# ECMA31000: Introduction to Empirical Analysis
# Hypothesis Testing; Linear Regression I

Joe Hardwick

University of Chicago

Autumn 2021

# Outline

- Last time:
    - Confidence Intervals, Hypothesis Testing.
- Today:
    - Non-testable hypotheses
    - Linear Regression

# Hypothesis Testing: Definitions

- The null hypothesis is a subset $\Theta_0 \subset \Theta$ of hypothesised values for $\theta_0$, written as:
$$H_0 : \theta_0 \in \Theta_0.$$
- If $\Theta_0$ is a singleton, the null hypothesis is called a simple hypothesis. If not, the hypothesis is composite.
- e.g. If $X \sim \mathcal{N}(\mu, 1)$, $H_0 : \mu = 0$ is a simple null, $H_0 : \mu \leq 0$ is a composite null.
- A test of $H_0$ is therefore a test of whether the data were generated by some $F_\theta$ such that $\theta \in \Theta_0$.
- The alternative hypothesis is $\Theta \backslash \Theta_0$.

# Hypothesis Testing: Definitions

- A test of $H_0$ is a function $\phi_n(X_1, \ldots, X_n) \to \{0, 1\}$.
- We reject $H_0$ at sample size $n$ iff $\phi_n = 1$.
- e.g. If $T_n(X_1, \ldots, X_n)$ is a sequence of statistics, and $c_n$ a sequence of real numbers,

$$\phi_n(X_1, \ldots, X_n) = \mathbf{1}(T_n > c_n)$$

is a test which rejects $H_0$ iff $T_n > c_n$.

# Two undesirable outcomes

- Suppose $\theta_0 \in \Theta_0$ but $\phi_n = 1$. This is called a Type I Error.
- Suppose $\theta_0 \notin \Theta_0$ but $\phi_n = 0$. This is called a Type II Error.
- It is customary to control the probability of a Type I Error first, and then minimize the probability of a Type II Error subject to this constraint.
- The power function associated with $\phi_n$ is the function

$$\beta_n(\theta) = \mathrm{P}_\theta\left(\phi_n(X_1, \ldots, X_n) = 1\right),$$

which is the probability that $\phi_n$ rejects $H_0$ if the true parameter is $\theta$.

# Properties of tests

- Given a significance level $\alpha$, we select a test $\phi_n$ such that

$$\beta_n(\theta) \leq \alpha \quad \text{for all } \theta \in \Theta_0,$$

  or, alternatively

$$\sup_{\theta \in \Theta_0} \beta_n(\theta) \leq \alpha.$$

- Similarly, given a test $\phi_n$, the size of $\phi_n$ with power function $\beta_n$ is

$$\alpha := \sup_{\theta \in \Theta_0} \beta_n(\theta).$$

- The probability that $\phi_n$ rejects when $H_0$ is true cannot exceed $\alpha$.

# Example: Normal Distribution

- Suppose $\{X_i\}_{i=1}^{n}$ is an iid sample from $\mathcal{N}(\mu, 1)$. We wish to test $H_0 : \mu \leq 0$ vs. $H_1 : \mu > 0$, so

$$\Theta_0 = \left\{\mu \in (-\infty, 0]; \sigma^2 \text{ known}\right\}.$$

.

- Let $c > 0$ and consider the test

$$\phi_n^{norm}(X_1, \ldots, X_n) = \mathbf{1}\left(\bar{X}_n > c\right).$$

- The power function is

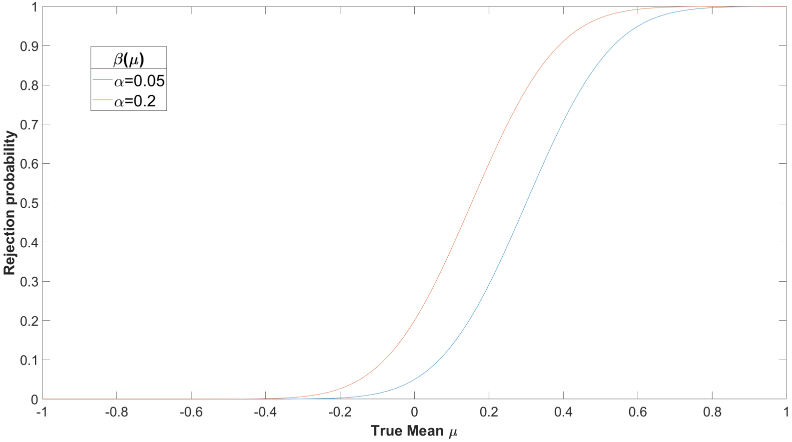
$$\begin{aligned}
\beta_n^{norm}(\mu) &= \mathrm{P}_\mu\left(\bar{X}_n > c\right) \\
&= \mathrm{P}_\mu\left(\sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma > \sqrt{n}(c - \mu)/\sigma\right) \\
&= 1 - \Phi\left(\sqrt{n}(c - \mu)/\sigma\right).
\end{aligned}$$

# Example: Normal Distribution

- The uniformly (in $\mu$) most powerful test of size $\alpha$ takes the form
$$\phi_n^{norm}\left(X_1,\ldots,X_n\right)=\mathbf{1}\left(\bar{X}_n>c^*\left(\alpha\right)\right),$$
where $c^*\left(\alpha\right)=z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$. Its power function is:
$$\begin{aligned}\beta_n^{norm}\left(\mu\right)&=1-\Phi\left(\sqrt{n}\left(c^*-\mu\right)/\sigma\right)\\&=1-\Phi\left(z_{1-\alpha}-\frac{\mu\sqrt{n}}{\sigma}\right).\end{aligned}$$

# Example: Normal Distribution

- Probability of rejecting $H_0$ is known for any true value of $\mu$ if, e.g. $X \sim \mathcal{N}(\mu, 1)$. This is the power curve:



# Nuisance Parameters

- Suppose $\{X_i\}_{i=1}^n$ is an iid sample from $\mathcal{N}\left(\mu, \sigma^2\right)$. We wish to test $H_0 : \mu \leq 0$ vs. $H_1 : \mu > 0$, so

$$\Theta_0 = \left\{\mu \in (-\infty, 0]; \sigma^2 > 0\right\}, \quad \Theta_1 = \left\{\mu \in (0, \infty); \sigma^2 > 0\right\}.$$

- $\sigma^2$ is unobserved, and is called a nuisance parameter because it isn't of immediate interest but changing it may alter the distribution of our test statistic under the null and alternative hypotheses.
- One solution: Choose a test statistic whose distribution under $H_0$ controls size by accounting for changing $\sigma^2$ (e.g. the t-test).
- Problem: We've now added a restriction to the class of tests, so the most powerful one will be less powerful.

# Nuisance Parameters

- Let $c > 0$ and consider the t-test:

$$\phi_n^t(X_1, \ldots, X_n) = \mathbf{1}\left(\frac{\sqrt{n}\left(\bar{X}_n - 0\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}} > c\right).$$

- If $\mu \in \mathbb{R}$, $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_{n-1}$ and $Z$ is independent of $V$, then

$$\frac{Z + \mu}{\sqrt{V/(n-1)}} \sim t_{\mu,n-1}$$

has a non-central $t-$distribution with $n-1$ degrees of freedom and non-centrality parameter $\mu$, where $t_{\mu,n-1}$ denotes its CDF.

# Power of the t-test

- Let

$$T_n = \frac{\sqrt{n}\left(\bar{X}_n - 0\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}} = \frac{\sqrt{n}\left(\bar{X}_n - 0\right)/\sigma}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2/\sigma^2}}.$$

- Therefore:

$$T_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma + \sqrt{n}\mu/\sigma}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2/\sigma^2}} \sim t_{\frac{\sqrt{n}\mu}{\sigma}, n-1}.$$

- The power function is

$$\begin{aligned} \beta_n^t(\mu) &= \mathrm{P}_{\mu,\sigma}\left(T_n > c\right) \\ &= 1 - t_{\frac{\sqrt{n}\mu}{\sigma}, n-1}(c). \end{aligned}$$

# Power of the t-test

- When $\alpha \geq 0.5$, the uniformly (in $(\mu, \sigma)$) most powerful test of size $\alpha$ takes the form

$$\phi_n^t(X_1, \ldots, X_n) = \mathbf{1}\left(T_n > t^*_{n-1,1-\alpha}\right),$$

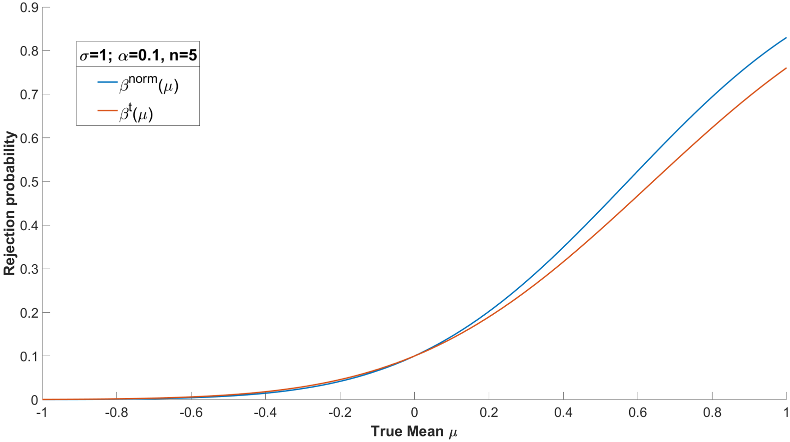
where $t^*_{n-1,1-\alpha}$ is the $1-\alpha$ quantile of the (central) $t_{n-1}$distribution. Its power function is:

$$\beta_n^t(\mu) = 1 - t_{\frac{\sqrt{n}\mu}{\sigma},n-1}\left(t^*_{n-1,1-\alpha}\right).$$

- (When $\alpha < 0.5$, the most powerful test at the point $(\mu, \sigma)$ varies with $(\mu, \sigma)$, so there is no UMP test).

# Power of the t-test

- Power curves for the t-test and UMP test when it is known that $\sigma = 1$:



# Power of the t-test

- Expanding the statistical model to allow $\sigma^2 \neq 1$ made it harder to satisfy
$$\sup_{\theta \in \Theta_0} \beta_n(\theta) \leq \alpha.$$
- The t-test is less powerful than $\phi_n^{norm}(X_1, \ldots, X_n)$ when $\sigma$ is known, but has significance level $\alpha$ for any value of $\sigma$ when it is unknown.
- Moreover, $\phi_n^{norm}(X_1, \ldots, X_n)$ requires us to specify a value $\sigma = \sigma_0$, which, if incorrect, may cause the test to "over-reject" when $H_0$ is true:
$$\beta_n^{norm}(\mu) = 1 - \Phi\left(z_{1-\alpha} \cdot \frac{\sigma_0}{\sigma} - \frac{\mu\sqrt{n}}{\sigma}\right),$$
which roughly equals 0.5 if $\mu = 0$ and $\sigma_0/\sigma \approx 0$.

# Testing the mean in general

- Now suppose we drop the restriction to normal distributions. The statistical model is greatly expanded:

$$\mathcal{F} = \left\{ F \text{ is a CDF and } 0 < \sigma^2(F) < \infty \right\}.$$

- This is a 'non-parametric' class of distributions.

**Theorem**

*Let $\mathcal{F}_0 = \{F \in \mathcal{F} : \mu(F) = 0\}$ be the null hypothesis. Any test $\phi$ which has size $\alpha$ has power $\leq \alpha$ for any $F \in \mathcal{F} \setminus \mathcal{F}_0$. Any test which has power $\beta$ against some alternative $F \in \mathcal{F} \setminus \mathcal{F}_0$ has size $\geq \beta$.*

# Example: RDD

# Linear Regression: Definitions

- Let $(Y, X, U)$ be a random vector such that $Y$ and $U$ are scalar random variables and $X \in \mathbb{R}^{k+1}$.
- Assume the first component of $X$ equals 1:

$$X = (X_0, X_1 \ldots, X_k),$$

  where $X_0 = 1$.
- Let $\beta = (\beta_0, \ldots, \beta_k) \in \mathbb{R}^{k+1}$ be a constant vector of unknown parameters such that

$$Y = X'\beta + U.$$

- $\beta_0$ is the *intercept* and the remaining $\beta_j$ are the *slope* parameters.

# Interpretations

- Note that given $(Y, X)$, and any $\beta$, we may construct $U$ such that the representation

$$Y = X'\beta + U$$

  holds without loss of generality (simply take $U = Y - X'\beta$).
- We now consider additional restrictions on this relationship which lead us to interpret the quantity $X'\beta$ as a "best predictor" of $Y$, and also (uniquely) determine $\beta$.

# Linear Conditional Expectation

- Consider
$$Y = X'\beta + U.$$
- Suppose $\mathrm{E}\left(Y|X\right) = X'\beta$ and define $U = Y - \mathrm{E}\left(Y|X\right)$.
- Then $\mathrm{E}\left(U|X\right) = 0$ and so $\mathrm{E}\left(U\right) = \mathrm{E}\left(UX\right) = 0$, which means $Cov\left(U, X\right) = 0$. This model is often written equivalently as
$$Y = X'\beta + U; \qquad \mathrm{E}\left(U|X\right) = 0.$$
- The vector $\beta$ is a feature of the joint distribution of $\left(Y, X\right)$ which represents the conditional expectation.

# Linear Conditional Expectation: Example I

- This model merely posits that if $X_j$ increases by 1 unit, the best predictor of $Y$ under square loss increases/decreases by $\beta_j$ units.
- It does NOT claim that an increase in $X_j$ *causes* a change in $Y$.
- For example, suppose we observe an iid sample of $\{Y_i, D_i\}_{i=1}^n$ where $Y_i$ is the scrap rate of factory $i$, and

$$D_i = \begin{cases} 1 & \text{if factory } i \text{ receives job training grant,} \\ 0 & \text{otherwise.} \end{cases}$$

# Linear Conditional Expectation: Example I

- Without loss of generality (see problem set 5):

$$\mathrm{E}\left(Y_i | D_i\right) = \beta_0 + \beta_1 D_i.$$

- Hence, $\beta_0 = \mathrm{E}\left(Y | D = 0\right)$ and $\beta_1 = \mathrm{E}\left(Y | D = 1\right) - \mathrm{E}\left(Y | D = 0\right)$.
- Suppose $\beta_1 < 0$, so the mean scrap rate of factories without a grant is higher than for those with a grant.
- This does NOT imply receiving a grant negatively impacts productivity, especially if grants are first come first served and the less efficient factories applied first.
- We lack a model of how $Y$ is determined as a function of $D$ and other variables (like prior scrap rate).

# Linear Conditional Expectation: Example I

- A more complete model would consider *potential outcomes*:

$$Y_i = D_i Y_{i1} + (1 - D_i)\, Y_{i0},$$

where $Y_{i1}$ is the scrap rate of firm $i$ if they receive the grant and $Y_{i0}$ is their scrap rate without. Note that

$$\begin{aligned}\beta_1 &= \mathrm{E}\left(Y_1 | D = 1\right) - \mathrm{E}\left(Y_0 | D = 0\right) \\ &= \mathrm{E}\left(Y_1 | D = 1\right) - \mathrm{E}\left(Y_0 | D = 1\right) \\ &\quad + \mathrm{E}\left(Y_0 | D = 1\right) - \mathrm{E}\left(Y_0 | D = 0\right).\end{aligned}$$

- It may also hold that $\beta_1 = 0$ but offering training grants is highly effective, if the baseline productivity of firms that don't enrol is much higher.

# Linear Conditional Expectation: Example II

- If $(Y, X)$ have a multivariate normal distribution, with $Y \in \mathbb{R}, X \in \mathbb{R}^k$, the assumption

$$\mathrm{E}(Y|X) = \beta_0 + X'\beta_1 = \left(1, X'\right)\beta$$

  holds for some vector $\beta = (\beta_0, \beta_1')' \in \mathbb{R}^{k+1}$.
- $\beta$ is unique if $X$ is non-degenerate (no component of $X$ is a linear function of the others).
- If $X$ is degenerate, $\beta$ is not uniquely determined because multiple values of $\beta$ lead to the same value of $(1, X')\,\beta$. To see this, suppose for some constants $a_1, \ldots, a_k$ that

$$1 = a_1 X_1 + \cdots + a_k X_k.$$

# Linear Conditional Expectation: Example II

- Then:

$$
\begin{aligned}
\left(1, X'\right)\beta &= \beta_0 + X_1\beta_1 + \cdots + X_k\beta_k \\
&= \left(a_1X_1 + \ldots a_kX_k\right)\beta_0 + \cdots + X_k\beta_k \\
&= X_1\left(a_1\beta_0 + \beta_1\right) + \cdots + X_k\left(a_k\beta_0 + \beta_k\right) \\
&= \left(1, X'\right)\tilde{\beta},
\end{aligned}
$$

  where

$$\tilde{\beta} = \left(0, a_1\beta_0 + \beta_1, \ldots, a_k\beta_0 + \beta_k\right).$$

- Therefore, for any $m > 0$, $\left(1, X'\right)\beta = \left(1, X'\right)\beta_m^*$, where

$$\beta_m^* = \beta + m\left(\tilde{\beta} - \beta\right).$$

# Questions?

# Best Linear approximation to $\mathrm{E}(Y|X)$

- In the absence of joint normality, or if $X$ can take more than two values, $\mathrm{E}(Y|X)$ is generally non-linear.
- Suppose $\mathrm{E}\left(Y^2\right) < \infty$ and $\mathrm{E}\left(X_j^2\right) < \infty$ for each $j = 1, \ldots, k$.
- Consider the problem

$$\min_{b \in \mathbb{R}^{k+1}} \mathrm{E}\left(\mathrm{E}(Y|X) - X'b\right)^2 .$$

- The solution is the best linear approximation to $\mathrm{E}(Y|X)$ under square loss.

# Best Linear predictor of $Y$

- Any solution to this problem is also represents a best linear predictor of $Y$, since

$$\begin{aligned}
\mathrm{E}\left[\left(\mathrm{E}(Y|X) - X'b\right)^2\right] &= \mathrm{E}\left(\mathrm{E}(Y|X) - Y + Y - X'b\right)^2 \\
&= \mathrm{E}(Y - \mathrm{E}(Y|X))^2 \\
&+ \mathrm{E}\left(Y - X'b\right)^2 \\
&- 2\mathrm{E}\left[(Y - \mathrm{E}(Y|X))\left(Y - X'b\right)\right].
\end{aligned}$$

- The law of iterated expectation implies

$$\mathrm{E}\left[(Y - \mathrm{E}(Y|X))\left(Y - X'b\right)\right] = \mathrm{E}(Y(Y - \mathrm{E}(Y|X))).$$

- Therefore,

$$\mathrm{E}\left[\left(\mathrm{E}(Y|X) - X'b\right)^2\right] = \mathrm{E}\left(Y - X'b\right)^2 + Constant.$$

# Best Linear predictor of $Y$

- Since $\mathrm{E}\,(Y - X'b)^2$ is convex in $b$, any minimizer $b^*$ is characterized by the FOC

$$\frac{\partial}{\partial b}\mathrm{E}\left(Y - X'b^*\right)^2 = 2\mathrm{E}\left(XX'\right)b^* - 2\mathrm{E}\,(XY) = 0.$$

- Hence $b^*$ must satisfy

$$\mathrm{E}\left(X\left(Y - X'b^*\right)\right) = 0.$$

- Defining $U = Y - X'b^*$ yields

$$\mathrm{E}\,(XU) = 0.$$

# Causal Model

- Suppose we assume that $Y$ is determined by the equation

$$Y = X'\beta + U,$$

  where $X$ is observed and $U$ is not.
- The *ceteris paribus* effect of $X_j$ on $Y$ holding the other elements of $X$ and $U$ constant is $\beta_j$.
- We may assume $\mathrm{E}(U) = 0$ WLOG, by shifting $\beta_0$ accordingly, but $\mathrm{E}(XU), \mathrm{E}(U|X)$ etc. are not necessarily equal to 0.
- The statement $\mathrm{E}(UX) = 0$ is therefore an assumption about the joint distribution of $(Y, X)$. If this assumption uniquely determines $\beta$, it implies that the best linear predictor of $Y$ given $X$ also represents the causal effect of $X$ on $Y$.

# Causal Model: Example I revisited

- We observe an iid sample of $\{Y_i, D_i\}$ where $Y_i$ is the scrap rate of factory $i$, and

$$D_i = \begin{cases} 1 & \text{if factory } i \text{ receives a job training grant,} \\ 0 & \text{otherwise.} \end{cases}$$

- The observed outcome $Y_i$ is a function of the potential outcomes $Y_{i0}, Y_{i1}$ and treatment:

$$Y_i = D_i Y_{i1} + (1 - D_i)\, Y_{i0}.$$

- WLOG we write

$$\mathrm{E}\left(Y_i | D_i\right) = \beta_0 + \beta_1 D_i.$$

# Causal Model: Example I revisited

- We derived

$$\beta_1 = \underbrace{\mathrm{E}\left(Y_1|D=1\right) - \mathrm{E}\left(Y_0|D=1\right)}_{ATT} + \underbrace{\mathrm{E}\left(Y_0|D=1\right) - \mathrm{E}\left(Y_0|D=0\right)}_{\text{Selection Bias}}$$

- Suppose that assignment of grants is now independent of the factory's potential outcomes $Y_0, Y_1$. This means that factories are not applying for the grant based on $Y_0$. We say assignment to treatment is "randomized".
- The Selection Bias term vanishes, and

$$\beta_1 = \mathrm{E}\left(Y_1|D=1\right) - \mathrm{E}\left(Y_0|D=1\right) = \mathrm{E}\left(Y_1 - Y_0\right),$$

  where the second equality holds by random assignment.
- $\beta_1$ now represents the mean effect (often called Average Treatement Effect or ATE) of receiving a grant in the population of firms.

# Causal Model: Example I revisited

- It is still not the case that the model

$$Y_i = \beta_0 + \beta_1 D_i + U_i$$

has a causal interpretation, even though $\beta_1$ may be equal to a parameter we are interested in (namely the ATE).

- The reason is that the grants may not impact firms equally, and $\beta_1$ represents the average effect.
- If we go one step further, and assume

$$Y_{i0} = \beta_0 + U_i,$$
$$Y_{i1} = \beta_0 + \beta_1 + U_i,$$

then $\beta_1 = Y_{i1} - Y_{i0}$ represents an (homogeneous) additive treatment effect and our model has a causal interpretation.

# Linear regression when $\mathrm{E}(XU) = 0$

- Let $(Y, X, U)$ be a random vector such that $Y$ and $U$ are scalar random variables and $X \in \mathbb{R}^{k+1}$.
- Assume the first component of $X$ equals 1:

$$X = (X_0, X_1 \ldots, X_k),$$

where $X_0 = 1$.
- Let $\beta = (\beta_0, \ldots, \beta_k) \in \mathbb{R}^{k+1}$ be a constant vector of unknown parameters such that

$$Y = X'\beta + U.$$

- Suppose $\mathrm{E}(XU) = 0$, justified according to how we interpret the model.
- Suppose also that $\mathrm{E}\left(X_j^2\right) < \infty$ for $j \leq k$, so $\mathrm{E}(XX')$ exists.

# Linear regression when $\mathrm{E}(XU) = 0$

- There is perfect collinearity in $X$ if there exists a constant vector $a \neq 0$ such that

$$\mathrm{P}\left(a'X = 0\right) = 1.$$

- We assume there is no perfect collinearity in $X$. This assumption is equivalent to the condition that $\mathrm{E}(XX')$ is invertible.
- Since $\mathrm{E}(XX')$ is positive semidefinite, it is invertible iff it is positive definite.

# Linear regression when $\mathrm{E}(XU) = 0$

## Lemma

*Suppose $X$ is a $(K \times 1)$ random vector and $\mathrm{E}(XX')$ exists. Then $\mathrm{E}(XX')$ is invertible iff there is no perfect collinearity in $X$.*

## Proof.

If there is perfect collinearity $X$, then there exists a vector $a \neq 0$ such that $\mathrm{P}(X'a = 0) = 1$. For this vector,

$$\mathrm{E}\left(XX'\right) a = \mathrm{E}\left(X\left(X'a\right)\right) = \mathrm{E}\left(X \cdot 0\right) = 0.$$

Therefore, $\mathrm{E}(XX')$ is not full column rank and so not invertible. Now suppose there is no perfect collinearity in $X$. For any vector $c \in \mathbb{R}^{k+1} \setminus \{0\}$,

$$c'\mathrm{E}\left(XX'\right) c = \mathrm{E}\left(\left(X'c\right)^2\right) > 0,$$

since the expectation equals 0 if and only if $\mathrm{P}(X'c = 0) = 1$ which is ruled out by assumption. $\square$

# Linear regression when $\mathrm{E}(XU) = 0$

- Since $\mathrm{E}(XU) = \mathrm{E}(X(Y - X'\beta)) = 0$, we obtain

$$\mathrm{E}(XY) = \mathrm{E}\left(XX'\right)\beta.$$

- Since there is no perfect collinearity in $X$, $\mathrm{E}(XX')$ is invertible, so we can solve for a unique $\beta$:

$$\beta = \mathrm{E}\left(XX'\right)^{-1}\mathrm{E}(XY).$$

- In this case, $\beta$ is point identified, since it is uniquely determined by $\mathrm{E}(XX'), \mathrm{E}(XY)$.

# Linear regression when $\mathrm{E}(XU) = 0$

- If $\mathrm{E}(XX')$ is not invertible, there are multiple solutions to

$$\mathrm{E}(XY) = \mathrm{E}(XX')\,\beta.$$

- If $\tilde{\beta}$ satisfies $\mathrm{E}(XX')\,\tilde{\beta} = \mathrm{E}(XY)$, then $\mathrm{E}(XX')\left(\tilde{\beta} - \beta\right) = 0$, so

$$\left(\tilde{\beta} - \beta\right)' \mathrm{E}(XX')\left(\tilde{\beta} - \beta\right) = \mathrm{E}\left(\left[X'\left(\tilde{\beta} - \beta\right)\right]^2\right) = 0,$$

which implies $\mathrm{P}\left(X'\tilde{\beta} = X'\beta\right) = 1$.

- If we interpret $X'\beta$ as a best linear predictor of $Y$, this says there are multiple best predictors.
- If the model is interpreted causally, however, different values of $\beta$ imply different ceteris paribus effects, holding $U$ and the other components of $X$ fixed.

# Questions?