Econometrics A HW3

Xinyu Liu

January 30, 2021

1.

1. Explain in your own words the conditional independence assumption and why it permits to provide a causal interpretation to your regression results. If the model above verifies the CIA, explain what this entails with respect to the matrix of regressors x above.

The description about CIA in class is given as: conditioning for observed characteristics, $\boldsymbol{Z}$, selection bias disappears. As here in the model we don't have such control variables, when applied to this situation, CIA can be interpreted as that the conditional mean of prediction error u given $\boldsymbol{x}$ is 0, i.e. $E(u|\boldsymbol{x}) = 0$. This permits causal interpretation of the regression coefficient because it is equivalent to have a randomized experiment. In the next question I will rigorously prove that the OLS estimator is a consistent estimator of the population parameter. In the case of CIA, we can say that $\boldsymbol{x}$ is strictly exogenous, and to get the unique estimation, it must hold that $E(\boldsymbol{x}'\boldsymbol{x})$ is non-singular.

2. Assume that you have some data and you have obtained the OLS estimator of $\beta, \hat{\beta}_N$. Explain the meaning of consistency, clearly state the conditions that you need for the OLS estimator to be consistent and sketch the proof of consistency.

Consistency means that when sample is large enough and provided $\boldsymbol{x}$ are exogeneous and not perfectly collinear (OLS1, OLS2), the OLS estimator will be close to the true value β. It will eventually converge in probability to the true value as $N \to \infty$. I can express the OLS estimator using sample moments based on method of moments:

$$\hat{\beta}_N = [\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i' y_i]$$

By law of large numbers, $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{x}_i \xrightarrow{p} E[\boldsymbol{x}'\boldsymbol{x}]$, and $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i' y_i \xrightarrow{p} E[\boldsymbol{x}'y]$. By continuous mapping theorem I can conclude that:

$$\begin{aligned}\hat{\beta}_N &\xrightarrow{p} E[\boldsymbol{x}'\boldsymbol{x}]^{-1}E[\boldsymbol{x}'y] \\ &= E[\boldsymbol{x}'\boldsymbol{x}]^{-1}E[\boldsymbol{x}'(\boldsymbol{x}\beta + u)] \\ &= \beta + E[\boldsymbol{x}'\boldsymbol{x}]^{-1}E[\boldsymbol{x}'u]\end{aligned}$$

Because OLS2 holds, $E[\boldsymbol{x}'\boldsymbol{x}]^{-1}$ exists. Because OLS1 holds, $E(\boldsymbol{x}'u) = E[E(\boldsymbol{x}'u|\boldsymbol{x})] = E[\boldsymbol{x}'E(u|\boldsymbol{x})] = 0$. Substitute this in and I get:

$$\hat{\beta}_N \xrightarrow{p} \beta$$

3. Without introducing additional restrictions, could you compute the exact distribution of $\hat{\beta}_N$ in finite samples? Explain how you could construct an approximation to that distribution if N, the sample size, is large enough. Derive the asymptotic distribution of $\hat{\beta}_N$ under the assumptions above (Hint: notice the assumption about the conditional variance of y and what this implies for the conditional variance of u.) Using the analogy principle, provide an estimator for the variance-covariance matrix of $\hat{\beta}_N$.

Without additional restrictions, it is impossible to give the finite sample distribution of the estimator. As for asymptotic property, this part is mainly based on CLT:

$$\begin{aligned}\hat{\beta}_N &= [\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'y_i] \\ \text{substitute in the population model} &= [\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'(x_i\beta + u_i)] \\ &= \beta + [\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i\end{aligned}$$

Rewriting it gives:

$$\begin{aligned}\sqrt{N}(\hat{\beta}_N - \beta) &= [\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i \\ &\xrightarrow{p} N(0, Q^{-1}\Omega Q^{-1} = \sigma^2 E(\boldsymbol{x}'\boldsymbol{x})^{-1})\end{aligned}$$

where $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i \xrightarrow{p} E(\boldsymbol{x}'\boldsymbol{x}) = Q$, $Var(x_i'u_i) = \sigma^2 E(\boldsymbol{x}'\boldsymbol{x}) = \Omega$. Note that here I use the fact that $Var(y|\boldsymbol{x}) = Var(u|\boldsymbol{x}) = \sigma^2$. As long as the sample size is large enough, we can use the asymptotic distribution as an approximation.

To estimate the variance-covariance matrix of $\hat{\beta}_N$, I can calcualte sample matrix counterpart of Q, Ω:

$$\begin{aligned}\hat{Q}_N &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i \\ \hat{\Omega}_N &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\hat{u_i}\end{aligned}$$

where $\hat{u_i}$ is the OLS residual. Therefore, the estimator for the variance-covariance matrix is:

$$\hat{Q}_N^{-1}\hat{\Omega}_N\hat{Q}_N^{-1} = \frac{1}{N}\sum_{i=1}^{N}\hat{u_i}^2[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}$$

4. Explain the difference between the standard error that you've derived in c) and the heteroscedasticity robust standard errors. explain what's the correct standard error associated to $\hat{\beta}_N$ and why.

The difference is that the conditional variance of the prediction error given $\boldsymbol{x}$ is no longer homogeneous. This will require a modification of the variance. In the presence of heteroscedasticity, the Ω in the variance-covariance matrix of $\hat{\beta}_N$ needs to be modified as:

$$\Omega_{White} = E(\boldsymbol{x}'\boldsymbol{x}u^2)$$

and the estimator of this variance-covariance matrix becomes:

$$\hat{Q}_N^{-1}\hat{\Omega}_{White,N}\hat{Q}_N^{-1} = [\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\hat{u_i}^2[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i]^{-1}$$

5. Explain the conditions under which the OLS estimator is BLUE and also explain the meaning of this term.

BLUE stands for best linear unbiased estimator. Best means that it has the smallest sample variance among all estimators. Linear estimators are those that can be computed as linear combinations of the data on the dependent variable. Unbiasness requires that the expectation of the estimator equals its true value.

In order to be BLUE, the model must satisfy: 1. independence, i.e. samples are draw from iid distributions; 2. conditional mean zero, i.e. $E(u|\boldsymbol{x}) = 0$; 3. non-colinearity, i.e. $E(\boldsymbol{x}'\boldsymbol{x})$ must be invertable; 4. homoscedasticity, i.e. $Var(u|\boldsymbol{x}) = \sigma^2$

2.

1. What's the mean of the estimated residuals $\hat{u}$?

By construction the in sample mean of $\hat{u} = 0$, because this regression includes a constant term. And the coefficients are estimated from the sample moment conditions:

$$
\begin{aligned}
\hat{\beta} &= [X'X]^{-1}[X'Y] \\
\Rightarrow [X'X]\hat{\beta} &= [X'Y] \\
\Rightarrow X'(X\hat{\beta} - Y) &= 0 \\
\text{remember: } \hat{u} &= Y - X\hat{\beta} \\
\Rightarrow X'\hat{u} &= 0
\end{aligned}
$$

Since the first row of X' is all 1, therefore I can get $\sum_{i=1}^{N} \hat{u}_i = 0$, which is equivalent to say its in sample mean is 0.

2. By construction, the OLS estimator is computed in a such a way that $(x_1\ x_2)'\hat{u} = 0$. Does this imply that $x'u = 0$? (where u is the population error term)

No it does not imply. The condition $E(x'u) = 0$ in the population is fundamentally untestable, because u is unobservable. By contrast, in OLS, we use this as a condition to estimate coefficients. Therefore, OLS estimate will be unbiased if and only if our assumption holds.

3. Assume that you've obtained an estimate for heroscedasticity-robust variance-covariance matrix of $\hat{\beta}$ given by the matrix below. Explain the meaning of the elements of the main diagonal and the elements off the main the diagonal.

The diagonals are the estimated variance of the regression coefficients, and the off diagonals are the estimated covariance across regression coefficients.

4. Explain the difference between point estimation and interval estimation. Using the data above, provide a confidence interval at the level of confidence of 95% for β_1 and β_2. Provide an interpretation. Are these estimates statistically significant?.

Point estimation are the regression coefficients given at the beginning of the question. They are chosen to minimize the sample sum of squared errors. However, in many cases it is not enough to just have the point estimates, especially when we would like to know more about the confidence of these estimates. Interval estimation is constructed to provide this information. In particular, for a given confidence level (say 95%), the confidence interval is constructed such that if drawing samples repeatedly, the way of interval construction will have 95% chance to contain the true parameter.
For a two-sided test and assume large enough sample size, the confidence intervals are:

$$
\beta_1 : (2 - 1.96 * \sqrt{0.3}, 2 + 1.96 * \sqrt{0.3}) = [0.93, 3.07]
$$
$$
\beta_2 : (3.5 - 1.96 * \sqrt{0.1}, 3.5 + 1.96 * \sqrt{0.1}) = [2.88, 4.12]
$$

Because 0 is not in either of these two intervals, I can conclude that these estimates are statistically significant.

5. Compute the p-value associated to β_2 and interpret its meaning.

Note that the null is $\beta_2 = 0$, therefore, $p = \frac{\hat{\beta_2}-0}{\sqrt{Var(\hat{\beta_2})}} = \frac{3.5}{\sqrt{0.1}} = 11.07$, which is much greater than 1.96. It is essentially the probability of seeing results more extreme than the current one given the null is true.

3. Computer Practise

Table 1: **OLS regression with heteroskedastic standard error**

	log(wage)	log(wage)	log(wage)	log(wage)	log(wage)	log(wage)	log(wage)
years of education	0.060***	0.031***	0.034***	0.046***	0.046***	0.048***	0.049***
	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
IQ score		0.004***	0.005***	0.005***	0.005***	0.005***	0.003***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
knowledge of world work score		0.010***	0.006***	0.006***	0.007***	0.006***	0.004*
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
age in years			0.017***	0.008	0.008	0.007	0.008
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
years of work experience				0.013***	0.013***	0.012***	0.012***
				(0.004)	(0.004)	(0.004)	(0.004)
average weekly hours					–0.005**	–0.005***	–0.006***
					(0.002)	(0.002)	(0.002)
=1 if married						0.193***	0.200***
						(0.041)	(0.039)
=1 if black							–0.148***
							(0.040)
=1 if live in south							–0.086***
							(0.028)
=1 if live in SMSA							0.174***
							(0.027)
Constant	5.973***	5.563***	5.011***	5.012***	5.216***	5.086***	5.232***
	(0.082)	(0.093)	(0.180)	(0.177)	(0.190)	(0.189)	(0.188)
No. of Observations	935	935	935	935	935	935	935
R^2	0.097	0.155	0.168	0.178	0.186	0.206	0.263
Adjusted R^2	0.096	0.152	0.164	0.174	0.181	0.200	0.255

Notes: Standard errors in parentheses.
for p<0.1, ** for p<0.05, and *** for p<0.01.

Assume that we are interested in learning the casual effect of years of education on the percentage change of wage. This table puts together 7 different regressions based on the data set that is given. The dependent variable is the log of wage, and the independent variables are individual characteristics.

Provide an interpretation of the direction of the relationship of educ and wages, as well as of the magnitude of the effect.

One additional year of education increases about 5% wage, which is significant under 1% significance level.

Provide an interpretation for the R^2 statistic and for the F-test of joint statistical significance. Do you reject the null hypothesis of that test?

The adjusted R^2 for the last regression is 0.255, which means that 25.5% of the in sample variation of *lwage* is explained by the in sample variation of independent variables. This is a measure of prediction, not causality. F test is a joint test of all the regression coefficients. , for F test, the last regression gives $F(10, 924) = 38.79$, which rejects the null that all coefficients are 0 under 1% significance level.

Test the hypothesis that black, south, urban are all equal to zero using an F-test.

First of all the fact that their t-tests are all significant makes me believe that the F test will also be rejected. The result is that $F(3, 924) = 24.21$, which indeed rejects the null under 1% significance level.

Test the hypothesis that the coefficient of education is 0.5 **using a two-tailed t-test,** $\alpha = 0.05$**.**

When there is only one coefficient for test, the F statistics is the squared t statistics, I get $F = 3326.77$, the t statistics is therefore -57.68 (note that their is a negative sign). This is substantially large and the null $H_0 : \beta = 0.5$ can be rejected. For the one tailed test with $H_1 : \beta > 0.5$, the 5% cut-off test statistics is 1.69. Because $-57 << 1.69$. we can not reject the null.

Use the command predict to obtain the predicted values for lwage. Also obtain the residuals

See the code for details.

Use the command estat hettest to test for heteroskedasticity in the residuals. What do you conclude from this test.

I first run a regression with homoskedasticity and then conduct the test (H_0: Constant variance). I get $\chi^2_1 = 1.76$, with $p = 0.18$. Therefore, we fail to reject the null hypothesis.

Introduce conditions in your regression: compute again the regression in column 7 in two different cases: for married men and for unmarried men (to do this, use the command if at the end of your regression, before the ",r" option). What do you observe?

Table 2: **OLS regression with heteroskedastic standard error**

	Married	Unmerried
years of education	0.052***	0.032
	(0.008)	(0.029)
IQ score	0.003***	0.004
	(0.001)	(0.003)
knowledge of world work score	0.004*	0.002
	(0.002)	(0.007)
age in years	0.008	0.005
	(0.006)	(0.016)
years of work experience	0.013***	0.008
	(0.004)	(0.013)
average weekly hours	–0.006***	–0.009*
	(0.002)	(0.005)
=1 if black	–0.131***	–0.244**
	(0.044)	(0.105)
=1 if live in south	–0.086***	–0.113
	(0.030)	(0.082)
=1 if live in SMSA	0.164***	0.255***
	(0.028)	(0.094)
Constant	5.363***	5.683***
	(0.200)	(0.505)
No. of Observations	835	100
R^2	0.249	0.251
Adjusted R^2	0.241	0.176

Notes: Standard errors in parentheses.
for p<0.1, ** for p<0.05, and *** for p<0.01.

It seems that under 1% significance level, the positive effect of education on lwage for married men is significant, However, we cannot reject the null that for unmarried men years of education is insignificant.

But this may have to do with the size of these two subsamples, noting that there are much more married men therefore the estimation is more precise.

Use the binscatter command to produce a plot of lwage versus education, controlling by all the remaining variables. Plot the scatter plot of lwage versus education. Also compute the binscatter of the same variables. Finally, plot a new binscatter, controlling this time for all the regressors in column 7.

Below I show the scatter plot, the binscatter plot without controls, and the binscatter plot with controls. It seems that scatter plot is best at capturing the variation of the independent variables. Then the binscatter without control gives simple prediction of the dependent variable, as it nonparametrically estimate the conditional expectations. The one with control and more bins is still in support of a positive relations between education and lwage, and it is not clear whether conditional variance is constant.

Figure 1

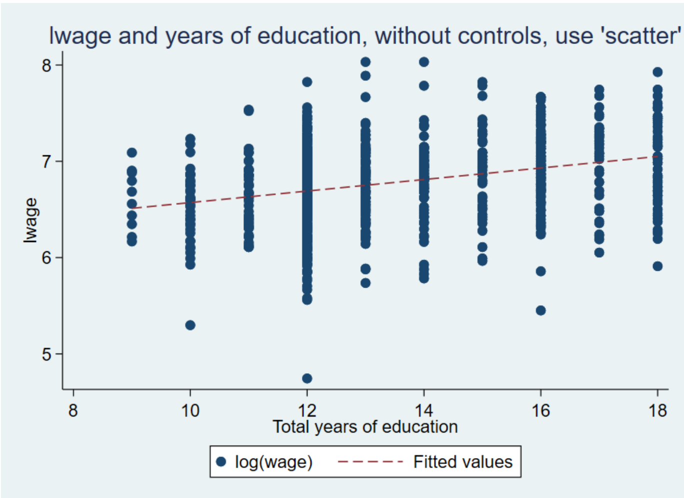


Figure 2

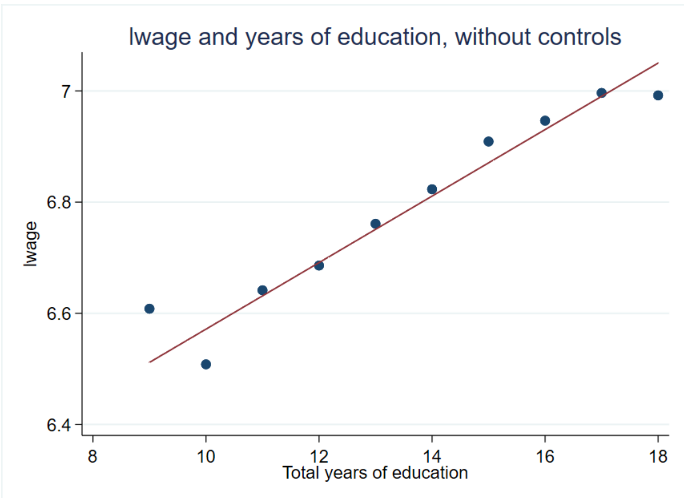
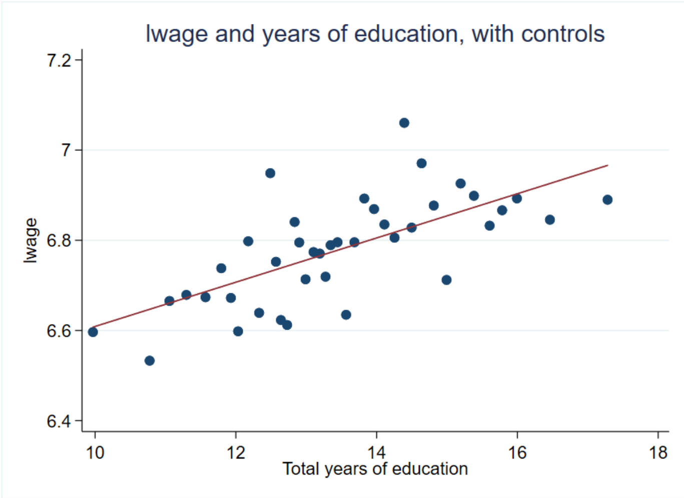
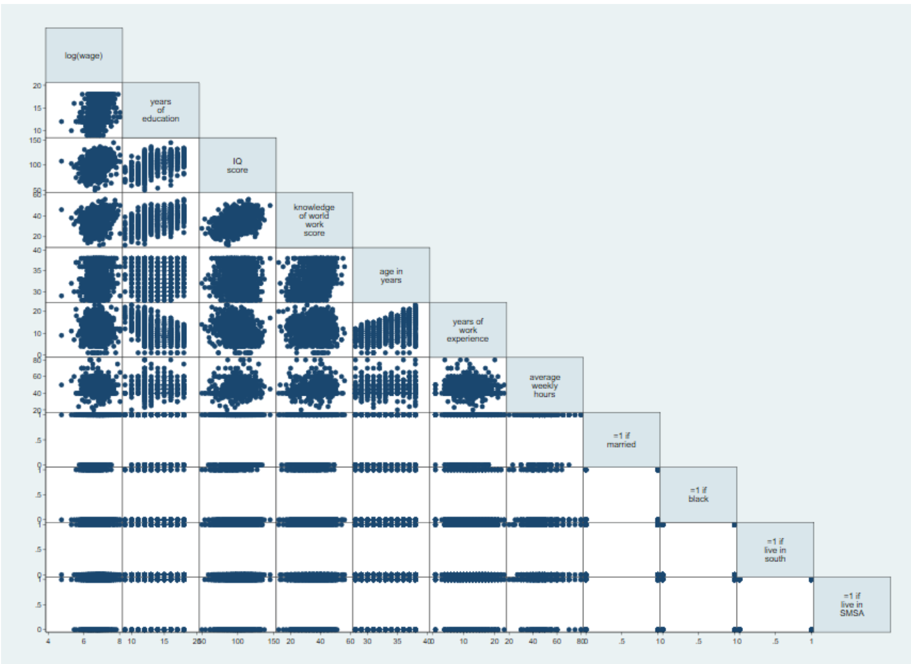


Figure 3



Use the commands pwcorr and graphic matrix to compute the correlation matrix of the regressors in column 7.

Figure 4



Apparently this plot is quite messy in that many of its variables are either discrete or categorical. Roughly speaking the most apparent relations are the 'IQ' and 'kww' score pair, which makes sense as they are closely related.

Appendix

```
capture log close
capture drop _all

pwd
cd "D:\INSEAD\Course\P3\Econometrics A\Econometrics-A-2021\HW3"
set logtype text
log using hw3.txt, replace

use NLS80, clear
eststo clear

eststo: regress lwage educ, r
eststo: regress lwage educ iq kww, r
eststo: regress lwage educ iq kww age, r
eststo: regress lwage educ iq kww age exper, r
eststo: regress lwage educ iq kww age exper hours, r
eststo: regress lwage educ iq kww age exper hours married, r
eststo: regress lwage educ iq kww age exper hours married black south urban, r

*Latex output

esttab using q4_tab1e.tex, replace fragment cells(b(star fmt(3)) se(par fmt(3))) collabels(none) stats(N r2 r2_a, fmt(%9.0f 3) label("No. of Observations" "\$R^2\$" "Adjusted \$R^2\$")) nonumber msign(--) nonotes star(* 0.10 ** 0.05 *** 0.01) label title("OLS regression with heteroskedastic standard error") prehead("\begin{center}" "\begin{threeparttable}\caption{\textbf{@title}}\label{q3tab1}" "\begin{tabular}{l*{@M}{cc}}" "\toprule") posthead(\hline) prefoot(\hline) postfoot("\bottomrule" "\end{tabular}" "\begin{tablenote}""\small{\textit{Notes:} {Standard errors in parentheses.\\* for p$<$0.1, ** for p$<$0.05, and *** for p$<$0.01.}" "\end{tablenote}" "\end{threeparttable}""\end{center}")


test black=south=urban=0

test educ=0.5

predict wage_predicted

predict residual, residual


regress lwage educ iq kww age exper hours married black south urban
estat hettest

eststo clear
eststo: regress lwage educ iq kww age exper hours black south urban if (married==1), r
eststo: regress lwage educ iq kww age exper hours black south urban if (married==0), r

#delimit
esttab using q4_tab1e2.tex, replace fragment cells(b(star fmt(3)) se(par fmt(3)))
 collabels(none) stats(N r2 r2_a, fmt(%9.0f 3) label("No. of Observations" "\$R^2\$" "Adjusted \$R^2\$"))
 nonumber msign(--) nonotes star(* 0.10 ** 0.05 *** 0.01) label  title("OLS regression with heteroskedastic standard error")
 mlabels("Married" "Unmerried",span prefix(\multicolumn{@span}{c}{) suffix(}))
 prehead("\begin{center}" "\begin{threeparttable}\caption{\textbf{@title}}\label{q3tab1}" "\begin{tabular}{l*{@M}{cc}}" "\toprule") posthead(\hline) prefoot(\hline) postfoot("\bottomrule" "\end{tabular}" "\begin{tablenote}""\small{\textit{Notes:} {Standard errors in parentheses.\\* for p$<$0.1, ** for p$<$0.05, and *** for p$<$0.01.}" "\end{tablenote}" "\end{threeparttable}""\end{center}")
;
twoway (scatter lwage educ) (lfit lwage educ, lpatt(dash)), title("lwage and years of education, without controls, use ’scatter’") scheme(s2color) plotregion(style(none)) ylabel(,angle(0)) xtitle("Total years of education") ytitle("lwage")
graph export "D:\INSEAD\Course\P3\Econometrics A\Econometrics-A-2021\HW3\q4p1.png", as(png) replace
binscatter lwage educ,nq(40) title("lwage and years of education, without controls") scheme(s2color) plotregion(style(none)) ylabel(,angle(0)) xtitle("Total years of education") ytitle("lwage")
graph export "D:\INSEAD\Course\P3\Econometrics A\Econometrics-A-2021\HW3\q4p2.png", as(png) replace
```

```
binscatter lwage educ,nq(40) title("lwage and years of education, with controls")
    controls(iq kww age exper hours married black south urban) scheme(s2color)
    plotregion(style(none)) ylabel(,angle(0)) xtitle("Total years of education") ytitle
    ("lwage")
graph export "D:\INSEAD\Course\P3\Econometrics A\Econometrics-A-2021\HW3\q4p3.png", as(
    png) replace

pwcorr lwage educ iq kww age exper hours married black south urban
graph matrix lwage educ iq kww age exper hours married black south urban, half
graph export "D:\INSEAD\Course\P3\Econometrics A\Econometrics-A-2021\HW3\q4p4.png", as(
    png) replace
```