

# ECON3150/4150: Introductory Econometrics – Exam Spring 2024

Be brief and to the point. Always motivate your answers. All sub-questions have equal weight.

The following study

*Altonji, J.G., Elder, T.E, Taber, C.R. (2005). An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling. Journal of Human Resources 40(4), 791-821.*

investigated the effect of Catholic schooling on school outcomes. Below you find definitions of some variables in their dataset as well as descriptive statistics:

```
# id:      person identifier
# read12:  reading standardized score
# math12:  mathematics standardized score
# female:  =1 if female
# asian:   =1 if Asian
# hispan:  =1 if Hispanic
# black:   =1 if black
# motheduc: mother's years of education (8,11,11.5,12,14,16,17,18)
# fatheduc: father's years of education (8,11,11.5,12,14,16,17,18)
# lfaminc: log of annual family income (USD)
# hsgrad:  =1 if graduated from high school by 1994
# cathhs:  =1 if attended Catholic high school
# parcatch: =1 if a parent reports being Catholic
```

| ##          | mean             | SD              | min           | max            | N    |
|-------------|------------------|-----------------|---------------|----------------|------|
| ## id       | 4648285.08190955 | 2721625.8314156 | 124902.000000 | 7979086.000000 | 5970 |
| ## read12   | 51.54847571      | 9.3952552       | 29.150000     | 68.090000      | 5970 |
| ## math12   | 51.85017923      | 9.4783118       | 30.140000     | 71.370000      | 5970 |
| ## female   | 0.52763819       | 0.4992774       | 0.000000      | 1.000000       | 5970 |
| ## asian    | 0.06281407       | 0.2426486       | 0.000000      | 1.000000       | 5970 |
| ## hispan   | 0.11139028       | 0.3146412       | 0.000000      | 1.000000       | 5970 |
| ## black    | 0.07755444       | 0.2674916       | 0.000000      | 1.000000       | 5970 |
| ## motheduc | 13.24154104      | 2.0021641       | 8.000000      | 18.000000      | 5970 |
| ## fatheduc | 13.54396985      | 2.2630167       | 8.000000      | 18.000000      | 5970 |
| ## lfaminc  | 10.31288253      | 0.8019290       | 6.214608      | 12.34583       | 5970 |
| ## hsgrad   | 0.93031826       | 0.2546312       | 0.000000      | 1.000000       | 5970 |
| ## cathhs   | 0.07202680       | 0.2585539       | 0.000000      | 1.000000       | 5970 |
| ## parcatch | 0.36264657       | 0.4808043       | 0.000000      | 1.000000       | 5970 |

```
# Results to Question 1.
```

```
##                reg1                reg2                reg3
## Dependent Var.:    hsgrad                hsgrad                hsgrad
##
## cathhs            0.060 (0.007)  0.039 (0.006)  0.029 (0.007)
## fatheduc = 11      -0.051 (0.027) -0.052 (0.027)
## fatheduc = 11.5    0.023 (0.036)  0.005 (0.036)
## fatheduc = 12      0.058 (0.024)  0.042 (0.023)
## fatheduc = 14      0.070 (0.023)  0.043 (0.023)
## fatheduc = 16      0.110 (0.023)  0.069 (0.023)
## fatheduc = 17      0.103 (0.024)  0.055 (0.024)
## fatheduc = 18      0.111 (0.024)  0.049 (0.024)
## black              -0.025 (0.014) -0.003 (0.014)
## asian              0.027 (0.009)  0.031 (0.009)
## hispan             -0.018 (0.013) -0.008 (0.013)
## lfaminc            0.045 (0.006)
## Constant          0.926 (0.004)  0.870 (0.023)  0.429 (0.064)
##
## _____|_____|_____|_____
## S.E. type      Heteros--rob. Heterosk--rob. Heterosk--rob.
## SSR            385.57          371.59          365.68
## Observations   5,970           5,970           5,970
```

```
# Results to Question 1j.
```

```
cor(predict(reg3), catholic$hsgrad)^2
cor(residuals(reg3), catholic$cathhs)
```

1. (60%) We are interested in estimating the effect of attending a Catholic high school on high school graduation. To answer this question we estimated the models in the R-output above:
  - a. Interpret the coefficient on `cathhs` in the 1st results column (`reg1`).
  - b. Interpret the coefficient on `fatheduc = 11` in the 2nd column (`reg2`). Is this estimate significant at the 1% level?
  - c. Interpret the coefficient on `cathhs` in the 2nd column. Explain why the size dropped compared to the first column and use the omitted-variable-bias formula to motivate your answer.
  - d. What is the interpretation of the intercept in the 2nd column (`reg2`)? And in the 3rd column (`reg3`)?
  - e. The third column (`reg3`) adds the logarithm of family income to the specification. What is the predicted value of high-school graduation for an asian student who attended a catholic high-school, whose father has 16 years of schooling, and with a family income of 100,000 USD? What do you conclude?
  - f. Compute the R-squared of the regression in the 3rd column (`reg3`).
  - g. Interpret the coefficient on `lfaminc` in the 3rd column (`reg3`). How would the coefficients in that regression change if family income was not measured in US dollars but in 1000s of US dollars?
  - h. Perform a joint two-sided F-test (assuming homoskedasticity) of the null-hypothesis that the coefficient on the race variables and father's education in the 2nd column (`reg2`) are zero. Do you reject the null hypothesis at the 1% level?
  - i. Explain under what assumption the coefficient on `cathhs` in the 3rd column (`reg3`) estimates the causal effect of Catholic high-school attendance on high-school graduation. Do you believe this assumption?
  - j. What two numbers (rounded to 3 decimals) does the R-code above in the output (marked Question 1j.) produce?
  - k. Using the dataset above and starting from the specification in the first column (`reg1`) you want to jointly estimate how i) high-school completion differs for kids with a mother who has at least a high-school education (ie at least 12 years of schooling), and ii) how this gap is different in Catholic high-schools. What is the regression specification that you would estimate?
  - l. Write the R-code that implements the estimation proposed in 1k.

```
# Results to Question 2.  
t(aggregate( . ~ paracath, catholic, mean))
```

```
##           [,1]           [,2]  
## paracath    0.00000000    1.00000000  
## id 4707410.34218134 4544372.09561201  
## read12    51.45292510    51.71640647  
## math12    51.55578712    52.36757506  
## female     0.53850197     0.50854503  
## asian      0.06806833     0.05357968  
## hispan     0.03679369     0.24249423  
## black      0.11222076     0.01662818  
## motheduc   13.29566360    13.14642032  
## fatheduc   13.53679369    13.55658199  
## lfaminc    10.28120838    10.36855001  
## hsgrad     0.92115637     0.94642032  
## cathhs     0.01340342     0.17505774
```

2. (40%) Some studies have used religious affiliation as exogenous source of variation to estimate the causal effect of Catholic high-school attendance on outcomes. Taking inspiration from this approach we plan to use whether a parent reports being Catholic as a so-called instrumental variable. The output below reports sample averages separately for children without a catholic parent (`parcath=0`) and with a catholic parent (`parcath=1`).
- The output above shows that kids with a catholic parent are about 2.5 percentage points more likely to graduate from high school than those without a catholic parent. Explain what we need to assume to for this to be a causal effect.
  - Given the output above, do you think the assumption in 2a. is valid?
  - Maintain the assumption in 2a. What is the (estimated) causal effect of having a Catholic parent on attending a catholic high-school?
  - Use the Wald estimator to estimate the effect of attending a Catholic high-school on high-school graduation.
  - What do you need to assume for the estimator in 2c. to be consistent? Discuss in the context of this application.
  - Explain how you can improve on your estimate in 2c. using the data above.

## Critical Values for the $F_{m,\infty}$ Distribution

## Rows denote degrees of freedom (m), and columns significance level (%)

| ##      | 10%    | 5%     | 1%     |
|---------|--------|--------|--------|
| ##      |        |        |        |
| ## 1 :  | 2.7055 | 3.8415 | 6.6349 |
| ## 2 :  | 2.3026 | 2.9957 | 4.6052 |
| ## 3 :  | 2.0838 | 2.6049 | 3.7816 |
| ## 4 :  | 1.9449 | 2.3719 | 3.3192 |
| ## 5 :  | 1.8473 | 2.2141 | 3.0173 |
| ## 6 :  | 1.7741 | 2.0986 | 2.8020 |
| ## 7 :  | 1.7167 | 2.0096 | 2.6393 |
| ## 8 :  | 1.6702 | 1.9384 | 2.5113 |
| ## 9 :  | 1.6315 | 1.8799 | 2.4073 |
| ## 10 : | 1.5987 | 1.8307 | 2.3209 |
| ## 11 : | 1.5705 | 1.7886 | 2.2477 |
| ## 12 : | 1.5458 | 1.7522 | 2.1847 |
| ## 13 : | 1.5240 | 1.7202 | 2.1299 |
| ## 14 : | 1.5046 | 1.6918 | 2.0815 |
| ## 15 : | 1.4871 | 1.6664 | 2.0385 |
| ## 16 : | 1.4714 | 1.6435 | 2.0000 |
| ## 17 : | 1.4570 | 1.6228 | 1.9652 |
| ## 18 : | 1.4439 | 1.6038 | 1.9336 |
| ## 19 : | 1.4318 | 1.5865 | 1.9048 |
| ## 20 : | 1.4206 | 1.5705 | 1.8783 |
| ## 21 : | 1.4102 | 1.5557 | 1.8539 |
| ## 22 : | 1.4006 | 1.5420 | 1.8313 |
| ## 23 : | 1.3916 | 1.5292 | 1.8104 |
| ## 24 : | 1.3832 | 1.5173 | 1.7908 |
| ## 25 : | 1.3753 | 1.5061 | 1.7726 |
| ## 26 : | 1.3678 | 1.4956 | 1.7554 |
| ## 27 : | 1.3608 | 1.4857 | 1.7394 |
| ## 28 : | 1.3541 | 1.4763 | 1.7242 |
| ## 29 : | 1.3478 | 1.4675 | 1.7099 |
| ## 30 : | 1.3419 | 1.4591 | 1.6964 |