

Instrumental variables (IV)

Eduardo Fe

01/07/2019

Introduction.

Does childbearing hinder women's labour supply? The last 80 years have seen an increase in the proportion of women joining the work force. This trend has gone hand in hand with a reduction in fertility rates. One might wonder if the decline in fertility is the cause of the sharp increase in female labour supply. Unfortunately, this regurality, which is typically found in most data sources, is masking two potential effects: the effect of female labour supply on child-bearing, on the one hand, and the effect of child-bearing on labour supply, on the other. This is the problem of selection all over again; we cannot observe a woman's labour supply under different number of children (and vice-versa); as a result, we cannot use data (on their own) to understand why some women settled on a particular number of childre and how this decision affected their labour market status.

Humans are, however, interesting *machines*. Since early work in the 1960s by Charles Westoff and his collaborators, research has documented a parental preference for mixed sibling-sex composition. Specifically, parents of same-sex siblings are **significantly** more likely to have additional children. This has been observed for the US, the UK, Finland, Sweden, Argentina, Mexico and a vast array of other countries. For example, data from Sweden suggests that women with two boys or two girls have a 1.3 times higher risk of having a thrid child than women with a boy and a girl¹. Angrist and Evans² (1998) and Iacovou³ (1996; 2001) quickly noted the potential for an instrumental variable strategy...

SUPpose you want to estimate the effect of children of women's labour supply. For the reasons above, this relationship is confounded. Now, think of women with two children. It has been argued that the sex mix is virtually random. Under that premise, When a woman gives birth to two first siblings of the same sex, the mixed-sibling-sex preferences act as an *unconfounded* predictor of fertility. In other words, for women with 2 children, a dummy variable indicating in the sex of their children is the same can be used as an instrumental variable for further childbearing for these women.

To study this question, we are going to use data from Cruces and Galiani⁴ (2007), who study the question for Mexico and Argentina. We focus on the latter country for the illustration. Part of the task will be estimate the LATE for Mexico as well.

Their data comes from the 1991 Argentinian Census, which include 16,023,180 observations (about 50% of the population). They restrict the sample to 21-35 yera old women with at least two children. They limit the study further to women whose oldest child is younger than 18 at the time of census and whose younger child is older than one year of age. This results in a final sample of 599,941 observations.

You can load the data and do a few descriptive analyses now. The key variables for the analysis are **morethan2** (the treatment variable), **samesex** (the instrumental variable) and **ocupado** (which is Spanish, for employed, and is the woman's labour market status -the outcome variable). Another interesting variables are marital status (**s3**), age at first birth (**agefb**), current age (**age**).

As a first step, compute a simple regression of the outcome variable on the treatment. The coefficient of the treatment variable in this regression is numerically identical to a difference in sample means by treatment group. We know that the results from this regression will not have a causal interpretation (why?).

¹Andersson, G., K. Hank and A. Vikat (2006). Understanding parental gender preferences in advanced societies: Lessons from Sweden and Finland

²Angrist, J. and W. N. Evans (1998) Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. American Economic Review, 88 (3), 450 - 477

³Iacovou, M. (2001) Fertility and female labour supply. ISER Working Paper Series 2001-19.

⁴Cruces, G. and S. Galiani (2007) Fertility and female labor supply in Latin America: New causal evidence.

Once you have done that, you can move on to estimate the LATE. Remember this LATE is a ratio of the difference in means by assignment group (defined by the instrumental variable) to the difference in means in the treatment by assignment group.

The main data set is `ARG_21-35_pub.dta`

After some work, you will be able to show that the LATE equals

```
late
```

```
## argentina$samesex  
##          -0.09054618
```

Next, you can explore variation by marital status, age at first birth. When you have done that, load the Mexican data and repeat the analysis. The goal is to try to replicate the results in Cruces and Galiani.

When calculating LATE, you can use the ratio of differences in means discussed in the class. However, you will also want to obtain standard errors for the estimator. Therefore, as an additional exercise try to figure out how implemente of the method of Two Stage Least Squares using regression (`1m`); for the latter task, Lab 1 should be useful.