The Causal Effect of Education on Age at First Birth - the Use of Danish Twin Data

1 Introduction

1.1 Background and Motivation

Numbers of studies have documented strong correlation between women's educational attainment and fertility. In particular, there seems to be a positive correlation between the age at first birth (AFB) and the length of education. One possible explanation for this result is that the opportunity costs of children for highly educated women are higher than for less educated, and therefore women with high education have fewer children and they enter into motherhood at a later age (Lappegård and Rønsen (2005); Hoem et al. (2006)). On the other hand, such an effect might be due to selection arising because of unobserved heterogeneity or confounding factors affecting both fertility and educational choice (e.g. preferences for children or possibly certain genetic factors) and hence not represent a *causal* effect of education on fertility, e.g. age at first birth.

In this study, we will use Danish register data, in particular the Danish Twin Registry (Skytthe et al. (2002)), to estimate the causal effect of educational attainment on AFB by means of a *fixed effects duration analysis*. This modeling approach exploits the fact that monozygotic (MZ) twins share all of their genes along with their early childhood environment and therefore, by making intra-pair comparisons of the effect of education on AFB, unobserved endowments possibly influencing both education and fertility (in this case AFB) are taken into account. Hence, the comparison of a *standard* model with a fixed effects model will provide evidence as to whether such selection mechanisms are in play. Furthermore, since dizygotic (DZ) twins share on average half of their genes along with their early childhood environment, if a similar comparison for DZ twins gives rise to a smaller degree of attenuation than the comparison between models for MZ twins, this suggests that genetic confounding factors play a more important role than environmental confounding factors (Kohler et al. (2010)). Previous studies have used twin data to study fertility (Kohler et al. (1999, 2002); Rodgers et al. (2008)). However, to our knowledge this is the first study to use a fixed effects duration analysis to study the relationship between education and fertility.

1.2 Data and Descriptives

The study is based on data from Statistics Denmark and from the Danish Twin Registry from which the female twin cohorts born 1961-1975 are extracted. The available information comprises links from the twins to their all children (in 2001) along with the birth dates of these children. Also, yearly updated information on education was drawn from the Demographic Database of Statistics Denmark. Attention is restricted to twins who are alive and living in Denmark at the age of 20 and who have not yet given birth to their first child at that age. This gives rise to a study population of 6006 individuals, 2842 monozygotic (MZ) and 3164 dizygotic (DZ) twins. The women are censored at death, emigration or end-of-study (31 December 2001), whichever comes first.

In order to examine the relationship between education and age at first birth, the study population is grouped according to whether the women have obtained an upper secondary education no later than age 20 or not. There are 41.2% women who meet this criterion, the share being sligthly larger among MZ than DZ twins (44.7% and

38.0%, respectively). The number of education discordant pairs among the 1421 intact MZ twin couples, i.e. the number of twin couples in which one has an upper secondary education is 184 (12.9%) whereas the corresponding number for the 1582 DZ twin couples is 404 (25.5%). This is important from a modeling perspective, since only the discordant pairs provide information on the effect of education in the fixed effects model.

2 Statistical Models

Two different models are employed for both MZ and for DZ twins, respectively. Both models are versions of the Cox Proportional Hazards Model:

$$\lambda_{ij}(t) = \lambda_0(t) \cdot \exp(\beta \cdot X_{ij}), \quad i = 1, \ldots, n; j = 1, 2 \tag{1}$$
$$\lambda_{ij}(t) = \lambda_{0j}(t) \cdot \exp(\beta \cdot X_{ij}), \quad i = 1, \ldots, n; j = 1, 2. \tag{2}$$

Model (1) is referred to as a *marginal model* in which $\exp(\beta)$ is to be interpreted as a comparison of two individuals who differ with respect to their educational attainment and who do not necessarily belong to the same twin pair. In Model (2), on the other hand, $\exp(\beta)$ should be interpreted as an *intra-pair* comparison (Holt and Prentice (1974)). This model is referred to as the *fixed-effects model*. The comparison of the results from these two models in twins provides evidence of whether the education-AFB relationship is confounded by unobserved endowments which are shared within twin pairs, as described above.

3 Results

3.1 Monozygotic Twins

Among the MZ twins with an upper secondary education in the study population, the median AFB 30.2 years, whereas for MZ twins without an upper secondary education, the median waiting time is 28.0 years. Fitting the marginal model (1) for the MZ twins gives the following results:

```
                                   (Std. Err. adjusted for 1508 clusters in parnr)
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      stud20 |   .6935247   .0376205    -6.75   0.000     .6235743    .7713219
------------------------------------------------------------------------------
```

I.e. the hazard ratio for an MZ twin with an upper secondary education compared to an MZ twin with no upper secondary education is 0.694 with a 95%-confidence interval which is $[0.624; 0.771]$ which suggests a quite strong negative influence of education on AFB. However, as is evident from the output below, this effect completely vanishes when we employ the fixed effects model (2):

```
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      stud20 |          1   .1825742     0.00   1.000     .6991858    1.430235
------------------------------------------------------------------------------
                                                             Stratified by parnr
```

3.2 Dizyygotic Twins

For DZ twins with an upper secondary education, the median waiting time is 30.1 years, for DZ twins without upper secondary eduation, it is 27.4 years. Fitting the marginal model (1) gives a hazard ratio comparing DZ twins with and without an upper secondary education of 0.632 with a 95%-confidence interval of $[0.576; 0.692]$ as the following output shows:

(Std. Err. adjusted for 1705 clusters in parnr)

_t	Haz. Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
stud20	.6317789	.0295392	-9.82	0.000	.5764567	.6924102

When the fixed effects model is employed to the sample of DZ twins, the effect of education is attenuated, since the HR changes to 0.809 and it is now only borderline significant.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stud20	.8092105	.0981417	-1.75	0.081	.6380095	1.026351

4 Preliminary Conclusion

The results in this study show a negative effect of having an upper secondary education on AFB for MZ twins. However, this effect is attenuated in the fixed effects model in which it vanishes. This attenuation seems to be higher for MZ than for DZ twins (even if the confidence intervals overlap to a certain extent). Now, MZ twins share all of their genes while DZ twins share on average half of their genes. If we furthermore make the *equal environment assumption*, i.e. assume that MZ and DZ twins share their childhood environment to the same degree, the possibly larger attenuation for MZ twins could suggest that unobserved heterogeneity is more likely to arise due to genetic factors than due to environmental factors. Put differently, the results suggest that *nature* might play a more important confounding factor than *nurture*.

In this paper we will - apart from the results shown above - discuss possible other and more detailed measures of education along with a comparison of the results obtained from different modeling approaches such as a random effects model which exploits the information in the data differently than the fixed effects approach used here.

References

Hoem, J. M., Neyer, G., and Andersson, G. (2006). Education and childlessness. The relationship between educational field, educational level, and childlessness among Swedish women born in 1955-59. *Demographic Research*, 14:331–380.

Holt, J. D. and Prentice, R. L. (1974). Survival analyses in twin studies and matched pair experiments. *Biometrika*, 61:17–30.

Kohler, H. P., Behrman, J., and Schnittker, J. (2010). Social Science Methods for Twins Data: Integrating Causality, Endowments and Heritability. PSC Working Paper Series 21, Population Studies Center, University and Pennsylvania.

Kohler, H. P., Rodgers, J. L., and Christensen, K. (1999). Is Fertility Behavior in Our Genes? Findings from a Danish Twin Study. *Population and Development Review*, 25:253–288.

Kohler, H. P., Rodgers, J. L., and Christensen, K. (2002). Between Nurture and Nature: The Shifting Determinants of Female Fertility in Danish Twin Cohorts. *Social Biology*, 49(3/4):218–248.

Lappegård, T. and Rønsen, M. (2005). The Multifaceted Impact of Education on Entry into Motherhood. *European Journal of Population*, 21:31–49.

Rodgers, J. L., Kohler, H. P., McGue, M., Behrman, J. R., Petersen, I., Bingley, P., and Christensen, K. (2008). Education and cognitive ability as direct, mediating, or spurious influences on female age at first birth: Behavior genetic models fit to danish twin data. *American Journal of Sociology*, 114:202–232.

Skytthe, A., Kyvik, K., Holm, N. V., Vaupel, J. W., and Christensen, K. (2002). The Danish Twin Registry: 127 Birth Cohorts of Twins. *Twin Research*, 5:352–357.