P218 Econometrics I

TA Session 2

Gabriel Simões Gaspar

London Business School

Fall 2022

Roadmap

Gauss-Markov Theorem

Gauss-Markov Theorem

In lecture you saw the famous Gauss-Markov Theorem. Consider the linear regression model described by

$$
\begin{aligned}
Y &= X\beta + \varepsilon \\
\mathbb{E}[\varepsilon \mid X] &= 0 \\
\mathrm{Var}(\varepsilon \mid X) &= \mathbb{E}[\varepsilon\varepsilon' \mid X] = \sigma^2\Sigma < \infty
\end{aligned}
$$

where Y is an $n \times 1$ random vector, X is an $n \times m$ full-rank matrix of regressors such that $m < n$ and ε is an $n \times 1$ vector of regression errors.

If we add the assumption that the variance-covariance matrix $\Sigma = I_n$, then you saw that the OLS estimator has **minimum variance** among the estimators that (**conditional** on regressors) are **linear** and **unbiased**. This is famously known as the best linear conditionally unbiased estimator, i.e. BLUE.

Gauss-Markov Assumptions

We will summarise the conditions that make the OLS estimator BLUE as follows:

$$
\begin{aligned}
&(GM0) \qquad Y = X\beta + \varepsilon \\
&(GM1) \qquad \operatorname{rank}(X) = m \\
&(GM2) \qquad \mathbb{E}[Y \mid X] = X\beta \iff \mathbb{E}[\varepsilon \mid X] = 0 \\
&(GM3) \qquad \operatorname{Var}(Y \mid X) = \operatorname{Var}(\varepsilon \mid X) = \sigma^2 I
\end{aligned}
$$

These are enough for us to prove the Gauss-Markov theorem. Let's quickly go over them.

Unbiasedness

To show that the OLS is conditionally unbiased, simply note that

$$\hat{\beta}_{OLS} \overset{GM1}{=} (X'X)^{-1}X'Y$$
$$\hat{\beta}_{OLS} \overset{GM0}{=} \beta + (X'X)^{-1}X'\varepsilon$$

We can take the expectation on both sides of the equation above noting that $\mathbb{E}[\beta \mid X] = \beta$

$$\mathbb{E}[\hat{\beta}_{OLS} \mid X] = \beta + \mathbb{E}[(X'X)^{-1}X'(\varepsilon) \mid X]$$
$$\mathbb{E}[\hat{\beta}_{OLS} \mid X] = \beta + (X'X)^{-1}X'\mathbb{E}[\varepsilon \mid X]$$

which finally shows us that:

$$\mathbb{E}[\hat{\beta}_{OLS} \mid X] \overset{GM2}{=} \beta$$

Minimum Variance

Imagine now a general-form **unbiased linear** estimator $\tilde{\beta}$ for β that follows $(GM0)$ - $(GM3)$ defined as follows:

$$\tilde{\beta} = AY \overset{GM0}{=} A(X\beta + \varepsilon) \implies \mathbb{E}[\tilde{\beta} \mid X] \overset{GM2}{=} AX\beta = \beta \implies AX = I_n$$

The conditional variance of this estimator is simply:

$$\text{Var}(\tilde{\beta} \mid X) = A\,\text{Var}(Y \mid X)A' \overset{GM3}{=} \sigma^2 AA'$$

We can decompose the general matrix A by adding and subtracting another matrix:

$$A = A - \underbrace{(X'X)^{-1}X' + (X'X)^{-1}X'}_{=0} = W + (X'X)^{-1}X'$$

where we have defined $W \equiv A - (X'X)^{-1}X'$.

Minimum Variance

Note that:

$$W \equiv A - (X'X)^{-1}X' \implies WX = \underbrace{AX}_{=I_n} - \underbrace{(X'X)^{-1}X'X}_{I_n} = 0$$

Plug in the decomposed A we derived in the previous slide in the conditional variance:

$$\begin{aligned}
\text{Var}(\tilde{\beta} \mid X) &= \sigma^2 AA' \\
&= \sigma^2 (W + (X'X)^{-1}X')(W' + X(X'X)^{-1}) \\
&= \sigma^2 [WW' + \underbrace{WX}_{=0}(X'X)^{-1} + (X'X)^{-1} \underbrace{X'W'}_{(WX)'=0} + (X'X)^{-1}] \\
&= \sigma^2 WW' + \sigma^2 (X'X)^{-1} \\
&> \sigma^2 (X'X)^{-1} \\
&= \text{Var}(\hat{\beta}_{OLS} \mid X)
\end{aligned}$$

General Var-Cov Matrix

Let's now make a slight change and consider a model such that:

$$(GM3') \qquad \text{Var}(\varepsilon \mid X) = \mathbb{E}[\varepsilon\varepsilon' \mid X] = \Omega < \infty$$

so we are not (necessarily) considering the homoskedastic case anymore. What is the variance of the OLS estimator in this case?

$$\text{Var}(\hat{\beta}_{OLS} \mid X) \overset{GM1}{=} \text{Var}((X'X)^{-1}X'Y \mid X)$$

$$\text{Var}(\hat{\beta}_{OLS} \mid X) \overset{GM0}{=} (X'X)^{-1}X' \text{Var}(\varepsilon \mid X) X(X'X)^{-1}$$

$$\text{Var}(\hat{\beta}_{OLS} \mid X) \overset{GM3'}{=} (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Is OLS best in this case? We will see this later in the course.

Question 1

Part (a)

We have that:

$$Y = -X^2 + (2+\varepsilon)X$$

The effect of a marginal change of X in Y is simply:

$$\frac{\partial Y}{\partial X} = -2X + (2+\varepsilon)$$

These effects are heterogeneous. Farms with good soil will have $\varepsilon = 1$, and so the marginal effect is $-2X + 3$. Farms with bad soil have $\varepsilon = -1$ and so the marginal effect is $-2X + 1$.

Part (b)

To find the *ACE* of X on Y as a function of X we have to integrate out ε. This can be done in two steps. First, find the marginal density of X:

$$
\begin{aligned}
f_X(x) &= \int_{-1}^{1} f_{X\varepsilon}(x,z)dz \\
&= \int_{-1}^{1} \frac{1+x-z}{8} dz \\
&= \varepsilon\left(\frac{1+x}{8}\right) - \frac{z^2}{16}\Bigg|_{z=-1}^{z=1} \\
&= \frac{1+x}{4}
\end{aligned}
$$

Part (b)

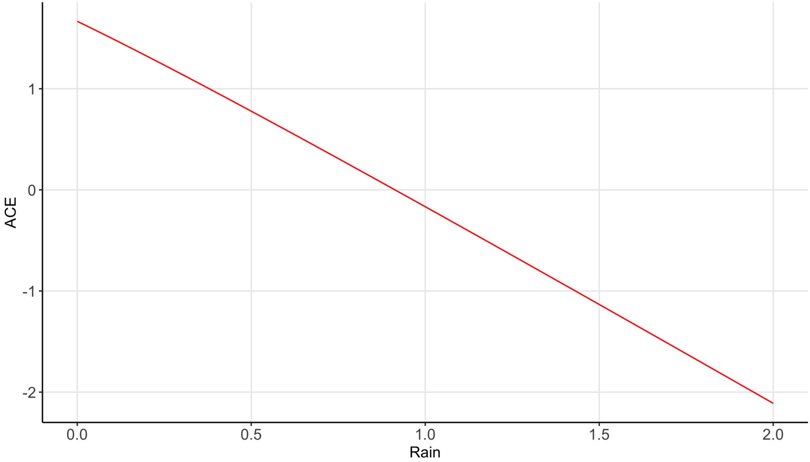
Now that we have the marginal density of X, we can calculate the conditional density:

$$\begin{aligned} f_{\varepsilon|X}(z \mid x) &= \frac{f_{X\varepsilon}(x, z)}{f_X(x)} \\ &= \frac{1 + x - z}{2 + 2x} \end{aligned}$$

Finally, the ACE is then:

$$\begin{aligned} ACE(X) &= \int_{-1}^{1} (-2x + 2 + z)\left(\frac{1 + x - z}{2 + 2x}\right) dz \\ &= \frac{5 - 6X^2}{3X + 3} \end{aligned}$$

Part (b)



Part (c)

To find the unconditional expectation just proceed as usual for a continuous variable:

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[-X^2 + (2+\varepsilon)X] \\
&= \int_0^2 \int_{-1}^1 (-x^2 + (2+z)x) f_{X\varepsilon}(x,z) dz dx \\
&= \int_0^2 \int_{-1}^1 (-x^2 + (2+z)x) \frac{1+x-z}{8} dz dx \\
&= \frac{1}{2}
\end{aligned}$$

Part (c)

And the conditional mean:

$$\begin{aligned}
\mathbb{E}[Y \mid X] &= \mathbb{E}[-X^2 + (2+\varepsilon)X \mid X] \\
&= -X^2 + X\mathbb{E}[(2+\varepsilon) \mid X] \\
&= -X^2 + X\int_{-1}^{1}(2+z)f_{\varepsilon|X}(z \mid x)dz \\
&= -X^2 + X\int_{-1}^{1}(2+z)\left(\frac{1+X-z}{2+2X}\right)dz \\
&= -X^2 + X\frac{6X+5}{3X+3}
\end{aligned}$$

Part (c)

In case you were skeptical about the Law of Iterated Expectations:

$$
\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y \mid X]] \\
&= \int_0^2 f_X(x)\mathbb{E}[Y \mid X]dx \\
&= \int_0^2 \left(\frac{1+x}{4}\right)\left(-X^2 + X\frac{6X+5}{3X+3}\right)dx \\
&= \frac{1}{2}
\end{aligned}
$$

Part (d)

Note that

$$\frac{\partial}{\partial X}\mathbb{E}[Y \mid X] = -2X + \frac{6X^2 + 12X + 5}{3(X+1)^2} > \frac{5 - 6x^2}{3(x+1)} = ACE(X)$$

We can see that:

$$\begin{aligned} \frac{\partial}{\partial X}\mathbb{E}[Y \mid X] - ACE(X) &= -2X + \frac{6X^2 + 12X + 5}{3(X+1)^2} - \frac{5 - 6x^2}{3(x+1)} \\ &= \frac{X}{3(x+1)^2} > 0 \end{aligned}$$

This means that the slope of the regression overstates the causal effect of rain X. Why does this happen? What is the CEF capturing?

Part (e)

Let's do this by parts. We can calculate the variance of X as:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \int_0^2 x^2 f_X(x)dx - \left(\int_0^2 x f_X(x)dx\right)^2 \\
&= \int_0^2 x^2\left(\frac{1+x}{4}\right)dx - \left(\int_0^2 x\left(\frac{1+x}{4}\right)dx\right)^2 \\
&= \frac{5}{3} - \left(\frac{7}{6}\right)^2 \\
&= \frac{11}{36}
\end{aligned}$$

Part (e)

And the co-variance of X and Y is:

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[X^2(-X + 2 + \varepsilon)] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \int_0^2 \int_{-1}^1 x^2(-x + 2 + z) f_{X\varepsilon}(x, z) dz dx - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \int_0^2 \int_{-1}^1 x^2(-x + 2 + z) \left(\frac{1 + x - z}{8} \right) dz dx - \frac{7}{12} \\
&= \frac{23}{45} - \frac{7}{12} \\
&= -\frac{13}{180}
\end{aligned}
$$

Part (e)

Recall that the BLP of Y given X is $Y = \beta_0 + \beta_1 X$ where:

$$\beta_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = -\frac{13}{55}$$
$$\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X] = \frac{128}{165}$$

So we have:

$$BLP(Y \mid X) = \frac{128}{165} - \frac{13}{55} X$$

If this is true and causal, an increase in rain X will increase production Y by $-\frac{13}{55}$

Question 2

Part (a)

You are told the the estimator is:

$$\tilde{\beta} = \frac{\bar{Y}}{\bar{X}} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} = \frac{\mathbf{1}'Y}{\mathbf{1}'X}$$

where $\mathbf{1}$ is a $n \times 1$ column vector of ones. The expression above is linear in Y. We can take the expectation on both sides to show that this estimator is unbiased:

$$\begin{aligned} \mathbb{E}[\tilde{\beta} \mid X] &\overset{GM0}{=} \mathbb{E}\left[\frac{\mathbf{1}'X\beta}{\mathbf{1}'X} \mid X\right] + \mathbb{E}\left[\frac{\mathbf{1}'\varepsilon}{\mathbf{1}'X} \mid X\right] \\ &\overset{GM3}{=} \beta \end{aligned}$$

Part (a)

To get the conditional variance we can look at

$$\begin{aligned}\mathrm{Var}(\tilde{\beta} \mid X) &= \mathrm{Var}\left(\frac{\mathbf{1}'Y}{\mathbf{1}'X} \mid X\right) \\ &= \frac{1}{(\mathbf{1}'X)^2}\mathrm{Var}(\mathbf{1}'Y \mid X) \\ &= \frac{\mathbf{1}'\mathrm{Var}(Y \mid X)\mathbf{1}}{(\mathbf{1}'X)^2}\end{aligned}$$

Given the assumptions we have made, we have:

$$\mathrm{Var}(\tilde{\beta} \mid X) \overset{GM3}{=} \sigma^2 \frac{n}{(\mathbf{1}'X)^2}$$

Part (a)

Remember that the conditional variance of the OLS estimator under our assumptions $GM0$ - $GM3$ is

$$\text{Var}(\hat{\beta}_{OLS} \mid X) = \sigma^2 (X'X)^{-1}$$

But since X in this case is a $n \times 1$ vector and we know that for a general $X_{n\times m}$ the square symmetric matrix $X'X$ is $m \times m$, this means that $X'X = \sum_{i=1}^{n} x_i^2$, which is a scalar. So we can conclude that:

$$\text{Var}(\hat{\beta}_{OLS} \mid X) = \sigma^2 \frac{1}{X'X} \leq \sigma^2 \frac{n}{(\mathbf{1}'X)^2} = \text{Var}(\tilde{\beta} \mid X)$$

Part (a)

Why is that last part true?

$$
\begin{aligned}
X'X &= \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n\bar{x} \\
&> n\bar{x} \\
&= \frac{(\mathbf{1}'X)^2}{n}
\end{aligned}
$$

Check Appendix A of Wooldridge if you need a refresher. Does this make sense given our assumptions and what we know about the Gauss-Markov Theorem?

Part (b)

You are now told to use the first $m < n$ observations and compute the OLS estimator. Let X_m and Y_m the the vectors of the first m observations. Then

$$\hat{\beta}_{m,OLS} \stackrel{GM1}{=} (X_m' X_m)^{-1} X_m' Y_m$$

which is linear in Y_m. Since $GM0$ - $GM3$ are met in this case, we know that it is linearly unbiased (but you have to prove in the PS!). To test if it has minimum variance note that:

$$\text{Var}(\hat{\beta}_{m,OLS} \mid X) = \sigma^2 \frac{1}{(X_m' X_m)} \geq \sigma^2 \frac{1}{(X'X)} = \text{Var}(\hat{\beta}_{OLS} \mid X)$$

since $X_m' X_m = \sum_{i=1}^{m} x_i^2 \leq \sum_{i=1}^{n} x_i^2 = X'X$ for $m < n$.

Part (c)

Can we find another estimator with a smaller conditional variance? We can find an infinite amount of them. The Gauss-Markov Theorem says that the OLS estimator is BLUE, but we can relax the assumptions to find something with a smaller variance. For example, for $k \in \mathbb{N}$, consider the estimator

$$\check{\beta} = k \implies \mathrm{Var}(\check{\beta}) = 0 < \mathrm{Var}(\hat{\beta}_{OLS})$$

This estimator is not unbiased though.

Question 3

Part (a)

We can still compute the OLS estimator as always since our assumptions $GM0$ - $GM3$ still hold. Since we only have one observation (y, x), we can plug this into our usual formula.

$$\hat{\beta}_{OLS} \overset{GM0}{=} \frac{xy}{x^2} = \frac{y}{x}$$

The unconditional mean of this estimator equals β. What assumption do we need in order to say that?

$$\mathbb{E}[\hat{\beta}_{OLS}] = \beta$$

You need to use LIE to prove this.

Law of Total Variance

To find the unconditional variance of this estimator, let's look at the Law of Total Variance. We know that the unconditional variance for a random variable Y is given by:

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \overset{LIE}{=} \mathbb{E}\Big[\mathbb{E}[Y^2 \mid X]\Big] - \mathbb{E}\Big[\mathbb{E}[Y \mid X]\Big]^2$$

But note that:

$$\begin{aligned}\mathbb{E}[Y^2] \overset{LIE}{=} \mathbb{E}\Big[\mathbb{E}[Y^2 \mid X]\Big] &= \mathbb{E}\Big[\text{Var}(Y \mid X) + \mathbb{E}[Y \mid X]^2\Big] \\ &= \mathbb{E}\Big[\text{Var}(Y \mid X)\Big] + \mathbb{E}\Big[\mathbb{E}[Y \mid X]^2\Big]\end{aligned}$$

since the expectation of the sum is the sum of the expectation.

Law of Total Variance

We can plug this back into the formula for the unconditional variance:

$$\begin{aligned}\text{Var}(Y) &= \mathbb{E}\Bigg[\text{Var}(Y \mid X)\Bigg] + \overbrace{\mathbb{E}\Bigg[\mathbb{E}[Y \mid X]^2\Bigg] - \mathbb{E}\Bigg[\mathbb{E}[Y \mid X]\Bigg]^2}^{=\text{Var}\left(\mathbb{E}[Y|X]\right)} \\ &= \mathbb{E}\Bigg[\text{Var}(Y \mid X)\Bigg] + \text{Var}\Bigg(\mathbb{E}[Y \mid X]\Bigg)\end{aligned}$$

Part (a)

Back to our problem. To find the unconditional variance of our OLS estimator, we can apply the Law of Total Variance:

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}_{OLS}) &= \mathbb{E}\left[\mathrm{Var}(\hat{\beta}_{OLS} \mid X)\right] + \mathrm{Var}\left(\mathbb{E}[\hat{\beta}_{OLS} \mid X]\right) \\
&= \mathbb{E}\left[\sigma^2 (X'X)^{-1}\right] + \underbrace{\mathrm{Var}(\beta)}_{=0} \\
&= \sigma^2 \mathbb{E}\left[\frac{1}{x^2}\right] \\
&= \Pr(x = 1/5) \times \left(\frac{5}{1}\right)^2 + \Pr(x = 7/5)\left(\frac{5}{7}\right)^2 \\
&= \frac{625}{49}
\end{aligned}
$$

Part (b)

We are now considering the estimator $\tilde{\beta} = xy$. This is an unconditionally unbiased estimator for β

$$\begin{aligned}\mathbb{E}[\tilde{\beta}] &\overset{GM0}{=} \mathbb{E}[xy \mid x] \\ &\overset{LIE}{=} \mathbb{E}[x^2\beta] + \mathbb{E}[x\varepsilon] \\ &\overset{GM2}{=} \beta\mathbb{E}[x^2]\end{aligned}$$

But $\mathbb{E}[x^2] = \Pr(x = 1/5) \times \left(\frac{1}{5}\right)^2 + \Pr(x = 7/5)\left(\frac{7}{5}\right)^2 = 1$. This means that $\tilde{\beta}$ is conditionally unbiased for β.

Part (b)

We can just plug in the estimator to find the variance:

$$\begin{aligned}
\text{Var}(\tilde{\beta}) &= \text{Var}(xy) \\
&= \text{Var}(\beta x^2 + x\varepsilon) \\
&= \beta^2 \,\text{Var}(x^2) + \text{Var}(x\varepsilon) \\
&= \beta^2 \left(\mathbb{E}[x^4] - \mathbb{E}[x^2]^2 \right) + \underbrace{\mathbb{E}[x^2\varepsilon^2]}_{\mathbb{E}[x^2]\mathbb{E}[\varepsilon^2]} - \underbrace{\mathbb{E}[x\varepsilon]^2}_{(\mathbb{E}[x]\mathbb{E}[\varepsilon])^2=0} \\
&= \beta^2 \left(\frac{1}{2} \times \frac{1}{625} + \frac{1}{2} \times \frac{2401}{625} - 1 \right) + \sigma^2 \mathbb{E}[x^2] \\
&= \beta^2 \frac{576}{625} + 1
\end{aligned}$$

Part (c)

If the true $\beta = 0$ our variances will be:

$$\text{Var}(\tilde{\beta}) = 1 < \frac{625}{49} = \text{Var}(\hat{\beta}_{OLS})$$

The OLS estimator does not have minimum variance in this case. What is going on? Does the Gauss-Markov Theorem not hold?

Question 4

Question 4

This question is a lot of algebra and we won't go over all of it during this session. But keep in mind these formulas

$$\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}$$
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Note that

$$\widehat{\mathrm{Var}}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = s^2 \begin{bmatrix} n & n\bar{x} \\ n\bar{x} \sum_{i=1}^{n} x_i^2 & \end{bmatrix}^{-1}$$
$$= \frac{SSR}{(n-2)\sum_{i=1}^{n}(x_i - \bar{x})^2} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

Question 4

Moreover,

$$
\begin{aligned}
R^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{x}_i - \bar{x})^2} \\
&= \hat{\beta} \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= 1 - \frac{SSR/(n-2)}{\sum_{i=1}^n (y_i - \bar{y})^2/(n-1)} \\
SSR &= (1 - R^2) \sum_{i=1}^n (y_i - \bar{y})^2
\end{aligned}
$$

Question 5

Part (a)

Table: Results

	Dependent variable:
	Log(wage)
Education	0.144***
	(0.013)
Experience	0.026
	(0.028)
Experience^2	0.001
	(0.001)
Constant	7.385***
	(0.280)
Observations	1,500
R^2	0.085
Adjusted R^2	0.083
Residual Std. Error	0.812 (df = 1496)
F Statistic	46.277*** (df = 3; 1496)
Note:	*p<0.1; **p<0.05; ***p<0.01

Part (b)

Table: Results

	Dependent variable:
	Log(wage)
Education	−0.000
	(0.200)
Experience	−0.000
	(0.064)
Fitted Values	1.000
	(1.382)
Constant	−0.000
	(10.078)
Observations	1,500
R^2	0.085
Adjusted R^2	0.083
Residual Std. Error	0.812 (df = 1496)
F Statistic	46.277*** (df = 3; 1496)
Note:	*p<0.1; **p<0.05; ***p<0.01

Part (c)

Table: Results

	Dependent variable:
	Residual of Log(wage) on Education and Experience^2
Residual of Experience on Education and Experience^2	0.026
	(0.028)
Constant	-0.000
	(0.021)
Observations	1,500
R^2	0.001
Adjusted R^2	-0.0001
Residual Std. Error	0.812 (df = 1498)
F Statistic	0.869 (df = 1; 1498)
Note:	*p$<$0.1; **p$<$0.05; ***p$<$0.01

Part (d)

Table: Results

	Dependent variable:
	Log(wage)
Residual of Experience on Education and Experience^2	0.026
	(0.030)
Constant	9.645***
	(0.022)
Observations	1,500
R^2	0.001
Adjusted R^2	-0.0001
Residual Std. Error	0.848 (df = 1498)
F Statistic	0.796 (df = 1; 1498)
Note:	*p<0.1; **p<0.05; ***p<0.01