Motivation

Economists often want to understand what drives decisions at the family or household level, such as decisions on how to allocate time, human capital or financial resources. In particular, a widely studied topic are decisions on fertility and investment in children: that is, how many children a family chooses to have, and how many resources the family decides to invest in these children.

The hypothesis investigated in this empirical work is called the "quantity-quality trade-off", which is detailed in the theoretical work by Becker and Lewis (1973) and Becker and Tomes (1976). It states that families face a trade-off between the number of children in the household (or "quantity"), and the children's outcomes (or "quality"). In this model, parents can increase the quality of their children by making costly investments in the children, such as investments in human capital or living conditions. Quality expenditures are proportional to the number of children[1], so an increase in quantity of children raises the marginal cost of children quality; similarly, a higher children quality raises the marginal cost of an additional child.

This could generate a non-monotonic relationship between the number of children and *exogeneous* changes in income. For example, imagine low-income families have many children and do not invest much in quality. With an exogenous increase in income, these families may invest more in quality, implying that the "marginal cost" of a child increases. Therefore, quantity may fall with such increase in income. With further increases in income, then quantity may start to increase again, generating an U-shaped relationship between quantity of children and exogeneous changes in income (as in Figure 1 of Becker and Tomes (1976)).

Another implication of this theory is that exogeneous changes in the number of children should decrease the quality of children. This theory has been tested in many different settings, using different empirical strategies. See, for example, Black et al. (2005) who estimate the effect of quantity of children on education attainment, and Millimet and Wang (2011) who estimate the effect of quantity of children on height and Body Mass Index (BMI).

One of the strategies used in these papers rely on the use of Instrumental Variables (IVs). They use two IVs for family size that have been used in a seminal paper by Angrist and Evans (1998) to study the effect of fertility on parents' labor supply. One instrument is the occurrence of twins at birth, and the other instrument is an indicator if the first two children are of the same sex.

---

[1]Put simply, providing a quality investment $\bar{q}$ for two children is twice as costly as providing the same investment $\bar{q}$ for one child.

Questions

Read the motivation text above to understand the discussion on the relationship between family size and children quality. The papers cited are important references for this assignment; however, we do not expect you to read them in detail. You should only understand what each paper does in a broad sense, and consult the papers for specific details when they are useful for your work. Specifically, we hope that our motivation is enough to understand the intuition behind Becker's models and conclusions in the two papers.

You will have to choose a dataset that has enough information to answer the empirical questions below. IPUMS has data from a number of countries in many different periods, and is a good option to find a dataset that is suitable to do this empirical work. But you are free to use alternative datasets. This could be particularly interesting in case, for example, you find a dataset that allows you to construct more interesting outcome measures. Each student must choose a dataset that has not been chosen by anyone else. Once you choose your dataset, go to this spreadsheet, and fill in you choice (after checking that no one else had chosen that). You are free to change your choice.

Also, pay attention on how these papers present their results. In particular, avoid simply copying & pasting directly the outputs from R or Stata. Also, be sure that tables have explanatory notes, and that the reader understands the name of the variables in the tables (for example, do not simply put the variable name "Inc98").

1. (10 points) Let I be family income, and S be the total number of siblings. Considering different economic stories that could affect the relationship between these two variables, discuss different shapes that the function $\mathbb{E}[S|I = a]$ might have.

2. (10 points) Using the dataset that you chose, estimate this relationship using an OLS estimator. Each observation should be a family. Consider different models, so that you allow for more flexibility on the approximation of the CEF $\mathbb{E}[S|I = a]$. For example, you may divide the income variable in q bins to do that. Present your results in a professional way, and discuss them.

3. (10 points) Do you think the CEF you estimated has a causal interpretation? Discuss why or why not. You may also try to provide evidence from the data to support your claims.

4. (5 points) Now let Y be a measure of "quality" of the oldest child, and again S be the number of children. Discuss again different economic stories that could affect the shape of $\mathbb{E}[Y|S = s]$.

5. (5 points) Discuss how different environments might change $\mathbb{E}[Y|S = s]$? That is, why and how you should expect this function to have different shapes in different settings (for example, but not limited to, different countries or historical periods)?

6. (15 points) Using the dataset that you chose, define a measure of children "quality". A disadvantage we have relative to, for example, Black et al. (2005) is that we will not have long-run

outcomes for the children. Still, you can think about some measures of quality with the dataset that you have. For example, it could be school enrollment, school grade vs age distortion, enrollment in private school, and so on. Bonus points in case you think about an interesting measure!

Then, using the measure you constructed, estimate an OLS regression of Y on S. Again, each observation should be a family. Consider also alternative regressions that allow for a more flexible approximation of the CEF. Present your results in a professional way, and discuss them.

Note: You can restrict your sample to families such that the oldest child is 14 years old, so your regressions are easier to interpret.

What if you included in your regression all families with children? That is, your sample would have families where the oldest child is 1 year old, and families where the oldest child is 14 years old. Explain why it would be hard to interpret the parameters that the OLS estimator would recover in this case.

7. (10 points) Do you think the CEF you estimated has a causal interpretation? Discuss why or why not. You may also try to provide evidence from the data to support your claims.

8. (10 points) Re-estimate the specification from Question 6, including other control variables that might help you justify that your estimators have a causal interpretation. What are the characteristics that these control variables must satisfy? Would family income be a good candidate? Discuss the plausibility that these regressions identify a causal effect. Present your results in a professional way, and discuss them.

9. (10 points) Another approach that has been commonly used to estimate the effects of family size on children "quality" is to use instrumental variables, such as same-sex siblings and twins. These instruments have been used in many different settings as a source of exogenous variation in family size, you can see Angrist and Evans (1998) for an example. In this question you should address the following topics, in any order you want to:

   - Discuss the validity of these instruments in your setting (if possible, provide evidence in favor or against their validity).

   - Suppose the conditions for validity of the instrument hold, so the IV estimand identifies a causal parameter. How do you interpret this causal parameter in this context?

10. (15 points) Implement this IV strategy in your setting (even if your conclusion in the previous question is that this is not a reasonable approach). Present your results in a professional way, and discuss them. How do they compare to Black et al. (2005)?

References

Angrist, J. D. and W. N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review 88*(3), 450–477.

Becker, G. S. and H. G. Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2), S279–S288.

Becker, G. S. and N. Tomes (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy 84*(4), S143–S162.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics 120*(2), 669–700.

Millimet, D. L. and L. Wang (2011). Is the quantity-quality trade-off a trade-off for all, none, or some? *Economic Development and Cultural Change 60*(1), 155–195.