Properties of OLS in the multiple regression model

Paul Schrimpf

UBC
Economics 326

February 1, 2018

References

- Wooldridge (2013) chapters 3 and 4
- Stock and Watson (2009) chapter 7 and 18
- Angrist and Pischke (2014) chapter 2
- Kasahara's slides
- Bierens (2010)
- Angrist and Pischke (2009) pages 48-69
- Baltagi (2002) chapter 4

Section 1

Setup

Model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \epsilon_i \tag{1}$$

Assumptions:

- **MLR.1** (linear model) equation 1 holds
- **MLR.2** (independence) $\{(x_{1,i}, x_{2,i}, y_i)\}_{i=1}^n$ is an independent random sample
- **MLR.3** (rank condition) no multicollinearity: no $x_{j,i}$ is constant and there is no exact linear relationship among the $x_{j,i}$
- **MLR.4** (exogeneity) $\mathrm{E}[\epsilon_i | x_{1,i}, \ldots, x_{k,i}] = 0$

Discussion of assumptions

- Assumptions MLR.1 (linear model), MLR.2 (independence), and MLR.4 (exogeneity) are the same as in bivariate regression
- MLR.1 (linear model)
  - Only by writing down a model can we talk about bias and exogeneity
  - Linearity is a convenient approximation
- MLR.2 (independence)
  - OLS generally still unbiased with non-independent observations, but variance different
  - Forms of dependence: time series, clustering, spatial
- MLR.4 (exogeneity)
  - Key assumption for OLS to be unbiased

Discussion of assumpptions

- MLR.3 (rank condition / no perfect collinearity): Wooldridge "None of the independent variables (x's) is constant, and there are no exact linear relationships among the independent variables"
  - Ensures that there is are unique values for $\hat{\beta}_0$, ..., $\hat{\beta}_k$ that solve the first order conditions,

$$
\begin{aligned}
\sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i}\right) & = 0 \\
\sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i}\right) x_{1,i} & = 0 \\
\vdots \qquad\qquad\qquad\qquad & \vdots = \vdots \\
\sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i}\right) x_{k,i} & = 0
\end{aligned}
$$

Discussion of assumpptions

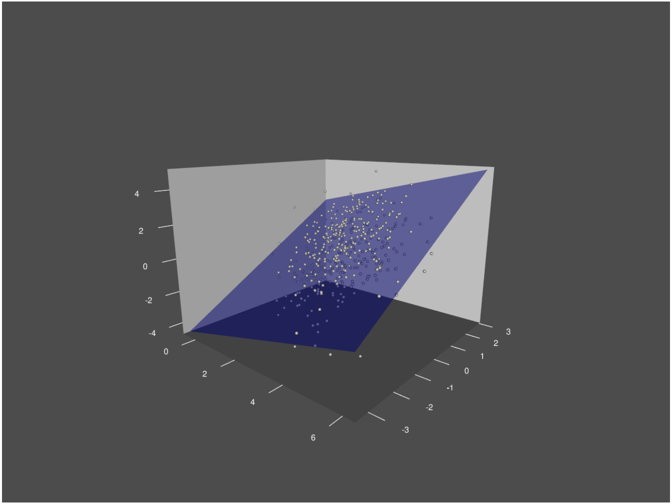
- In the language of linear algebra[1], this condition says that rank of the following $n \times (k+1)$ matrix of x's,

$$\begin{pmatrix} 1 & x_{1,1} & \cdots & x_{k,1} \\ 1 & x_{1,2} & \cdots & x_{k,2} \\ \cdots & & & \vdots \\ 1 & x_{1,n} & \cdots & x_{k,n} \end{pmatrix}$$

must be $k+1$

[1] If you have not heard of matrices and their rank before, this bullet point can safely be ignored.

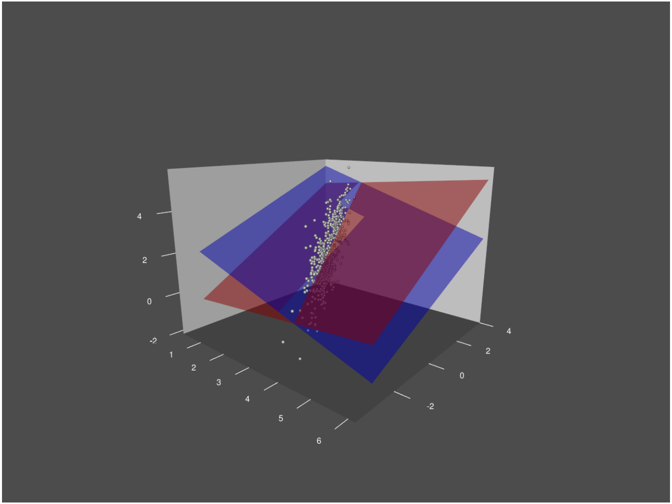
No collinearity



Cloud of points, so there's a unique plane that minimizes squared residuals

[1]Code

Perfect collinearity



All points lie in one plane, so there's many planes that minimize squared residuals

[1]Code

Section 2

Expected value

OLS is unbiased

Theorem

Under assumptions MLR.1-4, OLS is unbiased,

$$\mathrm{E}[\hat{\beta}_j] = \beta_j \textit{ for } j = 0, 1, \ldots, k.$$

Proof that OLS is unbiased

- Use partitioned regression formula,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{x}_{1,i} y_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2} \tag{2}$$

where $\tilde{x}_{1,i}$ is the OLS residual from regressing $x_{1,i}$ on the other controls,

$$x_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2,i} + \cdots + \hat{\gamma}_k x_{k,i} + \tilde{x}_{1,i}$$

- Substitute $y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$ in (2)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{x}_{1,i} \left(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \epsilon_i\right)}{\sum_{i=1}^n \tilde{x}_{1,i}^2}$$

- Rearrange and use the following properties of residuals $\tilde{x}_{1,i}$ to simplify

Proof that OLS is unbiased

1. $\sum_{i=1}^n \tilde{x}_{1,i} = 0$
2. $\sum_{i=1}^n \tilde{x}_{1,i} x_{j,i} = 0$ for $j = 2, ..., k$
3. $\sum_{i=1}^n \tilde{x}_{1,i} x_{1,i} = \sum_{i=1}^n \tilde{x}_{1,i}^2$

and get

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n \tilde{x}_{1,i}\epsilon_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2} \tag{3}$$

- Take expectations of (3) and use iterated expectations

$$\begin{aligned} \mathrm{E}[\hat{\beta}_1] =& \mathrm{E}\left[\mathrm{E}[\hat{\beta}_1|X]\right] \\ =& \beta_1 + \mathrm{E}\left[\mathrm{E}\left[\left.\frac{\sum_{i=1}^n \tilde{x}_{1,i}\epsilon_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2}\right| X\right]\right] \end{aligned}$$

- Use MLR.4 to conclude

$$\mathrm{E}[\hat{\beta}_1] = \beta_1$$

- Identical argument works for $j = 2, ..., k$

Section 3

Variance

Variance

MLR.5 (homoskedasticity) $\mathrm{Var}(\epsilon_i | X) = \sigma_\epsilon^2$

Theorem

Under assumptions MLR.1-5, the variance of OLS conditional on the controls is

$$\mathrm{Var}(\hat{\beta}_j | X) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n \tilde{x}_{j,i}^2}.$$

Also, the covariance of $\hat{\beta}_j$ and $\hat{\beta}_\ell$ conditional on the controls is

$$\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_\ell | X) = \sigma_\epsilon^2 \frac{\sum_{i=1}^n \tilde{x}_{j,i} \tilde{x}_{\ell,i}}{\left(\sum_{i=1}^n \tilde{x}_{j,i}^2\right)\left(\sum_{i=1}^n \tilde{x}_{\ell,i}^2\right)}$$

Proof of OLS variance

- Use

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^n \tilde{x}_{j,i}\epsilon_i}{\sum_{i=1}^n \tilde{x}_{j,i}^2}$$

- So

$$\hat{\beta}_j - \mathrm{E}[\hat{\beta}_j] = \hat{\beta}_j - \beta_j = \frac{\sum_{i=1}^n \tilde{x}_{j,i}\epsilon_i}{\sum_{i=1}^n \tilde{x}_{j,i}^2}$$

- Then,

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_\ell) =& \mathrm{E}\left[(\hat{\beta}_j - \mathrm{E}[\hat{\beta}_j])(\hat{\beta}_\ell - \mathrm{E}[\hat{\beta}_\ell])|X\right] \\ =& \mathrm{E}\left[\left.\frac{\sum_{i=1}^n \tilde{x}_{j,i}\epsilon_i}{\sum_{i=1}^n \tilde{x}_{j,i}^2}\frac{\sum_{i=1}^n \tilde{x}_{\ell,i}\epsilon_i}{\sum_{i=1}^n \tilde{x}_{\ell,i}^2}\right| X\right]\end{aligned}$$

- Use MLR.5 to get desired result

Alternative expression for the variance

- From above we have

$$\mathrm{Var}(\hat{\beta}_j|X) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n \tilde{x}_{j,i}^2}.$$

  and

$$\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_\ell|X) = \sigma_\epsilon^2 \frac{\sum_{i=1}^n \tilde{x}_{j,i}\tilde{x}_{\ell,i}}{\left(\sum_{i=1}^n \tilde{x}_{j,i}^2\right)\left(\sum_{i=1}^n \tilde{x}_{\ell,i}^2\right)}$$

- The dominator, $\sum_{i=1}^n \tilde{x}_{j,i}^2$, is the sum of squared residuals from regression $x_{j,i}$ on the other x's
- Recall $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$, so

$$\begin{aligned} R_j^2 =& 1 - \frac{\sum_{i=1}^n \tilde{x}_{j,i}^2}{\sum_{i=1}^n (x_{j,i} - \bar{x}_j)^2} \\ \sum_{i=1}^n \tilde{x}_{j,i}^2 =& (1 - R_j^2)\sum_{i=1}^n (x_{j,i} - \bar{x}_j)^2 \\ =& (1 - R_j^2) n \widehat{\mathrm{Var}}(x_j) \end{aligned}$$

- So,

$$\mathrm{Var}(\hat{\beta}_j|X) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n \tilde{x}_{j,i}^2} = \frac{\sigma_\epsilon^2}{(1 - R_j^2) n \widehat{\mathrm{Var}}(x_j)}$$

Variance and omitted variables

- Suppose
$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i$$
and MLR1-5 hold
- What are the bias of $\hat{\beta}_1^s$ and $\hat{\beta}_1$?
- What are $\mathrm{Var}(\hat{\beta}_1^s)$ and $\mathrm{Var}(\hat{\beta}_1)$? Which is larger?
- What is the mean square error (MSE) of $\hat{\beta}_1^s$ and $\hat{\beta}_1$?
- If we want to minimize MSE, what is better $\hat{\beta}_1^s$ or $\hat{\beta}_1$?

Section 4

Efficiency

Gauss-Markov theorem

Theorem

Under assumptions MLR.1-5, OLS is the best linear unbiased estimator

- By best linear unbiased estimator, we mean that OLS has the lowest variance for any linear combination of the coefficients,

$$\mathrm{Var}\left(\sum_{j=1}^{k} \lambda_j \hat{\beta}_j\right) \leq \mathrm{Var}\left(\sum_{j=1}^{k} \lambda_j \tilde{\beta}_j\right)$$

  where $\tilde{\beta}_j$ are any other linear unbiased estimators
  - Linear: $\tilde{\beta}_j = \sum_{i=1}^{n} w_i y_i$
  - Unbiased: $\mathrm{E}[\tilde{\beta}_j] = \beta_j$

Section 5

Distribution

Distribution

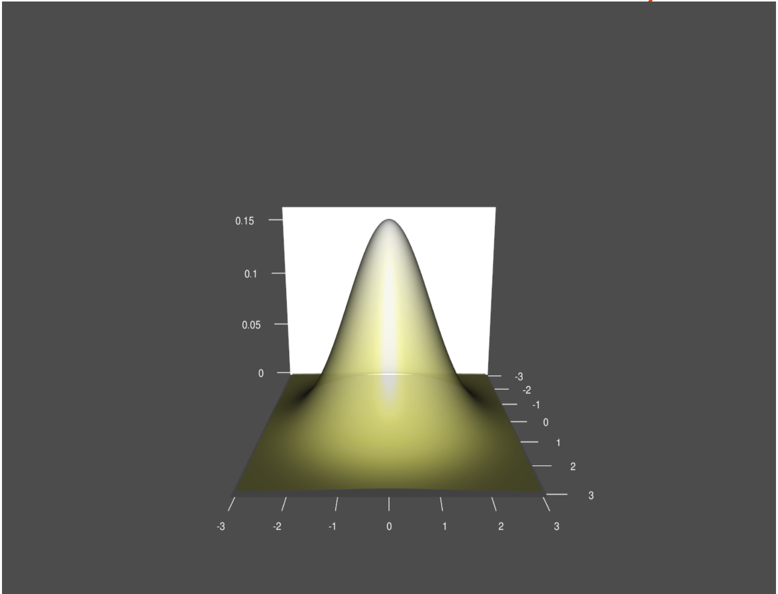
MLR.6 $\epsilon_i | X \sim N(0, \sigma_\epsilon^2)$

Theorem

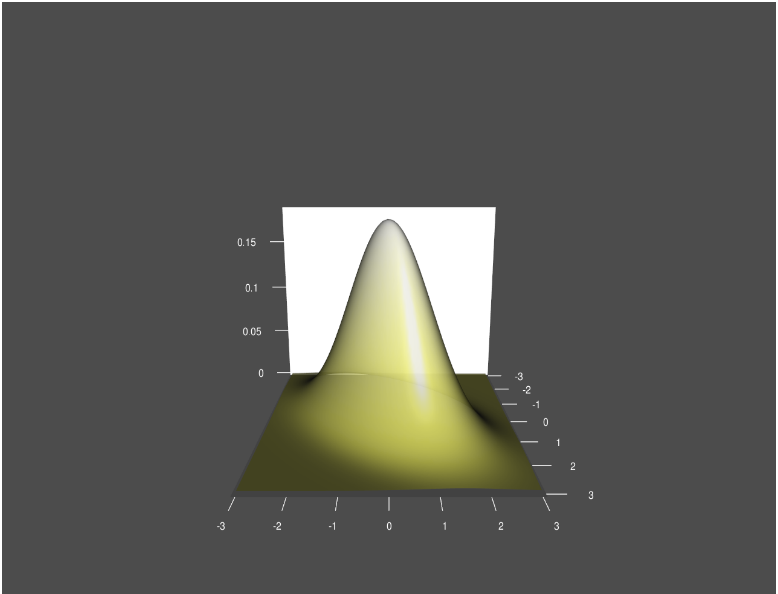
Under assumptions MLR.1-6, OLS is normally distributed

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} | X \sim N \left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} , \begin{pmatrix} \mathrm{Var}(\hat{\beta}_0 | X) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1 | X) & \cdots & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_k | X) \\ \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0 | X) & \mathrm{Var}(\hat{\beta}_1 | X) & \cdots & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_k | X) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{\beta}_k, \hat{\beta}_0 | X) & \mathrm{Cov}(\hat{\beta}_k, \hat{\beta}_1 | X) & \cdots & \mathrm{Var}(\hat{\beta}_k | X) \end{pmatrix} \right)$$

Standard bivariate normal density



Bivariate normal density with correlation



Section 6

Inference

Inference with normal errors

- Regression estimates depend on samples, which are random, so the regression estimates are random
    - Some regressions will randomly look “interesting” due to chance
- Logic of hypothesis testing: figure out probability of getting an interesting regression estimate due solely to change
- Null hypothesis, H_0 : the regression is uninteresting
    - If we mainly care about the jth control, then

$$H_0 : \beta_j = 0$$

    - If we care about all the regressors, then maybe

$$H_0 : \beta_1 = 0, \beta_2 = 0, \ldots, \beta_k = 0$$

Inference with normal errors

- With assumptions MLR.1-MLR.6 and under $H_0 : \beta_j = \beta_j^*$, we know

$$\hat{\beta}_j \sim N\left(\beta_j^*, \frac{\sigma_\epsilon^2}{\sum_{i=1}^n \tilde{x}_{j,i}^2}\right)$$

or equivalently,

$$z \equiv \frac{\hat{\beta}_j - \beta_j^*}{\sigma_\epsilon / \sqrt{\sum_{i=1}^n \tilde{x}_{j,i}^2}} \sim N(0,1)$$

- We do not know σ_ϵ^2, so estimate it using residuals,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-(k+1)} \sum_{i=1}^n \hat{\epsilon}_i^2$$

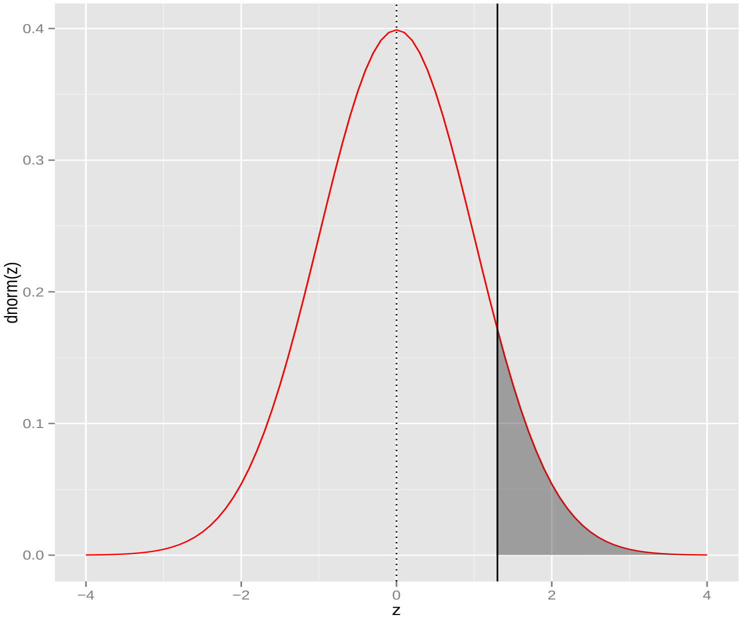
Inference with normal errors

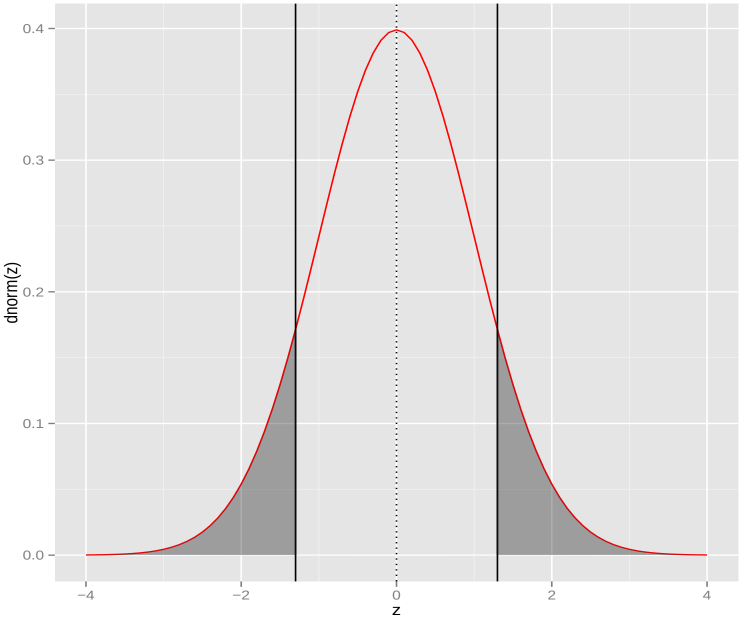
- We divided by $n-(k+1)$ instead of n because $\hat{\epsilon}_i$ depends on $k+1$ estimated parameters ($\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$)

- Because of estimated $\hat{\sigma}^2_\epsilon$, test statistic has t distribution instead of normal,

$$t \equiv \frac{\hat{\beta}_j - \beta_j^*}{\hat{\sigma}_\epsilon / \sqrt{\sum_{i=1}^n \tilde{x}_{j,i}^2}} \sim t\left(n-(k+1)\right)$$

- P-value: the probability of getting a regression estimate as or more "interesting" than the one we have
  - As or more interesting $=$ as far or further away from β_j^*
  - If we are only interested when $\hat{\beta}_j$ is on one side of β_j^*, then we have a one sided alternative, e.g. $H_a : \beta_j > \beta_j^*$
  - If we are equally interested in either direction, then $H_a : \beta_j \neq \beta_j^*$

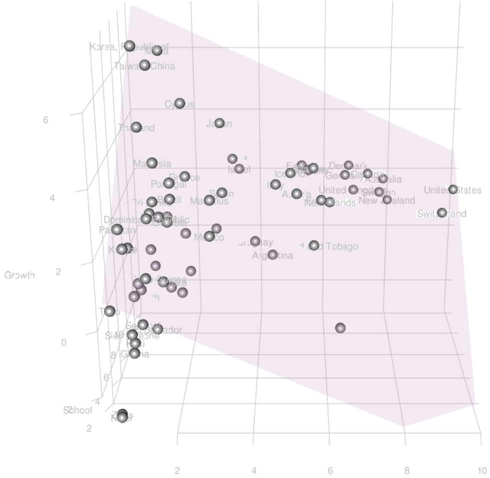

0.4
0.3
0.2
0.1
0.0
dnorm(z)
−4
−2
0
2
4
z


0.4
0.3
0.2
0.1
0.0
dnorm(z)
−4
−2
0
2
4
z

Inference with normal errors

- One-sided p-value: $p = F_{t,n-k-1}(-|t|) = 1 - F_{t,n-k-1}(|t|)$
- Two-sided p-value:
  $p = 2F_{t,n-k-1}(-|t|) = 2(1 - F_{t,n-k-1}(|t|))$
- Interpretation:
  - The probability of getting an estimate as strange as the one we have if the null hypothesis is true.
  - It is *not* about the probability of β_j being any particular value. β_j is not a random variable. It is some unknown number. The data is what is random. In particular, the p-value is *not* the probability that that H_0 is false given the data.
- Hypothesis testing: we must make a decision (usually reject or fail to reject H_0)
  - Choose significance level α (usually 0.05 or 0.10)
  - Construct procedure such that if H_0 is true, we will incorrectly reject with probability α
  - Reject null if p-value less than α

Example: growth, GDP, and schooling



[1] Code

Example: growth, GDP, and schooling

	Model 1	Model 2	Model 3
(Intercept)	1.796^{***}	0.958^{*}	0.895^{*}
	(0.378)	(0.418)	(0.389)
rgdp60	0.047		-0.485^{**}
	(0.095)		(0.146)
yearsschool		0.247^{**}	0.640^{***}
		(0.089)	(0.144)
R^2	0.004	0.110	0.244
Adj. R^2	-0.012	0.095	0.219
Num. obs.	65	65	65
RMSE	1.908	1.804	1.676

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Growth and GDP and education in 1960

Confidence intervals

- A $1-\alpha$ confidence interval, $CI_{1-\alpha} = [LB_{1-\alpha}, UB_{1-\alpha}]$ is an interval estimator for β_j such that

$$\mathrm{P}\left(\beta_j \in CI_{1-\alpha}\right) = 1-\alpha$$

  ($CI_{1-\alpha}$ is random; β_j is not)

- $1-\alpha$ confidence interval

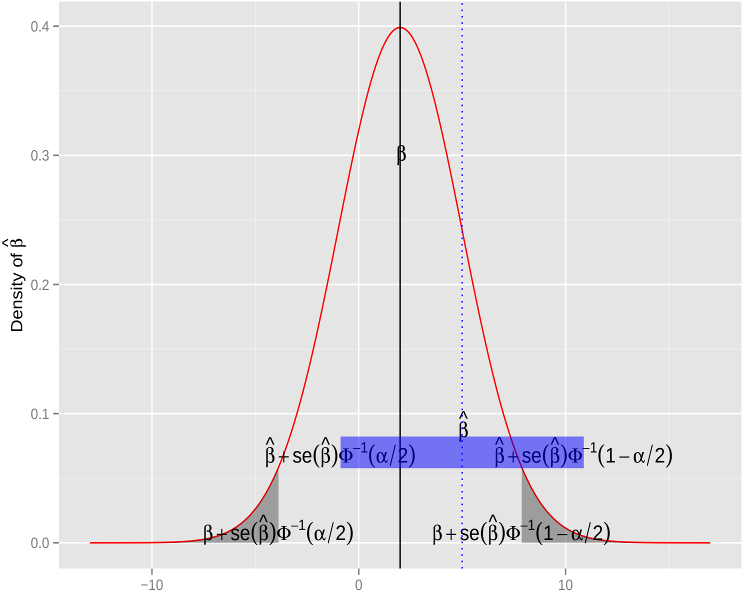
$$\hat{\beta}_j \pm \sqrt{\mathrm{Var}(\hat{\beta}_j)}\Phi^{-1}(\alpha/2)$$

- With estimated $\hat{\sigma}_\epsilon^2$, use t distribution instead of normal

$$\hat{\beta}_j \pm \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_j)}F_{t,n-2}^{-1}(\alpha/2)$$

  $F_{t,n-2}^{-1}$ = inverse CDF of $t(n-2)$ distribution

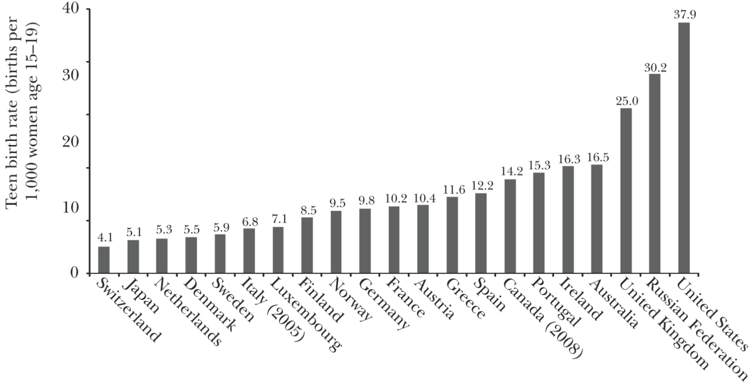
Confidence intervals



Example: Kearney and Levine (2012)

Figure 1

International Comparison of Teen Birth Rates, 2009

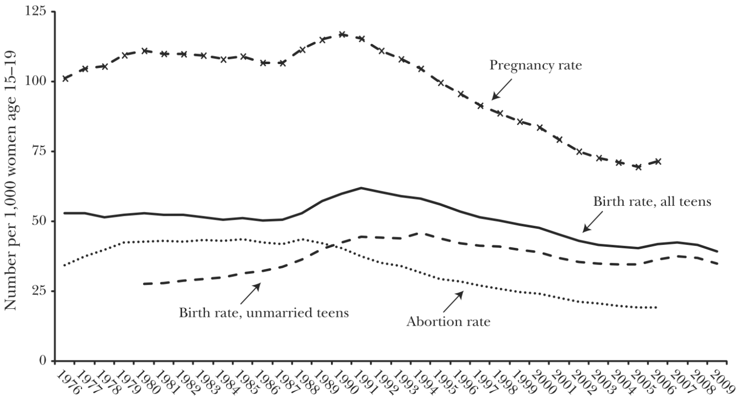


Sources: UNECE Statistical Database and *United Nations Demographic Yearbook, 2009–2010.*

Example: Kearney and Levine (2012)

Figure 2

Trends in the Teen Pregnancy, Abortion, and Birth Rate



Example: Kearney and Levine (2012)

Table 1

Teen Birth Rates (Birth per 1,000 Females Age 15–19), 2009

State	*Teen birth rate*	*State*	*Teen birth rate*	*State*	*Teen birth rate*
Low		*Moderate*		*High*	
New Hampshire	16.4	Michigan	32.7	North Carolina	44.9
Vermont	17.4	Oregon	33.1	Wyoming	45.0
Massachusetts	19.6	Nebraska	34.6	Nevada	47.4
Connecticut	21.0	Delaware	35.3	Washington, DC	47.7
New Jersey	22.7	Idaho	35.9	Georgia	47.7
Minnesota	24.3	Illinois	36.1	South Carolina	49.1
Maine	24.4	California	36.6	West Virginia	49.8
New York	24.4	South Dakota	38.4	Arizona	50.6
Rhode Island	26.8	Colorado	38.5	Tennessee	50.6
North Dakota	27.9	Montana	38.5	Alabama	50.7
Pennsylvania	29.3	Ohio	38.9	Kentucky	51.3
Wisconsin	29.4	Florida	39.0	Louisiana	52.7
Utah	30.7	Hawaii	40.9	Arkansas	59.2
Virginia	31.0	Missouri	41.6	Oklahoma	60.1
Maryland	31.3	Indiana	42.5	Texas	60.7
Washington	31.9	Kansas	43.8	New Mexico	63.9
Iowa	32.1	Alaska	44.5	Mississippi	64.2

Example: Kearney and Levine (2012)

Table 2

Rates of Pregnancy, Birth, and Abortion across Countries and States in the United States

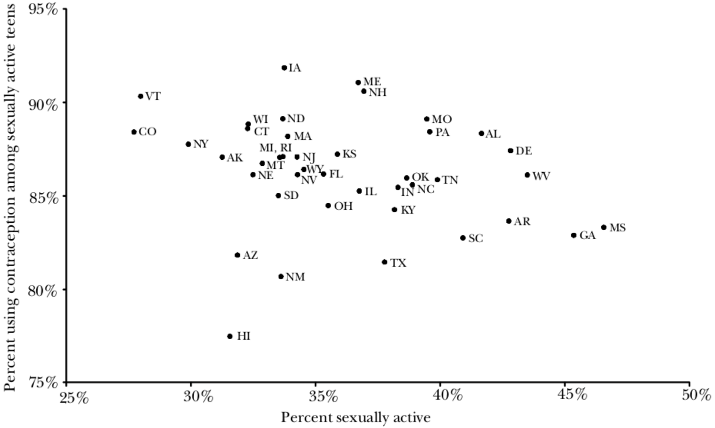
	Pregnancies (per 1,000)	*Births (per 1,000)*	*Abortions (per 1,000)*	*% of pregnancies aborted*
Denmark (2003)	24	5	15	63.2
Germany (2003)	23	12	7	31.1
New Hampshire (2005)	33	18	11	33.3
United Kingdom (2003)	59	27	23	38.8
United States (2005)	70	40	19	27.1
Mississippi (2005)	85	61	11	12.9

Sources: State data are from Guttmacher Institute (2010). International birth data are from the UNECE statistical database. International abortion data are from Sedgh, Henshaw, Singh, Bankole, and Drescher (2007).

Example: Kearney and Levine (2012)

Figure 4

Rates of Sexual Activity and Contraceptive Use among School-Aged (14–18) Girls



Source: Authors calculations from the 2007 and 2009 Youth Risky Behavior Surveillance survey state microdata.

Example: Kearney and Levine (2012)

Table 3

Mechanical Correlations with Teen Fertility

(standard errors in parentheses)

Sexual activity and use of any contraception		*Sexual activity and specific forms of contraception*	
% Any sexual activity in past 3 months	0.162 (0.017)	% Any sexual activity in past 3 months	0.151 (0.019)
% Used any contraception if sexually active	–0.186 (0.030)	% Used pill if sexually active	–0.156 (0.028)
		% Used condom if sexually active	–0.120 (0.023)
		% Used Depo or other if sexually active	–0.018 (0.043)
		% Used withdrawal if sexually active	0.022 (0.062)
R^2	0.64	R^2	0.71
Number of states/years	167	Number of states/years	167

Source: Authors using data from the Youth Risky Behavior Surveillance survey and Vital Statistics natility data.

Notes: We estimate regression models of the state-year teen birth rate as a function of measures of sexual activity and contraceptive use by state-year. The dependent variable in each model is the probability of giving birth as a teen in a year. The independent variables are share of teenagers in a state in a given year who have engaged in sexual activity in the previous three months along with measures of alternative contraceptive choices. All regressions are weighted by the population of women age 15 to 19 in each state-year. Withdrawal is counted as a form of contraception.

Hypothesis tests for multiple coefficients

- Sometimes we want to test a hypothesis that involves multiple coefficients
  1. Single restriction on a linear combinations of coefficients, e.g.
     $$H_0 : \beta_1 = \beta_2$$
     same as
     $$H_0 : \beta_1 - \beta_2 = 0$$
  2. Multiple restrictions, e.g.
     $$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$

Hypothesis tests for linear combination of coefficients

- $H_0 : \lambda_1\beta_1 + \cdots + \lambda_k\beta_k = 0$
- To simplify notation, focus on $H_0 : \beta_1 - \beta_2 = 0$
- We know that

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} \mathrm{Var}(\hat{\beta}_1) & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \mathrm{Var}(\hat{\beta}_2) \end{pmatrix} \right)$$

so

$$\hat{\beta}_1 - \hat{\beta}_2 \sim N\left(\beta_1 - \beta_2, \mathrm{Var}(\hat{\beta}_1) + \mathrm{Var}(\hat{\beta}_2) - 2\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2)\right)$$

and under $H_0 : \beta_1 = \beta_2$,

$$t \equiv \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1) + \widehat{\mathrm{Var}}(\hat{\beta}_2) - 2\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}} \sim t(n-k-1)$$

Hypothesis tests for linear combination of coefficients

- Example: Kearney and Levine (2012) test $H_0 : \beta_{sex} + \beta_{protection} = 0$
  - $$\begin{aligned} t = & \frac{\hat{\beta}_{sex} + \hat{\beta}_{protection}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1) + \widehat{\mathrm{Var}}(\hat{\beta}_2) + 2\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}} \\ = & \frac{0.162 + -0.186}{\sqrt{0.017^2 + 0.03^2 + \mathrm{Cov}}} \end{aligned}$$
  - Cov not reported, but we can consider the possible range
    - $\mathrm{Cov} = 0$: $t = \frac{-0.024}{0.034} = 0.68$
    - $\mathrm{Cov} = +\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)\widehat{\mathrm{Var}}(\hat{\beta}_2)}$: $t = \frac{-0.024}{0.047} = 0.51$
    - $\mathrm{Cov} = -\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)\widehat{\mathrm{Var}}(\hat{\beta}_2)}$: $t = \frac{-0.024}{0.013} = 1.85$

Hypothesis tests for linear combination of coefficients

- Another approach to testing linear combinations of coefficients is to re-specify the model so that the null hypothesis is about a single coefficient
  - E.g. instead of

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i$$

with $H_0 : \beta_1 - \beta_2 = 0$, write

$$\begin{aligned} y_i = & \beta_0 + \underbrace{\beta_1}_{=\theta_1}(x_{1,i} + x_{2,i}) + \underbrace{(\beta_2 - \beta_1)}_{=\theta_2} x_{2,i} + \epsilon_i \\ = & \theta_0 + \theta_1(x_{1,i} + x_{2,i}) + \theta_2 x_{2,i} + \epsilon_i \end{aligned}$$

then test $H_0 : \theta_2 = 0$.

Testing multiple restrictions

- What if we have

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i$$

  and want to test $H_0 : \beta_1 = 0$ and $\beta_2 = 0$?
- Cannot test one at time:
  - Suppose test $H_0 : \beta_1 = 0$ and $H_0' : \beta_2 = 0$ separately at 5% significance level
  - If H_0 true and $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 0$, then $\mathrm{P}(\text{fail to reject both}) = 0.95 \times 0.95 = 0.9025$, so doing separate tests we will reject too often
- Can use joint normal distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$
  - p-value $= \mathrm{P}(\text{estimates as far or further from null hypothesis})$
  - Need to take into account correlation of $\hat{\beta}_1$ and $\hat{\beta}_2$

Testing multiple restrictions

- F-statistic

$$F = \frac{1}{2}\begin{pmatrix}\hat\beta_1 \\ \hat\beta_2\end{pmatrix}^T \begin{pmatrix} \widehat{\mathrm{Var}}(\hat\beta_1) & \widehat{\mathrm{Cov}}(\hat\beta_1,\hat\beta_2) \\ \widehat{\mathrm{Cov}}(\hat\beta_1,\hat\beta_2) & \widehat{\mathrm{Var}}(\hat\beta_2)\end{pmatrix}^{-1}\begin{pmatrix}\hat\beta_1 \\ \hat\beta_2\end{pmatrix}$$

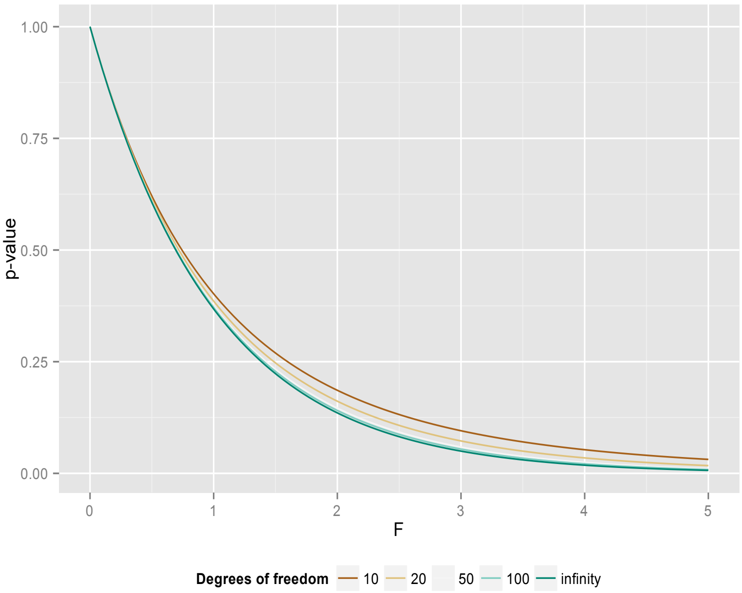
$$= \frac{1}{2}\frac{\hat\beta_1^2\widehat{\mathrm{Var}}(\hat\beta_2) + \hat\beta_2^2\widehat{\mathrm{Var}}(\hat\beta_1) - 2\widehat{\mathrm{Cov}}(\hat\beta_1,\hat\beta_2)\hat\beta_1\hat\beta_2}{\widehat{\mathrm{Var}}(\hat\beta_1)\widehat{\mathrm{Var}}(\hat\beta_2) - \widehat{\mathrm{Cov}}(\hat\beta_1,\hat\beta_2)^2}$$

has an $F(2, n-2)$ distribution

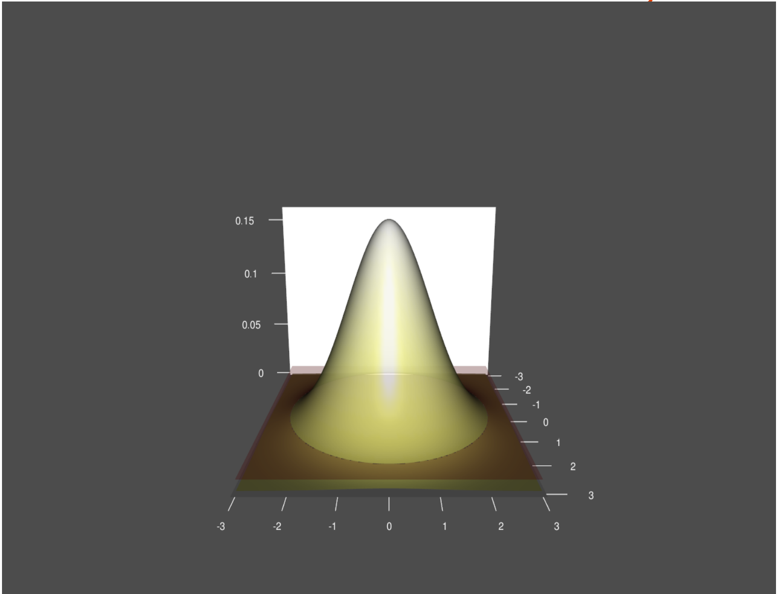
  - $kF(k,\infty) = \chi^2(k) =$ distribution of the sum of k independent standard normal random variables each squared

- This F is valid even with heteroskedasticity and dependent observations as long as the variances are calculated correctly

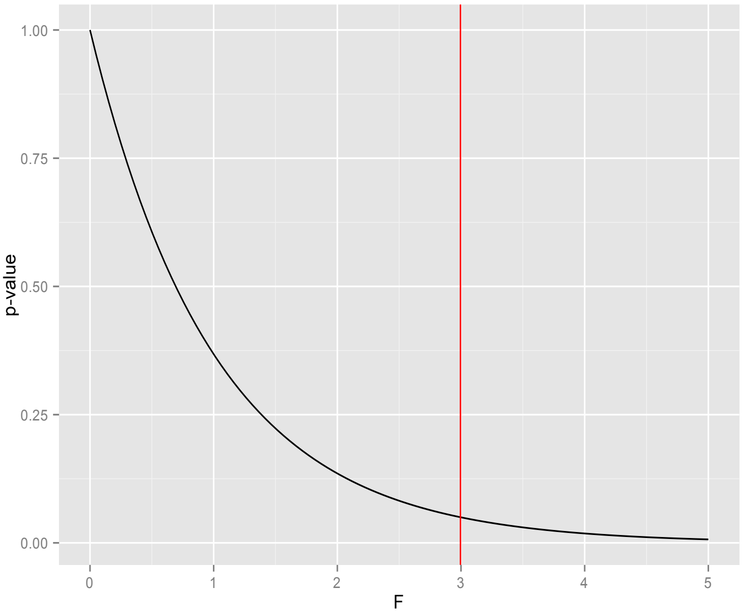
$F(2, n-2)$ distributions



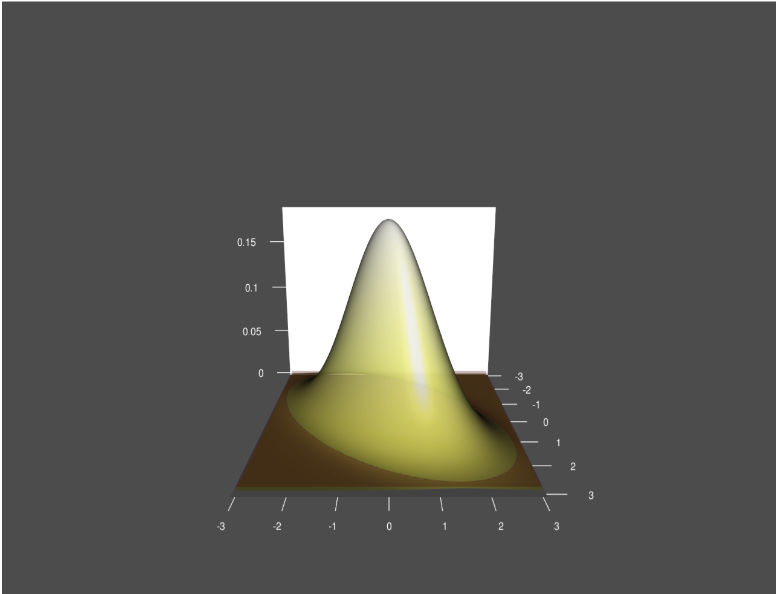
Standard bivariate normal density



$F(2,\infty)$ distribution



Bivariate normal density with correlation



Testing multiple restrictions: alternative form of F statistic

- We have an unrestricted model:
$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$$
and a restricted model that imposes $H_0 : \beta_{k-q+1} = 0, ..., \beta_k = 0$
- Estimate both models, compute sum of squared residuals call SSR_{ur} and SSR_r
- Calculate F statistics
$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$
- $F \sim F(q, n-k-1)$, use to calculate p-values
- This form of F is only valid with homoskedasticity and independence

Aron-Dine, Einav, and Finkelstein (2013)

Plans' Effects on Utilization

	Total spending[a]		*Inpatient spending*		*Outpatient spending*	
	Share with any (1)	*Spending in \$ (2)*	*Share with any (3)*	*Spending in \$ (4)*	*Share with any (5)*	*Spending in \$ (6)*
Constant (Free Care Plan, $N = 6{,}840$)	0.931 (0.006)	2,170 (78)	0.103 (0.004)	827 (60)	0.930 (0.006)	1,343 (35)
25% Coinsurance ($N = 2{,}361$)	−0.079 (0.015)	−648 (152)	−0.022 (0.009)	−229 (116)	−0.078 (0.015)	−420 (62)
Mixed Coinsurance ($N = 1{,}702$)	−0.053 (0.015)	−377 (178)	−0.018 (0.009)	21 (141)	−0.053 (0.016)	−398 (70)
50% Coinsurance ($N = 1{,}401$)	−0.100 (0.019)	−535 (283)	−0.031 (0.009)	4 (265)	−0.100 (0.019)	−539 (77)
Individual Deductible ($N = 4{,}175$)	−0.124 (0.012)	−473 (121)	−0.006 (0.007)	−67 (98)	−0.125 (0.012)	−406 (52)
95% Coinsurance ($N = 3{,}724$)	−0.170 (0.015)	−845 (119)	−0.024 (0.007)	−217 (91)	−0.171 (0.016)	−629 (50)
p-value: all differences from Free Care = 0	< 0.0001	< 0.0001	0.0008	0.1540	< 0.0001	< 0.0001

References

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein. 2013. "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives* 27 (1):197–222. URL http://www.aeaweb.org/articles.php?id=10.1257/jep.27.1.197.

Baltagi, BH. 2002. *Econometrics*. Springer, New York. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001086635&title=Econometrics.

References

Bierens, Herman J. 2010. "Multivariate Linear Regression." URL http://personal.psu.edu/hxb11/LINREG3.PDF.

Kearney, Melissa S. and Phillip B. Levine. 2012. "Why Is the Teen Birth Rate in the United States So High and Why Does It Matter?" *Journal of Economic Perspectives* 26 (2):141–63. URL http://www.aeaweb.org/articles?id=10.1257/jep.26.2.141.

Stock, J.H. and M.W. Watson. 2009. *Introduction to Econometrics, 2/E*. Addison-Wesley.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.