Эконометрика — Совбак ВШЭ и РЭШ, 2020 final

Совбак ВШЭ и РЭШ | Эконометрика | 2020 | final

Рисунки пока рендерятся в тестовом режиме и могут отличаться от исходных материалов.

QUESTION 1

Part I: Female labor supply — 21 points

Harvard economist Claudia Goldin attributes much of the rise of professional women in the U.S. labor force to their ability to engage in family planning after the introduction of the birth-control pill. In developing countries, early childbearing is associated with lower education and greater dependence on husbands' earnings.

This part studies the effect of family size on female labor supply using $n = 254{,}654$ married women aged 21-35 from the 1980 U.S. Census. The data refer to calendar year 1979.

Table 1. Variables

Variable	Definition
Wife's weeks worked	Number of weeks the wife worked for pay in 1979
Husband's weeks worked	Number of weeks the husband worked for pay in 1979
Same sex	1 if the first two children have the same sex, 0 otherwise
2 boys	1 if the first two children are boys, 0 otherwise
2 girls	1 if the first two children are girls, 0 otherwise
Kids > 2	1 if the family has more than two children, 0 otherwise
Boy first	1 if the first child is a boy, 0 otherwise
Current age of mother	Mother's age in 1979
Age of mother at first birth	Mother's age when her first child was born
Black	1 if Black, 0 otherwise
Hispanic	1 if Hispanic, 0 otherwise
Other race	1 if nonwhite, non-Black, and non-Hispanic, 0 otherwise

Table 2. Child sex composition, family size, and labor supply

Robust standard errors are in parentheses. All regressions include an intercept, not reported. ** denotes 1% significance and * denotes 5% significance.

Regressor/statistic	(1) OLS: Kids > 2	(2) OLS: Kids > 2	(3) OLS: wife's weeks	(4) TSLS: wife's weeks	(5) TSLS: wife's
Instrument(s)	Same sex	2 boys, 2 girls	—	Same sex	2 boys,
Same sex	0.0694** (0.0018)	—	—	—	
2 boys	—	0.0599** (0.0026)	—	—	
2 girls	—	0.0789** (0.0026)	—	—	
Kids > 2	—	—	-8.04** (0.09)	-5.40** (1.21)	-5.16**<br
Boy first	-0.0011 (0.0019)	-0.0015 (0.0026)	-0.05 (0.08)	-0.02 (0.08)	-0.02<br
Current age of mother	0.0304** (0.0003)	0.0304** (0.0003)	1.33** (0.01)	1.25** (0.04)	1.25**<br
Age at first birth	-0.0436** (0.0003)	-0.0436** (0.0003)	-1.36** (0.17)	-1.24** (0.05)	-1.24**<br
Black	0.0680** (0.0042)	0.0680** (0.0042)	10.83** (0.19)	10.66** (0.21)	10.64**<br
Hispanic	0.1260** (0.0039)	0.1260** (0.0039)	-0.04 (0.18)	-0.38 (0.23)	-0.41<br
Other race	0.0480** (0.0044)	0.0480** (0.0044)	2.82** (0.20)	2.70** (0.21)	2.69**<br

Regressor/statistic	(1) OLS: Kids > 2	(2) OLS: Kids > 2	(3) OLS: wife's weeks	(4) TSLS: wife's weeks	(5) TSLS: wife's
N	254,654	254,654	254,654	254,654	2
First-stage F statistic	1413.0	725.9	—	—	
J statistic	—	—	—	—	

1 — 3 points

Give the best reason why the OLS estimator of the coefficient on Kids > 2 in column (3) may be biased.

2 — 3 points

Consider the hypothesis that, on average, U.S. parents want children of both sexes. Does Table 2 provide evidence for this hypothesis, against it, or neither? Explain.

3 — 6 points

Consider each proposed instrument for Kids > 2 in regression (3). Is it arguably valid? Explain.

1. **(3 points)** Whether the wife came from a large family.
2. **(3 points)** The teenage-pregnancy rate in the wife's city or town.

4 — 6 points

Using judgment and the empirical results in Table 2:

1. **(3 points)** Is Same sex a valid instrument in regression (4)?
2. **(3 points)** Are 2 boys and 2 girls a valid instrument set in regression (5)?

5 — 3 points

The estimated coefficient on Kids > 2 is more negative in OLS regression (3) than in TSLS regression (4). Give a real-world interpretation that could explain this difference.

QUESTION 2

Part II: Female labor supply, continued — 19 points

Consider the hypothetical regression

$$WifeWeeks_i = \beta_0 + \beta_1 (Kids > 2)_i + u_i,$$

estimated by TSLS using Same sex as the instrument. For this question, assume Same sex is valid in regression (4) and is independent of every control in regression (4), so that, for example,

$$E(BoyFirst \mid SameSex) = E(BoyFirst),$$

and analogously for the other controls.

1 — 7 points

1. **(3.5 points)** Explain why Same sex is a valid instrument in regression (7).

2. **(3.5 points)** Despite its validity in regression (7), why might regression (4) still be preferable?

2 — 4 points

Suppose the labor-supply effect of having a large family differs across women: the more professionally ambitious a woman is, the smaller the effect, and the most ambitious women work whether or not they have a large family. How does this affect interpretation of regressions (4) and (5)?

Use Table 2 to assess each statement.

3 — 4 points

Families with many children may be unusual because of religious or ethnic background. Therefore, regressions (4) and (5) do not estimate the effect of family size on labor supply; they only capture religious or ethnic effects. Agree or disagree, with a specific explanation.

4 — 4 points

Even if large families reduce female labor-force participation, husbands work more to compensate for the loss of the wife's earnings. Agree or disagree using Table 2.

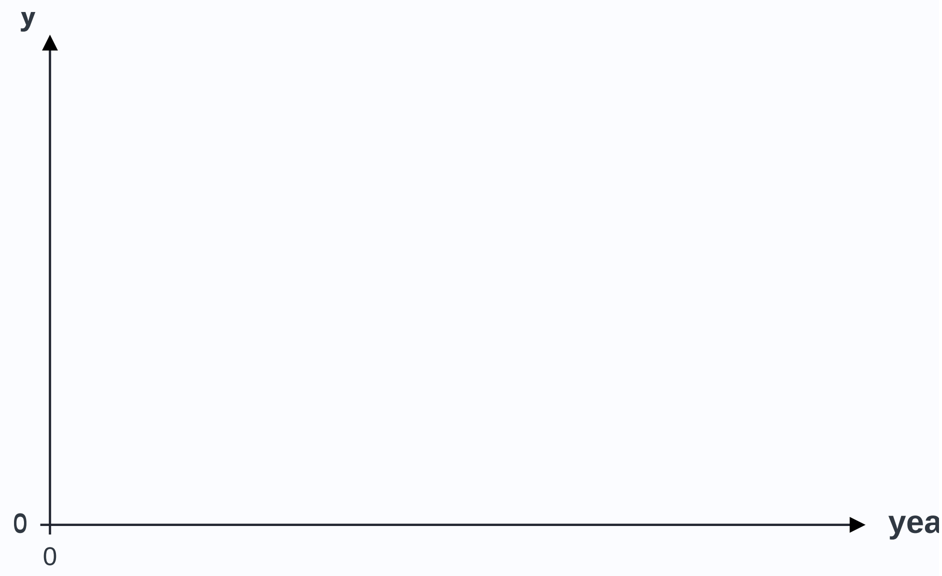
QUESTION 3

Part III: Public smoking bans and smoking habits — 19 points

Do smoking bans in bars reduce smoking? The data are a panel of 50 U.S. states observed from 2001 through 2009, for $50 \times 9 = 450$ state-year observations.

Table 3. Variable definitions and summary statistics

Variable	Definition	Mean	Std. dev.
smokingrate	Fraction of adults who currently smoke	0.242	0.044
statebarban	1 if a bar-smoking ban is in effect	0.202	0.402
staterestban	1 if a restaurant-smoking ban is in effect	0.248	0.422
stateworkban	1 if a workplace-smoking ban is in effect	0.182	0.375
all3bans	1 if bans apply in bars, restaurants, and workplaces	0.129	0.335
drinkingrate	Fraction of adults who drink	0.596	0.098
somehs	Fraction with less than a high-school diploma	0.068	0.028
hsgrad	Fraction with a high-school diploma and no further education	0.269	0.046
somecollege	Fraction with some college but no college degree	0.287	0.035
collegegrad	Fraction with a college degree	0.376	0.073
white	Fraction white	0.755	0.143
black	Fraction Black	0.098	0.098
Hispanic	Fraction Hispanic	0.081	0.092
other	Fraction neither white, Black, nor Hispanic	0.066	0.078



Number of states with a bar-smoking ban and average smoking rate, 2001-2009

Table 4. Smoking rates and public smoking bans

Dependent variable: smokingrate. Standard errors are in parentheses. Regressions (1)-(2) use 2009 only; regressions (3)-(6) use all years. Regressions (3)-(6) include state fixed effects. Regressions (4)-(6) also include year fixed effects. Standard errors are heteroskedasticity-robust in (1)-(2) and clustered by state in (3)-(6).

Regressor/statistic	(1)	(2)	(3)	(4)	
statebarban	-0.0494** (0.0097)	-0.0306** (0.0077)	-0.0187** (0.0045)	-0.0120** (0.0033)	-0.0133**<l
statebarban $\times$ drinkingrate	—	—	—	—	
staterestban	—	—	-0.0003 (0.0044)	0.0034 (0.0040)	0.0040<l
stateworkban	—	—	-0.0075* (0.0029)	-0.0032 (0.0030)	-0.0041<l
all3bans	—	—	—	—	0.0018<l
drinkingrate	—	0.229** (0.052)	0.015 (0.036)	0.014 (0.036)	0.018<
somehs	—	-0.693** (0.236)	0.209 (0.127)	0.256** (0.092)	0.256**<
somecollege	—	-0.926** (0.209)	0.005 (0.119)	-0.046 (0.079)	-0.046<
collegegrad	—	-0.642** (0.111)	-0.374** (0.067)	-0.204** (0.049)	-0.203**<
black	—	-0.027 (0.045)	-0.029 (0.037)	-0.028 (0.037)	-0.028<
Hispanic	—	-0.193** (0.044)	-0.207** (0.030)	-0.208** (0.030)	-0.208**<
other	—	0.272** (0.087)	0.169* (0.070)	0.169* (0.070)	0.166*<
N	50	50	450	450	
F: statebarban and interaction	—	—	—	—	
F: education variables	—	12.32, $p = 0.000$	32.06, $p = 0.000$	24.88, $p = 0.000$	24.97
F: race variables	—	10.63, $p = 0.000$	23.73, $p = 0.000$	23.11, $p = 0.000$	23.96

1

Using regression (2):

1. **(2 points)** Interpret the coefficient on statebarban.

2. **(2 points)** Construct a 95% confidence interval for the population coefficient.

2 — 2.5 points

Give a reason why the statebarban coefficient changes between regressions (1) and (2), including the direction of the change.

3 — 2.5 points

Give a reason why the statebarban coefficient changes between regressions (3) and (4), including the direction of the change.

4

Using regression (4):

1. **(2 points)** Test at the 5% level whether all coefficients on educational-achievement variables are zero.

2. **(2 points)** Are the estimated education-related differences in smoking rates large or small in a real-world sense?

5 — 2 points

Regression (5) includes all3bans, which equals the product of statebarban, staterestban, and stateworkban. Does regression (5) suffer from perfect multicollinearity? Explain.

6

Using regression (6):

1. **(2 points)** Compute the predicted effect of a bar-smoking ban when drinkingrate is 0.70.

2. **(2 points)** Explain precisely how to construct a 95% confidence interval for that predicted effect. A numerical interval is not required.

QUESTION 4

Part IV: Miscellaneous questions — 41 points

1 — 5 points

Consider

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_i t + u_{it},$$

where $t = 1, \ldots, T$, $i = 1, \ldots, n$, and $\alpha_i + \lambda_i t$ is an unobserved entity-specific time trend. How would you estimate β_1 ?

2. Linear probability model — 8 points

Consider

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad E(u_i \mid X_i) = 0.$$

1. **(1 point)** Show that $P(Y_i = 1 \mid X_i) = \beta_0 + \beta_1 X_i$.

2. **(2 points)** Show that

$$\text{Var}(u_i \mid X_i) = (\beta_0 + \beta_1 X_i)\big[1 - (\beta_0 + \beta_1 X_i)\big].$$

3. **(1 point)** Is u_i heteroskedastic? Explain.

4. **(4 points)** Derive the likelihood function.

3. Two instruments — 5 points

A model has one endogenous regressor X_i and two instruments Z_{1i} and Z_{2i}. There is a strong theoretical reason for Z_{1i} to be exogenous because it results from a random lottery, but Z_{1i} alone is weak. Instrument Z_{2i} is strongly relevant but less likely to be exogenous. With both instruments, the overidentification statistic is

$$J = 7.5.$$

1. **(2.5 points)** Does this suggest $E(u_i \mid Z_{1i}, Z_{2i}) \neq 0$? Explain.

2. **(2.5 points)** Does it suggest $E(u_i \mid Z_{2i}) \neq 0$? Explain.

4. Omitted controls in IV — 5 points

One student estimates

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$$

using Z_i as an instrument. Another estimates the same relationship but omits W_i.

1. **(2.5 points)** The first student says that if Z_i and W_i are correlated, the second student's IV estimator is inconsistent. Is this correct?

2. **(2.5 points)** The second student says that if the true $\beta_2 = 0$, the IV estimator remains consistent. Is this correct?

5. Forecasting an AR(1) — 5 points

Consider the stationary model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t,$$

where u_t is i.i.d. with mean zero and variance σ_u^2. Using observations $t = 1, \ldots, T$, the forecast is

$$\widehat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T.$$

1. **(1 point)** Show that

$$Y_{T+1} - \widehat{Y}_{T+1|T} = u_{T+1} - \big[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) Y_T\big].$$

2. **(1 point)** Show that u_{T+1} is independent of Y_T.

3. **(1 point)** Show that u_{T+1} is independent of $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

4. **(2 points)** Show that

$$\mathrm{Var}(Y_{T+1} - \widehat{Y}_{T+1|T}) = \sigma_u^2 + \mathrm{Var}\big[(\widehat{\beta}_0 - \beta_0) + (\widehat{\beta}_1 - \beta_1)Y_T\big].$$

6. Random walk — 6 points

Suppose

$$Y_t = Y_{t-1} + u_t, \qquad Y_0 = 0,$$

where u_t is i.i.d. with mean zero and variance σ_u^2.

1. **(2 points)** Compute the mean and variance of Y_t.

2. **(2 points)** Compute $\mathrm{Cov}(Y_t, Y_{t-k})$.

3. **(2 points)** Use the results to show that Y_t is nonstationary.

7. OLS with serially correlated regressors and errors — 7 points

Consider

$$Y_t = \beta_0 + \beta_1 X_t + u_t,$$

where

$$u_t = \phi_1 u_{t-1} + \tilde{u}_t, \qquad |\phi_1| < 1,$$

and

$$X_t = \gamma_1 X_{t-1} + e_t, \qquad |\gamma_1| < 1.$$

The innovations $\tilde{u}_t$ and e_t are i.i.d. with variances $\sigma_{\tilde{u}}^2$ and σ_e^2, and e_t is independent of $\tilde{u}_{t+i}$ for all t and i.

1. **(1 point)** Show that

$$\mathrm{Var}(u_t) = \frac{\sigma_{\tilde{u}}^2}{1 - \phi_1^2}, \qquad \mathrm{Var}(X_t) = \frac{\sigma_e^2}{1 - \gamma_1^2}.$$

2. **(1 point)** Show that

$$\mathrm{Cov}(u_t, u_{t-j}) = \phi_1^j \,\mathrm{Var}(u_t), \qquad \mathrm{Cov}(X_t, X_{t-j}) = \gamma_1^j \,\mathrm{Var}(X_t).$$

3. **(1 point)** Show that the corresponding correlations are ϕ_1^j and γ_1^j.

4. **(4 points)** Find the asymptotic variance of $\widehat{\beta}_1^{OLS}$.