Leaders and democratic institutions

Jessica Leight

Williams Department of Economics

February 8, 2016

Leadership and economic outcomes

- One feature of the political landscape that clearly affects economic outcomes is the quality of political leaders.
- North Korea is clearly suffering under the rule of the Kim family.
- By contrast, other rapid growth countries are argued to have benefited from better "quality" leaders, where quality could be knowledge, administrative capacity, integrity, etc. Examples: Singapore, Hong Kong, Chile.
- How can we learn about the impact of a better quality political leader on economic outcomes?

Leadership and economic outcomes

- One feature of the political landscape that clearly affects economic outcomes is the quality of political leaders.
- North Korea is clearly suffering under the rule of the Kim family.
- By contrast, other rapid growth countries are argued to have benefited from better "quality" leaders, where quality could be knowledge, administrative capacity, integrity, etc. Examples: Singapore, Hong Kong, Chile.
- How can we learn about the impact of a better quality political leader on economic outcomes?

Leadership and economic outcomes

- One feature of the political landscape that clearly affects economic outcomes is the quality of political leaders.
- North Korea is clearly suffering under the rule of the Kim family.
- By contrast, other rapid growth countries are argued to have benefited from better "quality" leaders, where quality could be knowledge, administrative capacity, integrity, etc. Examples: Singapore, Hong Kong, Chile.
- How can we learn about the impact of a better quality political leader on economic outcomes?

Leadership and economic outcomes

- One feature of the political landscape that clearly affects economic outcomes is the quality of political leaders.
- North Korea is clearly suffering under the rule of the Kim family.
- By contrast, other rapid growth countries are argued to have benefited from better "quality" leaders, where quality could be knowledge, administrative capacity, integrity, etc. Examples: Singapore, Hong Kong, Chile.
- How can we learn about the impact of a better quality political leader on economic outcomes?

Identification challenge

- One possibility would be to somehow measure quality of leaders (again, in terms of education, intelligence, integrity, etc.), and run a regression of economic growth on quality of leader.
- What are the potential problems with this strategy?
- In this module we will first review some basic concepts of causal inference; then we will discuss two papers by the same authors that analyze the impact of leadership transitions on economic and political outcomes.

Identification challenge

- One possibility would be to somehow measure quality of leaders (again, in terms of education, intelligence, integrity, etc.), and run a regression of economic growth on quality of leader.
- What are the potential problems with this strategy?
- In this module we will first review some basic concepts of causal inference; then we will discuss two papers by the same authors that analyze the impact of leadership transitions on economic and political outcomes.

Identification challenge

- One possibility would be to somehow measure quality of leaders (again, in terms of education, intelligence, integrity, etc.), and run a regression of economic growth on quality of leader.
- What are the potential problems with this strategy?
- In this module we will first review some basic concepts of causal inference; then we will discuss two papers by the same authors that analyze the impact of leadership transitions on economic and political outcomes.

Simple linear regression

- Linear regression is the fundamental building block of empirical economics.
- We want to understand the relationship between two variables X and Y in a population, believing that X helps to explain Y, but the two variables do not have a fully deterministic relationship.
- An example of a fully deterministic relationship: how your birthday affects your age. You know one, you know the other; doesn't require much exploration.
- An example of a relationship that is not fully deterministic: how studying affects your GPA.

Simple linear regression

- Linear regression is the fundamental building block of empirical economics.
- We want to understand the relationship between two variables X and Y in a population, believing that X helps to explain Y, but the two variables do not have a fully deterministic relationship.
- An example of a fully deterministic relationship: how your birthday affects your age. You know one, you know the other; doesn't require much exploration.
- An example of a relationship that is not fully deterministic: how studying affects your GPA.

Simple linear regression

- Linear regression is the fundamental building block of empirical economics.
- We want to understand the relationship between two variables X and Y in a population, believing that X helps to explain Y, but the two variables do not have a fully deterministic relationship.
- An example of a fully deterministic relationship: how your birthday affects your age. You know one, you know the other; doesn't require much exploration.
- An example of a relationship that is not fully deterministic: how studying affects your GPA.

Simple linear regression

- Linear regression is the fundamental building block of empirical economics.
- We want to understand the relationship between two variables X and Y in a population, believing that X helps to explain Y, but the two variables do not have a fully deterministic relationship.
- An example of a fully deterministic relationship: how your birthday affects your age. You know one, you know the other; doesn't require much exploration.
- An example of a relationship that is not fully deterministic: how studying affects your GPA.

Linear regression equation

- A linear regression equation is written as follows.

$$Y = \beta_0 + \beta_1 X + u$$

- Note it looks like a line (hence linear), with intercept β_0 and slope β_1, plus the additional u term.
- To quickly review some general terminology: we normally refer to Y as the dependent or explained variable, X as the independent or explanatory variable, u as the error term, β_0 as the constant or intercept, and β_1 as the slope.

Linear regression equation

- A linear regression equation is written as follows.

$$Y = \beta_0 + \beta_1 X + u$$

- Note it looks like a line (hence linear), with intercept β_0 and slope β_1, plus the additional u term.
- To quickly review some general terminology: we normally refer to Y as the dependent or explained variable, X as the independent or explanatory variable, u as the error term, β_0 as the constant or intercept, and β_1 as the slope.

Linear regression equation

- A linear regression equation is written as follows.

$$Y = \beta_0 + \beta_1 X + u$$

- Note it looks like a line (hence linear), with intercept β_0 and slope β_1, plus the additional u term.
- To quickly review some general terminology: we normally refer to Y as the dependent or explained variable, X as the independent or explanatory variable, u as the error term, β_0 as the constant or intercept, and β_1 as the slope.

Error term

- A crucial assumption made in linear regression is that the error term is independent of x, $E[u|X] = 0$; in other words, simply because we know X does not mean we know anything additional about u.
- Just because we know how many hours you studied doesn't mean we know anything about the likelihood that you will have the flu during final exams and unfortunately knock down your GPA!
- What is the expected value of Y conditional on X, or $E[Y|X]$?

$$
\begin{aligned}
E[Y|X] &= E[\beta_0|X] + E[\beta_1 X|X] + E[u|X] \\
&= \beta_0 + \beta_1 X
\end{aligned}
$$

Error term

- A crucial assumption made in linear regression is that the error term is independent of x, $E[u|X] = 0$; in other words, simply because we know X does not mean we know anything additional about u.
- Just because we know how many hours you studied doesn't mean we know anything about the likelihood that you will have the flu during final exams and unfortunately knock down your GPA!
- What is the expected value of Y conditional on X, or $E[Y|X]$?

$$\begin{aligned} E[Y|X] &= E[\beta_0|X] + E[\beta_1 X|X] + E[u|X] \\ &= \beta_0 + \beta_1 X \end{aligned}$$

Error term

- A crucial assumption made in linear regression is that the error term is independent of x, $E[u|X] = 0$; in other words, simply because we know X does not mean we know anything additional about u.
- Just because we know how many hours you studied doesn't mean we know anything about the likelihood that you will have the flu during final exams and unfortunately knock down your GPA!
- What is the expected value of Y conditional on X, or $E[Y|X]$?

$$\begin{aligned} E[Y|X] &= E[\beta_0|X] + E[\beta_1 X|X] + E[u|X] \\ &= \beta_0 + \beta_1 X \end{aligned}$$

Understanding intuitively

- This statement about expectations can be understood as follows: holding all else equal, by how much does a one unit increase in X change Y? By the amount β_1.
- Mathematically, u corresponds to "all else": all other factors that might affect your GPA that we are ignoring.

Understanding intuitively

- This statement about expectations can be understood as follows: holding all else equal, by how much does a one unit increase in X change Y? By the amount β_1.
- Mathematically, u corresponds to "all else": all other factors that might affect your GPA that we are ignoring.

Estimating the regression line

- Up to now I've just described what a regression is in theory; if we actually want to estimate a regression given a sample of data, how would we proceed?
- If we have a sample, we can define for every observation the sample residual, denoted $\hat{u}_i$.

$$\hat{u}_i = y_i - \beta_0 - \beta_1 x_i$$

- The "least squares" method seeks to find estimators for β_0 and β_1 that minimize the sum of squared residuals.
- How do we do this mathematically?

Estimating the regression line

- Up to now I've just described what a regression is in theory; if we actually want to estimate a regression given a sample of data, how would we proceed?
- If we have a sample, we can define for every observation the sample residual, denoted $\hat{u}_i$.

$$\hat{u}_i = y_i - \beta_0 - \beta_1 x_i$$

- The "least squares" method seeks to find estimators for β_0 and β_1 that minimize the sum of squared residuals.
- How do we do this mathematically?

Estimating the regression line

- Up to now I've just described what a regression is in theory; if we actually want to estimate a regression given a sample of data, how would we proceed?
- If we have a sample, we can define for every observation the sample residual, denoted $\hat{u}_i$.

$$\hat{u}_i = y_i - \beta_0 - \beta_1 x_i$$

- The "least squares" method seeks to find estimators for β_0 and β_1 that minimize the sum of squared residuals.
- How do we do this mathematically?

Estimating the regression line

- Up to now I've just described what a regression is in theory; if we actually want to estimate a regression given a sample of data, how would we proceed?
- If we have a sample, we can define for every observation the sample residual, denoted $\hat{u}_i$.

$$\hat{u}_i = y_i - \beta_0 - \beta_1 x_i$$

- The "least squares" method seeks to find estimators for β_0 and β_1 that minimize the sum of squared residuals.
- How do we do this mathematically?

Least-squares method

- We want to minimize $\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2$.
- How do we normally find the minimum point of a function? Differentiate.
- This leads to a formula that is hopefully familiar.

$$\beta_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{N}(x_i - \bar{X})^2}$$

- This is known as the OLS estimator.

Least-squares method

- We want to minimize $\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2$.
- How do we normally find the minimum point of a function? Differentiate.
- This leads to a formula that is hopefully familiar.

$$\beta_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{N}(x_i - \bar{X})^2}$$

- This is known as the OLS estimator.

Least-squares method

- We want to minimize $\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2$.
- How do we normally find the minimum point of a function? Differentiate.
- This leads to a formula that is hopefully familiar.

$$\beta_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{N}(x_i - \bar{X})^2}$$

- This is known as the OLS estimator.

Least-squares method

- We want to minimize $\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2$.
- How do we normally find the minimum point of a function? Differentiate.
- This leads to a formula that is hopefully familiar.

$$\beta_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{N}(x_i - \bar{X})^2}$$

- This is known as the OLS estimator.

Evaluating the quality of an estimator

- We can generate this estimator of β_0 and β_1, but how do we know that it's "any good"? I.e., it approximates the true relationship?
- Under a certain set of assumptions, OLS is unbiased.
- An unbiased estimator means that if we analyzed many different samples, the expected value of the parameters over these many samples would equal the true values β_0 and β_1.
- Clearly, this is not necessarily true for any particular sample.

Evaluating the quality of an estimator

- We can generate this estimator of β_0 and β_1, but how do we know that it's "any good"? I.e., it approximates the true relationship?
- Under a certain set of assumptions, OLS is unbiased.
- An unbiased estimator means that if we analyzed many different samples, the expected value of the parameters over these many samples would equal the true values β_0 and β_1.
- Clearly, this is not necessarily true for any particular sample.

Evaluating the quality of an estimator

- We can generate this estimator of β_0 and β_1, but how do we know that it's "any good"? I.e., it approximates the true relationship?
- Under a certain set of assumptions, OLS is unbiased.
- An unbiased estimator means that if we analyzed many different samples, the expected value of the parameters over these many samples would equal the true values β_0 and β_1.
- Clearly, this is not necessarily true for any particular sample.

Evaluating the quality of an estimator

- We can generate this estimator of β_0 and β_1, but how do we know that it's "any good"? I.e., it approximates the true relationship?
- Under a certain set of assumptions, OLS is unbiased.
- An unbiased estimator means that if we analyzed many different samples, the expected value of the parameters over these many samples would equal the true values β_0 and β_1.
- Clearly, this is not necessarily true for any particular sample.

Key assumption for unbiasedness

- Perhaps the most important assumption is that the expected value of u conditional on X is zero for every X, i.e. $E[u|X] = 0$.
- What does this mean, intuitively? Let's return to our hours of studying and GPA example.
- There are some unobserved factors that affect GPA (e.g., having flu during exam week); we're assuming that the average probability that you have the flu is unrelated to how much you study.
- This might be true, or it might be the students who study more also take more prophylatic measures against the flu and they're less likely to have it.
- There are many other unobserved characteristics for which it is probably *not* plausible that $E[u|X] = 0$.

Key assumption for unbiasedness

- Perhaps the most important assumption is that the expected value of u conditional on X is zero for every X, i.e. $E[u|X] = 0$.
- What does this mean, intuitively? Let's return to our hours of studying and GPA example.
- There are some unobserved factors that affect GPA (e.g., having flu during exam week); we're assuming that the average probability that you have the flu is unrelated to how much you study.
- This might be true, or it might be the students who study more also take more prophylatic measures against the flu and they're less likely to have it.
- There are many other unobserved characteristics for which it is probably *not* plausible that $E[u|X] = 0$.

Key assumption for unbiasedness

- Perhaps the most important assumption is that the expected value of u conditional on X is zero for every X, i.e. $E[u|X] = 0$.
- What does this mean, intuitively? Let's return to our hours of studying and GPA example.
- There are some unobserved factors that affect GPA (e.g., having flu during exam week); we're assuming that the average probability that you have the flu is unrelated to how much you study.
- This might be true, or it might be the students who study more also take more prophylatic measures against the flu and they're less likely to have it.
- There are many other unobserved characteristics for which it is probably *not* plausible that $E[u|X] = 0$.

Key assumption for unbiasedness

- Perhaps the most important assumption is that the expected value of u conditional on X is zero for every X, i.e. $E[u|X] = 0$.
- What does this mean, intuitively? Let's return to our hours of studying and GPA example.
- There are some unobserved factors that affect GPA (e.g., having flu during exam week); we're assuming that the average probability that you have the flu is unrelated to how much you study.
- This might be true, or it might be the students who study more also take more prophylatic measures against the flu and they're less likely to have it.
- There are many other unobserved characteristics for which it is probably *not* plausible that $E[u|X] = 0$.

Key assumption for unbiasedness

- Perhaps the most important assumption is that the expected value of u conditional on X is zero for every X, i.e. $E[u|X] = 0$.
- What does this mean, intuitively? Let's return to our hours of studying and GPA example.
- There are some unobserved factors that affect GPA (e.g., having flu during exam week); we're assuming that the average probability that you have the flu is unrelated to how much you study.
- This might be true, or it might be the students who study more also take more prophylatic measures against the flu and they're less likely to have it.
- There are many other unobserved characteristics for which it is probably *not* plausible that $E[u|X] = 0$.

Multiple regression

- What to do in response to this problem? The easiest approach is to control for the other characteristics that may be relevant, if we can measure them.
- We expand the simple regression model to include more variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Now, β_0 is the intercept, β_1 is the "partial effect" of X_1 on Y holding X_2 constant, and β_2 is the "partial effect" of X_2 on Y holding X_1 constant.
- Adding additional variables can help, but it is not always a panacea.
- We'll return to the question of omitted variable bias and the challenge it poses for causal inference soon.

Multiple regression

- What to do in response to this problem? The easiest approach is to control for the other characteristics that may be relevant, if we can measure them.
- We expand the simple regression model to include more variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Now, β_0 is the intercept, β_1 is the "partial effect" of X_1 on Y holding X_2 constant, and β_2 is the "partial effect" of X_2 on Y holding X_1 constant.
- Adding additional variables can help, but it is not always a panacea.
- We'll return to the question of omitted variable bias and the challenge it poses for causal inference soon.

Multiple regression

- What to do in response to this problem? The easiest approach is to control for the other characteristics that may be relevant, if we can measure them.
- We expand the simple regression model to include more variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Now, β_0 is the intercept, β_1 is the "partial effect" of X_1 on Y holding X_2 constant, and β_2 is the "partial effect" of X_2 on Y holding X_1 constant.
- Adding additional variables can help, but it is not always a panacea.
- We'll return to the question of omitted variable bias and the challenge it poses for causal inference soon.

Multiple regression

- What to do in response to this problem? The easiest approach is to control for the other characteristics that may be relevant, if we can measure them.
- We expand the simple regression model to include more variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Now, β_0 is the intercept, β_1 is the "partial effect" of X_1 on Y holding X_2 constant, and β_2 is the "partial effect" of X_2 on Y holding X_1 constant.
- Adding additional variables can help, but it is not always a panacea.
- We'll return to the question of omitted variable bias and the challenge it poses for causal inference soon.

Multiple regression

- What to do in response to this problem? The easiest approach is to control for the other characteristics that may be relevant, if we can measure them.
- We expand the simple regression model to include more variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Now, β_0 is the intercept, β_1 is the "partial effect" of X_1 on Y holding X_2 constant, and β_2 is the "partial effect" of X_2 on Y holding X_1 constant.
- Adding additional variables can help, but it is not always a panacea.
- We'll return to the question of omitted variable bias and the challenge it poses for causal inference soon.

Inference

- We know that OLS estimators are unbiased, but in each sample we also know that we don't find the true population parameter.
- Sometimes, because of bad luck (or a small sample), our estimate is very far from the population parameter.
- How confident, then, can we be about our estimates?
- Hypothesis testing provides us with a structure to understand how confident we can be.

Inference

- We know that OLS estimators are unbiased, but in each sample we also know that we don't find the true population parameter.
- Sometimes, because of bad luck (or a small sample), our estimate is very far from the population parameter.
- How confident, then, can we be about our estimates?
- Hypothesis testing provides us with a structure to understand how confident we can be.

Inference

- We know that OLS estimators are unbiased, but in each sample we also know that we don't find the true population parameter.
- Sometimes, because of bad luck (or a small sample), our estimate is very far from the population parameter.
- How confident, then, can we be about our estimates?
- Hypothesis testing provides us with a structure to understand how confident we can be.

Inference

- We know that OLS estimators are unbiased, but in each sample we also know that we don't find the true population parameter.
- Sometimes, because of bad luck (or a small sample), our estimate is very far from the population parameter.
- How confident, then, can we be about our estimates?
- Hypothesis testing provides us with a structure to understand how confident we can be.

Introduction to hypothesis testing

- Suppose there is a true population parameter θ, and given a random sample from the population, we'd like to test whether $\theta = \gamma$, where γ is any value. (In reality, γ is often zero).
- The null hypothesis, denoted H_0, is that $\theta = \gamma$; the alternate hypothesis is $\theta \neq \gamma$ for a two-tailed test, or $\theta > \gamma$ or $\theta < \gamma$ for a one-tailed test.
- For a given sample, we will construct a test statistic, and given the value of the test statistic we will reject the null or fail to reject the null (never "accept", just "fail to reject").

Introduction to hypothesis testing

- Suppose there is a true population parameter θ, and given a random sample from the population, we'd like to test whether $\theta = \gamma$, where γ is any value. (In reality, γ is often zero).
- The null hypothesis, denoted H_0, is that $\theta = \gamma$; the alternate hypothesis is $\theta \neq \gamma$ for a two-tailed test, or $\theta > \gamma$ or $\theta < \gamma$ for a one-tailed test.
- For a given sample, we will construct a test statistic, and given the value of the test statistic we will reject the null or fail to reject the null (never "accept", just "fail to reject").

Introduction to hypothesis testing

- Suppose there is a true population parameter θ, and given a random sample from the population, we'd like to test whether $\theta = \gamma$, where γ is any value. (In reality, γ is often zero).
- The null hypothesis, denoted H_0, is that $\theta = \gamma$; the alternate hypothesis is $\theta \neq \gamma$ for a two-tailed test, or $\theta > \gamma$ or $\theta < \gamma$ for a one-tailed test.
- For a given sample, we will construct a test statistic, and given the value of the test statistic we will reject the null or fail to reject the null (never "accept", just "fail to reject").

Error in hypothesis testing

- Of course no process is perfect; there is the potential for error here.
- Type I error is rejecting the null when the null is true; type II error is failing to reject the null when the null is false.
- We set the significance level to be the probability of Type I error; i.e., $\alpha = Pr(reject\ H_0 \mid H_0\ true)$.
- This is normally set to be 5%.

Error in hypothesis testing

- Of course no process is perfect; there is the potential for error here.
- Type I error is rejecting the null when the null is true; type II error is failing to reject the null when the null is false.
- We set the significance level to be the probability of Type I error; i.e., $\alpha = Pr(\textit{reject } H_0 \mid H_0 \textit{ true})$.
- This is normally set to be 5%.

Error in hypothesis testing

- Of course no process is perfect; there is the potential for error here.
- Type I error is rejecting the null when the null is true; type II error is failing to reject the null when the null is false.
- We set the significance level to be the probability of Type I error; i.e., $\alpha = Pr(reject\ H_0 \mid H_0\ true)$.
- This is normally set to be 5%.

Error in hypothesis testing

- Of course no process is perfect; there is the potential for error here.
- Type I error is rejecting the null when the null is true; type II error is failing to reject the null when the null is false.
- We set the significance level to be the probability of Type I error; i.e., $\alpha = Pr(\textit{reject } H_0 \mid H_0 \textit{ true})$.
- This is normally set to be 5%.

Understanding significance

- When we discuss empirical economic research, there is often a lot of emphasis on a "significant" effect.
- What does this mean? If we have estimated an effect and deem it significant, that means we have rejected the null of **zero effect**; i.e., there is a less than 5% (or 10%, or 1%) probability that the observed relationship between X and Y is due to chance alone, and that the true effect of X on Y is zero.
- I.e., there is a less than 5% probability that the relationship between X and Y is due to chance alone (and hence we have rejected the null in error).

Understanding significance

- When we discuss empirical economic research, there is often a lot of emphasis on a "significant" effect.
- What does this mean? If we have estimated an effect and deem it significant, that means we have rejected the null of **zero effect**; i.e., there is a less than 5% (or 10%, or 1%) probability that the observed relationship between X and Y is due to chance alone, and that the true effect of X on Y is zero.
- I.e., there is a less than 5% probability that the relationship between X and Y is due to chance alone (and hence we have rejected the null in error).

Understanding significance

- When we discuss empirical economic research, there is often a lot of emphasis on a "significant" effect.
- What does this mean? If we have estimated an effect and deem it significant, that means we have rejected the null of **zero effect**; i.e., there is a less than 5% (or 10%, or 1%) probability that the observed relationship between X and Y is due to chance alone, and that the true effect of X on Y is zero.
- I.e., there is a less than 5% probability that the relationship between X and Y is due to chance alone (and hence we have rejected the null in error).

Constructing a test statistic

- Given a true population parameter β_j and an estimated parameter $\hat{\beta}_j$, how do we decide whether $\hat{\beta}_j$ is far enough from the hypothesized (null) value a_j – far enough so that we reject the null?
- Again, a_j will almost always be zero in practice, though in theory there is no reason that it must be.
- We construct the test statistic (T-statistic) as follows.

$$T_{\hat{\beta}j} = \frac{\hat{\beta}_j - a_i}{se(\hat{\beta}_j)}$$

- We do not know the standard deviation, but we can estimate it using the standard error.

Constructing a test statistic

- Given a true population parameter β_j and an estimated parameter $\hat{\beta}_j$, how do we decide whether $\hat{\beta}_j$ is far enough from the hypothesized (null) value a_j – far enough so that we reject the null?
- Again, a_j will almost always be zero in practice, though in theory there is no reason that it must be.
- We construct the test statistic (T-statistic) as follows.

$$T_{\hat{\beta}j} = \frac{\hat{\beta}_j - a_i}{se(\hat{\beta}_j)}$$

- We do not know the standard deviation, but we can estimate it using the standard error.

Constructing a test statistic

- Given a true population parameter β_j and an estimated parameter $\hat{\beta}_j$, how do we decide whether $\hat{\beta}_j$ is far enough from the hypothesized (null) value a_j – far enough so that we reject the null?
- Again, a_j will almost always be zero in practice, though in theory there is no reason that it must be.
- We construct the test statistic (T-statistic) as follows.

$$T_{\hat{\beta}j} = \frac{\hat{\beta}_j - a_i}{se(\hat{\beta}_j)}$$

- We do not know the standard deviation, but we can estimate it using the standard error.

Constructing a test statistic

- Given a true population parameter β_j and an estimated parameter $\hat{\beta}_j$, how do we decide whether $\hat{\beta}_j$ is far enough from the hypothesized (null) value a_j – far enough so that we reject the null?
- Again, a_j will almost always be zero in practice, though in theory there is no reason that it must be.
- We construct the test statistic (T-statistic) as follows.

$$T_{\hat{\beta}j} = \frac{\hat{\beta}_j - a_i}{se(\hat{\beta}_j)}$$

- We do not know the standard deviation, but we can estimate it using the standard error.

Evaluating a test statistic

- We can reject the null when the absolute value of the T statistic is sufficiently large.
- When is the absolute value of the T statistic large? When $\hat{\beta}_j$ is very far from a_j, or when $se(\hat{\beta}_j)$ is smaller (i.e., when the coefficient is very precisely estimated).
- Under distributional assumptions that I won't review here, this test statistic is distributed according to the T distribution.
- Given the level of significance we want (often 95%), the T distribution provides us with a critical value: if the absolute value of the test statistic is above this value, then we reject the null hypothesis.

Evaluating a test statistic

- We can reject the null when the absolute value of the T statistic is sufficiently large.
- When is the absolute value of the T statistic large? When $\hat{\beta}_j$ is very far from a_j, or when $se(\hat{\beta}_j)$ is smaller (i.e., when the coefficient is very precisely estimated).
- Under distributional assumptions that I won't review here, this test statistic is distributed according to the T distribution.
- Given the level of significance we want (often 95%), the T distribution provides us with a critical value: if the absolute value of the test statistic is above this value, then we reject the null hypothesis.

Evaluating a test statistic

- We can reject the null when the absolute value of the T statistic is sufficiently large.
- When is the absolute value of the T statistic large? When $\hat{\beta}_j$ is very far from a_j, or when $se(\hat{\beta}_j)$ is smaller (i.e., when the coefficient is very precisely estimated).
- Under distributional assumptions that I won't review here, this test statistic is distributed according to the T distribution.
- Given the level of significance we want (often 95%), the T distribution provides us with a critical value: if the absolute value of the test statistic is above this value, then we reject the null hypothesis.

Evaluating a test statistic

- We can reject the null when the absolute value of the T statistic is sufficiently large.
- When is the absolute value of the T statistic large? When $\hat{\beta}_j$ is very far from a_j, or when $se(\hat{\beta}_j)$ is smaller (i.e., when the coefficient is very precisely estimated).
- Under distributional assumptions that I won't review here, this test statistic is distributed according to the T distribution.
- Given the level of significance we want (often 95%), the T distribution provides us with a critical value: if the absolute value of the test statistic is above this value, then we reject the null hypothesis.

Rule of two

- If we're testing with a null hypothesis of zero, as we often are, we can check to see if the coefficient is twice the standard error. If it is (or more specifically, if it is more than 1.96 * SE), then the coefficient is significant at the 5% level.
- The critical value of significance at the 10% level is 1.64; for the 1% level, it is 2.58.
- These are all two-sided tests; in practice we don't use one-sided tests very often, because in almost all cases we have no reason ex ante to believe the parameter estimate is higher or lower than the null.

Rule of two

- If we're testing with a null hypothesis of zero, as we often are, we can check to see if the coefficient is twice the standard error. If it is (or more specifically, if it is more than 1.96 * SE), then the coefficient is significant at the 5% level.
- The critical value of significance at the 10% level is 1.64; for the 1% level, it is 2.58.
- These are all two-sided tests; in practice we don't use one-sided tests very often, because in almost all cases we have no reason ex ante to believe the parameter estimate is higher or lower than the null.

Rule of two

- If we're testing with a null hypothesis of zero, as we often are, we can check to see if the coefficient is twice the standard error. If it is (or more specifically, if it is more than 1.96 * SE), then the coefficient is significant at the 5% level.
- The critical value of significance at the 10% level is 1.64; for the 1% level, it is 2.58.
- These are all two-sided tests; in practice we don't use one-sided tests very often, because in almost all cases we have no reason ex ante to believe the parameter estimate is higher or lower than the null.

Statistical significance vs. economic significance

- Great emphasis is placed on statistical significance, but it is not all that matters.
- The t-statistic can be large because the estimate is large or because the standard error is small; if the latter is the case, the independent variable can have a precisely estimated but small effect on Y, that we might not care much about.
- We can also fail to reject the null because the standard error is large - meaning we don't have much precision.
- This shouldn't be interpreted as conclusive evidence that X doesn't have an effect on Y; we might want to collect more data or use some other strategy.
- **Most important**, significance doesn't imply causality. More about causal inference shortly.

Statistical significance vs. economic significance

- Great emphasis is placed on statistical significance, but it is not all that matters.
- The t-statistic can be large because the estimate is large or because the standard error is small; if the latter is the case, the independent variable can have a precisely estimated but small effect on Y, that we might not care much about.
- We can also fail to reject the null because the standard error is large - meaning we don't have much precision.
- This shouldn't be interpreted as conclusive evidence that X doesn't have an effect on Y; we might want to collect more data or use some other strategy.
- **Most important**, significance doesn't imply causality. More about causal inference shortly.

Statistical significance vs. economic significance

- Great emphasis is placed on statistical significance, but it is not all that matters.
- The t-statistic can be large because the estimate is large or because the standard error is small; if the latter is the case, the independent variable can have a precisely estimated but small effect on Y, that we might not care much about.
- We can also fail to reject the null because the standard error is large - meaning we don't have much precision.
- This shouldn't be interpreted as conclusive evidence that X doesn't have an effect on Y; we might want to collect more data or use some other strategy.
- **Most important**, significance doesn't imply causality. More about causal inference shortly.

Statistical significance vs. economic significance

- Great emphasis is placed on statistical significance, but it is not all that matters.
- The t-statistic can be large because the estimate is large or because the standard error is small; if the latter is the case, the independent variable can have a precisely estimated but small effect on Y, that we might not care much about.
- We can also fail to reject the null because the standard error is large - meaning we don't have much precision.
- This shouldn't be interpreted as conclusive evidence that X doesn't have an effect on Y; we might want to collect more data or use some other strategy.
- **Most important**, significance doesn't imply causality. More about causal inference shortly.

Statistical significance vs. economic significance

- Great emphasis is placed on statistical significance, but it is not all that matters.
- The t-statistic can be large because the estimate is large or because the standard error is small; if the latter is the case, the independent variable can have a precisely estimated but small effect on Y, that we might not care much about.
- We can also fail to reject the null because the standard error is large - meaning we don't have much precision.
- This shouldn't be interpreted as conclusive evidence that X doesn't have an effect on Y; we might want to collect more data or use some other strategy.
- **Most important**, significance doesn't imply causality. More about causal inference shortly.

Causality

- Let Y_i be wage of person i and let $X_i \in \{0, 1\}$ be whether an individual went to college.
- How do you measure the effect of going to college on wage for person i?

$$\begin{aligned} Y_i(X_i = 1) &= \beta_0 + \beta_1 \times 1 + \epsilon_i = \beta_0 + \beta_1 + \epsilon_i \\ Y_i(X_i = 0) &= \beta_0 + \beta_1 \times 0 + \epsilon_i = \beta_0 + \epsilon_i \end{aligned}$$

- What is the expected effect of going to school? Denote the estimate of interest T^*.

$$\begin{aligned} T^* &= E[Y_{1i} - Y_{0i} | X = 1] \\ T^* &= E[Y_{1i} | X = 1] - [Y_{0i} | X = 1] \end{aligned}$$

Causality

- Let Y_i be wage of person i and let $X_i \in \{0, 1\}$ be whether an individual went to college.
- How do you measure the effect of going to college on wage for person i?

$$
\begin{aligned}
Y_i(X_i = 1) &= \beta_0 + \beta_1 \times 1 + \epsilon_i = \beta_0 + \beta_1 + \epsilon_i \\
Y_i(X_i = 0) &= \beta_0 + \beta_1 \times 0 + \epsilon_i = \beta_0 + \epsilon_i
\end{aligned}
$$

- What is the expected effect of going to school? Denote the estimate of interest T^*.

$$
\begin{aligned}
T^* &= E[Y_{1i} - Y_{0i} | X = 1] \\
T^* &= E[Y_{1i} | X = 1] - [Y_{0i} | X = 1]
\end{aligned}
$$

Causality

- Let Y_i be wage of person i and let $X_i \in \{0, 1\}$ be whether an individual went to college.
- How do you measure the effect of going to college on wage for person i?

$$\begin{aligned} Y_i(X_i = 1) &= \beta_0 + \beta_1 \times 1 + \epsilon_i = \beta_0 + \beta_1 + \epsilon_i \\ Y_i(X_i = 0) &= \beta_0 + \beta_1 \times 0 + \epsilon_i = \beta_0 + \epsilon_i \end{aligned}$$

- What is the expected effect of going to school? Denote the estimate of interest T^*.

$$\begin{aligned} T^* &= E[Y_{1i} - Y_{0i} | X = 1] \\ T^* &= E[Y_{1i} | X = 1] - [Y_{0i} | X = 1] \end{aligned}$$

Fundamental problem of causal inference

- How can we estimate

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Problem: we observe (Y_i, X_i) implies observing **either** $E[Y_{1i}(X_i = 1)]$ or $E[Y_{0i}(X_i = 0)]$.
  - Observe Kobe's wage, given he did not go to college: $E[Y_{0i}(X_i = 0)]$
  - Observe Duncan's wage, given he went to college: $E[Y_{1i}(X_i = 1)]$
- Note X denotes what was actually observed; Y is an outcome that can be either hypothetical or observed.
- We can observe and calculate $E[Y_{1i}(X_i = 1)] - E[Y_{0i}(X_i = 0)]$.

Fundamental problem of causal inference

- How can we estimate

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Problem: we observe (Y_i, X_i) implies observing **either** $E[Y_{1i}(X_i = 1)]$ or $E[Y_{0i}(X_i = 0)]$.
  - Observe Kobe's wage, given he did not go to college: $E[Y_{0i}(X_i = 0)]$
  - Observe Duncan's wage, given he went to college: $E[Y_{1i}(X_i = 1)]$
- Note X denotes what was actually observed; Y is an outcome that can be either hypothetical or observed.
- We can observe and calculate $E[Y_{1i}(X_i = 1)] - E[Y_{0i}(X_i = 0)]$.

Fundamental problem of causal inference

- How can we estimate

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Problem: we observe (Y_i, X_i) implies observing **either** $E[Y_{1i}(X_i = 1)]$ or $E[Y_{0i}(X_i = 0)]$.
  - Observe Kobe's wage, given he did not go to college: $E[Y_{0i}(X_i = 0)]$
  - Observe Duncan's wage, given he went to college: $E[Y_{1i}(X_i = 1)]$
- Note X denotes what was actually observed; Y is an outcome that can be either hypothetical or observed.
- We can observe and calculate $E[Y_{1i}(X_i = 1)] - E[Y_{0i}(X_i = 0)]$.

Fundamental problem of causal inference

- How can we estimate

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Problem: we observe (Y_i, X_i) implies observing **either** $E[Y_{1i}(X_i = 1)]$ or $E[Y_{0i}(X_i = 0)]$.
  - Observe Kobe's wage, given he did not go to college: $E[Y_{0i}(X_i = 0)]$
  - Observe Duncan's wage, given he went to college: $E[Y_{1i}(X_i = 1)]$
- Note X denotes what was actually observed; Y is an outcome that can be either hypothetical or observed.
- We can observe and calculate $E[Y_{1i}(X_i = 1)] - E[Y_{0i}(X_i = 0)]$.

Fundamental problem of causal inference

- How can we estimate

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Problem: we observe (Y_i, X_i) implies observing **either** $E[Y_{1i}(X_i = 1)]$ or $E[Y_{0i}(X_i = 0)]$.
  - Observe Kobe's wage, given he did not go to college: $E[Y_{0i}(X_i = 0)]$
  - Observe Duncan's wage, given he went to college: $E[Y_{1i}(X_i = 1)]$
- Note X denotes what was actually observed; Y is an outcome that can be either hypothetical or observed.
- We can observe and calculate $E[Y_{1i}(X_i = 1)] - E[Y_{0i}(X_i = 0)]$.

Fundamental problem of causal inference

- How can we estimate

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Problem: we observe (Y_i, X_i) implies observing **either** $E[Y_{1i}(X_i = 1)]$ or $E[Y_{0i}(X_i = 0)]$.
  - Observe Kobe's wage, given he did not go to college: $E[Y_{0i}(X_i = 0)]$
  - Observe Duncan's wage, given he went to college: $E[Y_{1i}(X_i = 1)]$
- Note X denotes what was actually observed; Y is an outcome that can be either hypothetical or observed.
- We can observe and calculate $E[Y_{1i}(X_i = 1)] - E[Y_{0i}(X_i = 0)]$.

Fundamental problem of causal inference part II

- Remember, we would like to estimate the following:

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Notice that

$$E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 0] = \underbrace{E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 1]}_{Goal} + \underbrace{E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0]}_{Bias}$$

- The observational estimate, naively comparing the treated and untreated, is equal to the estimate of interest, plus a bias term.

Fundamental problem of causal inference part II

- Remember, we would like to estimate the following:

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Notice that

$$\begin{aligned} E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 0] &= \underbrace{E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 1]}_{Goal} \\ &+ \underbrace{E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0]}_{Bias} \end{aligned}$$

- The observational estimate, naively comparing the treated and untreated, is equal to the estimate of interest, plus a bias term.

Fundamental problem of causal inference part II

- Remember, we would like to estimate the following:

$$T^* = E[Y_{1i} - Y_{0i}]$$

- Notice that

$$\begin{aligned} E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 0] &= \underbrace{E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 1]}_{Goal} \\ &+ \underbrace{E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0]}_{Bias} \end{aligned}$$

- The observational estimate, naively comparing the treated and untreated, is equal to the estimate of interest, plus a bias term.

Bias

- The bias term can be written as follows:

$$E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0] = E[\epsilon_i|X_i = 1] - E[\epsilon_i|X_i = 0]$$

- What do we need to assume?

$$E[\epsilon_i|X_i = 1] = E[\epsilon_i|X_i = 0]$$

- Intuitively: unobservable characteristics are the same across groups.

Bias

- The bias term can be written as follows:

$$E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0] = E[\epsilon_i|X_i = 1] - E[\epsilon_i|X_i = 0]$$

- What do we need to assume?

$$E[\epsilon_i|X_i = 1] = E[\epsilon_i|X_i = 0]$$

- Intuitively: unobservable characteristics are the same across groups.

Bias

- The bias term can be written as follows:

$$E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0] = E[\epsilon_i|X_i = 1] - E[\epsilon_i|X_i = 0]$$

- What do we need to assume?

$$E[\epsilon_i|X_i = 1] = E[\epsilon_i|X_i = 0]$$

- Intuitively: unobservable characteristics are the same across groups.

Bias

- The bias term can be written as follows:

$$E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0] = E[\epsilon_i|X_i = 1] - E[\epsilon_i|X_i = 0]$$

- What do we need to assume?

$$E[\epsilon_i|X_i = 1] = E[\epsilon_i|X_i = 0]$$

- Intuitively: unobservable characteristics are the same across groups.

Returning to the returns to education

- Given a population of students who went to college and a population of students who did not go to college, do we expect that their other characteristics will be the same? For example: family income, ability, interest in education?
- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?

Returning to the returns to education

- Given a population of students who went to college and a population of students who did not go to college, do we expect that their other characteristics will be the same? For example: family income, ability, interest in education?
- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?

Returning to the returns to education

- Given a population of students who went to college and a population of students who did not go to college, do we expect that their other characteristics will be the same? For example: family income, ability, interest in education?
- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?

Omitted Variable Bias

- There is an alternate approach that can be useful in understanding the sources of bias in regressions induced by differences in unobservable characteristics.
- Recall from the definition of linear regression that the coefficient of interest can be calculated as follows.

$$\hat{\beta} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$$

- Assume that $Y_i = X_i\beta_0 + \epsilon_i = X_i\beta_0 + \gamma W_i + \eta_i$ where $Cov(W, X) \neq 0$ and $\eta \perp X$.
- Now we regress Y on X.
  - Assume W is an omitted variable.
  - This is a source of omitted variable bias.

Omitted Variable Bias

- There is an alternate approach that can be useful in understanding the sources of bias in regressions induced by differences in unobservable characteristics.
- Recall from the definition of linear regression that the coefficient of interest can be calculated as follows.

$$\hat{\beta} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$$

- Assume that $Y_i = X_i\beta_0 + \epsilon_i = X_i\beta_0 + \gamma W_i + \eta_i$ where $Cov(W, X) \neq 0$ and $\eta \perp X$.
- Now we regress Y on X.
  - Assume W is an omitted variable.
  - This is a source of omitted variable bias.

Omitted Variable Bias

- There is an alternate approach that can be useful in understanding the sources of bias in regressions induced by differences in unobservable characteristics.
- Recall from the definition of linear regression that the coefficient of interest can be calculated as follows.

$$\hat{\beta} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$$

- Assume that $Y_i = X_i \beta_0 + \epsilon_i = X_i \beta_0 + \gamma W_i + \eta_i$ where $Cov(W, X) \neq 0$ and $\eta \perp X$.
- Now we regress Y on X.
  - Assume W is an omitted variable.
  - This is a source of omitted variable bias.

Omitted Variable Bias

- There is an alternate approach that can be useful in understanding the sources of bias in regressions induced by differences in unobservable characteristics.
- Recall from the definition of linear regression that the coefficient of interest can be calculated as follows.

$$\hat{\beta} = \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)}$$

- Assume that $Y_i = X_i\beta_0 + \epsilon_i = X_i\beta_0 + \gamma W_i + \eta_i$ where $Cov(W,X) \neq 0$ and $\eta \perp X$.
- Now we regress Y on X.
  - Assume W is an omitted variable.
  - This is a source of omitted variable bias.

Defining the bias tern

- Observe that

$$\frac{Cov(X,Y)}{Var(X)} = \frac{Cov(X, X\beta_0 + \epsilon)}{Var(X)} = \beta_0 + \frac{Cov(X,\epsilon)}{Var(X)}$$

- Let's look at the final term.

$$\frac{Cov(X,\epsilon)}{Var(X)} = \frac{Cov(X,\gamma W + \eta)}{Var(X)} = \frac{\gamma Cov(X,W)}{Var(X)} + \underbrace{\frac{Cov(X,\eta)}{Var(X)}}_{=0,\text{ why??}}$$

- How does this affect our estimation?

Defining the bias tern

- Observe that

$$\frac{Cov(X,Y)}{Var(X)} = \frac{Cov(X, X\beta_0 + \epsilon)}{Var(X)} = \beta_0 + \frac{Cov(X,\epsilon)}{Var(X)}$$

- Let's look at the final term.

$$\frac{Cov(X,\epsilon)}{Var(X)} = \frac{Cov(X, \gamma W + \eta)}{Var(X)} = \frac{\gamma Cov(X,W)}{Var(X)} + \underbrace{\frac{Cov(X,\eta)}{Var(X)}}_{=0,\ \text{why??}}$$

- How does this affect our estimation?

Defining the bias tern

- Observe that

$$\frac{Cov(X,Y)}{Var(X)} = \frac{Cov(X, X\beta_0 + \epsilon)}{Var(X)} = \beta_0 + \frac{Cov(X,\epsilon)}{Var(X)}$$

- Let's look at the final term.

$$\frac{Cov(X,\epsilon)}{Var(X)} = \frac{Cov(X,\gamma W + \eta)}{Var(X)} = \frac{\gamma Cov(X,W)}{Var(X)} + \underbrace{\frac{Cov(X,\eta)}{Var(X)}}_{=0,\text{ why??}}$$

- How does this affect our estimation?

Signing the bias term

- The direction of the bias thus depends on the correlation between X_i, the independent variable of interest, and the unobserved variable.
- If $Cov(X, W) > 0$, then $\hat{\beta}_0 > \beta_0$: our estimate is biased upwards.
- If $Cov(X, W) < 0$, then $\hat{\beta}_0 < \beta_0$: our estimate is biased downwards.
- Let's return to the two cases we considered earlier in estimating the return to college education.

Signing the bias term

- The direction of the bias thus depends on the correlation between X_i, the independent variable of interest, and the unobserved variable.
- If $Cov(X, W) > 0$, then $\hat{\beta}_0 > \beta_0$: our estimate is biased upwards.
- If $Cov(X, W) < 0$, then $\hat{\beta}_0 < \beta_0$: our estimate is biased downwards.
- Let's return to the two cases we considered earlier in estimating the return to college education.

Signing the bias term

- The direction of the bias thus depends on the correlation between X_i, the independent variable of interest, and the unobserved variable.
- If $Cov(X, W) > 0$, then $\hat{\beta}_0 > \beta_0$: our estimate is biased upwards.
- If $Cov(X, W) < 0$, then $\hat{\beta}_0 < \beta_0$: our estimate is biased downwards.
- Let's return to the two cases we considered earlier in estimating the return to college education.

Signing the bias term

- The direction of the bias thus depends on the correlation between X_i, the independent variable of interest, and the unobserved variable.
- If $Cov(X, W) > 0$, then $\hat{\beta}_0 > \beta_0$: our estimate is biased upwards.
- If $Cov(X, W) < 0$, then $\hat{\beta}_0 < \beta_0$: our estimate is biased downwards.
- Let's return to the two cases we considered earlier in estimating the return to college education.

Signing the bias term: an example

- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?

Signing the bias term: an example

- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?

Signing the bias term: an example

- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?

Signing the bias term: an example

- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?

Signing the bias term: an example

- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?

Signing the bias term: an example

- If students who go to college are on average wealthier than students who do not go to college ex ante, how will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?
- What if students who go to college are on average less motivated to work (i.e., more lazy)? How will this affect our estimate of returns to education?
  - What is W?
  - Is $Cov(X, W)$ positive or negative?

Jones and Olken, Hit or Miss

- This paper takes advantage of the fact that sometimes the identity of political leaders change for reasons that are not linked to underlying political conditions.
- Specifically: periodically, assassinations lead to sudden changes in political leadership.
- Should we then compare countries where a political leader was assassinated to countries where a leader was not? E.g., compare the U.S. in 1963 to Canada in 1963?
- What is (intuitively) the problem with this strategy?

Jones and Olken, Hit or Miss

- This paper takes advantage of the fact that sometimes the identity of political leaders change for reasons that are not linked to underlying political conditions.
- Specifically: periodically, assassinations lead to sudden changes in political leadership.
- Should we then compare countries where a political leader was assassinated to countries where a leader was not? E.g., compare the U.S. in 1963 to Canada in 1963?
- What is (intuitively) the problem with this strategy?

Jones and Olken, Hit or Miss

- This paper takes advantage of the fact that sometimes the identity of political leaders change for reasons that are not linked to underlying political conditions.
- Specifically: periodically, assassinations lead to sudden changes in political leadership.
- Should we then compare countries where a political leader was assassinated to countries where a leader was not? E.g., compare the U.S. in 1963 to Canada in 1963?
- What is (intuitively) the problem with this strategy?

Jones and Olken, Hit or Miss

- This paper takes advantage of the fact that sometimes the identity of political leaders change for reasons that are not linked to underlying political conditions.
- Specifically: periodically, assassinations lead to sudden changes in political leadership.
- Should we then compare countries where a political leader was assassinated to countries where a leader was not? E.g., compare the U.S. in 1963 to Canada in 1963?
- What is (intuitively) the problem with this strategy?

Core empirical intuition

- This paper uses a different approach: it compares countries and years in which an assassination attempt was **successful** to countries and years in which it was **unsuccessful**.
- Underlying intuition: underlying political conditions in countries and years in which assassinations were attempted are similar.
- **And**, whether a given attempt succeeds is (more or less) random.
- E.g., in 1939 there was an attempt on Hitler's life; he was saved by the fact that bad weather led to the cancellation of his flight, and he left early for a train.

Core empirical intuition

- This paper uses a different approach: it compares countries and years in which an assassination attempt was **successful** to countries and years in which it was **unsuccessful**.
- Underlying intuition: underlying political conditions in countries and years in which assassinations were attempted are similar.
- **And**, whether a given attempt succeeds is (more or less) random.
- E.g., in 1939 there was an attempt on Hitler's life; he was saved by the fact that bad weather led to the cancellation of his flight, and he left early for a train.

Core empirical intuition

- This paper uses a different approach: it compares countries and years in which an assassination attempt was **successful** to countries and years in which it was **unsuccessful**.
- Underlying intuition: underlying political conditions in countries and years in which assassinations were attempted are similar.
- **And**, whether a given attempt succeeds is (more or less) random.
- E.g., in 1939 there was an attempt on Hitler's life; he was saved by the fact that bad weather led to the cancellation of his flight, and he left early for a train.

Core empirical intuition

- This paper uses a different approach: it compares countries and years in which an assassination attempt was **successful** to countries and years in which it was **unsuccessful**.
- Underlying intuition: underlying political conditions in countries and years in which assassinations were attempted are similar.
- **And**, whether a given attempt succeeds is (more or less) random.
- E.g., in 1939 there was an attempt on Hitler's life; he was saved by the fact that bad weather led to the cancellation of his flight, and he left early for a train.

Primary empirical specification

- Key empirical specification is the following:

$$y_i = \beta SUCCESS_i + \gamma X_i + \epsilon_i$$

- Note i denotes a country-year in which there is an assassination attempt; y_i is an outcome of interest (institutional change or change in war status), and X_i is a set of other control variables.

Primary empirical specification

- Key empirical specification is the following:

$$y_i = \beta SUCCESS_i + \gamma X_i + \epsilon_i$$

- Note i denotes a country-year in which there is an assassination attempt; y_i is an outcome of interest (institutional change or change in war status), and X_i is a set of other control variables.

Key definition

- In this lecture, and in future lectures, I'm going to discuss identification assumptions frequently.
- What is an identification assumption? It's an assumption that we have to make to interpret the results as causal, rather than a correlation.
- For example, in our hypothetical comparison of students who go to college and students who don't, the identification assumption would be that students who do and don't attend college are otherwise alike.
- Good assumption? Probably not - so we can't interpret these results as causal.
- Economists (almost) always want to interpret results causally - if we don't believe the identification assumption, then we can't conclude causality, and generally that's bad.

Key definition

- In this lecture, and in future lectures, I'm going to discuss identification assumptions frequently.
- What is an identification assumption? It's an assumption that we have to make to interpret the results as causal, rather than a correlation.
- For example, in our hypothetical comparison of students who go to college and students who don't, the identification assumption would be that students who do and don't attend college are otherwise alike.
- Good assumption? Probably not - so we can't interpret these results as causal.
- Economists (almost) always want to interpret results causally - if we don't believe the identification assumption, then we can't conclude causality, and generally that's bad.

Key definition

- In this lecture, and in future lectures, I'm going to discuss identification assumptions frequently.
- What is an identification assumption? It's an assumption that we have to make to interpret the results as causal, rather than a correlation.
- For example, in our hypothetical comparison of students who go to college and students who don't, the identification assumption would be that students who do and don't attend college are otherwise alike.
- Good assumption? Probably not - so we can't interpret these results as causal.
- Economists (almost) always want to interpret results causally - if we don't believe the identification assumption, then we can't conclude causality, and generally that's bad.

Key definition

- In this lecture, and in future lectures, I'm going to discuss identification assumptions frequently.
- What is an identification assumption? It's an assumption that we have to make to interpret the results as causal, rather than a correlation.
- For example, in our hypothetical comparison of students who go to college and students who don't, the identification assumption would be that students who do and don't attend college are otherwise alike.
- Good assumption? Probably not - so we can't interpret these results as causal.
- Economists (almost) always want to interpret results causally - if we don't believe the identification assumption, then we can't conclude causality, and generally that's bad.

Key definition

- In this lecture, and in future lectures, I'm going to discuss identification assumptions frequently.
- What is an identification assumption? It's an assumption that we have to make to interpret the results as causal, rather than a correlation.
- For example, in our hypothetical comparison of students who go to college and students who don't, the identification assumption would be that students who do and don't attend college are otherwise alike.
- Good assumption? Probably not - so we can't interpret these results as causal.
- Economists (almost) always want to interpret results causally - if we don't believe the identification assumption, then we can't conclude causality, and generally that's bad.

Discussion questions

- What is the key identifying assumption here? How do Jones and Olken evaluate whether or not it is true?
- What are the primary effects of assassinations, and are they evident for all polities?
- What are potential sources of bias in this paper?
- What do the results about the impact of leadership transitions on conflict suggest?
- What are the policy implications? Have we learned anything useful from this paper, or is it a vanity exercise?

Discussion questions

- What is the key identifying assumption here? How do Jones and Olken evaluate whether or not it is true?
- What are the primary effects of assassinations, and are they evident for all polities?
- What are potential sources of bias in this paper?
- What do the results about the impact of leadership transitions on conflict suggest?
- What are the policy implications? Have we learned anything useful from this paper, or is it a vanity exercise?

Discussion questions

- What is the key identifying assumption here? How do Jones and Olken evaluate whether or not it is true?
- What are the primary effects of assassinations, and are they evident for all polities?
- What are potential sources of bias in this paper?
- What do the results about the impact of leadership transitions on conflict suggest?
- What are the policy implications? Have we learned anything useful from this paper, or is it a vanity exercise?

Discussion questions

- What is the key identifying assumption here? How do Jones and Olken evaluate whether or not it is true?
- What are the primary effects of assassinations, and are they evident for all polities?
- What are potential sources of bias in this paper?
- What do the results about the impact of leadership transitions on conflict suggest?
- What are the policy implications? Have we learned anything useful from this paper, or is it a vanity exercise?

Discussion questions

- What is the key identifying assumption here? How do Jones and Olken evaluate whether or not it is true?
- What are the primary effects of assassinations, and are they evident for all polities?
- What are potential sources of bias in this paper?
- What do the results about the impact of leadership transitions on conflict suggest?
- What are the policy implications? Have we learned anything useful from this paper, or is it a vanity exercise?

Another question about leaders' performance

- The previous paper analyzed the impact of (violent) leadership change on political outcomes.
- Leaders can also affect economic outcomes, particularly growth.
- Developing countries frequently see sharp reversals in growth; this paper asks whether leadership changes might account for this.
- Clearly, however, there is an underlying identification challenge: leadership transitions are themselves driven by underlying economic conditions.

Another question about leaders' performance

- The previous paper analyzed the impact of (violent) leadership change on political outcomes.
- Leaders can also affect economic outcomes, particularly growth.
- Developing countries frequently see sharp reversals in growth; this paper asks whether leadership changes might account for this.
- Clearly, however, there is an underlying identification challenge: leadership transitions are themselves driven by underlying economic conditions.

Another question about leaders' performance

- The previous paper analyzed the impact of (violent) leadership change on political outcomes.
- Leaders can also affect economic outcomes, particularly growth.
- Developing countries frequently see sharp reversals in growth; this paper asks whether leadership changes might account for this.
- Clearly, however, there is an underlying identification challenge: leadership transitions are themselves driven by underlying economic conditions.

Another question about leaders' performance

- The previous paper analyzed the impact of (violent) leadership change on political outcomes.
- Leaders can also affect economic outcomes, particularly growth.
- Developing countries frequently see sharp reversals in growth; this paper asks whether leadership changes might account for this.
- Clearly, however, there is an underlying identification challenge: leadership transitions are themselves driven by underlying economic conditions.

Summary statistics

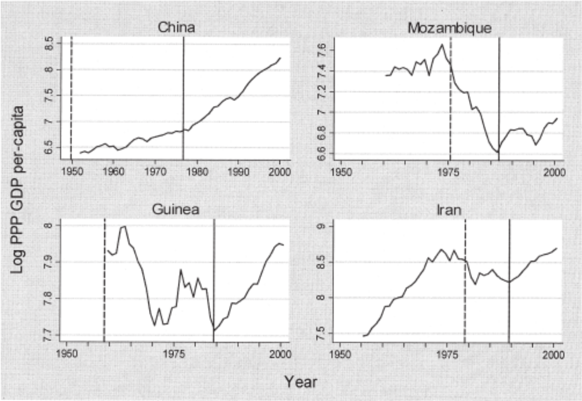

FIGURE I
Growth and Leader Deaths

Identification strategy

- The authors identify cases where leaders died, either from natural causes or in an accident, from 1945 to 2000.
- Empirical strategy is more complex, but I will summarize it here.
- Authors estimate the following equation.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- PRE and POST initially refer to 5 years before and after the death.

Identification strategy

- The authors identify cases where leaders died, either from natural causes or in an accident, from 1945 to 2000.
- Empirical strategy is more complex, but I will summarize it here.
- Authors estimate the following equation.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- PRE and POST initially refer to 5 years before and after the death.

Identification strategy

- The authors identify cases where leaders died, either from natural causes or in an accident, from 1945 to 2000.
- Empirical strategy is more complex, but I will summarize it here.
- Authors estimate the following equation.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- PRE and POST initially refer to 5 years before and after the death.

Identification strategy

- The authors identify cases where leaders died, either from natural causes or in an accident, from 1945 to 2000.
- Empirical strategy is more complex, but I will summarize it here.
- Authors estimate the following equation.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- PRE and POST initially refer to 5 years before and after the death.

Identification strategy

- The authors postulate the following growth process

$$g_{it} = v_i + \theta l_{it} + \epsilon_{it}$$

where g_{it} is growth in country i and time t, v_i is a fixed effect for country i, and ϵ_{it} is a normally distributed error term.

- Leaders are selected as follows.

$$l_{it} = \begin{cases} l_{it-1} & P(\delta_0 g_{it} + \delta_1 g_{i,t-1} + \ldots) \\ l' & 1 - P(\delta_0 g_{it} + \delta_1 g_{i,t-1} + \ldots) \end{cases}$$

Identification strategy

- The authors postulate the following growth process

$$g_{it} = v_i + \theta l_{it} + \epsilon_{it}$$

where g_{it} is growth in country i and time t, v_i is a fixed effect for country i, and ϵ_{it} is a normally distributed error term.

- Leaders are selected as follows.

$$l_{it} = \begin{cases} l_{it-1} & P(\delta_0 g_{it} + \delta_1 g_{i,t-1} + \ldots) \\ l' & 1 - P(\delta_0 g_{it} + \delta_1 g_{i,t-1} + \ldots) \end{cases}$$

Identification strategy, cont.

- The general structure is as follows: growth may depend on leadership quality l_{it}, but we also know that leadership quality l_{it} depends on past growth.
- The hypothesis of interest is whether $\theta = 0$, and the authors devise a Wald test to test this hypothesis - the econometric details of which are beyond our scope here.
- In order to implement this test, they estimate the following regression.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- Note this equation estimates separate coefficients α and β for each leader death z. PRE and POST initially refer to 5 years before and after the death.

Identification strategy, cont.

- The general structure is as follows: growth may depend on leadership quality l_{it}, but we also know that leadership quality l_{it} depends on past growth.
- The hypothesis of interest is whether $\theta = 0$, and the authors devise a Wald test to test this hypothesis - the econometric details of which are beyond our scope here.
- In order to implement this test, they estimate the following regression.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- Note this equation estimates separate coefficients α and β for each leader death z. PRE and POST initially refer to 5 years before and after the death.

Identification strategy, cont.

- The general structure is as follows: growth may depend on leadership quality l_{it}, but we also know that leadership quality l_{it} depends on past growth.
- The hypothesis of interest is whether $\theta = 0$, and the authors devise a Wald test to test this hypothesis - the econometric details of which are beyond our scope here.
- In order to implement this test, they estimate the following regression.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- Note this equation estimates separate coefficients α and β for each leader death z. PRE and POST initially refer to 5 years before and after the death.

Identification strategy, cont.

- The general structure is as follows: growth may depend on leadership quality l_{it}, but we also know that leadership quality l_{it} depends on past growth.
- The hypothesis of interest is whether $\theta = 0$, and the authors devise a Wald test to test this hypothesis - the econometric details of which are beyond our scope here.
- In order to implement this test, they estimate the following regression.

$$g_{yt} = \alpha_z PRE_z + \beta_z POST_z + v_i + v_t + \epsilon_{it}$$

- Note this equation estimates separate coefficients α and β for each leader death z. PRE and POST initially refer to 5 years before and after the death.

Discussion questions for this paper

- What impact do the authors find of leader quality on growth?
- What about the possibility that it is the dying leader in office who matters (e.g., because he starts to lose control), rather than the actual leadership transition?
- What heterogeneous effects do the authors find? Do you believe these heterogeneous effects to be plausible?
- Again, is there anything useful to be gleaned from this paper in terms of policy implications? (No is an acceptable answer, but please justify!)

Discussion questions for this paper

- What impact do the authors find of leader quality on growth?
- What about the possibility that it is the dying leader in office who matters (e.g., because he starts to lose control), rather than the actual leadership transition?
- What heterogeneous effects do the authors find? Do you believe these heterogeneous effects to be plausible?
- Again, is there anything useful to be gleaned from this paper in terms of policy implications? (No is an acceptable answer, but please justify!)

Discussion questions for this paper

- What impact do the authors find of leader quality on growth?
- What about the possibility that it is the dying leader in office who matters (e.g., because he starts to lose control), rather than the actual leadership transition?
- What heterogeneous effects do the authors find? Do you believe these heterogeneous effects to be plausible?
- Again, is there anything useful to be gleaned from this paper in terms of policy implications? (No is an acceptable answer, but please justify!)

Discussion questions for this paper

- What impact do the authors find of leader quality on growth?
- What about the possibility that it is the dying leader in office who matters (e.g., because he starts to lose control), rather than the actual leadership transition?
- What heterogeneous effects do the authors find? Do you believe these heterogeneous effects to be plausible?
- Again, is there anything useful to be gleaned from this paper in terms of policy implications? (No is an acceptable answer, but please justify!)