**Topic 2: Asymptotic Properties of Various Regression Estimators**

- Our results to date apply for any *finite* sample size \((n)\).
- In more general models we often can’t obtain *exact* results for estimators’ properties.
- In this case, we might consider their properties as \(n \to \infty\).
- A way of “approximating” results.
- Also of interest in own right – inferential procedures should “work well” when we have lots of data

**Definition:** An estimator, \(\hat{\theta}\), for \(\theta\), is said to be (weakly) **consistent** if

\[
\lim_{n \to \infty} \{Pr. \left[ \left| \hat{\theta}_n - \theta \right| < \epsilon \right] \} = 1.
\]

Note: A *sufficient* condition for this to hold is that *both*

(i) \(\text{Bias}(\hat{\theta}_n) \to 0\) ; as \(n \to \infty\).

(ii) \(V(\hat{\theta}_n) \to 0\) ; as \(n \to \infty\).

We call this “Mean Square Consistency”. (Often useful for checking.)

If \(\hat{\theta}\) is weakly consistent for \(\theta\), we say that “the probability limit of \(\hat{\theta}\) equals \(\theta\).

We denote this by using “\(plim\)” operator, and we write

\[
plim(\hat{\theta}_n) = \theta \quad \text{or,} \quad \hat{\theta}_n \stackrel{p}{\to} \theta
\]

**Example** \(x_i \sim [\mu, \sigma^2]\) \((i.i.d)\)

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
E[\bar{x}] = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = \frac{1}{n} (n\mu) = \mu \quad \text{(unbiased, for all } n)\]

\[
\text{var.} [\bar{x}] = \frac{1}{n^2} \text{var.} [\sum_{i=1}^{n} x_i] = \frac{1}{n^2} \sum_{i=1}^{n} \text{var.} (x_i)
\]

\[
= \frac{1}{n^2} (n\sigma^2) = \sigma^2 / n
\]

So, \(\bar{x}\) is an unbiased estimator of \(\mu\), and \(\lim_{n \to \infty} \{\text{var.} [\bar{x}]\} = 0.\)
This implies that $\bar{x}$ is both a mean-square consistent, and weakly consistent estimator of $\mu$.

Note:

- If an estimator is inconsistent, then it is a pretty useless estimator!

**Slutsky’s Theorem**

Let $\text{plim}(\hat{\theta}_n) = c$, and let $f(.)$ be any continuous function.

Then, $\text{plim}[f(\hat{\theta}_n)] = f(c)$.

For example –

\[
\text{plim} \left( \frac{1}{\hat{\beta}} \right) = \frac{1}{c} \quad ; \quad \text{scalars}
\]

\[
\text{plim}(e^{\hat{\theta}}) = e^c \quad ; \quad \text{vectors}
\]

\[
\text{plim}(\hat{\theta}^{-1}) = C^{-1} \quad ; \quad \text{matrices}
\]

A very useful result – the “\text{plim}” operator can be used very flexibly.

**Asymptotic Properties of LS Estimator(s)**

- Consider LS estimator of $\beta$ under our standard assumptions, in the “large $n$” asymptotic case.
- Can relax some assumptions:
  (i) Don’t need Normality assumption for the error term of our model
  (ii) Columns of $X$ can be random – just assume that $\{x_i', \varepsilon_i\}$ is a random and independent sequence; $i = 1, 2, 3, \ldots$.
  (iii) Last assumption implies $\text{plim}[n^{-1}X'\varepsilon] = 0$. (Greene, pp. 64-65.)
- Amend (extend) our assumption about $X$ having full column rank – assume instead that $\text{plim}[n^{-1}X'X] = Q$; positive-definite & finite
Note that $Q$ is $(k \times k)$, symmetric, and \textit{unobservable}.

What are we assuming about the elements of $X$, which is $(n \times k)$, as $n$ increases without limit?

\textbf{Theorem:} The LS estimator of $\beta$ is \textit{weakly consistent}.

\textbf{Proof:} 
\begin{align*}
 b &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \epsilon) \\
 &= \beta + (X'X)\epsilon \\
 &= \beta + \left[\frac{1}{n}(X'X)\right]^{-1}\left[\frac{1}{n}X'\epsilon\right].
\end{align*}

If we now apply Slutsky’s Theorem repeatedly, we have:

$$plim(b) = \beta + Q^{-1}.0 = \beta.$$ 

We can also show that $s^2$ is a consistent estimator for $\sigma^2$.

\textbf{An Issue}

- Suppose we want to compare the (large $n$) asymptotic behaviour of our LS estimators with those of other potential estimators.
- These other estimators will presumably also be \textit{consistent}.
- This means that \textit{in each case} the sampling distributions of the estimators collapse to a “spike”, located exactly at the true parameter values.
- So, how can we compare such estimators when $n$ is very large – aren’t they \textit{indistinguishable}?
- If the limiting density of any consistent estimator is a degenerate “spike”, it will have zero variance, in the limit.
- Can we still compare large-sample variances of consistent estimators?

\textit{In other words, is it meaningful to think about the concept of asymptotic efficiency?}
Asymptotic Efficiency

- The key to asymptotic efficiency is to “control” for the fact that the distribution of any consistent estimator is “collapsing”, as $n \to \infty$.
- The rate at which the distribution collapses is crucially important.
- This is probably best understood by considering an example.
  - $\{x_i\}_{i=1}^n$ : random sampling from $[\mu, \sigma^2]$.
  - $E[\bar{x}] = \mu$ ; $var.[\bar{x}] = \sigma^2/n$
  - Now construct: $y = \sqrt{n}(\bar{x} - \mu)$.
  - Note that $E(y) = \sqrt{n}(E(\bar{x}) - \mu) = 0$.
  - Also, $var.[y] = (\sqrt{n})^2 var.(\bar{x} - \mu) = n var.(\bar{x}) = \sigma^2$.
  - The scaling we’ve used results in a finite, non-zero, variance.
  - $E(y) = 0$, and $var.[y] = \sigma^2$ ; unchanged as $n \to \infty$.
  - So, $y = \sqrt{n}(\bar{x} - \mu)$ has a well-defined “limiting” (asymptotic) distribution.
  - The asymptotic mean of $y$ is zero, and the asymptotic variance of $y$ is $\sigma^2$.
  - Question – Why did we scale by $\sqrt{n}$, and not (say), by $n$ itself ?
  - In fact, because we had independent $x_i$’s (random sampling), we have the additional result that $y = \sqrt{n}(\bar{x} - \mu) \xrightarrow{d} N[0, \sigma^2]$, the Lindeberg-Lévy Central Limit Theorem.
  - Now we can define “Asymptotic Efficiency” in a meaningful way.

**Definition:** Let $\hat{\theta}$ and $\tilde{\theta}$ be two consistent estimator of $\theta$ ; and suppose that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} [0, \sigma^2] \quad \text{and} \quad \sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} [0, \varphi^2].$$

Then $\hat{\theta}$ is “asymptotically efficient” relative to $\tilde{\theta}$ if $\sigma^2 < \varphi^2$.

In the case where $\theta$ is a vector, $\hat{\theta}$ is “asymptotically efficient” relative to $\tilde{\theta}$ if

$$\Delta = \text{asy}.V(\hat{\theta}) - \text{asy}.V(\tilde{\theta})$$

is positive definite.
Asymptotic Distribution of the LS Estimator:

Let's consider the full asymptotic distribution of the LS estimator, \( b \), for \( \beta \) in our linear regression model.

We'll actually have to consider the behaviour of \( \sqrt{n}(b - \beta) \):

\[
\sqrt{n}(b - \beta) = \sqrt{n}[(X'X)^{-1}X'e]
\]

\[
= \left[\frac{1}{n}(X'X)\right]^{-1}\left(\frac{1}{\sqrt{n}}X'e\right).
\]

It can be shown, by the Lindeberg-Feller Central Limit Theorem, that

\[
\left(\frac{1}{\sqrt{n}}X'e\right) \xrightarrow{d} N[0, \sigma^2 Q],
\]

where

\[
Q = \text{plim}\left[\frac{1}{n}(X'X)\right].
\]

So, the asymptotic covariance matrix of \( \sqrt{n}(b - \beta) \) is

\[
\text{plim}\left[\frac{1}{n}(X'X)\right]^{-1}(\sigma^2 Q)\text{plim}\left[\frac{1}{n}(X'X)\right]^{-1} = \sigma^2 Q^{-1}.
\]

In full, the asymptotic distribution of \( b \) is correctly stated by saying that:

\[
\sqrt{n}(b - \beta) \xrightarrow{d} N[0, \sigma^2 Q^{-1}]
\]

The asymptotic covariance matrix is unobservable, for two reasons:

1. \( \sigma^2 \) is typically unknown.
2. \( Q \) is unobservable.

- We can estimate \( \sigma^2 \) consistently, using \( s^2 \).
- To estimate \( \sigma^2 Q^{-1} \) consistently, we can use \( ns^2(X'X)^{-1} \):

\[
\text{plim}[ns^2(X'X)^{-1}] = \text{plim}(s^2)\text{plim}\left[\frac{1}{n}(X'X)\right]^{-1} = \sigma^2 Q^{-1}.
\]
Loosely speaking, the asymptotic covariance matrix for \( b \) itself is \( s^2 (X'X)^{-1} \); and the square roots of the diagonal elements of this matrix are the asymptotic std. errors for the \( b_i \)'s themselves.

**Instrumental Variables**

- We have been assuming either that the columns of \( X \) are non-random; or that the sequence \( \{x_i', \epsilon_i\} \) is independent. Often, neither of these assumptions is tenable.
- This implies that \( plim \left( \frac{1}{n} X' \epsilon \right) \neq 0 \), and then the LS estimator is inconsistent (prove this).
- In order to motivate a situation where \( \{x_i', \epsilon_i\} \) are dependent, consider an omitted, or unobservable variable.
We will consider a situation where the unobservable variable is correlated with one of the regressors, and correlated with the dependent variable.

Consider the population model:

\[ y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon_1. \]  \hspace{1cm} [1]

Consider that \( \text{cov}(X_1, X_2) \neq 0 \). For example, \( X_2 \) causes \( X_1 \):

\[ X_1 = X_2 \gamma + \varepsilon_2. \]  \hspace{1cm} [2]

Now consider that \( X_2 \) is unobservable, so that the observable model is:

\[ y = X_1 \beta_1 + \varepsilon_3. \]  \hspace{1cm} [3]

- Notice that in [3], \( \varepsilon_3 \) contains \( \beta_2 X_2 \), so that \( X_1 \) and \( \varepsilon_3 \) are not independent (\( X_1 \) is endogenous)
- OLS will be biased, since \( E[\varepsilon_3 | X_1] \neq 0 \)
- Note that when estimating from [3], \( E[b_1] = \beta_1 + \gamma^{-1} \beta_2 \)
- OLS will be inconsistent, since \( \text{plim} \left( \frac{1}{n} X_1 \varepsilon_3 \right) \neq 0 \)
- In such cases we want a safe way of estimating \( \beta_1 \).
- We just want to ensure that we have an estimator that is (at least) consistent.
- One general family of such estimators is the family of Instrumental Variables (I.V.) Estimators.

An instrumental variable, \( Z \), must be:

1. Correlated with the endogenous variable(s) \( X_1 \)
   - Sometimes called the “relevance” of an I.V.
   - This condition can be tested
2. Uncorrelated with the error term, or equivalently, uncorrelated with the dependent variable other than through its correlation with \( X_1 \)
   - Sometimes called the “exclusion” restriction
   - This restriction cannot be tested directly
Suppose now that we have a variable $Z$ which is

- Relevant: $\text{cov}(Z, X_1) \neq 0$
- Satisfies exclusion restriction: $\text{cov}(Z, \varepsilon) = 0$. In the above D.G.P.s ([1] - [3]), it is sufficient for the instrument to be uncorrelated with the unobservable variable: $\text{cov}(Z, X_2) = 0$.

Relevance means that [2] becomes:

$$X_1 = Z\delta + X_2\gamma + \varepsilon_4$$ \hspace{1cm} [4]

Substituting [4] into [1]:

$$y = X_2\gamma\beta_1 + Z\delta\beta_1 + X_2\beta_2 + \varepsilon_5.$$ \hspace{1cm} [5]

$X_2$ is still unobservable, but is uncorrelated with $Z$! The observable population model is now:

$$y = Z\delta\beta_1 + \varepsilon_6.$$ \hspace{1cm} [6]

Now, we have a population model involving $\beta_1$, and where $\text{cov}(Z, \varepsilon_6) = 0$. So, $(\delta\beta_1)$ can be estimated by OLS. But we need $\beta_1$!

By Slutsky’s Theorem, if $\text{plim}(\delta\beta_1) = \delta\beta_1$, and if $\text{plim}(\delta) = \delta$, then $\text{plim}(\delta^{-1}\delta\beta_1) = \beta_1$. So if we can find a consistent estimator for $\delta$, we can find one for $\beta_1$. How to estimate $\delta$?

Recall [4]. Since $X_2$ and $Z$ are uncorrelated, we can estimate $\delta$ by an OLS regression of $X_1$ on $Z$:

$$\hat{\delta} = (Z'Z)^{-1}Z'X_1$$

Now solve for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \delta^{-1}\delta\hat{\beta}_1 = [(Z'Z)^{-1}Z'X_1]^{-1}(Z'Z)^{-1}Z'y$$

If $Z$ and $X_1$ have the same number of columns, then:

$$\hat{\beta}_1 = (Z'X_1)^{-1}Z'Z(Z'Z)^{-1}Z'y = (Z'X_1)^{-1}Z'y$$
In this example we had one endogenous variable \((X_1)\) and one instrument \((Z)\). In this case, the I.V. estimate may be found by dividing the OLS estimate from a regression of \(y\) on \(Z\) by the OLS estimates of a regression of \(X_1\) on \(Z\).

In more general models, we will have more explanatory variables. As long as there is one instrument per endogenous variable, I.V. is possible and the simple I.V. estimator is:

\[
b_{IV} = (Z'X)^{-1}Z'y
\]

In general, this estimator is biased. We can show it’s consistent, however:

\[
y = X\beta + \epsilon
\]

\[
plim \left( \frac{1}{n} X'X \right) = Q \quad ; \quad \text{p.d. and finite}
\]

\[
plim \left( \frac{1}{n} X'\epsilon \right) = y \neq 0
\]

Find a (random) \((n \times k)\) matrix, \(Z\), such that:

1. \( plim \left( \frac{1}{n} Z'Z \right) = Q_ZZ \quad ; \quad \text{p.d. and finite.} \)
2. \( plim \left( \frac{1}{n} Z'X \right) = Q_XZ \quad ; \quad \text{p.d. and finite.} \)
3. \( plim \left( \frac{1}{n} Z'\epsilon \right) = 0 \).

Then, consider the estimator: \( b_{IV} = (Z'X)^{-1}Z'y \). This is a consistent estimator of \(\beta\).

\[
b_{IV} = (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta + \epsilon)
\]

\[
= (Z'X)^{-1}Z'X\beta + (Z'X)^{-1}Z'\epsilon
\]

\[
= \beta + (Z'X)^{-1}Z'\epsilon
\]

\[
= \beta + \left( \frac{1}{n} Z'X \right)^{-1} \left( \frac{1}{n} Z'\epsilon \right).
\]

So, \( plim(b_{IV}) = \beta + [plim \left( \frac{1}{n} Z'X \right)]^{-1} plim \left( \frac{1}{n} Z'\epsilon \right) \)

\[
= \beta + Q_{ZZ}^{-1}0 = \beta \quad (\text{consistent})
\]
Choosing different Z matrices generates different members of I.V. family.

Although we won’t derive the full asymptotic distribution of the I.V. estimator, note that it can be expressed as:

$$\sqrt{n}(b_{IV} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q_{XX}^{-1} Q_{ZZ} Q_{XZ}^{-1})$$

where $Q_{XZ} = Q_{ZX}'$. [How would you estimate Asy. Covar. Matrix?]

Interpreting I.V. as two-stage least squares (2SLS)

1st stage: Regress $X$ on $Z$, get $\hat{X}$.

- $\hat{X}$ contains the variation in $X$ due to $Z$ only
- $\hat{X}$ is not correlated with $\epsilon$

2nd stage: Estimate the model $y = \hat{X}\beta + \epsilon$

From 1st stage: $\hat{X} = Z(Z'Z)^{-1}Z'X$

From 2nd stage: $b_{IV} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y$

In fact, this is the Generalized I.V. estimator of $\beta$. We can actually use more instruments than regressors (the “Over-Identified” case).

Note that if $X$ and $Z$ have the same dimensions, then the generalized estimator collapses to the simple one.

Let’s check the consistency of the I.V. estimator. Let $M_Z = Z(Z'Z)^{-1}Z'$. Then the generalized I.V. estimator is:

$$b_{IV} = [X'M_ZX]^{-1}X'M_Zy$$

$$b_{IV} = [X'M_ZX]^{-1}X'M_Zy = [X'M_ZX]^{-1}X'M_Z(X\beta + \epsilon)$$

$$= [X'M_ZX]^{-1}X'M_ZX\beta + [X'M_ZX]^{-1}X'M_Z\epsilon$$

$$= \beta + [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\epsilon$$
So,

\[ \mathbf{b}_{IV} = \beta + \left[ \left( \frac{1}{n} \mathbf{X}' \mathbf{Z} \right) \left( \frac{1}{n} \mathbf{Z}' \mathbf{Z} \right)^{-1} \left( \frac{1}{n} \mathbf{Z}' \mathbf{X} \right) \right]^{-1} \left( \frac{1}{n} \mathbf{X}' \mathbf{Z} \right) \left( \frac{1}{n} \mathbf{Z}' \mathbf{Z} \right)^{-1} \left( \frac{1}{n} \mathbf{Z}' \mathbf{\varepsilon} \right). \]

**Modify our assumptions:**

We have a (random) \((n \times L)\) matrix, \(\mathbf{Z}\), such that:

1. \(\text{plim} \left( \frac{1}{n} \mathbf{Z}' \mathbf{Z} \right) = Q_{\mathbf{Z}\mathbf{Z}} ; \ (L \times L)\), p.d.s. and finite.
2. \(\text{plim} \left( \frac{1}{n} \mathbf{Z}' \mathbf{X} \right) = Q_{\mathbf{Z}\mathbf{X}} ; \ (L \times k)\), rank = \(k\), and finite.
3. \(\text{plim} \left( \frac{1}{n} \mathbf{Z}' \mathbf{\varepsilon} \right) = \mathbf{0} \ ; \ (L \times 1)\)

So,

\[ \text{plim} (\mathbf{b}_{IV}) = \beta + \left[ Q_{\mathbf{X}\mathbf{Z}} Q_{\mathbf{Z}\mathbf{Z}}^{-1} Q_{\mathbf{Z}\mathbf{X}} \right]^{-1} Q_{\mathbf{X}\mathbf{Z}} Q_{\mathbf{Z}\mathbf{Z}}^{-1} \mathbf{0} = \beta \ ; \ \text{consistent} \]

Similarly, a *consistent estimator* of \(\sigma^2\) is

\[ s_{IV}^2 = (\mathbf{y} - \mathbf{X} \mathbf{b}_{IV})'(\mathbf{y} - \mathbf{X} \mathbf{b}_{IV})/n \]

- Recall that each choice of \(\mathbf{Z}\) leads to a different I.V. estimator.
- \(\mathbf{Z}\) must be chosen in way that ensures consistency of the I.V. estimator.
- How might we choose a suitable set of instruments, *in practice*?

We need to find instruments that are uncorrelated with the errors, but highly correlated with the regressors – *asymptotically*.

This is not easy to do!

- **Time series data** -
  1. Often, we can use lagged values of the regressors as suitable instruments.
  2. This will be fine as long as the errors are serially uncorrelated.
• Cross-section data –
  1. Geography, weather, biology.
  2. Various “old” tricks – e.g., using “ranks” of the data as instruments.

Testing if I.V. estimation is needed

• Why does LS fail, and when do we need I.V.?
• If \( \text{plim} \left( \frac{1}{n} X' \epsilon \right) \neq 0 \).
• We can test to see if this is a problem, & decide if we should use LS or I.V.

The Hausman Test

We want to test \( H_0 : \text{plim} \left( \frac{1}{n} X' \epsilon \right) = 0 \) vs. \( H_A : \text{plim} \left( \frac{1}{n} X' \epsilon \right) \neq 0 \)

• If we reject \( H_0 \), we will use I.V. estimation.
• If we cannot reject \( H_0 \), we’ll use LS estimation.
• Hausman test is a general “testing strategy” that can be applied in many situations – not just for this particular situation!
• Basic idea – construct 2 estimators of \( \beta \):
  1. \( b_E \): estimator is both consistent and asymptotically efficient if \( H_0 \) true.
  2. \( b_I \): estimator is at least consistent, even if \( H_0 \) false.
• In our case here, \( b_E \) is the LS estimator; and \( b_I \) is the I.V. estimator.
• If \( H_0 \) is true, we’d expect \( (b_I - b_E) \) to be “small”, at least for large \( n \), as both estimators are consistent in that case.
• The test statistic is, \( H = (b_I - b_E)' \left[ \hat{V}(b_I) - \hat{V}(b_E) \right]^{-1} (b_I - b_E) \).
• \( H \xrightarrow{d} \chi_J^2 \), if \( H_0 \) is true.
• Here, \( J \) is the number of columns in \( X \) which may be correlated with the errors, & for which we need instruments.
**The Durbin-Wu Test**

This test is *specific* to testing

\[ H_0 : \text{plim} \left( \frac{1}{n} X' \varepsilon \right) = 0 \quad \text{vs.} \quad H_A : \text{plim} \left( \frac{1}{n} X' \varepsilon \right) \neq 0 \]

Again, an asymptotic test.

**Testing the exogeneity of Instruments**

The key assumption that ensures the consistency of I.V. estimators is that

\[ \text{plim} \left( \frac{1}{n} Z' \varepsilon \right) = 0. \]

This condition involves the *unobservable* \( \varepsilon \). In general, it cannot be tested.

“**Weak Instruments**” – Problems arise if the instruments are *not* well correlated with the regressors (not relevant).

- These problems go beyond loss of asymptotic efficiency.
- Small-sample bias of I.V. estimator can be greater than that of LS!
- Sampling distribution of I.V. estimator can be bi-modal!
- Fortunately, we can again *test* to see if we have these problems.
Empirical Example: Using geographic variation in college proximity to estimate the return to schooling

- Have data on wage, years of education, and demographic variables
- Want to estimate the return to education
- Problem: ability (intelligence) may be correlated with (cause) both wage and education
- Since ability is unobservable, it is contained in the error term
- The education variable is then correlated with the error term (endogenous)
- OLS estimation of the returns to education may be inconsistent

First, let’s try OLS.

```r
library(AER)
attach(CollegeDistance)
lm(wage ~ urban + gender + ethnicity + unemp + education)
```

Note that the returns to education are not statistically significant.

Now let’s try using distance from college (while attending high school) as an instrument for education. For the instrument to be valid, we require that distance and education be correlated:

```r
summary(lm(education ~ distance))
```

---

While \textit{distance} appears to be statistically significant, this isn’t quite enough to test for validity (a testing problem we won’t address here).

From the 2SLS interpretation, we know that we can get the IV estimator by:

1.) getting the predicted values from a regression of \textit{education} on \textit{distance}

\begin{verbatim}
educfit <- predict(lm(education ~ distance))
\end{verbatim}

2.) regressing \textit{wage} on the same variables, but using \textit{educfit} instead of \textit{education}

\begin{verbatim}
lm(wage ~ urban + gender + ethnicity + unemp + educfit)
\end{verbatim}

Note that \textit{educfit} is the variation in \textit{education} as it can be explained by \textit{distance}. These fitted values are uncorrelated with \textit{ability}, since \textit{distance} is uncorrelated with \textit{ability} (by assumption).

Results of IV estimation:

\begin{verbatim}
Regression Estimates
-0.5
0.0
0.5
1.0
urban female black hispanic unemp educfit
-0.5
0.0
1.0

The estimate for the return to education is now positive, and significant.