ECON 121, Applied Econometrics and Data Analysis **Prof. Tom Vogl**

Spring 2021 **University of California San Diego**

FINAL EXAM

You will analyze data from India's most recent *National Family Health Survey* to answer two questions:

- What is the effect of birth order on infant mortality? In other words, does being second-born rather than first-born in the family cause a baby to be more or less likely to die?
- What is the effect of children on maternal labor supply? In other words, does having more children cause women to be less likely to engage in market work?

The dataset https://github.com/tomvogl/econ121/raw/main/data/nfhs5.rds contains data on mothers aged 15-49 and their children, with one observation per child. The survey followed a clustered sampling design which first randomly sampled primary sampling units (villages or city block) and then sampled mothers within them. The variable `psu_id` is a unique number for each primary sampling unit, and the variable `mom_id` is a unique number for each mother. The other `mom_` variables are mother-level characteristics and outcomes, while the `child_` variables are child-level characteristics and outcomes.

You may refer to books, internet resources, notes, and code from problem sets, solutions, and classroom examples. You may **not** communicate with anyone but Prof. Vogl during the exam; evidence of such communication will result in failure. Questions that require coding have a **C** next to them. Questions that require words have a **W** next to them. (The **W** questions may require you to do some calculations "by hand," but no coding is necessary.) Every question is worth the same number of points.

Part A: Effect of Birth Order on Infant Mortality

1. Estimate a linear probability model in which the outcome is infant mortality and the covariate is the child's birth order. Make sure to take the sample design into account. **C**
2. Interpret the coefficient on birth order in the above regression. Is it statistically significant? **W**
3. If instead of a linear probability model you estimated a logit model and computed an odds ratio, do you think the odds ratio would be negative, between 0 and 1, or greater than 1? Explain. **W**
4. A concern is that children of higher birth order come from larger families, which may have different infant mortality risk. To address it, rerun the regression with mother fixed effects. **C**
5. Interpret the coefficient on birth order in the above regression. Comparing the results with and without fixed effects, do you think infant mortality risk is higher or lower in larger families? Explain. **W**
6. Another concern is that children of higher birth order tend to be born in later years, and mortality risk may depend on birth year. Add the child's birth year as a covariate in the fixed effects regression. **C**

7. Interpret the coefficient on birth order in the above regression. Do you think this coefficient represents a causal effect of birth order? If so, explain why. If not, give a concrete reason for remaining bias. **W**
8. Families may treat boys and girls differently. Estimate a new regression with one or more interaction terms to ask whether the effect of birth order on infant mortality differs by gender. **C**
9. Policymakers are also interested in how infant mortality risk varies with maternal characteristics. Run a linear probability model of infant mortality on mother's rural status, education, and age. Compute the **difference** in predicted infant mortality risk between a 15-year-old rural mother with 5 years of education (mother A) and a 40-year-old urban mother with 12 years of education (mother B). Also compute the **ratio** of the predicted risk for mother A to the predicted risk for mother B.[1] **C**
10. Interpret the results of your analyses in the previous question. How much higher or lower is infant mortality risk for mother A than for mother B? Provide answers in absolute terms and proportional terms, and describe the statistical significance of both answers. **W**

Part B: Effect of Children on Maternal Labor Supply

1. Drop observations with birth order greater than 1, so your dataset consists of one observation per mother. Estimate a linear probability model of the mother's work status on her number of children. **C**
2. Describe one reason the estimated coefficient on the number of children may be biased upward. **W**
3. Because the number of children may be endogenous, we rely on an instrument: whether the mother's first-born child was a boy. Many Indian parents desire a boy, so having a first-born boy reduces the subsequent number of children. Do you think this variable is a valid instrument? Explain. **W**
4. Estimate a regression of the number of children on the gender of the first-born child. Also estimate a regression of work status on the gender of the first-born child. **C**
5. In the language of instrumental variables, what are each of the above regressions called? W
6. Using the results of the last two regressions, calculate ("by hand") the instrumental variables estimate of the effect of children on maternal labor supply. Interpret your result. **W**
7. Estimate the effect of children on maternal labor supply using two-stage least squares. **C**
8. If the assumptions for instrumental variables with heterogeneous treatment effects were met, whose average treatment effect would we be estimating using this strategy? **W**
9. To assess instrument validity, estimate a probit regression of the gender of the first-born child on the mother's rural status, education, and age. Report marginal effects. **C**
10. Interpret the marginal effect on mother's years of education. Is this association a concern for any of the assumptions for instrumental variables with heterogeneous treatment effects? If so, which one(s)? **W**

[1] To access the intercept from a model estimated using `feols()`, type `` `(Intercept)` `` including backticks and parentheses.