

Completed Fertility and Age at First Birth: Comparative Evidence from a Sample Selection Model

*Età al Primo Figlio e Discendenza Finale:
uno Studio Comparativo mediante un Modello di Selezione*

Francesco C. Billari, Riccardo Borgoni

Max Planck Institute for Demographic Research, Doberaner Str. 114,
D-18057 Rostock, Germany. E-mail: billari@demogr.mpg.de, borgoni@demogr.mpg.de

Riassunto: Nei paesi a bassissima fecondità è importante valutare l'impatto in termini di discendenza finale del rinvio nell'età al primo figlio. In questo lavoro si utilizza un modello di selezione alla Heckman come semplice soluzione per tenere conto del fatto che coloro che hanno un figlio sono una parte selezionata della popolazione. Il modello è applicato a dati individuali FFS. Si mostra come nei paesi a bassissima fecondità rinviare la nascita del primo figlio pesi maggiormente in termini di discendenza finale, pur controllando per la selettività nell'avere almeno un figlio.

Keywords: total fertility, age at first birth, lowest-low fertility, FFS, Heckman sample selection model.

1. The demographic importance of the relationship

In countries with so-called lowest-low fertility (Kohler et al., 2001a), the postponement of first births is an important mechanism that contributes to decreasing the quantum of fertility. More specifically, postponing first births has a particularly important impact on the quantum fertility in those societies where institutional arrangements such as labor market regulations, child care, and gender relationships make it more problematic for women to have both a career and a family.. First births have been severely postponed in countries such as Italy and Spain, while in Eastern European lowest-low fertility countries, this postponement is less severe. It has also been shown that the postponement of union formation, including the low prevalence of cohabitation, is a key factor in the postponement of first births (Baizán et al., 2001).

The negative relationship between age at first birth and completed fertility, which has been previously documented may also be spurious (see i.e. Marini and Hodsdon, 1981). According to Kohler et al. (2001b), there may be three classes of unobserved factors behind the spurious components. 1) Heterogeneity in the preference or desire for children. Wishing a larger family size may influence the timing of first births. The impact of age at first birth on completed fertility may thus be overestimated using standard data that do not allow to control for (initial) childbearing desires. 2) Heterogeneity in fecundability. All other factors being equal, couples who conceive first births earlier may be more fecund than couples who conceive later, and that will show

up in completed fertility. This also contributes to an overestimate of the true impact of postponing first births. 3) Heterogeneity in ability that affects the incentives to invest in education or labor market skills and age-related costs of postponing first births. Higher ability in human capital accumulation leads to a later age at first birth and higher fertility conditional on age at first birth. Higher age-related or human-capital related costs of children lead to an earlier onset of fertility, and to a lower fertility conditional on age at first birth. The impact of postponing first births is underestimated if the latter factors prevail. There are no *a priori* theoretical reasons to assume that one of the two directions prevails.

The study of birth intervals and thus of the key components of the quantum at the individual level has been addressed during the last few decades using event history analysis. One of the disadvantages of such methods, which is basically the most appropriate for this study, is the inherent complexity and difficulty to interpret results in terms of completed fertility. Especially for comparative purposes, it is useful to have access to relatively simple procedures, possibly applicable with standard software packages to ordinary survey data. Kohler et al. (2001b) used Danish data on monozygotic twins to estimate the postponement effect in a linear model for the logarithm of the quantum, controlling for unobserved factors influencing age at first birth. Kohler et al. show that in not taking into account such factors, one may underestimate the effect of postponing first births by about 10-25%. Of course, as far as the link between age at first birth and subsequent fertility are specific to a society, such results cannot be generalized. Twin data such as the ones used exist only in very few nations. To draw a larger picture, other methods are needed. We shall keep the inherent simplicity of the approach by Kohler (the linear specification in particular) and colleagues but control for the selection into first births. We address the following questions: how can we account for the fact that this postponement effect can only be assessed for those who had a first birth? And how does the impact of postponing first births on completed fertility differ among societies?

2. An application of Heckman sample selection model to FFS Data

We propose to apply the well-known Heckman (1979) sample selection model to estimate the individual-level impact of postponing the first birth on the quantum of fertility. We assume that selection into having or not giving first birth depends on unobserved factors that do also affect the transition to higher births, for instance in the ways mentioned in Section 1. Controlling for such factors, we are more confident in building cross-country comparisons which take into account the heterogeneity within national populations. The model is structured as follows:

$$Y = \beta X + \varepsilon \text{ where } Y \text{ is observed only if } I=1 \quad (1)$$

$$I^* = \delta' Z + \eta \quad (2)$$

$$I=1 \text{ if } I^* \geq 0, \text{ and } I=0 \text{ if } I^* < 0 \quad (3)$$

Similarly to Kohler et al. (2001), in (1) Y represents the natural logarithm of completed fertility. X is the age at first birth (observed only for those who had a first birth), and β

measures the impact of postponing first births on completed fertility. I indicates whether an individual has ever had a child. I^* is a latent propensity to have at least one child. Z is a set of variables influencing the propensity to have at least one birth. ε is normally distributed with variance σ^2 , η is normally distributed with variance 1, and $\text{corr}(\varepsilon, \eta) = \rho$. When X and Z coincide, the identification of the model is based on arbitrary distributional assumptions, and a full parametric model does not require exclusion restrictions (Vella, 1998; Moffitt, 1999). In our analysis, even if we assume a fully parametric model, we use exclusion restrictions. We apply the model to individual-level FFS data on 5 lowest-low fertility countries (Italy, Spain, Czech Republic, Hungary, and Bulgaria) as well as on a relatively high fertility country (Sweden). We use only data on women who, prior to age 38, have given birth. We exclude right-censored cases and we focus on cohorts born between 1945 and 1958.

We use a maximum likelihood estimation procedure. In the equation (1) the only explanatory variable is age at first birth, while in the probit selection equation (2) we used a dummy variable for high/low education and a series of dummy variables for cohorts. In Table 1, we show that the correction for sample selection, according to the Wald test for independent equations, is necessary in all Western countries and in Bulgaria. Now we briefly discuss our results on the postponement effect.

3. Results, comments and prospects for future research

In table 2, we show the estimates of the postponement effect in the sample selection model, and by comparison in an OLS model using only women who have at least one birth (for reasons of space, results concerning the selection equation are not reported here). Two important results emerge, each of which has implications for future research. First, the differences between OLS and Heckman-type estimates are of minor importance. In the Italian case, OLS overestimates the impact of postponing first births by about 4%. In the Swedish case, OLS underestimates the impact of postponing first births by 1.7%. All other cases are in between. As it was the case in Kohler et al. (2001b) when comparing twin-based estimates to OLS estimates, the bias in OLS estimates is not particularly important, which also suggests that OLS estimates may be used for comparative purposes. Of course, this result cannot be taken as a definitive answer. Future research needs a careful assessment of the statistical and econometric foundations of the model. In particular, we plan to assume different distributions both for the dependent variable and for the unobserved component. van der Klaauw and Koning (2001) for instance, discuss a flexible parametric specification that provides consistent parameter estimates.

The second important point concerns the demographic interpretation of the results. Postponing first births matters for completed fertility more in lowest-low fertility contexts than in countries with higher fertility, although the difference is not great. In Spain, where the postponement effect is the highest, a 5-year postponement of the first birth for a woman reduces her quantum by 0.23 children. In Sweden, the quantum is reduced by 0.14 children. This suggests the presence of interactions between the effect of postponing births and the quantum of fertility. In Kohler et al. (2001a) the low compatibility between the role of a mother and labor force participation is discussed as a key issue in this interaction. Further research on this point, possibly using more recent cohorts who exhibit lower fertility levels, is needed.

Table 1: *Estimated parameters of the Heckman sample selection model.*

	ρ	Std.Error.	p-value for the Wald test of independent equations
ITALY	-0.79280	0.03919	0.000
SPAIN	-0.77232	0.05191	0.000
CZECH	0.21329	0.16808	0.205
HUNGARY	0.16539	0.11746	0.167
BULGARIA	0.24158	0.10311	0.024
SWEDEN	-0.85598	0.03059	0.000

Table 2: *Results of the sample selection model: effect of a later age at first birth on log(total fertility). Women, FFS individual-level data.*

	Heckman selection model			OLS Model		
	Estimate β	Std.Error.	p-value	Estimate β	Std.Error.	p-value
ITALY	-0.03527	0.00229	0.000	-0.03679	0.00228	0.000
SPAIN	-0.04603	0.00316	0.000	-0.04569	0.00285	0.000
CZECH	-0.03618	0.00543	0.000	-0.03620	0.00498	0.000
HUNGARY	-0.03051	0.00386	0.000	-0.03048	0.00357	0.000
BULGARIA	-0.02871	0.00290	0.000	-0.02895	0.00302	0.000
SWEDEN	-0.02806	0.00229	0.000	-0.02758	0.00237	0.000

References

- Baizán P., Aassve A., Billari F.C. (2001). Cohabitation, marriage, first birth: The interrelationship of family formation events in Spain. Max Planck Institute for Demographic Research, Rostock, Working Paper WP 2001-036.
- Heckman J.J. (1979). Sample selection bias as a specification error. *Econometrica*, 47: 153-161.
- Kohler H.-P., Billari F.C., Ortega J.A. (2001a). Towards a Theory of Lowest Low Fertility. Max Planck Institute for Demographic Research, Rostock, Working Paper WP 2001-032.
- Kohler H.-P., Skytthe A., Christensen K. (2001b). The Age at First Birth and Completed Fertility Reconsidered: Findings from a Sample of Identical Twins. Max Planck Institute for Demographic Research, Rostock, Working Paper WP 2001-006.
- Marini M.M., Hodsdon P.J. (1981). Effects of the timing of first marriage and first birth on the spacing of subsequent births. *Demography* 18 (4): 529-548.
- Moffitt R.A. (1999). New Developments in Econometric Methods for Labor Market Analysis. In Ashenfelter O., Card D. (Eds), *Handbook of Labor Economics*, 1367-1397, Elsevier Science, Amsterdam.
- van der Klaauw B., Koning R.H. (2001). Testing the Normality Assumption in the Sample Selection Model with an Application to Travel Demand, *Journal of Business and Economic Statistics*, forthcoming.
- Vella F. (1998). Estimating Models with Sample Selection Bias: A Survey, *Journal of Human Resources*, 33 (1): 127-169.