alternative ideas from other fields

Demographers and sociologists have put forward ideas that extend and/or depart from the mainstream ideas based on Becker (1960)'s human capital theory.

An important idea put forward by demographers and sociologists is that, besides socio-economic factors, fertility is strongly influenced by social norms for family size/composition and uncertainty about the availability and costs of contraception. Effective contraception technology became widely available at the turn of the 20th century. Adoption of contraceptives, however, took decades and varied widely across the globe. Among other things, people of different countries and social backgrounds had, and still have, different levels of access to health services and, as a consequence, face widely different mortality risks let alone different religious beliefs and social/political institutions. In such contexts, some women may find it more difficult than others to get reliable information about the costs and risks of using contraceptives—many get information and opinions about family planning from family members, friends, and social contacts, who are not necessarily well informed themselves (see, for instance, Bongaarts and Watkins 1996). Under this perspective the demographic transition is seen as a dynamic process of diffusion of knowledge and adoption of new techniques of contraception and fertility 'norms'; see for example Montgomery and Casterline (1993), Rosero-Bixby and Casterline (1993), Bongaarts and Watkins (1996), Kohler (1997; 2000). In such a scenario, women, who are the decision unit, choose either to follow traditional fertility patterns or to adopt modern

contraceptives and reduce their lifetime fertility. In this context social-network effects are present so that an individual's costs or/and 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 a new fertility standard is generated (Kohler 1997, Montgomery and Casterline 1993, Ellison and Fudenberg 1995, Kapur 1995, Kirman 1993, Chwe 2000). This is, as a whole, an alternative micro-founded mechanism that explains the fertility transition.

2.2 Types and features of fertility data

In this section we illustrate some key features of fertility data that econometric models would be expected to account for. Two broad population categories are low-fertility-high-income and high-fertility-low-income. In 2012 total fertility rate (TFR) was 2.4 for the world, 1.7 for high income developed countries, 5.2 for sub-Saharan Africa, and 5.7 for Nigeria; see Fagbamigbe and Adebowale (2014). In high income populations birth counts cluster around a handful of values such as $(0,1,2,3)$ with a very short tail; in low-income populations both the mean and variance tend to be higher. Long time series (especially) European historical data show a rapid transition from high to low fertility regime as family incomes rise. At the risk of slight oversimplification one can summarize these features by treating the low TFR case as one involving underdispersed counts, and the high TFR case as one of overdispersed counts. Such a distinction takes the Poisson distribution, which has the property of mean-variance equality (equidispersion), as a benchmark.

There are two types, or flavours, of fertility data analyses: (a) using completed fertility data, and (b) using fertility history data.

Completed fertility dataset is by far the most widely used. It contains information on a cross-section of women who, at the time of the survey, are at the end of their childbearing life (normally aged 45 and over) and can report their 'completed fertility'—i.e. the total count of children ever born alive to a woman. Besides the number of children, completed fertility data typically have information on some characteristics of the mother, as measured at the time of

Table 1. Number of children: actual frequency distribution British data (N=5706)

Children	0	1	2	3	4	5	6	7	8	9	10	11
Freq.	1524	675	1779	1075	409	153	50	22	12	3	3	1
Percent	26.7	11.8	31.2	18.8	7.2	2.7	0.9	0.4	0.2	0.1	0.1	0.0

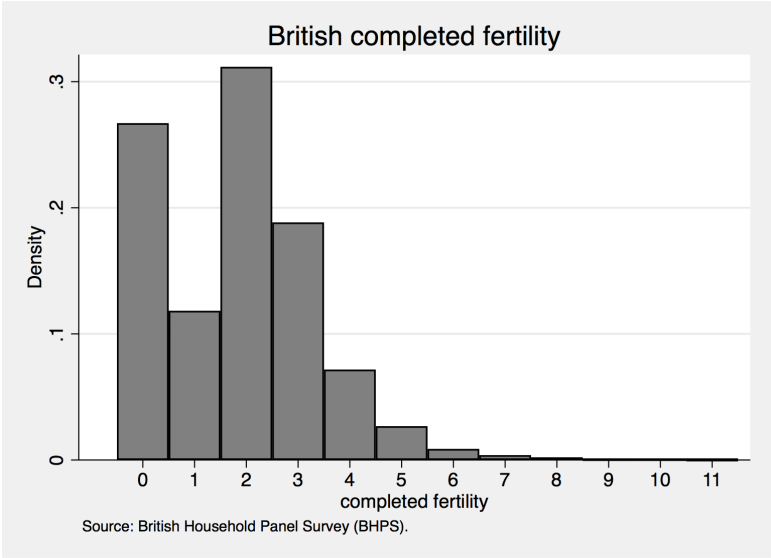
interview, including education, income, work status, occupation, and offspring sex composition. Measuring mother's characteristics at the end of fertile life may provide current or recent information but is less likely to provide information about the conditions (e.g. income, employment status, information about contraception) that prevailed at the time of each birth; this, in turn, is likely to lead to specification errors. So, the cross-section nature of the data becomes a serious drawback.

Because completed fertility comes as a count—i.e. a non-negative integer variable with values that accept a cardinal interpretation—the use of count data econometric models based on crossection data is a popular choice. We review those techniques in section 4.

Several features of completed fertility data are important and require careful consideration. We illustrate the relevant points with two examples. Table 1 reports data on completed fertility from the British Household Panel Survey (BHPS). This is a low-fertility-high-income setting ($N = 5706$). The sample mean number of children is 1.84 and the standard deviation is 1.5. There is some overdispersion with a sample variance is 1.23 times the mean. The distribution is clearly bimodal (see figure 1) with a half-mode at 0 and a more pronounced mode at 2. This latter feature has led Melkersson and Rooth (2000) to describe the distribution as being "inflated at zero and two" . Overdispersion and inflation at specific frequencies are features that motivates empirically important extensions of the Poisson regression that we cover in section 4.

From this examples it is possible to observe that completed fertility data in a low-fertility-high-income setting is characterized by limited support, in the sense that just a few event counts dominate the frequency distribution. A flexible econometric modeling strategy is required to capture such population differences in regression models.

Figure 1. Frequency distribution for British completed data



Data from high-fertility-low-income setting are different. Consider for example the case of the Mexican Family Life Survey (2002) ($N = 3,674$), which distribution of completed fertility is reported in table 2. In this case the sample mean is 3.19 and the standard deviation is 1.97, so the sample variance is 1.21 times the mean—a lower mean/variance ratio than what is typically observed in a low-fertility-high-income setting. Overdispersion, however, is not all that is different. Here there is not so much of a spike at 0 or 2. In fact, a 2-children outcome is almost as popular as a 3-children outcome. More importantly, the fertility distribution exhibits a much longer tail in the high-fertility-low-income setting than in the low-fertility-high-income setting. For the Mexican data outcomes 4 and 5 still carry a non-ignorable portion of the total probability mass. Miranda (2010) analyzes Mexican fertility data from the National Survey of Demographic Dynamics 1997 and reports a similar distribution. The author suggests that data from developing countries may show 'an excess of large counts' that requires explicit modelling—and understanding of the underlying behavioural drivers.[3]

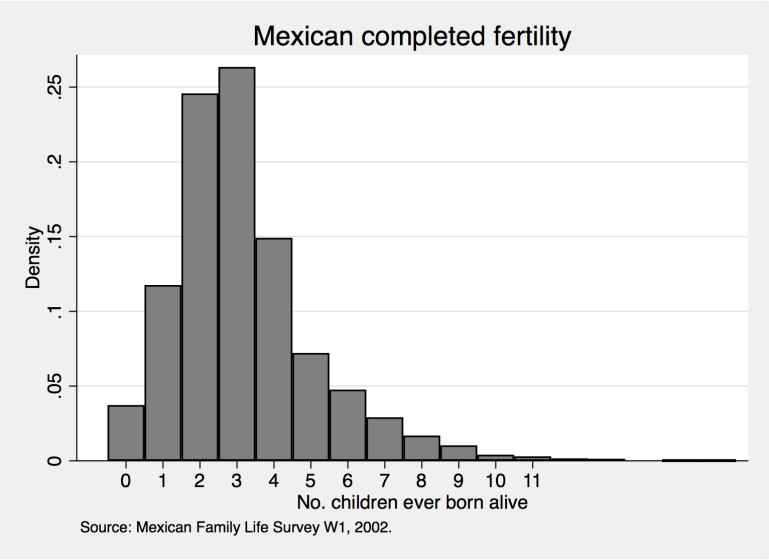
Regarding the timing of children data typically comes in the form of an event history,

[3]Over time, however, societies may transit from high to low fertility rates as social norms and economic choices favor smaller family units. Planned fertility studies that ask how many children a family would like to have generate samples with just a few realizations even in high-fertility-low-income settings (Miranda 2008).

Table 2. Number of children: actual frequency distribution for Mexican data (n=3674)

Children	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Frequency	137	432	903	968	548	265	175	107	62	38	15	11	6	5	0	1	1
Percent	3.7	11.8	24.6	26.4	14.9	7.2	4.8	2.9	1.7	1.0	0.4	0.3	0.2	0.1	0	0.0	0.0

Figure 2. Frequency distribution for Mexican completed fertility data



where a sample of women, not necessarily at the end of fertile life, provide exact birth dates for each child they have ever given birth by the time of the survey interview. An event history has a longitudinal design from the onset. However, data may be collected in a retrospective or prospective manner.

Retrospective data are collected in a single point of time—and from this point of view it is a cross-section—but involve asking people to look back and report events that occurred in the past; such as the date they first married or the date they left school. Using this technique, it is possible to 'rebuild' a fertility history retrospectively. The downside is that people may suffer substantial problems of "recall bias" or "telescoping". Recall bias arises when people selectively recalls better certain (positive) things or experiences than others (negative), while telescoping is present when an individual's perceived elapsed time since the occurrence of an

event is different depending on whether the experience was pleasant or not. In both cases, recall bias and telescoping, retrospective collection of an event history may introduce serious measurement error and bias (see, for instance, Gray 1955, Mathiowetz and Ouncan 1988, Bound et al. 2001, Pyy-Martikainen and Rendtel 2009). Such issues are worse for some variables than for other variables. For instance, most people can recall the exact birth dates of their children but have serious difficulties to recall their past salary.

In contrast, prospective longitudinal data ("panel data") surveys a sample of women over time and pregnancy and childbirth events are carefully recorded along many other individual characteristics. From a theoretical point of view, building an event history using prospective longitudinal data is better than building it retrospectively. It is not only that prospective data is less susceptible to recall bias and telescoping; the real advantage is that a number of important fertility determinants, such as income and employment status, get measured contemporaneously at each follow-up. Hence, besides a fertility history, panel data delivers a detailed history of the factors or/and variables that determine woman's fertility behaviour. From this point of view, analyses based on panel data are more promising. In practice, however, commonly available prospective longitudinal data may have gaps that impede identification.

In cohort studies, for example, children who are born in a given year/date are followed through time. For this type of study researchers may have to wait between 20 and 30 years to get interesting fertility data. Unfortunately, by the time cohort members start marrying and having children, most cohort studies may have to rely on data that are subject to substantial attrition bias due to drop-outs; see Cheng and Trivedi (2015) and chapter 19.9 in Wooldridge (2010).

A household panel study based on a random sample of the population also has its own problems. A major issue is that, by design, these studies follow a sample of women of different ages. As a consequence, many panel members enter the study years after they started having children. Discarding observations who have already entered motherhood by the start of the study is an option. However, in most cases, taking such a step results in small sample sizes that still require time to deliver full-blown fertility histories. Hence, as before, attrition becomes a

potentially serious problem unless it is purely random.

What is then the best option to build a fertility history? We suggest that the best option is to use a combination of prospective and retrospective data. The British Household Panel Data (BHPS) took such approach. Indeed, the BHPS follows prospectively all women who are panel members and records contemporaneously any births that occur along the study time. To complement, the BHPS introduced a retrospective fertility module in waves 2, 11 and 12. An example of the layout of a fertility history in the "long form" (ready for analysis) is given in table 3.

Table 3. Example of prospective and retrospective longitudinal fertility history data from the UK (long form)

pid	year	occ	resp	parity	clock	age	girl at $P = 1$	twins	same sex at P=2	income	mother's edu
1	1960	1	0	0	1	18	0	0	0	7.20	GCSE
1	1961	2	0	0	2	19	0	0	0	8.35	GCSE
1	1962	3	0	0	3	20	0	0	0	8.35	GCSE
1	1963	4	0	0	4	21	0	0	0	7.97	GCSE
1	1964	5	0	0	5	22	0	0	0	9.33	GCSE
1	1965	6	1	0	6	23	0	0	0	8.88	GCSE
1	1966	7	0	1	1	24	1	0	0	7.40	GCSE
1	1967	8	0	1	2	25	1	0	0	8.57	GCSE
1	1968	9	1	1	3	26	1	0	0	9.92	GCSE
1	1969	10	0	2	1	27	0	0	1	10.33	GCSE
1	1970	11	1	2	2	28	0	1	1	10.67	GCSE
1	1970	12	1	2	2	28	0	1	1	10.67	GCSE
1	1971	13	0	4	1	29	0	0	0	11.30	GCSE
1	1972	14	0	4	2	30	0	0	0	11.77	GCSE

In the example, the fertility history of a woman is recorded after age 18. Every row is constitutes a new measurement occasion (occ), which is not necessarily equivalent to the passage of calendar time, and a record of the status of all relevant variables gets included. The main response $(resp)$ is a $0/1$ binary variable that takes value one if on a given measurement occasion a new birth is registered. The variable *parity* measures the total number of children, or fertility, that the woman has had at each measurement occasion. So, parity increases by one unit every

time period $resp = 1$.[4] In our example, the survey follows the woman for 13 years (periods), has a child at age 23 and at age 26. At age 28 she has twins. The clock indicates, at each occasion, the time elapsed since the woman entered a particular parity and initiated a new duration spell to the next pregnancy/birth. With each new birth the clock gets restarted. Age, calendar time, and duration time are three concepts that vary in a different manner for each woman in the sample and, as a consequence, can be identified separately. To complete the picture, the data contains information on a set of control variables (regressors) observed during the fertility history. There are variables that are time-varying, such as income and age; variables that are time-fixed, such as (our individual) mother's education; and variables that change with parity, such as whether the first born was a girl or whether a two children of the same sex occurred at parity two. The example illustrates the richness that a longitudinal fertility history can offer to the analyst. These type of data are typically analysed with discrete hazard models, which are discussed in section 5.

Table 4. Example of longitudinal fertility history data from the UK (wide form)

pid	dur	fail	parity	girl at $P = 1$	age	income	same sex at $P = 2$
8	22	1	0	0	33	8.3518252	0
8	27	0	1	1	60	6.8155894	0
9	12	1	0	0	23	4.3997666	0
9	2	1	1	0	25	4.3997666	0
9	2	1	2	0	27	4.3997666	1
9	5	1	3	0	32	4.3997666	0
10	6	1	0	0	17	0	0
10	3	1	1	1	20	0	0
10	2	1	2	0	22	0	1
10	34	0	3	0	55	0	0

Another form in which fertility histories may come is presented in table 4. We call this 'the wide form' and each row contains data for one event (i.e. one birth) along with a measure of how long the duration spell or waiting time to event lasted (*dur*). In our context, an event is a

[4] In the demographic literature *parity* refers to the number of children previously born to a woman at a given point of time. Parity increases with the arrival of a newborn.

new birth. Along with the length of the duration spell, a dummy variable $fail$ indicates whether the spell was ended by the occurrence of a birth $fail = 1$, called a 'failure' in survival analysis, or ended as a censored observation $fail = 0$. Finally, we have data on various control variables which may include time-fixed, time-varying, and parity-varying variables. Notice that going from the long to the wide form of the fertility history we have 'thrown way' a substantive part of the covariate history. From that point of view, and despite the fact that both fertility histories are longitudinal, modelling using data that has a wide-form seems less promising. These type of data are typically analysed with continuous hazard models.

2.3 Dynamic inter-dependencies

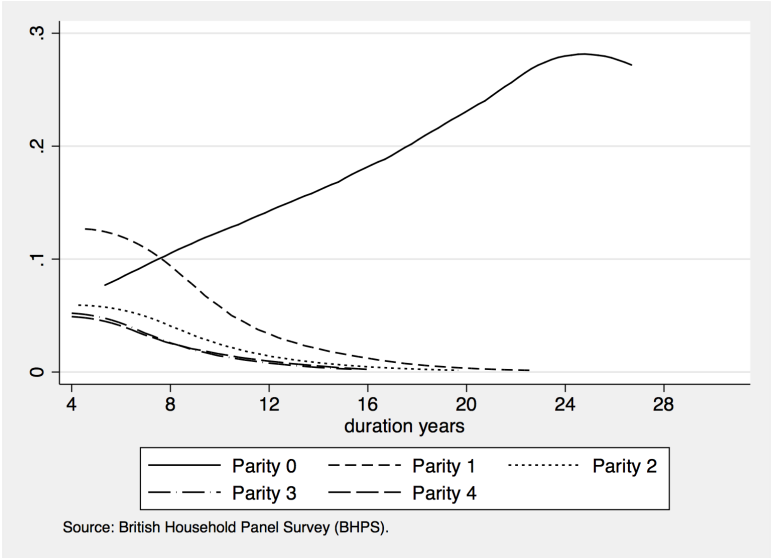
From the discussion in section 2.2 is clear that successful modeling of fertility needs start by recognising that each pregnancy is a decision in its own right.

Indeed, when deciding whether to become pregnant, women take into account all the information they have at the time (see, for instance, Barmby and Cigno 1990, Wolpin 1984). This includes the current number of her offspring (incentives and child benefit systems) (Barmby and Cigno 1990), their sex composition (due to gender preference) (Williamson 1976, Angrist and Evans 1998), the outcome of her last pregnancy (reduced fecundity after a c-section or miscarriage) (Kok et al. 2003, Hassan and Killick 2005, O'Neill et al. 2014, Sapra et al. 2014), her work status and salary (Bettio and Villa 1998, Mira and Ahn 2001), and the available child care support (see, for instance, Ermisch 1989, Boca 2002, Rindfuss et al. 2010). Many of such conditions change with time and can influence whether the same woman goes from childless to have her first child, but not whether she goes from having one child to having two or three children. That is, in general, factors that affect the transition from parity 0 to parity 1 may not play any role in the transition from parity 1 to parity 2. Dynamics are an essential feature of how a fertility history is generated.

To illustrate this point with the British data we present in figure 3 a smoothed kernel estimate of the probability of observing a birth at any point of time given that the birth has not yet

arrived — also known as the hazard function.

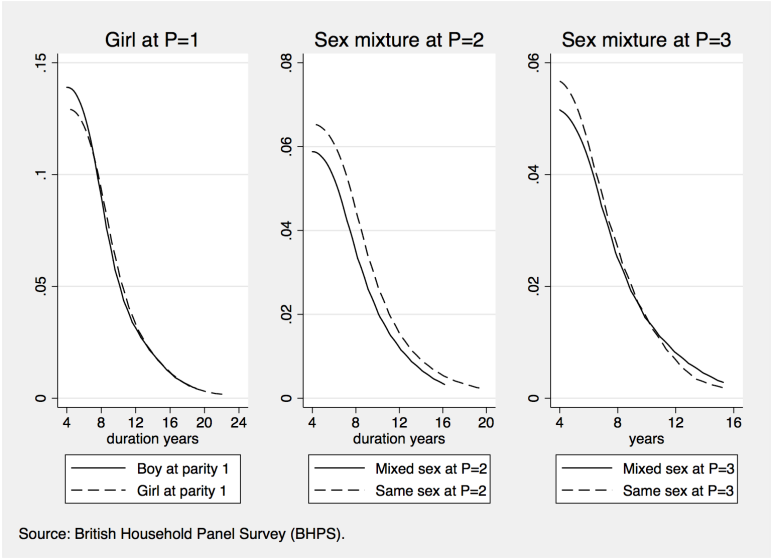
Figure 3. Hazard kernel estimate for British fertility data



From figure 3 we can conclude that the form of the hazard function is very different for each parity. While at parity 0 (no children), the hazard function exhibits a U-inverted form, the hazard at parities $1,2,3,4$ are monotonically decreasing. Moreover, though the hazard at parities $1,2,3,4$ is a non-increasing monotonic function, the hazard functions at parity 2 is clearly not a parallel shift of the hazard function at parity 1, nor is the hazard function of parity 3 a parallel shift of the hazard function at parity 2. These descriptive stylised facts have important implications for econometric modeling as the most popular methods cannot deal with non-monotonic hazard functions; let alone dealing at the same time with a whole fertility history which implies specifying a model that is flexible enough to accommodate for the special features of the data.

More to the challenge, figure 4 shows that even after fixing the parity the hazard function depends upon the outcome of the previous pregnancy. In particular, simple inspection shows that having two boys or two girls at $P = 2$ shifts outwards the hazard function—increasing the risk of a new pregnancy as well as decreasing the average duration time to it—but there is no descriptive evidence that sex composition plays a role at any other parity. In a similar

Figure 4. Hazard kernel estimate at selected parities



vein, figure 5 shows descriptive evidence that having twins at any parity shifts inwards the hazard function, decreasing dramatically the likelihood of observing any further births. These descriptive data examples illustrate that dynamics, other than duration dependance, play an important role in specifying an econometric model of fertility history.

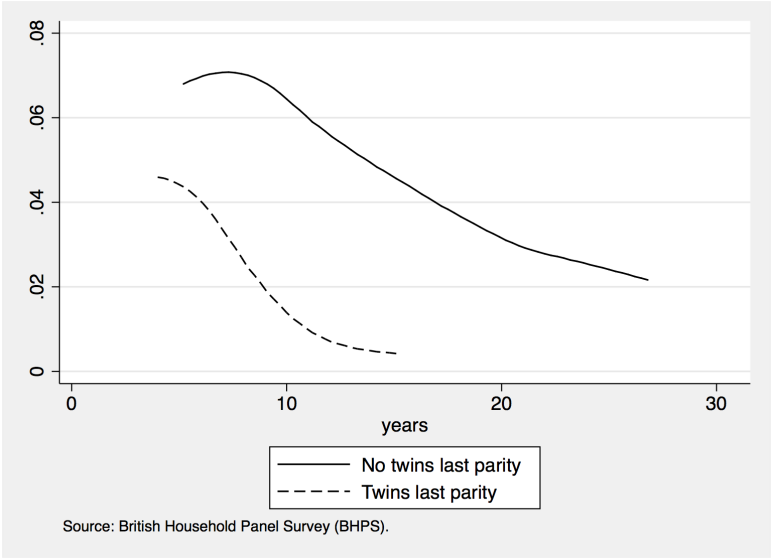
3 Reduced form vs. structural modeling

As in many fields of the social sciences, there are many possible modeling approaches available to the applied researcher when dealing with fertility data. Which approach is taken depends, among other things, on the research objectives and questions of interest, the intended use of the empirical findings, and the quality and coverage of the available data.

Broadly stated, the research objective is to understand the fertility process in specific populations. More specifically researchers try and shed light on the long running debate on the demographic versus economic perspectives on fertility behavior.

From a purist economic perspective, completed fertility is determined by economic optimizing behavior of a woman. A full specification of the optimization problem requires specifi-

Figure 5. Hazard kernel estimate for twins in last parity



cation of the utility function and all relevant constraints facing the woman which may depend upon government interventions such as those in welfare states. Even in a model which ignores dynamic inter-dependence between a sequence of birth events, the challenge of finding a theoretical optimum is compounded by issues such as whether the utility function reflects purely personal preferences or includes dependence on social norms of individual-specific reference groups. Economists, relative to demographers and sociologists, in the past have emphasized much less the role of social interactions and stress the role of individual preferences and economic constraints and opportunities. If dynamic interdependence is introduced in the model, and the issue of optimal spacing of preferred number of children is also considered, the goal of empirically identifying the parameters of the utility function, the parameters characterizing social interaction, and parameters that determine responses to interventions becomes even more complex. If such a goal is achievable, then the estimated parameters have a clear causal interpretation, given the fully specified model structure; for examples, see Hotz and Miller (1988), Manski and Mayshar (2003). Following standard practice, we will refer to such an approach as structural modeling as the goal is to identify structural or causal parameters.

Arroyo and Zhang (1997) have provided a useful taxonomy for dynamic models of fertility.

They distinguish between structural models explicitly based on relationships derived from solution of dynamic programming problem (e.g. Hotz and Miller 1993) and reduced form models that "may have a basis in some dynamic programming problems but do not rely heavily on that structure for specification of estimating equations", e.g. the hazard function approach used in Heckman and Walker (1990b).

Structural modeling requires strong a priori behavioral and functional form assumptions and, given typical data constraints, is difficult to implement fully. Many, if not most, empirical studies are of reduced form type, especially in demographic literature. They focus on regression models of variables such as completed fertility in which the regressors are socioeconomic individual or family specific factors. Where relevant they may also include incentive variables such as child support. Cross-section survey data are widely used and information from retrospective surveys may be incorporated. Within such models the mechanism that connects birth outcomes or (unobserved) individual gender and family size preferences with regressors is at best implicit, which makes unambiguous interpretation of results difficult. Despite this limitation reduced form models are useful descriptive models and can capture associations with predictive potential, fertility trends, and intergroup heterogeneity; we provided references in the next section.

4 Selected count regression models

The raw material of fertility data consists of the count of number of live births to a woman in a specific year-of-birth cohort and during a specific period which is called *exposure*. The completed fertility data consists of non-negative integers in a range that varies both over time and over socioeconomic groups—a feature which we explore further later in this section. Fertility *rate* is derived as a ratio of birth counts to corresponding population exposure. Age- and order-specific fertility rates measure childbearing intensity among women of specific age and parity (the number of previous live births). Such a fertility rate, a continuous variable, can be person-period specific, or age specific, or cohort specific.

4.1 Semiparametric regression

A simple starting point for modeling the total number of births (y), given cross-section or panel data, is a linear-in-parameters regression estimated by ordinary least squares (OLS). In that case one usually ignores the nonnegative integer-valued nature of the outcome variable, treats zeros and positive outcomes as generated by a common (but unspecified) process, and regressors ($\mathbf{x}$) are treated as exogenous. No specific probability distribution of outcomes is assumed.

Given a count of live births (denoted y) exposure (denoted t) and other information on socioeconomic and demographic information of surveyed households (denoted by $\mathbf{x}$), age and cohort specific fertility rates can be constructed as descriptive data, as is standard practice in demography. However, such an exercise will not adequately control for differences in observable characteristics of the individual and the household. Regression analysis based on $(y, t, \mathbf{x})$ provides a more informative framework for estimating fertility rates as well as studying the drivers of fertility. Poisson regression, based on the Poisson distribution and Poisson process, is a well-established benchmark regression model for birth events despite its strong underlying assumption of independence of events.

If one wishes to avoid distributional assumptions and ignore the integer-valued feature of births, a simple solution is to estimate a regression with an exponential mean by nonlinear least squares:

$$\mathrm{y}_i = \exp(\mathbf{x}_i'\beta) + u_i. \tag{15}$$

Standard Poisson-type regression models use the same functional form for the conditional mean, $E[y|\mathbf{x}] = \mu$, but estimation is by the more efficient maximum likelihood method. The marginal impact of a unit change in x_j is $\beta_j \mu$.

4.2 Poisson regression

When completed fertility is the non-negative outcome variable of interest, either Poisson regression or some extension of it, such as the negative binomial regression, is a popular starting point. Poisson regression and several of its extensions derive form the Poisson process and Poisson distribution which embody strong assumptions that are easily violated in practice.

Consider the cross-section regression of n independent observations, $(y_i, \mathbf{x}_i)$, where the dependent variable y_i denotes the number of occurrences of birth events, and $\mathbf{x}_i$ is the vector of linearly independent regressors. Poisson regression model conditions the distribution of y_i on covariates, $\mathbf{x}_i' = [x_{1i}, ..., x_{ki}]$ and parameters β through a continuous function $\mu(\mathbf{x}_i, \beta)$, such that $E[y_i|\mathbf{x}_i] = \mu(\mathbf{x}_i, \beta)$. Thus y_i given $\mathbf{x}_i$ is Poisson-distributed with probability mass function

$$f(y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots \tag{16}$$

This one-parameter distribution embodies the restriction of equality of mean μ and variance, also known as equidispersion, i.e. $E[y_i|\mathbf{x}_i] = \mu(\mathbf{x}_i, \beta) = Var[y_i|\mathbf{x}_i]$. That is, the inherent variability of birth events increases with the mean—the average number of births per period. In the standard version of the Poisson regression the mean parameter is parameterized as

$$\mu_i = \exp(\mathbf{x}_i'\beta), \tag{17}$$

to ensure $\mu > 0$. Estimation of the parameters β by maximum likelihood is straight-forward and feasible using widely available software. See Cameron and Trivedi (2013).

The direct application of this model to birth data assumes that all subjects have the same exposure period and are at equal 'risk' of experiencing the event. Often the subjects in the survey data have different exposure risk. This feature is handled by introducing an "offset" feature which decomposes the mean μ_i as a product of fertility rate λ_i and the length of exposure t_i, i.e. $\mu_i = t_i\lambda_i$ which implies log-linear mean function with an off-set term, i.e. $ln(\mu_i) = ln(t_i) + \mathbf{x}_i'\beta$.

Most software packages allow inclusion of the offset term.

The use of regression adds flexibility by combining features of descriptive data analysis with regression analysis, and also facilitating statistical inference on estimated parameters. Regression framework allows conditioning on relevant factors that affect the birth event, something that is not possible when a purely descriptive methodology is used in which the fertility rates are computed using cohort- and period-specific ratios. However, because such descriptive features are of interest per se, they can be added in a regression framework. For example, one can to do this by adding dummy (indicator) variables for each category of observation of interest, and then defining the reference group in analysis. That is the Poisson regression model now has the conditional mean function

$$\ln(\mu_i) = ln(t_i) + \sum \mathbf{x}_i \beta_i + \sum \mathbf{z}_j \gamma_j, \tag{18}$$

where $\mathbf{z}_j$ denotes the indicator variable, such as age-group or cohort. The estimated coefficient γ_j measures the difference in birth rates of a selected category j relative to the reference category, while controlling for other differences due to observed regressors $\mathbf{x}$; see Schoumaker and Hayford (2004).

If the conditional mean function is correctly specified, but the equidispersion assumption is suspect, then the literature often substitutes pseudo maximum likelihood (PML) in place of maximum likelihood. The main practical consequence is that a Huber-White type robust variance estimator is substituted in place of the standard maximum likelihood estimator. Perhaps the most frequent justification for this practice is to appeal to the overdispersion of data which one can test for formally; see Cameron and Trivedi (2013), chapter 3.4. Note, however, that, as previously mentioned, fertility data from high-income-low-fertility economies are likely to display *underdispersion*.

The Poisson regression model works well if the key assumptions underlying it, such as equidispersion, are satisfied. In practice count models often fail to fit fertility data well for several reasons. First, as was seen in the examples in previous sections, frequency distribution

has limited support with most of the mass concentrated on just three or four values. Second, count regression may identify sources of the average difference between high and low outcomes but may not be very informative about the underlying drivers of events, which makes interpretation of results difficult. Third, total birth counts involves aggregation of event information over time; and completed fertility typically span many years—a period over which many individual-specific observed and unobserved time-varying factors could potentially impact the decision to have a child. Fourth, when the event frequency distribution is bimodal, the underlying Poisson distribution assumptions are invalid. Finally, Poisson process assumption that events are independent may be unrealistic in the context of birth outcomes.

4.3 Negative binomial extension of the Poisson

A major limitation of the Poisson regression model is its property of equidispersion which restricts the mean and variance of outcome variable to be equal. Many data sets exhibit overdispersion (variance greater than the mean) or underdispersion (variance less than the mean). Overdispersion stretches the distribution of outcomes, and underdispersion squeezes it. If, for example, an actual count frequency distribution is trivially different from a binomial, data would show underdispersion. If, on the other hand, if the mean outcome is subject to unobserved heterogeneity with some well-defined properties, then the data would be overdispersed. Populations with low fertility often display underdispersion, and those with high fertility the opposite; later we will show data examples with these features. Other non-Poisson features of fertility data and methods for handling them will be discussed later in the paper.

A well-established alternative to the Poisson regression, and one supported in most software packages, is the negative binomial (NB) regression model which can be used to model overdispersed data. Overdispersed counts can be generated by replacing the constant mean parameter μ_i by (say) $\mu_i^* \varepsilon_i$ where ε denotes an individual-specific heterogeneity term with mean 1 and variance α. While there are at least two widely used variants of the NB regression, the most commonly used is the NB2 model, with mean μ and NB2 variance function $\mu + \alpha\mu^2$,

with density

$$f(y|\mu,\alpha) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y}, \quad \alpha \geq 0,\ y = 0,1,2,... \quad (19)$$

The function $\Gamma(\cdot)$ is the gamma function. Excess variance relative to Poisson is $\alpha\mu^2$ which reflects unobserved heterogeneity. NB2 reduces to the Poisson if $\alpha = 0$.

The specification of the conditional mean under NB2 is exactly as in the Poisson case. Estimating the Poisson regression by maximum likelihood when NB2 is the relevant model will still yield consistent estimates of the regression coefficients but their standard errors will be incorrect see Cameron and Trivedi (2013). Many software packages support maximum likelihood estimation of NB2. A test of the null hypothesis $H_0 : \alpha = 0$ can be implemented in using one of several different approaches (see Cameron and Trivedi 2013, chapter 5.5) and this often serves as a selection criterion between NB2 and Poisson. However, this test leaves open the possibility that even the selected model has other misspecification(s). A more precise indicator of model deficiency is a Pearson-type goodness of fit test, based on comparison of fitted and observed frequencies of events, which can point to potential misspecification that affects specific frequencies.

4.4 Modeling underdispersed counts

Although underdispersion gets less attention in the fertility literature, a number of authors have attempted to extend the parameterization of the Poisson regression to accommodate both over- and underdispersion via an additional parameter. For example, Winkelmann (1995) develops a duration model of events that directly leads to an underdispersed count model; Winkelmann and Zimmermann (1991; 1994) develop a parameterization based on the Katz family of distributions in which the variance function takes the form $\mu + (\sigma^2 - 1)\mu^{k+1}$ with an additional parameter k. Underdispersion corresponds to $\sigma^2 < 1$, and $\mu^k \leq 1/(1-\sigma^2)$. For additional details see Cameron and Trivedi (2013) , chapter 4.11.

4.5 Other non-Poisson features of fertility counts

Many studies that have used count data regressions to analyze the role of socioeconomic factors and social norms on fertility using cross-section data have pointed out non-Poisson features of fertility data that call for a modeling strategy that goes beyond the standard Poisson and negative binomial regressions. Poisson distribution is unimodal, displays equidispersion and is based on the assumption of independent events. These features are often at least partly absent in observed data. This is unsurprising because birth outcomes do not simply follow invariant biological stochastic processes but are responsive to social norms (e.g. popular notions of optimum number of children), heterogeneity of individual preferences (e.g. gender preferences), and economic constraints (e.g. absence of social security) and incentives (e.g. child allowances), as discussed in section 2. These factors affect the trends in fertility, as is evident in the extensive literature of fertility transitions, as populations experience improvement in living standards and governments provide substitutes for family support for ageing populations.

In a previous section we have alluded to differences in the distributional characteristics of births in high-income-low-fertility populations, such as Western industrialized countries, and low-income-high-fertility countries, such as developing countries in Africa and Asia. Birth distributions in the former characteristically show a low mean, underdispersion and departure from unimodality. The last feature appears in the data through a significant bump in frequencies $y = 0$ and $y = 2$ and a dip at $y = 1$, features that could indicate preference for no children and weaker preference for a single child; see ( ). Modifications of standard count data models are needed to capture these features of data and to provide a relatively better statistical fit to data, especially in the low-fertility populations.

We next discuss modified Poisson and NB regression models which attempt to deal with the problem of "zero inflation" which refers to the presence in the data of more zero outcomes than is consistent with the underlying parent distribution. One might also interpret this as a statistical solution of the problem of unobserved or unstated individual preferences for family

size and composition. That is, the refinements we next discuss would be superfluous if we had data on such family preferences.

4.5.1 Hurdle or two-part models

Hurdle or two-part models (Mullahy, 1986) specify a process for events that differs between zero and positive outcomes. Specifically, zero valued outcome constitutes a hurdle, which, once passed, transits to a different probability law. With probability $f_1(0)$ the threshold is not crossed, in which case we observe a count of 0. If the threshold is crossed we observe positive counts, with probabilities coming from the truncated density $f_2(y)/(1-f_2(0))$ (which is multiplied by $(1-f_2(0))$ to ensure probabilities sum to one. The first part of the model is a binary (usually logit or probit) model, while the second part models the positive outcomes. Here and in the next section we use $f(.)$ generically to denote distribution, which in the current context is most often Poisson or NB. This model is formally stated as follows:

$$Pr[y=j] = \begin{cases} f_1(0) & \text{if } j=0 \\ \dfrac{1-f_1(0)}{1-f_2(0)} f_2(j) & \text{if } j>0. \end{cases} \tag{20}$$

The standard model is a special case if $f_1(0) = f_2(0)$. If $f_1(0) > f_2(0)$, the model generates excess zeros, though in principle it can also model too few zeros if $f_1(0) < f_2(0)$. The density $f_2(\cdot)$ is a count density such as Poisson or NB2, while $f_1(\cdot)$ could also be a count data density or dichotomous density. The two-parts are independent so can be separately estimated by maximum likelihood.

4.5.2 Zero-inflated models

Suppose the base count density is now$f_2(y)$, using the same notation as for the hurdle model, but this under-predicts zeros. We can add a separate component that inflates the probability of

a zero by, say, π. Then the *zero-inflated model* specifies

$$\Pr[y=j] = \begin{cases} \pi + (1-\pi) f_2(0), & \text{if } j=0 \\ (1-\pi) f_2(j) & \text{if } j>0. \end{cases} \tag{21}$$

where the proportion of zeros, π, is added to the baseline distribution, and the probabilities from the base model $f_2(y)$ are decreased by the proportion$(1-\pi)$ to ensure that probabilities sum to one. The probability π may be set as a constant or may depend on regressors via a binary outcome model such as logit or probit. Such a specification will improve the statistical fit of the model while leaving open the question as to the interpretation of the inflation component. As mentioned by Melkersson and Rooth (2000) birth data for some European countries also shows 'excess 2s', a feature that is consistent with a high family preference for two children, indicating a contemporary social norm. This inflation can be modelled analogously to excess zeros. However, using a statistical device to capture the inflation at a particular cell frequency does not constitute an explanation.

4.5.3 Other modeling options

One further limitation of the count data models applied to cross-section birth data is that whereas the range of recorded outcomes can span as many as 10 or 12 values, most of the probability mass is likely concentrated on relatively few values. This implies as many as half or more cells will be thinly populated making it difficult to achieve a good fit. One way to deal with this problem is to aggregate (combine) the cells into fewer ordered categories and then use a multinomial model. For example, one might choose to aggregate cells with frequencies between 6 and higher into a single cell, or perhaps 2 less aggregated cells. As the cells are then ordered into about seven categories one could apply the ordinal probit or ordinal logit model which accommodates the natural ordering of discrete outcomes. The ordered outcomes model is different from the Poisson-process based models; in that it treats the data as generated by a continuous unobserved latent variable which on crossing a threshold increases the outcome

value by one unit. This framework, though not often used, is an attractive option for modeling data which are distributed over a small number of cells, as in the case of under-dispersed samples.

5 Event history models

Count data regressions are static models and hence necessarily uninformative about dynamic aspects of the fertility process that are captured by a complete events history of births. One alternative is to model the elapsed duration between successive transitions (births) conditional on the previous history of the event and on observed covariates. A second alternative is a discrete hazard model which is typically specified with the objective of modeling the transition between the status quo state $y = 0$ and one-additional event state $y = 1$, conditional on the previous history of transitions and on information about (usually) exogenous covariates. In a given time interval y has a binomial distribution. The conditional hazard of the transition between states can be derived from this property. We label this as Modified Dynamic Event History (MDEH) model.

5.1 Restructuring event history data for MDEH

The building blocks of the MDEH model are implicit in the description given in section 2.2 of how a "long form" panel data set may be constructed using event history data. A more formal description follows. For each sampled individual there is information over a historical period of calendar length T_i, $i = 1, ..., N$, which begins at $T_{b,i}$ and ends at $T_{e,i}$, $T_{e,i} > T_{b,i}$. T_i could be number of weeks, months or similar. In a standard event history analysis the irregular interval between events is left as it is. For established econometric methods, the irregular time interval between events poses a challenge. For example, the distribution of errors is difficult to specify. We propose a simple modification which would allow us to use "standard" panel data methods to estimate our models.

Our modified event history analysis takes irregularly spaced data as a starting point and then restructures the data as follows. The period $(T_{e,i} - T_{b,i})$ is divided into discrete periods of fixed-length common to all observations. The full data set spans a total of T observational periods, which are denoted by $\tau_1, \tau_2, ..., \tau_T$, respectively. In our data this fixed-length period will be just a calendar year, but this choice in general would be context dependent. Our choice is motivated by the consideration that for a binary outcome model we require that within each observational period at most one event/transition would be observed; this condition might be violated in some cases.

In general the sampled individuals will have event histories spanning overlapping observation periods with different beginnings and ends because individuals will enter or leave the survey at different times. Hence, a given individual will be observed over a sub-set of these T periods, which then constitutes the length of the observed event history for that individual. Different subsets of individuals may share particular beginnings or endings.

In this set-up a transition may occur for an individual from the current state to another state. In case there are S different states, there are $S-1$ different transitions that can occur. The simplest case is $S = 2$, in which case the occurrence of the event signals the change of state. Even when there are more than 2 states, by appropriate redefinition the set-up can be reduced to a simple dichotomy.

5.2 Dichotomous model

To formulate the standard dichotomous outcome model we distinguish between the $0/1$ observed binary response variable *resp* and an underlying continuous latent variable $resp^*$ that satisfies the single-index model

$$resp^* = \mathbf{x}'\beta + \mathbf{z}'\gamma + u, \tag{22}$$

where $\mathbf{x}$ denotes the vector of time-varying covariates, and $\mathbf{z}$ the vector of individual-specific time-invariant variables, and u is an i.i.d. error. As $resp^*$ is not observed, we specify an observability condition for its observable counterpart $resp$,

$$resp = 1[resp^* > 0], \tag{23}$$

where the zero threshold is a normalization. Define $resp_{i\tau}$ as follows:

$$resp_{i\tau} = \begin{cases} 0 \text{ if no transition is observed in period } \tau, \\ 1 \text{ if a transition from the current state is observed in period } \tau. \end{cases}$$

The connection with the count data model results from the fact that $\sum_\tau resp_{i\tau}$ is the number of events observed for i^{th} individual during the event history, but the length of the event history, also called exposure time in count data models, will not in general be the same for all individuals.

Then the event history data will consist of $(resp_{i\tau}, \mathbf{x}_{i\tau}, \mathbf{z}_i)$, $\tau = \tau_1, ..., \tau_P$, and τ refers to a time-period of selected length and the subscript P refers to parity, the number of events observed up to a particular point in time. However, once a fixed length of measurement is chosen, it is convenient to index τ as simply $\tau = 1, ..., T$. The elements of vector $\mathbf{x}_{i\tau}$ are time-varying regressors, including past outcomes $resp_{i\tau-j}$. Moreover, other observable features of past outcomes may also be included. This set up is more flexible for modelling dynamics than the autocorrelated conditional duration (ACD) model of Bauwens and Giot (2000) because it allows us to bring into our specification dynamic factors that are not satisfactorily reflected as lagged values of outcomes.

The restructured data set-up is now analogous to panel/longitudinal data in the "long form" with individual and period-specific subscripts. Because in practice we will have a full event history of different lengths on different individuals, the panel typically will be unbalanced. Notice that we end up with a layout that defines a multi-state recurrent process, as described by Steele and Goldstein (2004), that can be analysed as an instance of the multilevel multistate

competing risks model, with the exception that we put emphasis on dynamics and consistent fixed-effects estimation over random effects estimation.

As in the case of panel data, individual specific unobserved effects, denoted c_i, can be added to the model to capture unobserved heterogeneity. Also, as in the standard panel case, different assumptions can be made about c_i; it can be treated as an i.i.d. random variable (random effects assumption), or as correlated with $\mathbf{x}_{i\tau}$ (fixed effects assumption), or as a function of observed variables and an i.i.d. random error (conditional correlation assumption). Thus the event history data can be recast as unbalanced panel data model for a binary outcome.

5.3 Discrete hazards

The above dichotomous outcome model is a discrete hazard model for recurrent outcomes, analogous to a linear pooled panel data model in which the panel data is treated as a pooled cross-section. A general formulation of a pooled discrete-time transition model is

$$\begin{aligned} h_{i\tau} &\equiv \mathbb{P}\left[resp_{i\tau} = 1 \middle| \mathbf{x}_{i\tau}, \mathbf{z}_i, \sum_{j<\tau} resp_{ij} = 0 \right] \\ &= F\left(\lambda_\tau + \mathbf{x}'_{i\tau}\beta + \mathbf{z}_i\gamma\right), \quad \tau = 1, ..., T. \end{aligned} \tag{24}$$

where $h_{i\tau}$ represents the discrete hazard for the i-th individual, i.e. the probability of an event occurring (i.e. $resp = 1$) at time τ given that no event has occurred up to $\tau - 1$, and F denotes the c.d.f. This specification restricts the coefficients of regressors to be constant over time, while the intercept λ_τ, $\tau = 1, ..., T$, can vary over duration time. The only dynamics that allows this models is the one accounted for by λ_τ which models duration dependence. Dynamics related to state dependence and/or dependence of the hazard on lagged time-varying variables are not accounted for. This is not attractive for the study of fertility behaviour because a fertility history has an additional state dimension: the parity. Taking all three dimensions into account

generalizes the above transition model as follows:

$$
\begin{aligned}
h_{ip\tau} &= \mathbb{P}\left[resp_{ip\tau} = 1 \,\middle|\, \mathbf{x}_{ip\tau}, \mathbf{s}_{ip}, \mathbf{z}_i, \sum_{j<\tau} resp_{ipj} = 0 \right] \\
&= F\left(\lambda_{p\tau} + \mathbf{x}'_{ip\tau}\beta + \mathbf{s}_{ip}\delta + \mathbf{z}_i\gamma\right), \quad p = 1,...,P_i, \tau = 1,...,T,
\end{aligned} \tag{25}
$$

where explanatory variables may vary across individuals only ($\mathbf{z}$), across individuals and states ($\mathbf{s}$), or across individuals, states, and time ($\mathbf{x}$). Duration dependence within each state p is accounted for by a series of coefficients $\lambda_{p\tau} = \lambda_{p,1}, \ldots, \lambda_{p,T}$ that form a step-function that flexibly represents the baseline hazard.

For a parametric functional form of F two choices are popular: the standard normal c.d.f., or the logistic c.d.f. Then the parameters (λ, β, γ) can be estimated by a stacked logit or stacked probit model in which a separate intercept is permitted for each state and for each interval. The resulting likelihood function for a sequence of $\tau - 1$ zeros and a one at time τ is

$$
\begin{aligned}
L(\alpha, \lambda, \beta, \gamma) = \prod_{i=1}^{N} \prod_{p=1}^{P_i} &\left[\prod_{g=1}^{\tau-1} \left(1 - F\left(\lambda_{pg} + \mathbf{x}'_{ipg}\beta + w'_{ip}\delta + \mathbf{z}'_i\gamma\right)\right) \right] \\
&\times F\left(\lambda_{p\tau} + \mathbf{x}'_{ip\tau}\beta + w'_{ip}\delta + \mathbf{z}'_i\gamma\right).
\end{aligned} \tag{26}
$$

Extending this to allow for unobserved additive individual-specific heterogeneity term c_i leads to the conditional likelihood

$$
\begin{aligned}
L(\alpha, \lambda, \beta, \gamma | c_i) = \prod_{i=1}^{N} \prod_{p=1}^{P_i} &\left[\prod_{g=1}^{\tau-1} \left(1 - F\left(\lambda_{pg} + \mathbf{x}'_{ipg}\beta + \mathbf{w}'_{ip}\delta + \mathbf{z}'_i\gamma + c_i\right)\right) \right] \\
&\times F\left(\lambda_{p\tau} + \mathbf{x}'_{ip\tau}\beta + \mathbf{w}'_{ip}\delta + \mathbf{z}'_i\gamma + c_i\right).
\end{aligned} \tag{27}
$$

Estimation requires the choice of function F. This step causes no major difficulties for the pooled model underlying (27).

In the case of the fixed effects model, the nuisance parameters c_i are not straight-forward to eliminate for an arbitrary choice of F. For example, it cannot be eliminated by transformation if we choose F to be the Normal c.d.f. For the case of the static logit model, however, it

is possible to eliminate the fixed effects by conditioning on the total count $\sum_\tau resp_{i\tau}$ and perform inference on (λ, β, γ) on the basis of a conditional maximum likelihood strategy using the sample of individuals who experience at least one transition. A major disadvantage of this FE-Logit case is that it can only identify the parameters of time-varying regressors, whereas other parameters may also be of interest. Moreover, leaving out individuals who do not experience any transition is unattractive as the study's sample size can suffer importantly because in most samples many women stay childless, specially in low-fertility-high-income settings where nearly $1/4$ of women never have children—for instance in our BHPS example 26% of women remain childless by age 45. Beyond loss of efficiency, dropping childless observations is also unattractive because women who never have children may be systematically different from women who eventually enter motherhood. As a consequence, there may be important sample selection issues at play that need explicit modelling Heckman (1979), Heckman and Walker (1990a).

For the dynamic logit case, i.e. a model that includes $resp_{i,\tau-1}$ as a covariate, conditioning on a sufficient statistic for (c_i) is more problematic, see Honoré and Kyriazidou (2000). However, Bartolucci and Nigro (2010) have proposed a version of the quadratic exponential model of Cox (1972) and Cox and Wermuth (1994) that closely resembles the dynamic logit model and for which conditioning on a sufficient statistic is feasible as in a static panel logit model. We use the pseudo conditional maximum likelihood (PCMLE) estimation method of Bartolucci and Nigro (2010; 2012). In this approach the fixed effects are integrated out by conditioning on a sufficient statistic, which in the case of the binary logit model is the total number of events.This dynamic fixed effect model is used in our application.

We also estimate our MDEH specification under random effects assumptions. In this case F chosen to be either the probit or the logit as both are computationally manageable; however, both require a parametric assumption about the distribution of c_i, typically followed by numerical integration.

6 Modeling dynamics of transitions

There are two important motivations for a dynamic specification. First, in the context of fertility analysis, there is considerable theoretical and empirical work emphasizing the dependence of a new outcome on previous birth outcomes; see Bhalotra and Van Soest (2008) and references there cited. For example, in Wolpin (1984) dynamic stochastic model of life-cycle fertility, the model generates implications for the number, timing and spacing of children. And there is evidence in the literature that families have preferences over gender composition which imply that past birth outcomes will affect the desire for additional children (Arnold and Liu 1986).

A second motivation flows from wanting to distinguish between the effect of individual time-invariant unobserved heterogeneity and state dependence. Individual propensity to remain in the current state may be mainly a consequence of unobserved persistent characteristics of that individual. In such a case controlling for individual-specific effects would account for the state dependence. This is often referred to as spurious state dependence. If, however, simply continuing to remain in one state decreases the probability of transition, no matter what the individual-specific characteristics are, then such dynamic dependence is referred to as true state dependence, or duration dependence. Panel data potentially affords the possibility of distinguishing between the two alternatives, or even simultaneously controlling for the separate contributions of each.

Following Heckman and Borjas (1980), in a discrete hazard model for a binary outcome in which the lagged dependent outcome is a regressor, we may refer to dependence between the current and past outcomes as occurrence dependence; the term duration dependence refers to the case of continuous outcomes, and state dependence covers both types of dependence. Bhalotra and Van Soest (2008) consider some approaches for modeling fertility dynamics—from a perspective relevant to fertility studies of less developed economies.

6.1 Autoregressive dependence

A popular dynamic specification of a binary outcome model simply adds the lagged dichotomous variable $resp_{i\tau-1}$ as an additional control variable; specifically, we modify (25) thus:

$$\begin{aligned} h_{ip\tau} =& \mathbb{P}\left[resp_{ip\tau} = 1 \middle| \mathbf{x}_{ip\tau}, \mathbf{s}_{ip}, \mathbf{z}_i, \sum_{j<\tau} resp_{ipj} = 0, c_i \right] \\ =& F\left(\lambda_{p\tau} + \mathbf{x}'_{ip\tau}\beta + \mathbf{s}_{ip}\delta + \mathbf{z}_i\gamma + \rho resp_{i\tau-1} + c_i\right), \quad \tau = 1, ..., P, \end{aligned} \tag{28}$$

which now includes an individual-specific heterogeneity term c_i and the autoregressive term $resp_{i,\tau-1}$. The parameter ρ reflects state dependence, with $\rho = 0$ indicating zero occurrence dependence (or state dependence). The lagged regressor captures the effect of all unobserved or unmeasured factors that generate state dependence in the outcome. We call this specification Markovian in contrast to other (non-Markovian) specifications we discuss below. As previously indicated, given the presence in the equation of observed variables that capture individual heterogeneity, the lagged variable can potentially capture pure occurrence dependence.

Given the above specification, additional assumptions determine whether the random effects (RE), fixed effects (FE), or the flexible FE model in which the individual-specific parameter is parameterized as a function of exogenous variables plus a random component (often referred to as conditional correlation framework (CCR)) would identify the model. Whether one is in the large-N-small-τ framework common in microeconometric panels, or in the large-N-large-τ framework, is relevant to the choice of framework.

In the RE framework, the model will be augmented with an auxiliary assumption regarding the probability distribution of c_i; subsequent to which these nuisance parameters are (numerically) integrated out of the model—a task which is analytically more tractable than in the FE specification.

For the FE case, as previously noted, there are several proposed approaches (see Hsiao 2014) for estimating this FE model, including Honoré and Kyriazidou (2000), Bartolucci and

Nigro (2010; 2012), Al-Sadoon et al. (2017). In a large-N-large-τ setting, controlling for the fixed effects using a comprehensive set of dummy variables is theoretically an option but not much used in practice. In the more common small-τ setting of this paper conditional maximum likelihood has greater appeal. The conditional approach of Bartolucci and Nigro applied to the quadratic exponential specification, which closely mimics the dynamic logit, has been applied to estimate the empirical model in this paper. The discrete hazard is,

$$h_{ip\tau} = \mathbb{P}\left[resp_{ip\tau} = 1 \middle| \mathbf{x}_{ip\tau}, \mathbf{s}_{ip}, \mathbf{z}_i, \sum_{j<\tau} resp_{ipj} = 0, \sum_{p=1}^{P_i} \sum_{j=1}^{T_{ip}} resp_{ipj} = \kappa, c_i \right] \tag{29}$$

which is estimated by conditional maximum likelihood.

The use of the autoregressive form also poses problems of interpretation because it is based on a somewhat narrow view of the nature of dependence. In a *linear* autoregressive model, the lagged dependent variable captures the effect of all past changes in the exogenous drivers as the autoregressive model has a distributed lag representation. This is not so in a nonlinear model. More generally, capturing state dependence via a lagged dependent variable has a black-box character as it is not clear which mechanism is responsible for inertia in the transition from the current state.

6.2 Non-autoregressive dynamic dependence

Autoregressive specifications lead to a loss of observations and in some cases dynamic dependence can be captured without using the autoregressive form, e.g. through past values of exogenous regressors. Hence, using lagged exogenous variables as predictors of current outcomes instead of lagged outcomes is an appealing way of modeling dynamic dependence, especially because it avoids the complications due to lagged outcomes in fixed effect models.

To model dynamics without autoregression, following the discussion in section 2, we propose to use variables that characterize some observable qualitative features of past events that are potentially related to a woman's preferences and hence have predictive relevance. The di-

chotomous variable $resp_{i\tau}$ only indicates whether an event occurred. But many events vary in their nature and intensity. In such cases additional auxiliary (synthetic/constructed) regressors based partly on *qualitative features* of several past outcomes, rather than just the previous outcome, can be added because they are potentially useful predictors of future events. The relevance of such predictors may be easier to rationalize than that of the lagged outcome. Moreover, unlike the first-order autoregression which only takes into consideration the immediately preceding outcome, such constructed variables may reflect the role of a full or partial event history.

The proposed dynamic modeling framework has similarities with Moffit (1984) who has used "panel probit" and "pooled probit" to analyse fertility data from the National Longitudinal Survey of Young Women. The author refers to his key equation as "number-of-children" equation; however, "panel probit" equation has a binary dependent variable birth/no birth, and the dynamics are captured by including the total number of children as a regressor in the model, which is similar to our use of *parity* variable. Further, the framework we presented above incorporates unobserved heterogeneity by allowing for both random and fixed effects.

7 Application: A modified dynamic event history model for British fertility histories

In this section we discuss results from a modified dynamic event history (MDEH) model fitted to British longitudinal fertility histories data. We use data from the British Household Panel Survey (BHPS), a nationally representative United Kingdom longitudinal study that began in 1991 and ended in 2008.[5] The 1991 sample is composed by $8,167$ households and $10,264$

[5]The BHPS original sample issued in 1991 consisted of $8,166$ nationally representative addresses randomly drawn from the Postcode Address File (a UK comprehensive list of post codes). A three-stage clustered probability design was implemented, with postcode sectors sampled in the first stage, addresses sampled in the second stage, and households sampled in the third stage. Implicit stratification combined with systematic sampling was used to ensure that the sample is well balanced in key socio-economic characteristics at the sector level. The study design ensures that every household in the population had the same probability of entering the sample.

individuals. All individuals who belong to the original households were re-interviewed every year from 1991 to 2008.[6] Newborn children automatically become panel members and all children are interviewed individually once they reach age 16. On wave 9 (1999) a refreshment sample of $1,500$ households and $3,395$ individuals was taken in Scotland and Wales to allow for country level comparisons. In wave 11 (2001) a refreshment sample of approximately $2,000$ new households and $5,188$ individuals was taken in Northern Ireland.

In section 2.2 we briefly discussed the BHPS and pointed out that this survey takes the strategy of building fertility histories implementing a prospective fertility follow-up of all panel members, which is complemented by retrospective fertility data collected in Wave 2 (1992) for the original sample, and in waves 11 and 12 for the Scotland-Wales and Northern Ireland refresh samples. This approach delivers a detailed fertility history of all panel members along a complete history of the variables that are suggested by theory to determine the number and timing of children.

[6]Panel members are followed when they split from the original household, and all members of the new household are subsequently interviewed as long as they live with the original panel member.

Table 5. Descriptive statistics for longitudinal complemented with retrospective fertility history from BHPS ($N = 238,895$)

Variable	Mean	SD	Min	Max	Description
resp	0.08	0.27	0	1	Response variable
parity	1.36	1.41	0	15	Parity
clock	7.14	5.78	1	33	Clock
girlat1	0.09	0.28	0	1	girl at parity 1
ssex2	0.12	0.32	0	1	Same sex at $P = 2$
ssex3	0.03	0.18	0	1	Same sex at $P = 3$
ltwinpar	0.03	0.16	0	1	Last parity twins
age	32.44	9.20	18	50	Age
agesq	1137	618	324	2500	Age squared
lincome	1.22	1.10	0	6.87	Log-income (000s)
relgion_m	0.25	0.43	0	1	Religion: missing
norelgion	0.25	0.43	0	1	Religion: none
catholic	0.07	0.26	0	1	Religion: catholic
othrelgion	0.16	0.37	0	1	Religion:other
muslim	0.02	0.12	0	1	Religion: muslim
hindu	0.03	0.16	0	1	Religion: hindu
noedu	0.43	0.49	0	1	Highest qual: none
olevel	0.23	0.42	0	1	Highest qual: O level
alevel	0.13	0.34	0	1	Highest qual: A level
diploma	0.06	0.23	0	1	Highest qual: diploma
degree	0.09	0.28	0	1	Highest qual: 1st degree
postgrad	0.02	0.14	0	1	Highest qual: postgrad
borneu	0.02	0.15	0	1	Born: EU
bornoth	0.03	0.18	0	1	Norn: other
scend	15.11	3.42	0	26	School leaving age
scend_m	0.04	0.20	0	1	School leaving age missing
mnowk14	0.50	0.50	0	1	Mother not working when 14
mscl_m	0.61	0.49	0	1	Mother scl: missing
msclpromang	0.08	0.27	0	1	Mother scl: profesional/management/technical
msclskllnm	0.10	0.30	0	1	Mother scl: skilled non manual
msclpskll	0.09	0.29	0	1	Mother scl: partially skilled
mscluskll	0.07	0.25	0	1	Mother scl: unskilled

The analytical sample contains detailed fertility histories for $14,137$ women aged between 18 and 50. The data has a complex hierarchical structure explained at length in section 2.2. Table 3 illustrates the layout of the data and descriptive statistics of the analytical sample are in table 5. The response variable ($resp$) is a $0/1$ binary variable that takes value one when a

new birth is registered.

We estimate a reduced form model for the fertility history. Two specifications are considered: (a) a model with no autoregressive term, and (b) a model that includes an autoregressive term. We implement a number of alternative estimators by either maximum likelihood or conditional maximum likelihood (for FE models): (i) logit, (ii) logit random effects (RE logit), (iii) logit fixed effects (FE), and (iv) quadratic exponential models for binary panel data (cquad).

7.1 Regression results

We now selectively discuss several aspects of the empirical results that are of interest to economists, beginning with findings from non-autoregressive models and then going on to models with some sort of autoregressive inertia. All models include time-varying controls for age (quadratic function) and log-income (000s of 1992 constant pounds) as well as parity-varying controls for the birth of a girl at parity one, the occurrence of same sex siblings at parity two and three, and the occurrence of twins in the last partity. For models that implement random effects estimators we add additional time-fixed controls for maximum achieved qualification (GCSE / O level / A level / diploma / 1st degree / posgrad / control: no qualification), religion (no religion / Catholic / Muslim / Hindu / other religion / missing religion / control: Anglican), country of birth (born in a EU country / born in a non-EU country / control: born in the UK), age at which the woman left school, work status of the woman's mother when she was 14, and the the woman's mother social class (professional and manager / skilled non manual / skilled manual / partially skilled / social class missing / control: non skilled).

7.1.1 Non-autoregressive models

Table 6 present empirical results for the logit MDEH models. Here the dynamic event history is modelled by specifying a hazard function that gives the probability that the count of interest will be increased by one unit at time τ, given the whole history of a set of exogenous ex-

ogenous variables. The discrete hazard function models $\mathbb{P}[resp_{ip\tau} = 1|\mathbf{x}_{i\tau}, \mathbf{z}_i]$ and unobserved individual heterogeneity is explicitly allowed as discussed in section 5.3.

Because FE logit delivers a consistent estimator of (λ, β, γ) by conditioning on $\sum_\tau resp_{i\tau}$, which eliminates c_i from the conditional likelihood, no estimates for the fixed effects are obtained after fitting the model and, as a consequence, no valid average marginal effects are available. This is rather disappointing because each model has its own parametrisation and coefficients may not be directly comparable. To allow comparison, for the each of the $j = 1, \ldots, J$ control variables, we define $w_{ipj\tau} = \exp(x_{ipj\tau})$ and calculate the average (semi) elasticity of $x_{ipj\tau}$ as

$$\eta^w_{ipj\tau} = \left[\frac{\partial h_{ip\tau}}{\partial w_{ipj\tau}}\right]\left[\frac{w_{ipj\tau}}{h_{ip\tau}}\right]$$

with $h_{ip\tau} = \mathbb{P}\left[resp_{ip\tau} = 1 \left| \mathbf{x}_{ip\tau}, \mathbf{s}_{ip}, \mathbf{z}_i, \sum_{j<\tau} resp_{ij} = 0\right.\right]$. Kitazawa (2012) shows that η^w is computable without the fixed effects in the FE logit model and can be calculated without complication for RE logit and logit. Hence, using η^w a meaningful comparison is possible. Table 6 reports then average (semi) elasticities.

Table 6. Modified dynamic event history for BHPS fertility — Average (semi) elasticities of $Pr(resp = 1|x,u)$ from logit discrete hazard

	LOGIT	RE LOGIT	FE LOGIT
Girl at $P = 1$	-0.012	-0.012	0.462***
	(0.026)	(0.027)	(0.036)
Same sex at $P = 2$	0.144***	0.144***	0.211***
	(0.033)	(0.033)	(0.039)
Same sex at $P = 3$	0.018	0.018	-0.272***
	(0.060)	(0.060)	(0.068)
Last parity twins	-0.209***	-0.209***	-0.769***
	(0.063)	(0.064)	(0.083)
Age	0.316***	0.316***	0.777***
	(0.013)	(0.013)	(0.016)
Age squared	-0.006***	-0.006***	-0.010***
	(0.000)	(0.000)	(0.000)
Log(income)	-0.075***	-0.075***	-0.056
	(0.008)	(0.008)	(0.034)
Max qual: none	-0.006	-0.006	0.000
	(0.036)	(0.036)	(.)
Max qual: O level	-0.083**	-0.083**	0.000
	(0.035)	(0.035)	(.)
Max qual: A level	-0.179***	-0.179***	0.000
	(0.038)	(0.038)	(.)
Max qual: Diploma	-0.202***	-0.202***	0.000
	(0.046)	(0.046)	(.)
Max qual: First degree	-0.201***	-0.201***	0.000
	(0.042)	(0.043)	(.)
Max qual: Postgraduate	-0.228***	-0.228***	0.000
	(0.063)	(0.063)	(.)
ρ		.0000802	
$SE(\rho)$		.0105225	
N. of obs	238,895	238,895	211,639
N. of individuals	14,134	141,34	8,009

Note. Average (semi) elasticities of $Pr(resp = 1|x,u)$ reported (see Kitazawa (2012)). Clustered robust standard errors at individual level in parenthesis. *10% significant; **5% significant; ***1% significant. Except for FE logit, regressions include the following time-fixed controls: maximum qualification (6 levels), country of birth (EU vs non-EU), school leaving age, mother's social class, and mother's work status when respondent was 14. Note 2. For each parity a different baseline hazard is specified so that the form of the hazard function is fully flexible. For parity 0 the baseline hazard has steps at t: $1-2$, $3-4$, $5-6$, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 and larger (control is $t = 15$). For parity 1 the baseline hazard has steps for t at: 1, 3, 5, 6, 7, 8, 9, 10, 11 and larger (control is $t = 2$). For parity 2 the baseline hazard has steps at t: 1, 2, 4, 5, 6, 7 and larger (control is $t = 3$). For parity 3 the baseline hazard has steps at t: 1, 3, 4, 5, 6 and larger (control is $t = 2$). Finally, for parities 4 and larger we specify a common baseline hazard with steps at t: 2, 3,4,5, 6, 7 and larger (control is $t = 1$).

logit vs. RE logit. We find little evidence that unobserved individual heterogeneity is present in the RE logit, with a estimated $\rho = 0.001$, which is insignificant at 5%. As a consequence, average (semi) elasticities calculated on the basis of logit and RE logit are essentially the same.

RE logit vs. FE logit. In general RE estimators give substantially different results from FE estimators. Such a result may arise if age is correlated with the individual heterogeneity term c_i, possibly because the age at which a woman marries and enters motherhood for the first time can be a function of unobserved time-fixed factors such as general attachment to the labour market, taste for family size, fecundity, health frailty, etc. Such considerations favor the FE estimators and suggest that RE hazard specifications, which are the most popular method used in the literature, may deliver seriously biased and inconsistent estimators. In fact, our British fertility history example illustrates how even a correlated random effects formulation may not sufficiently control for relevant individual unobserved heterogeneity (see, for instance, Mundlak 1978, Chamberlain 1982).

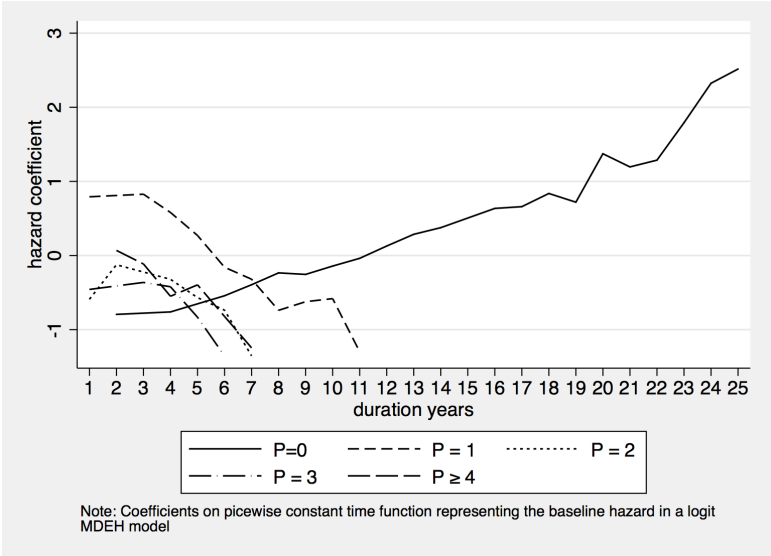
Baseline hazard and dynamics due to parity. Figure 6 presents estimates of the baseline hazard for the logit MDEH, which is similar to the estimates that are obtained from the RE logit and the FE logit. In our descriptive explorations in section 2.2, the kernel estimates of the (unconditional) hazard of entering motherhood (parity zero) showed a U-inverted form whereas the hazard for higher parities was decreasing monotonically. Moreover, descriptive kernel estimates suggested that the hazard for $P = 1$ is always larger than the hazard for $P = 2$, the hazard of $P = 2$ is always larger than the hazard of $P = 3$ and $P = 4$. Also the hazards for $P = 2$ through $P = 4$ are hardly a parallel shift of the hazard for $P = 1$. Most of these observations still hold when we look at the baseline hazard estimates from the logit MDEH in figure 6.

Note that once we condition on covariates, instead of a U-inverted shape the (conditional) hazard of entering motherhood ($P = 0$) increases monotonically with age. A possible reason

is that a U-inverted quadratic effect of age, which peaks at 26, accommodates the drop in the probability of entering motherhood (for the first time) as an individual grows older. Another interesting finding is a late peaking of the hazard at parity one ($P = 1$), which occurs between 8 and 10 years after the arrival of a woman's first child. This late surge for the second child is consistent with the timing theories of Happel et al. (1984), and the structural models of Rosenzweig and Wolpin (1980) and Heckman and Walker (1990b) that predict that women postpone motherhood as they accumulate human capital. Hence, it is likely that the late surge at $P = 1$ is driven by working mothers with high human capital who wait to the end of their fertile life to have their second child. In fact, our findings suggests that such postponement occurs mainly between the first and the second child. Notice also that for $P = 1$ and $P = 2$ the baseline hazard is increasing only for two or three years, which means that most women who plan to have two and three children schedule their pregnancies quite close once they enter motherhood; which is also predicted by Happel et al. (1984). In summary, the empirical evidence that we obtain suggest that parity is a major determinant of dynamics in the British fertility histories in a way that is consistent with the predictions from the theory of human capital of Becker (1960).

Offspring gender effects. We find also strong support that women prefer mixed-gender family composition and that they are prepared to increase family size to achieve such a mixture. This is clearly indicated in table 6 by the positive, and highly significant, coefficients on the dummy that indicates when a woman reaches parity 2 with all girls or boys. However, when a woman reaches parity 3 with three girls or three boys is likely to stop trying to achieve mixture. Interestingly, we do not find an statistically significant effect of the gender of the first child (girl at $P = 1$) for the logit and RE logit specification, but a positive effect for having a girl at parity 1 is reported by the FE logit specification. This effect is statistically significant at 1% and suggest that, after all, the probability of progressing to the second child increases when couples have a girl as their first child. Or to put it in other words, there is indeed preference for a boy a $P = 1$.

Figure 6. Estimated baseline hazard by parity from a modified dynamic event history model for British (BHPS) longitudinal fertility data.



Note that if we had kept the RE estimator, which is the estimator most used in the literature (see, for instance, Steele and Goldstein 2004), the inference for *girl at P = 1* would be that couples do not have preference for the gender of their first child. So, using a FE estimator is crucial, maybe because there is a time-fixed unobserved heterogeneity related to offspring sex preference that is correlated with the *girl at P = 1* control. The finding illustrates the relevance of implementing hazard models with fixed effects rather than the popular random effects specification.

The findings above are evidence that qualitative features of past outcomes are important and support the hypothesis that fertility is a dynamic sequential decision process. Controlling for such important observable aspects of past outcomes, even within the context of a non-autoregressive hazard model, it is possible approach to incorporate such dynamics in an easily interpretable way.

Family income. As expected, and suggested by the Beckerian human capital theory and the quantity-quality trade-off of children, the effect of income on the hazard in table 6 is negative and statistically significant at 1% in both the RE or the FE specifications. It is important to

say, however, that marginal effect calculations for logit and RE logit indicate that the marginal effect is rather small. In fact, an extra $1,000$ pounds reduce the probability of observing a second birth in less than -0.001 percentage points (p.p. hereafter) at the point of maximum probability of ever observing the arrival of a second child—i.e. at age 26 and three years after the birth of first child). Moreover, once a FE logit hazard is fitted the effect of income becomes statistically insignificant. These findings suggest that income incentives, such as the UK child tax credits program (replaced by universal credit in 2012), may have small impact on reproductive behavior.

Education. Consistent with the Beckerian theory of human capital, women with higher education have a lower hazard of having an additional child at any point of time than those who have compulsory education (the control group). The effect, however, is nonlinear. The stronger protective effect of education is felt when a woman progresses from no education to A levels—a marginal effect of -0.01 p.p. for the logit model. Progressing from A levels to Diploma reduces the logit hazard by $-.002$ p.p.; and progressing from A levels to first degree reduces the logit hazard by $-.0018$ p.p. Given that we already control for family labour income, the nonlinearity in the effect of education may reflect a wealth effect (or permanent income effect) that is not captured by income. That is, part of the smaller effect of education at the higher level may be explained by the fact that wealthier individuals tend to have larger families.

7.1.2 Autoregressive model: CQUAD

The above results show that nonautoregressive models can capture important dynamic features of the fertility outcomes. In this section we consider whether adding lagged outcome variables can further improve the model specification by capturing state dependence. Specifically, we employ the dynamic FE panel model of Bartolucci and Nigro (2010), hereafter referred to as the conditional quadratic exponential (CQUAD) specification.

The substantive difference between models in section 7.1.1 and CQUAD is that while in the

Table 7. Modified dynamic event history for BHPS fertility from CQUAD discrete hazard

	Coeff.	SE	t-stat	p-value
Girl at $P = 1$	0.558***	0.042	13.3	0.000
Same sex at $P = 2$	0.234***	0.044	5.3	0.000
Same sex at $P = 3$	-0.333***	0.076	-4.4	0.000
Last parity twins	-1.260***	0.099	-12.7	0.000
Age	0.778***	0.019	41.5	0.000
Age squared	-0.011***	0.000	-37.7	0.000
Log(income)	0.011	0.039	0.28	0.781
Lagged response	-0.394***	0.028	-14.2	0.000
$log - likelihood$				-36538.5
N. of obs				211,598
N. of individuals				8,007

Standard errors reported. *10% significant; **5% significant; ***1% significant. The following individual level controls: age, age squared, income. For each parity a different baseline hazard is specified so that the form of the hazard function is fully flexible. For parity 0 the baseline hazard has steps at: $t = 1-2$, $t = 3-4$, $t = 5-6$, $t = 7$, $t = 8$, $t = 9$, $t = 10$, $t = 11$, $t = 12$, $t = 13$, $t = 14$, $t = 16$, $t = 17$, $t = 18$, $t = 19$, $t = 20$, $t = 21$, $t = 22$, $t = 23$, $t = 24$, $t = 25$ plus (control is $t = 15$). For parity 1 the baseline hazard has steps at: $t = 1$, $t = 3$, $t = 5$, $t = 6$, $t = 7$, $t = 8$, $t = 9$, $t = 10$, $t = 11$ plus (control is $t = 2$). For parity 2 the baseline hazard has steps at: $t = 1$, $t = 2$, $t = 4$, $t = 5$, $t = 6$, $t = 7$ plus (control is $t = 3$). For parity 3 the baseline hazard has steps at: $t = 1$, $t = 3$, $t = 4$, $t = 5$, $t = 6$ plus (control is $t = 2$). Finally, for parities 4 and larger we specify a common baseline hazard with steps at: $t = 2$, $t = 3$,$t = 4$,$t = 5$, $t = 6$, $t = 7$ plus (control is $t = 1$).

former dynamics of the fertility history is accounted for by changes in explanatory variables, here dynamics are induced by changes in explanatory variables as well as the autoregressive effect of the lagged dependent variable; which is a genuine shifter of the current transition probability after controlling for individual unobserved heterogeneity in a fixed effects framework– just as in the dynamic logit model of Honoré and Kyriazidou (2000).

Because CQUAD implements a fixed effects approach, the model conditions on total scores $\sum_t resp_{it}$ and only uses the subset of observations in the sample for which at least one transition is recorded. Besides, like in any other FE approach, coefficients on time-fixed variables are not identified. However, unlike the dynamic logit of Honoré and Kyriazidou (2000), CQUAD allows for time dummies among the set of control variables because it does not impose conditions

on the support of the distribution of regressors. Table 7 shows results, reporting coefficients because average marginal effects or average (semi) elasticities are not available for this model. The sign-pattern of effects is similar to that in the logit fixed-effect hazard shown in Table 6, so we do not comment further.

Lagged outcome. Of special interest is the negative coefficient on the lagged response variable. This result has to be interpreted with care, following the analysis presented in section 3.1 of Bartolucci and Nigro (2010) who show that the coefficient of lagged outcome variable $resp_{i,\tau-1}$, say γ, can be shown to be

$$\gamma = \log\left\{\frac{\mathbb{P}(resp_{i\tau}=1\,|c_i,\mathbf{x}_{i\tau},resp_{i,\tau-1}=1\,)}{\mathbb{P}(resp_{i\tau}=0\,|c_i,\mathbf{x}_{i\tau},resp_{i,\tau-1}=1\,)}\right\} - \log\left\{\frac{\mathbb{P}(resp_{i\tau}=1\,|c_i,\mathbf{x}_{i\tau},resp_{i,\tau-1}=0\,)}{\mathbb{P}(resp_{i\tau}=0\,|c_i,\mathbf{x}_{i\tau},resp_{i,\tau-1}=0\,)}\right\} \tag{30}$$

which is the difference between the log-odds (of outcome 1 vs. 0) in τ given $resp_{i,\tau-1}=1$ and the log-odds in τ given $resp_{i,\tau-1}=0$. Thus the coefficient does not have an interpretation analogous to the standard linear autoregressive model for continuous outcomes. Clearly, in CQUAD the coefficient on the lagged outcome can be either positive or negative. While the statistical interpretation of the coefficient on the lagged outcome is clear, the economic interpretation is less clear. First, in our fertility application, it is not clear what should be considered the relevant "lagged outcome" as one could focus on $resp_{i,\tau-1}$, as CQUAD does, or on the effect of the outcome of the last pregnancy —live birth vs. still birth, for instance.

8 Discussion and conclusions

Our results support the position that specifying RE hazards, the most popular method used in applied work, may lead to biased and inconsistent estimators when there are control variables that are correlated with individual time-fixed unobserved characteristics. In such cases a FE hazard specification is preferred. The applied researcher should be aware that FE specifications are more robust than RE specifications to misspecification error not only in linear but also in

nonlinear models.

Results from our British fertility history data support the position that non-autoregressive models with lagged exogenous variables, such as gender composition and parity indicators, representing qualitative features of past fertility outcomes, can function as proxy variables for dynamic dependence. The role of these variables is easier to interpret than that of lagged dependent variables intended to model duration or occurrence dependence. When the lagged dependent variables are added to the non-autoregressive model, there are interaction effects between them and parity indicators which add to the difficulty of interpretation. Nevertheless, we detect some evidence of state dependence.

Table

8.1 Data Appendix

We use retrospective fertility data complemented with longitudinal data from the BHPS. The BHPS is a longitudinal study that began in 1991 and ended in 2008; in 2009 a large proportion of the sample of the BHPS became part of the new UK Household Longitudinal Study (UKHLS).

8.1.1 Coverage

The BHPS original sample issued in 1991 consisted $8,166$ nationally representative addresses randomly drawn from the Postcode Address File (a UK comprehensive list of post codes in the country). A three-stage clustered probability design was used, with postcode sectors sampled in the first stage, addresses sampled in the second stage, and households sampled in the third stage. In the first stage all postcode delivery points (addresses) listed in the frame were ordered so that an implicit stratification by region and socioeconomic status was created using information from the 1981 Census. Such implicit stratification allowed the use of systematic sampling for sample selection, setting a random start and a fixed sampling interval. This procedure ensured that every address in the frame had the same probability of selection and was more convenient than explicit stratification given the large number of existing strata. Postcode sectors were the Primary Sampling Units (PSUs), and have in average of $2,500$ delivery points.

A total of 250 postcode sectors were selected in the first stage using a selection probability proportional to size, guaranteed by the implicit stratification and the systematic sampling implemented. In the second stage 33 delivery points where selected within each selected postcode sector, using again systematic sampling. Finally, in the third-step non-residential addresses were excluded and all households in the selected delivery point were drawn into the sample unless the delivery point had more than three households, in which case a random sample of

three households was taken. Archived response rates were 95% for individuals and 75% for households. The original 1991 sample is composed by 8,167 households and 10,264 individuals. All individuals who belong to the original households are the Original Sample Members (OSM) and are re-interviewed every year.[7]

Newborn children automatically become panel members and all children are interviewed individually once they reach age 16. Around 13.6% of the original sample was lost due to attrition between Wave 1 and Wave 2, which amounts approximately to a 86% wave-on-wave response rate. After that, from Wave 2 on, wave-on-wave attrition becomes less onerous but still non-negligible, with a loss of around 2% to 3% of the original sample per year. By 2008 the BHPS had a sample of 4,411 OSMs, equivalent to 44.5% of the original sample. On wave 9 (1999) a refreshment sample was taken in Scotland and Wales to complement the survey and alleviate the small sample size (around 500 households in Scotland and 300 households in Wales) taken in each country in the original sample and to allow for country level comparisons, which became important due to the policy changes that followed the devolution of powers to Scotland and Wales during the 1990s. A total of 1,500 extra households and 3,395 individuals from Scotland and 1,500 households and 3,577 individuals from Wales were added to the BHPS sample. In the same vein, in wave 11 (2001) a refreshment sample of approximately 2,000 new households and 5,188 individuals was taken in Northern Ireland.

8.1.2 Incorporating retrospective information

The BHPS is one of the best suited existing surveys to analyze fertility histories in a developed country with a well established fertility transition. Indeed, the BHPS compiles histories using both retrospective and longitudinal information and carefully records the exact date of birth and gender of every child ever born to a woman even if they do not live currently in the mother's household.

Retrospective fertility data was collected in Wave 2 (1992) for the original sample, and in

[7] Panel members are followed when they split from the original household, and all members of the new household are subsequently interviewed as long as they live with the original panel member.

waves 11 and 12 for the Scotland-Wales and Northern Ireland refresh samples. All women aged 16 and over were asked to provide a detailed fertility history, including the number children ever born, the exact date of birth, the gender of the child, and the date of death if the child was not longer alive.

Besides retrospective fertility histories, the BHPS longitudinally logs the exact date of birth and the gender of any child born to an original sample member during the survey study time. Therefore, it is possible to build a detailed and complete fertility history for all adult women in the sample. This is particularly rich fertility data unlike other cross-section or even other longitudinal surveys. (The Swiss Household Panel (SHP), to give an example, only records date of birth and gender of resident children.) The role of some key factors, e.g. family preference for mixed-gender offsprings, cannot be analyzed without detailed history of birth outcomes, (see, for instance, Williamson 1976, Angrist and Evans 1998). The BHPS allows one to control for such factors.

8.1.3 The final dataset

The analytical sample contains information for $14,134$ women aged 15 and over and followed annually during the $1992-2008$ period. We have a hierarchical and longitudinal data structure, with newly reported children nested within years, years nested within parity, and parity nested within individuals.There are $87,311$ person-year records and, because each woman can report more than one birth per year, a total of $238,895$ person-year-parity entries are available. A woman can report more than one child in a year either because in that year she gave birth to multiple children (twins, triplets, etc.) or because that year she filled the BHPS retrospective fertility module and reported all children born before the start of the survey. This peculiarity of the data will be fully accounted for in our analysis.

Note that childless women contribute a single row per year and parity is set to zero in all entries. In contrast, women with a positive number of children can contribute one or more rows of data per year. Parity may vary within year. For instance, a woman who had no previous

children and gives birth to twins in a particular year will contribute two rows of data that year and parity will have gone from 1 to 2 the same year.

On average there are 1.36 children per woman, with a standard deviation of 1.4, a minimum of 0 and a maximum of 15. The distribution of number of children ever born is given in Figuree 1. Clearly, 95% of the probability mass is concentrated in the first 5 counts $\{0,1,2,3,4\}$. In fact, 70% of the probability mass falls in the $0,1,2$ counts. As a consequence, ignoring outliers in the tail, one could argue that the number of children is a count with limited support. This is typical of fertility data in developed countries and posses a challenge for analysis using count data techniques as standard count models do not fit well counts with limited support. In this context, modelling the whole event history seems to be more attractive.

8.1.4 Control variables

Coming back to a our fertility history data the response resp is a dichotomous variable that takes on 1 when a birth is reported and zero otherwise. There are a total of $189,999$ registered births in our data and various types of controls:

1. individual-specific and time-fixed variables;
2. individual-specific and time-varying variables;
3. parity-specific and time-invariant variables.

Type I controls include, for instance, gender, religion and education. Type II controls include age and family income. Finally, type III controls include whether the first child was a girl, whether there was a same-gender pair at parity one, or whether twins (multiple births) were born in the last pregnancy.

Table 5 presents a list of variables, their definition, and classifies variables in each control category along presenting some descriptive statistics.

Having type II and III control variables is critical for modelling and identifying the dynamic and sequential nature of the fertility decisions that women take over their lives. Clearly, each pregnancy is a decision in its own right and when deciding whether to become pregnant women

take into account all the information they have at the time (see, for instance, Barmby and Cigno 1990, Wolpin 1984). This includes the current number of her offspring (incentives and child benefit systems) (Barmby and Cigno 1990, Gonzalez 2013), their sex composition (due to gender preference) (Williamson 1976, Angrist and Evans 1998), the outcome of her last pregnancy (reduced fecundability after a c-section or miscarriage) (Kok et al. 2003, Hassan and Killick 2005, O'Neill et al. 2014, Sapra et al. 2014), her work status and salary (Bettio and Villa 1998, Mira and Ahn 2001), the available child care support (see, for instance, Ermisch 1989, Boca 2002, Rindfuss et al. 2010), etc. Many of such conditions change with time and can influence whether the same woman goes from childless to have her first child, but not whether she goes from having one child to having two or three children . That is, in general, factors that affect the transition from parity 0 to parity 1 may not play any role at all the transition from parity 1 to parity 2 (a similar argument is put forward in Miranda 2010; 2013). Here is where using an event history approach to model fertility outcomes becomes attractive, as variation in type II and type III variables will capture much of the dynamic nature of the counting process.

References

Adsera, A. (2006). An economic analysis of the gap between desired and actual fertility: The case of spain. *Review of Economics of the Household*, 4(1):75–95.

Agadjanian, V. and Prata, N. (2002). War, peace, and fertility in angola. *Demography*, 39(2):215–231.

Al-Sadoon, M. M., Li, T., and Pesaran, M. H. (2017). Exponential class of dynamic binary choice panel data models with fixed effects. *Econometric Reviews*, 36(6-9):898–927.

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

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence

from exogenous variation in family size. *The American Economic Review*, 88(3):pp. 450–477.

Arnold, F. and Liu, Z. (1986). Sex preference, fertility, and family planning in china. *Population and Development Review*, pages 221–246.

Arrow, K. J. (1972). Economic welfare and the allocation of resources for invention. In *Readings in industrial economics*, pages 219–236. Springer.

Arroyo, C. and Zhang, J. (1997). Dynamic microeconomic models of fertility choice: A survey. *Journal of Population Economics*, 10(1):23–65.

Barmby, T. and Cigno, A. (1990). A Sequential Probability Model of Fertility Patterns. *Journal of Population Economics*, 3(1):31–51.

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting an-consistent conditional estimator. *Econometrica*, 78(2):719–733.

Bartolucci, F. and Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, 170:102–116.

Bauwens, L. and Giot, P. (2000). The logarithmic acd model: an application to the bid-ask quote process of three nyse stocks. *Annales d'Economie et de Statistique*, (60):117–149.

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.

Bettio, F. and Villa, P. (1998). A mediterranean perspective on the breakdown of the relationship between participation and fertility. *Cambridge Journal of Economics*, 22:137–171.

Bhalotra, S. and Van Soest, A. (2008). Birth-spacing, fertility and neonatal mortality in india: Dynamics, frailty, and fecundity. *Journal of Econometrics*, 143(2):274–290.

Boca, D. D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in Italy. *Journal of Population Economics*, 15(3):549–573.

Boldrin, M., De Nardi, M., and Jones, L. E. (2015). Fertility and social security. *Journal of Demographic Economics*, 81(3):261–299.

Bongaarts, J. (1990). The measurement of wanted fertility. *Population and development review*, pages 487–506.

Bongaarts, J. (1992). Do reproductive intentions matter? *International Family Planning Perspectives*, pages 102–108.

Bongaarts, J. (2001). Fertility and reproductive preferences in post-transitional societies. *Population and development review*, 27:260–281.

Bongaarts, J. and Watkins, S. (1996). Social interactions and contemporary fertility transitions. *Population and Development Review*, 22(4):639–682.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics*, volume 5, pages 3705–3843. Elsevier.

Caldwell, J. C. (2006). Social upheaval and fertility decline. In *Demographic Transition Theory*, pages 273–299. Springer.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of econometrics*, 18(1):5–46.

Cheng, T. C. and Trivedi, P. K. (2015). Attrition bias in panel data: a sheep in wolf's clothing? a case study based on the mabel survey. *Health economics*, 24(9):1101–1117.

Chwe, M. (2000). Communication and coordination in social networks. *Review of Economic Studies*, 67:1–16.

Cox, D. and Wermuth, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika*, 81(2):403–408.

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied statistics*, pages 113–120.

Diamond, P. (1965). National debt in a neoclassical growth model. *The American Economic Review*, 55(5):1126–1150.

Doepke, M. and Tertilt, M. (2018). Women's empowerment, the gender gap in desired fertility, and fertility outcomes in developing countries. In *AEA Papers and Proceedings*, volume 108, pages 358–62.

Ellison, G. and Fudenberg, D. (1995). Word-of-mouth communication and social learning. *The Quarterly Journal of Economics*, 110(1):93–125.

Ermisch, J. (1989). Purchased childcare, optimal family size and mother's employment: theory and econometric analysis. *Journal of Population Economics*, 2(2):79–102.

Fagbamigbe, A. F. and Adebowale, A. S. (2014). Current and predicted fertility using poisson regression model: evidence from 2008 nigerian demographic health survey. *African journal of reproductive health*, 18(1):71–83.

Galor, O. and Weil, D. N. (1996). The Gender Gap, Fertility, and Growth. *American Economic Review*, 86(3):374–387.

Gonzalez, L. (2013). The effect of a universal child benefit on conceptions, abortions, and early maternal labor supply. *American Economic Journal: Economic Policy*, 5(3):160–88.

Gray, P. G. (1955). The memory factor in social surveys. *Journal of the American Statistical Association*, 50(270):344–363.

Günther, I. and Harttgen, K. (2016). Desired fertility and number of children born across time and space. *Demography*, 53(1):55–83.

Happel, S., Hill, J., and Low, S. (1984). An economic analysis of the timing of childbirth. *Population Studies*, 38(2):299–311.

Hassan, M. and Killick, S. (2005). Is previous aberrant reproductive outcome predictive of subsequently reduced fecundity? *Human Reproduction*, 20(3):657–664.

Heckman, J. and Walker, J. (1990a). Estimating fecundability from dara on waiting times to first conception. *Journal of The American Statistical Association*, 85(410):283–294.

Heckman, J. and Walker, J. (1990b). The relationship between wages and income and the timing and spacing of births: Evidence from swedish longitudinal data. *Econometrica*, 58(6):1411–1441.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, pages 153–161.

Heckman, J. J. and Borjas, G. J. (1980). Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica*, 47(187):247–283.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, pages 839–874.

Hotz, V. and Miller, R. (1988). An empirical analysis of life cycle fertility and female labor supply. *Econometrica*, 56(1):91–118.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529.

Hsiao, C. (2014). *Analysis of panel data*. Cambridge university press.

Jayaraman, A., Gebreselassie, T., and Chandrasekhar, S. (2009). Effect of conflict on age at marriage and age at first birth in rwanda. *Population Research and Policy Review*, 28(5):551.

Kapur, S. (1995). Technological diffusion with social learning. *The Journal of Industrial Economics*, 43(2):173–195.

Kirman, A. (1993). Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, 108(1):137–156.

Kitazawa, Y. (2012). Hyperbolic transformation and average elasticity in the framework of the fixed effects logit model. *Theoretical Economics Letters*, 2(02):192–199.

Knodel, J. and Prachuabmoh, V. (1973). Desired family size in thailand: Are the responses meaningful? *Demography*, 10(4):619–637.

Kohler, H. (1997). Learning in social networks and contraceptive choice. *Demography*, 34(3):369–383.

Kohler, H. (2000). Fertility decline as a coordination problem. *Journal of Development Economics*, 63:231–263.

Kok, H. S., Asselt, K. M., Schouw, Y. T. d., Grobbee, D. E., Velde, E. R., Pearson, P. L., and Peeters, P. H. (2003). Subfertility reflects accelerated ovarian ageing. *Human Reproduction*, 18(3):644–648.

Landry, A. (1909). Les trois théories principales de la population.

Manski, C. and Mayshar, J. (2003). Private incentives and social interactions: Fertility puzzles in israel. *Journal of the European economic association*, 1(1):181–211.

Mathiowetz, N. A. and Ouncan, G. J. (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *Journal of Business & Economic Statistics*, 6(2):221–229.

Melkersson, M. and Rooth, D. (2000). Modeling female fertility using inflated count data models. *Journal of Population Economics*, 13(2):189–203.

Mira, P. and Ahn, N. (2001). Job bust, baby bust?: Evidence from Spain. *Journal of Population Economics*, 14(3):505–521.

Miranda, A. (2008). Planned fertility and family background: a quantile regression for counts analysis. *Journal of Population Economics*, 21(1):67–81.

Miranda, A. (2010). A double-hurdle count model for completed fertility data from the developing world. DoQSS WP No. 10-01, Department of Quantitative Social Science, Institute of Education. Available at: https://repec.ucl.ac.uk/REPEc/pdf/qsswp1001.pdf.

Miranda, A. (2013). Un modelo de valla doble para datos de conteo y su aplicación en el estudio de la fecundidad en México. In Velázquez, A. M., editor, *Aplicaciones en Economía y Ciencias Sociales con Stata*, Stata Press book chapters. StataCorp LP.

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.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, pages 69–85.

Neher, P. (1971). Peasants, procreation, and pensions. *The American Economic Review*, 61(3):380–389.

Nelson, R. R. and Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *The American economic review*, 56(1/2):69–75.

Notestein, F. W. (1945). Population: The long view. In T. Schulz, E., editor, *Food for the World*.

Nugent, J. (1985). The old-age security motive for fertility. *Population and Development Review*, 11(1):75–97.

O'Neill, S., Khashan, A., Henriksen, T., Kenny, L., Kearney, P., Mortensen, P., Greene, R., and Agerbo, E. (2014). Does a caesarean section increase the time to a second live birth? a register-based cohort study. *Human Reproduction*.

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

Pyy-Martikainen, M. and Rendtel, U. (2009). Measurement errors in retrospective reports of event histories. a validation study with finnish register data. In *Survey Research Methods*, volume 3, pages 139–155.

Rindfuss, R. R., Guilkey, D. K., Morgan, S. P., and KRavdal, O. (2010). Child-care availability and fertility in norway. *Population and Development Review*, 36(4):725–748.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102.

Rosenzweig, M. and Wolpin, K. (1980). Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political economy*, 88(2):328–348.

Rosero-Bixby, L. and Casterline, J. (1993). Modelling diffusion effects in fertility transition. *Population Studies*, 47(1):147–167.

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.

Sánchez-Barricarte, J. J. (2017). The long-term determinants of marital fertility in the developed world (19th and 20th centuries): The role of welfare policies. *Demographic Research*, 36:1255–1298.

Sapra, K., McLain, A., Maisog, J., Sundaram, R., and Buck Louis, G. (2014). Successive time to pregnancy among women experiencing pregnancy loss. *Human Reproduction*, 29(11):2553–2559.

Schoumaker, B. and Hayford, S. (2004). A person-period approach to analysing birth histories. *Population*, 59(5):689–702.

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

Steele, F. and Goldstein, H. Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical modelling*, (4):145–159.

Stein, Z. and Susser, M. (1975). The dutch famine, 1944–1945, and the reproductive process. i. effects on six indices at birth. *Pediatric Research*, 9(2):70.

Thompson, W. (1929). Population. *American Journal of Sociology*, 34(6):959–975.

Torche, F. and Villarreal, A. (2014). Prenatal exposure to violence and birth weight in mexico: Selectivity, exposure, and behavioral responses. *American sociological review*, 79(5):966–992.

Torres, A. F. C. and Urdinola, B. P. (2019). Armed conflict and fertility in colombia, 2000–2010. *Population Research and Policy Review*, 38(2):173–213.

Uzawa, H. (1965). Optimum technical change in an aggregative model of economic growth. *International economic review*, 6(1):18–31.

Vandenbroucke, G. (2014). Fertility and wars: the case of world war i in france. *American Economic Journal: Macroeconomics*, 6(2):108–36.

Williamson, N. E. (1976). Sex preferences, sex control, and the status of women. *Signs*, 1(4):pp. 847–862.

Willis, R. (1973). A new approach to the economic theory of fertility behavior. *Journal of Political Economy*, 81(2):s14–s63.

Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. *Journal of business & economic statistics*, 13(4):467–474.

Winkelmann, R. and Zimmermann, K. (1994). Count data models for demographic data. *Mathematical Population Studies*, 4(3):205–221.

Winkelmann, R. and Zimmermann, K. F. (1991). A new approach for modeling economic count data. *Economics Letters*, 37(2):139–143.

Woldemicael, G. (2008). Recent fertility decline in eritrea: Is it a conflict-led transition? *Demographic Research*, 18:27–58.

Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy*, 92(5):pp. 852–874.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.