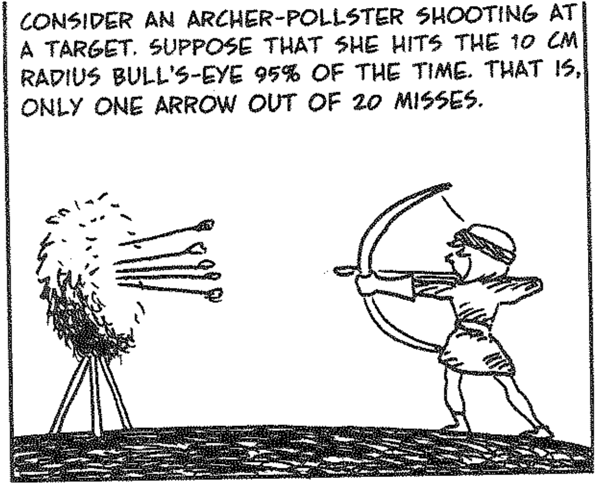
Lecture 7

Hypothesis testing and confidence intervals

Lectures in SDPE: *Econometrics I* on February 18, 2024

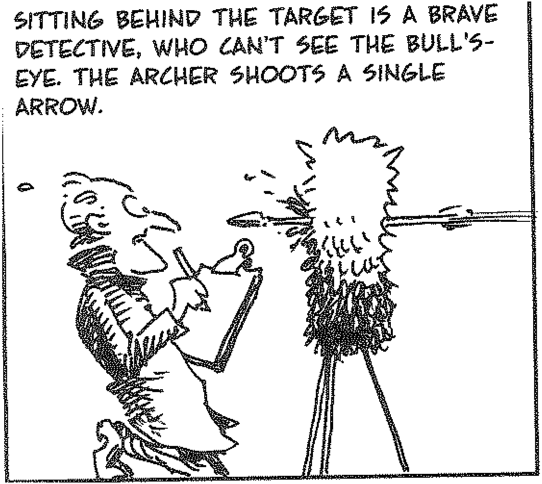
Markus Jäntti
Swedish Institute For Social Research
Stockholm University

Confidence intervals



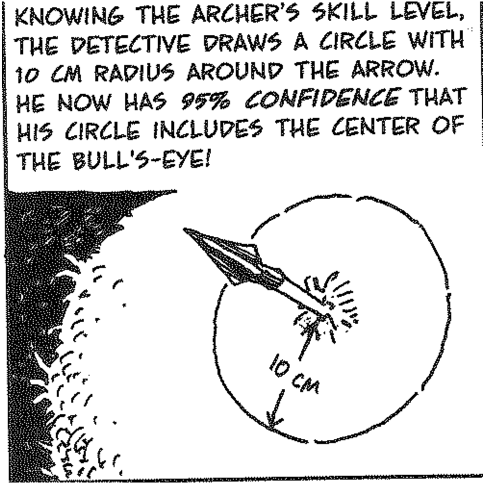
Source: Gonick and Smith (1993)

Confidence intervals



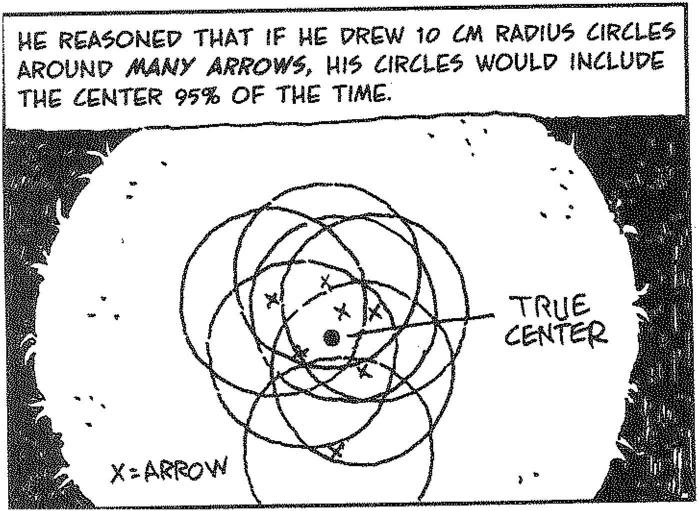
Source: Gonick and Smith (1993)

Confidence intervals



Source: Gonick and Smith (1993)

Confidence intervals



Source: Gonick and Smith (1993)

Confidence intervals

See Hansen (2021, sec 7.13)

- The LS estimate $\widehat{\beta}$ is a point estimate of β, giving the best "guess" at what point in $\mathbb{R}^k$ the true β is.
- A *set estimate* C_n is a collection of values in $\mathbb{R}^k$; an *interval estimate* is of the form $C_n = [L_n, U_n]$.
- As C_n is a function of random variables (the sample values) it is random with *coverage probability* $\mathbb{P}_\theta(\theta \in C_n)$.
- C_n are commonly called *confidence intervals*; C_n is a $100(1-\alpha)\%$ CI if $\inf_\theta \mathbb{P}_\theta(\theta \in C_n) = 1 - \alpha$.
- If $\widehat{\theta}$ is asymptotically normal with a standard error $s(\widehat{\theta})$, the CI is

$$C_n = [\widehat{\theta} - c \times s(\widehat{\theta}), \widehat{\theta} + c \times s(\widehat{\theta})]. \tag{1}$$

  The positive constant c is chosen to make the coverage probability $100(1-\alpha)\%$ or

$$C_n = \left\{\theta : -c \leq \frac{\widehat{\theta} - \theta}{s(\widehat{\theta})} \leq c\right\} \tag{2}$$

- In finite samples the true coverage probability is unknown, but as $n \longrightarrow \infty$, the *asymptotic coverage probability* is

$$\mathbb{P}_\theta(\theta \in C_n) \longrightarrow \mathbb{P}(|Z| \leq c) = 1 - \overline{\Phi}(c) \tag{3}$$

for a standard normal variate Z and the cdf of the standard normal distribution Φ.

- The "typical" choice of α is 5% for which $c = 1.96$, so

$$[\widehat{\theta} - 1.96 \times s(\widehat{\theta}), \widehat{\theta} + 1.96 \times s(\widehat{\theta})] \tag{4}$$

has an asymptotic coverage probability of $(1 - \alpha)\% = 95\%$.

Regression intervals

See Hansen (2021, sec 7.15)

- Interpreting regression as CEF, we have

$$m(\boldsymbol{x}) = \mathrm{E}[Y|\boldsymbol{X}] = \boldsymbol{x}'\beta. \tag{5}$$

- Let $\theta = h(\beta) = \boldsymbol{x}'\beta$ and, so $\widehat{m} = \widehat{\theta} = \boldsymbol{x}'\widehat{\beta}$, $\boldsymbol{H}_\beta = \boldsymbol{x}$ and we have

$$s(\widehat{\theta}) = \sqrt{\boldsymbol{x}'\widehat{\boldsymbol{V}}_{\widehat{\beta}}\boldsymbol{x}} \text{ and } \left[\boldsymbol{x}'\widehat{\beta} \pm 1.96\sqrt{\boldsymbol{x}'\widehat{\boldsymbol{V}}_{\widehat{\beta}}\boldsymbol{x}}\right] \tag{6}$$

- Note that the standard error depends on $\boldsymbol{x}$ quadratically.

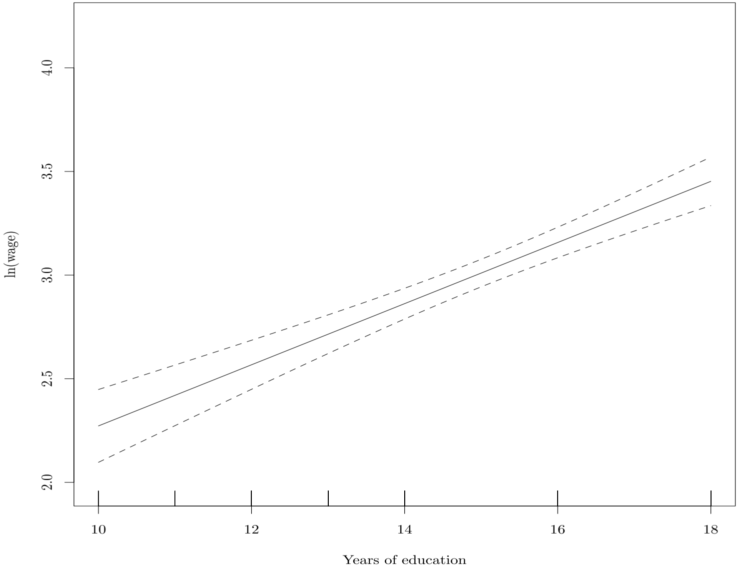
Regression intervals – illustration

CPS 2015, ln wage as function of education, then experience and education

	1	2
(Intercept)	0.80	0.99
	(0.23)	(0.04)
eduyears	0.15	0.12
	(0.02)	(0.00)
experience		0.02
		(0.00)
experience^2/100		−0.02
		(0.00)
n	516	20535
k	2	4
σ	0.76	0.7
R^2	0.15	0.16

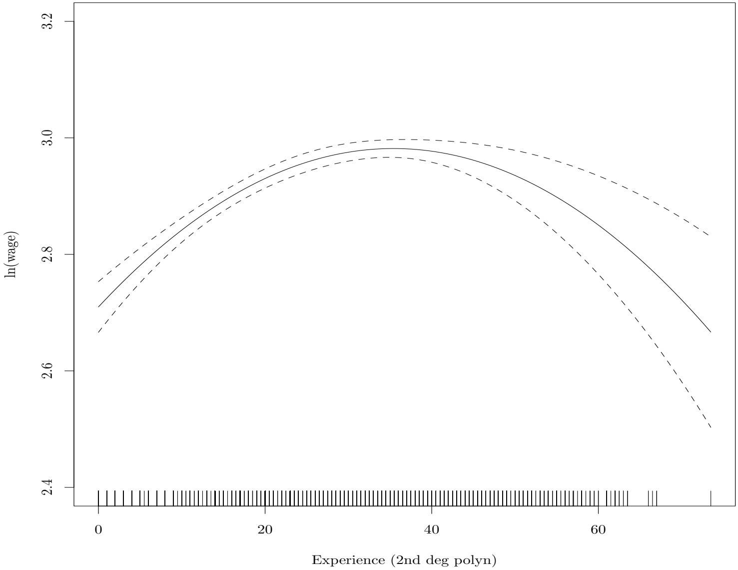
Regression intervals – illustration

CPS 2015



Regression intervals – illustration

CPS 2015; education set at mean



Forecast intervals

See Hansen (2021, sec 7.16)

- For out-of-sample forecasts at a point $\boldsymbol{x}_{n+1} = \boldsymbol{x}$, $\widehat{Y}_{n+1} = \boldsymbol{x}\widehat{\beta}$, we know that forecast error is
$$\tilde{e}_{n+1} = e_{n+1} - \boldsymbol{x}'(\widehat{\beta} - \beta). \tag{7}$$
- The variance of this is
$$\mathrm{E}[\tilde{e}_{n+1}^2 | \boldsymbol{x}_{n+1} = \boldsymbol{x}] = \sigma^2(\boldsymbol{x}) + \left(\boldsymbol{x}' \boldsymbol{V}_{\widehat{\beta}} \boldsymbol{x}\right) \tag{8}$$
(Note the use of the *conditional* variance $\sigma^2(\boldsymbol{x})$).
- A contender for the prediction interval is
$$[\boldsymbol{x}'\widehat{\beta} \pm 1.96\widehat{s}(\boldsymbol{x})] \tag{9}$$
with $\hat{s}(\boldsymbol{x}) = \sqrt{\widehat{\sigma^2}(\boldsymbol{x})}$, but the asymptotic distribution of
$$\frac{\tilde{e}_{n+1} + \boldsymbol{x}'(\widehat{\beta} - \beta)}{\widehat{s}(\boldsymbol{x})} \tag{10}$$
is unknown in general.

- Hypothesis testing is a game against nature. We need to decide if nature placed the true parameter $\theta \in \Theta$ in Θ_0 or Θ_1 ($\mathbb{H}_0 : \theta \in \Theta_0$ vs. $\mathbb{H}_1 : \theta \in \Theta_1$):

	Truth	
Decision	$\theta \in \Theta_0$	$\theta \in \Theta_1$
Accept $\mathbb{H}_0$		Error II
Reject $\mathbb{H}_0$	Error I	

- A hypothesis test consists of a real-valued *test statistic*

$$T = T((Y_1, X_1), \ldots, (Y_n, X_n)), \tag{11}$$

which we compare to a *critical value* c and the decision rule that

1. Accept $\mathbb{H}_0$ if $T \leq c$
2. Reject $\mathbb{H}_0$ if $T > c$

- For a good test
  - the probability of a Type I error is small if $\theta \in \Theta_0$ *and*
  - the probability of a Type II error is small if $\theta \in \Theta_1$.
- The first probability is called the *significance level* α or *size* of a test and $1-$ the second its *power* (denoted in Hansen (2021) as $\pi(\theta)$).
- The size and power of a test are in general related. Reducing the probability of a Type I error leads to a higher probability of a Type II error.

- We make an error or type II if we reject a true null hypothesis. We want to keep this probabibility

$$\mathbb{P}(\text{Reject } \mathbb{H}_0 | \mathbb{H}_0 \text{ true}) = \mathbb{P}(T > c | \mathbb{H}_0 \text{ true}) \tag{12}$$

  small.

- We tend to rely on test statistics for which we know that, *under the null hypothesis*,

$$T \xrightarrow{d} \xi \tag{13}$$

  where ξ has a known distribution G.

- If *asymptotic null distribution* G does not depend on θ (or other unknown parameters), we say T is *asymptotically pivotal*.

- The *asymptotic size* of our test is the asymptotic probabilitty

$$\mathbb{P}(T > c | \mathbb{H}_0 \text{ true}) = \mathbb{P}(\xi > c) = 1 - G(c) \tag{14}$$

  This is called the *significance level* of the test and determines the choice of c.

- While choosing α is a matter of judgement, judgement is rarely used and $\alpha = .05$ tends to be the default choice.
- If a test rejects the null for a given α, the statistic is often said to be *statistically significant*; in the opposite case we say *statistically insignificant*.
- Hypothesis testing is a binary activity (accept/reject); we may prefer to report the asymptotic p-value

$$p = 1 - G(T). \tag{15}$$

- When $p \leq \alpha$ we reject $\mathbb{H}_0$ (allowing for binary testing) but can convey more information than in the binary case.
- The asymptotic distribution of p under the null is uniform on $[0, 1]$:

$$\begin{aligned} \mathbb{P}(1 - G(\xi) \leq u) =& \mathbb{P}(1 - u \leq G(\xi) = 1 - \mathbb{P}(\xi \leq G^{-1}(1 - u)) \\ =& 1 - G(G^{-1}(1 - u)) = 1 - (1 - u) = u \end{aligned} \tag{16}$$

Hypotheses and tests

1. Choose significance level α.
2. Select a test statistic T whose asymptotic distribution G under $\mathbb{H}_0$ is known.
3. Set the critical value c from $1 - G(c) = \alpha$.
4. Calculate asymptotic p-value $p = 1 - G(c)$.
5. Reject $\mathbb{H}_0$ if $T > c$ or $p < \alpha$.
6. Accept $\mathbb{H}_0$ if $T \leq c$ or $p \geq \alpha$.
7. Report p.

Hypothesis testing under ML estimation

- $\mathbb{H}_0$ is chosen in general such that it places restrictions on θ.
  Let $\widehat{\theta}_R$ be the MLE of θ under the restrictions and $\widehat{\theta}_U$ the unrestricted MLE. $L(\widehat{\theta}_R)$ and $L(\widehat{\theta}_U)$ are the values of the likelihood function under the two cases.
- Econometricians use three basic strategies to construct test based on likelihood:
  1. likelihood-ratio tests
  2. Wald tests
  3. Lagrange-multiplier tests

- If the restrictions are true, $L(\widehat{\theta}_R)$ is close to $L(\widehat{\theta}_U)$, since a free estimate will be near the the true restricted parameter.
- A *Likelihood ratio* (LR) test statistic is

$$\lambda = \frac{L(\widehat{\theta}_R)}{L(\widehat{\theta}_U)} \in [0, 1] \tag{17}$$

- Under some technical conditions

$$-2 \ln \lambda \sim \chi^2(\text{number of restrictions}). \tag{18}$$

- The LR test has good properties, but its implementation requires that both the restricted and unrestricted likelihoods be maximised.
- It can be difficult to maximise either the restricted or the unrestricted likelihood. The two other test strategies apply to each of these cases.

Wald tests (in ML-estimation)

- A *Wald test* uses the *unrestricted* likelihood.
- Let $\mathbb{H}_0 : r(\theta) = c$ be the parameter restrictions. If $\mathbb{H}_0$ is true, the unrestricted MLE $\widehat{\theta}_U$ will "almost" fulfill the restrictions, but not exactly. The differences are random and $r(\hat{\theta}_U) - c \simeq 0$.
- The test statistic measures the difference between $r(\hat{\theta}_U)$ and c. It rejects the null if this distance is so great that it can not be viewed to be due to random fluctuation:

$$W = [r(\hat{\theta}_U) - c]'(\mathrm{Var}[r(\hat{\theta}_U) - c])^{-1}[r(\hat{\theta}_U) - c] \tag{19}$$

- The Wald statistic has a χ^2 distribution with degrees of freedom equal to the number of restrictions (the number of elements in c.
- The variance matrix $\mathrm{Var}[r(\hat{\theta}_U) - c]$ in in general unkown but a consistent estimate of it suffices.

Lagrange-multiplier tests

- The *Lagrange-multiplier* (LM) test uses the restricted likelihood.
- To estimate the restricted MLE, the log likelihood function is augmented by a term involving Lagrange multipliers:

$$\ln L^*(\theta) = \ln L(\theta) + \lambda' [r(\theta) - c] \tag{20}$$

- The restricted MLE is the solution to the problem:

$$\begin{aligned} \frac{\partial \ln L^*(\theta)}{\partial \theta} &= \frac{\partial \ln L(\theta)}{\partial \theta} + R'\lambda = 0 \\ \frac{\partial \ln L^*(\theta)}{\partial \lambda} &= r(\theta) - c = 0. \end{aligned} \tag{21}$$

- If the restrictions are true, the vector of Lagrange multipliers is small and

$$\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} = -\hat{R}'\hat{\lambda} = \hat{g}_R \simeq 0. \tag{22}$$

- The test statistic measures the distance of $\hat{\lambda}$ from zero and allows us to decide if that distance is too great to have occurred by chance:

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R}\right)' [I(\theta)]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R}\right). \tag{23}$$

- This statistic also follows a χ^2 distribution with degrees of freedom equal to the number of restrictions / elements in λ / c. A consistent estimator of the information matrix allows us to implement this test in practice.
- The three likelihood-based test approaches are *asymptotically equivalent*, i.e., in large samples they have equal power.

- The t-test is the most commonly performed (and abused) statistical test; of the hypotheses that for a single parameter

$$\mathbb{H}_0 : \theta = \theta_0 \tag{24}$$

- Most often (and unreflectingly) $\theta_0 = 0$.
- The t-statistic is

$$t_n(\theta) = \frac{\widehat{\theta} - \theta}{s(\widehat{\theta})}. \tag{25}$$

- As $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta)$ and $\hat{V}_\theta \xrightarrow{p} V_\theta$,

$$\begin{aligned} t_n(\theta) &= \frac{\widehat{\theta} - \theta}{s(\widehat{\theta})} = \frac{\sqrt{n}(\widehat{\theta} - \theta)}{\sqrt{\hat{V}_\theta}} \\ &\xrightarrow{d} \frac{N(0, V_\theta)}{\sqrt{V_\theta}} = N(0, 1). \end{aligned} \tag{26}$$

- The asymptotic distribution of t_n does not depend on θ, so it is asymptotically pivotal. In the normal regression model, in small samples t_n has a t-distribution with n degrees of freedom, in which case it is exactly pivotal.

- Note that in eq. 24, the alternative is not explicitly stated.
- In most cases, the alternative is $\mathbb{H}_1 : \theta \neq \theta_0$ (2-sided) but it could be $\mathbb{H}_1 : \theta > \theta_0$ or $\mathbb{H}_1 : \theta < \theta_0$ (1-sided)
- For 2-sided $\mathbb{H}_1$, the test-statistic of interest is $|t_n(\theta)|$ which by the *continuous mapping theorem* $|t_n(\theta)| \xrightarrow{d} |Z|$. Then

$$\begin{aligned}\mathbb{P}(|Z| < u) =& \mathbb{P}(-u \leq Z \leq u) = \mathbb{P}(Z \leq u) - \mathbb{P}(Z \leq -u) \\ &= \Phi(u) - \Phi(-u) = 2\Phi(u) - 1 := \overline{\Phi}(u)\end{aligned} \tag{27}$$

- By choosing the critical value $c = 1.96$, the asymptotic α in our 2-sided test is 5%.
- For c, α satisfying $\alpha = 2(1 - \Phi(c))$

$$\mathbb{P}(|t_n(\theta_0)| \geq c | \mathbb{H}_0) \longrightarrow \alpha \tag{28}$$

α is the asymptotic significance level.

- 1-sided test use the value (rather than absolute value) of $t_n(\theta_0)$.
- The critical value c is chosen to satisfy $\alpha = 1 - \Phi(c)$ (1.645 for $\mathbb{H}_1 : \theta > \theta_0$, -1.645 for $\mathbb{H}_1 : \theta < \theta_0$).
- 1-sided tests are rare, for the good reason that they are "incomplete" unless there are good substantive reason to ignore "the other half" of the real line.

Wald tests

- We can think of the q restrictions under $\mathbb{H}_0$, $\boldsymbol{r}(\beta) = \theta$. If $\mathbb{H}_0$ is true, the restrictions are a point θ_0.
- So we have a q vector of parameters θ, with consistent estimator $\widehat{\theta}$ with variance matrix $\boldsymbol{V}_\theta$, the weighted Euclidean distance

$$W_n = n(\widehat{\theta} - \theta_0)' \widehat{\boldsymbol{V}}_\theta^{-1} (\widehat{\theta} - \theta_0) \tag{29}$$

  is typically called a Wald statistic.
- When $q = 1$, $W_n = t_n^2$. (A 2-sided t-test and a Wald test are then equivalent.)
- As $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \boldsymbol{V}_\theta)$ and $\widehat{\boldsymbol{V}_\theta} \xrightarrow{p} \boldsymbol{V}_\theta$,

$$W_n(\theta) = \sqrt{n}(\widehat{\theta} - \theta_0)' \widehat{\boldsymbol{V}}_\theta^{-1} \sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \boldsymbol{Z}' \boldsymbol{V}_\theta^{-1} \boldsymbol{Z} \sim \chi_q^2 \tag{30}$$

  The asymptotic distribution of Wald statistics is thus the χ^2.

- With multiple restrictions the multiple testing problem could be posed as
$$\mathbb{H}_0 : \boldsymbol{r}(\beta) = \theta_0 \text{ against } \mathbb{H}_1 : \boldsymbol{r}(\beta) \neq \theta_0 \tag{31}$$
We often write up the restriction $\boldsymbol{r}()$ such that we can take $\theta_0 = \mathbf{0}$.
- For asymptotic results, the matrix of first derivatives
$$\boldsymbol{R} = \frac{\partial \boldsymbol{r}(\beta)'}{\partial \beta} \tag{32}$$
must have rank q.
- Linear restrictions have $\boldsymbol{r}(\beta) = \boldsymbol{R}'\beta - \theta_0$.

Wald tests

- In a given setting, *assuming the restrictions* $\theta = \boldsymbol{r}(\beta) = \theta_0 = \mathbf{0}$ *are true*,

$$W_n = n\widehat{\theta}'\widehat{\boldsymbol{V}}_\theta^{-1}\widehat{\theta} = n\boldsymbol{r}(\widehat{\beta})'(\widehat{\boldsymbol{R}}'\widehat{\boldsymbol{V}}_\beta\widehat{\boldsymbol{R}})^{-1}\boldsymbol{r}(\widehat{\beta}) \tag{33}$$

  converges to a χ^2_q distribution (denoted here by $G_q(u)$).
- For linear restrictions have $\boldsymbol{r}(\beta) = \boldsymbol{R}'\beta - \theta_0$,

$$W_n = n(\boldsymbol{R}'\widehat{\beta} - \theta_0)'(\boldsymbol{R}'\widehat{\boldsymbol{V}}_\beta\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\beta} - \theta_0). \tag{34}$$

- For α, c satisfying $\alpha = 1 - G_q(c)$,

$$\mathbb{P}(W_n \geq c|\mathbb{H}_0) \to \alpha \tag{35}$$

  so the asymptotic significance level is α.
- The asymptotic p-value is

$$p_n = 1 - G_q(W_n) \tag{36}$$

Criterion-based tests

- The standard Wald-test is based on the *value of the restrictions* evaluated at an *unrestricted estimate* $\widehat{\beta}$ (cf. the ML case).
- *Criterion-based* testing can be used when estimation is based on a minimizing a criterion function $J(\beta)$ for $\beta \in \boldsymbol{B}\ldots$
- $\ldots$ and makes use *both* an *unrestricted* estimator $\widehat{\beta}$ and one $\widetilde{\beta}$ that is *restricted* to lie in $\boldsymbol{B}_0$ (for $\mathbb{H}_0 : \boldsymbol{\beta} \in \boldsymbol{B}_0$):

$$\begin{aligned} \widehat{\beta} &= \arg\min_{\beta \in \boldsymbol{B}} J_n(\beta) \\ \widetilde{\beta} &= \arg\min_{\beta \in \boldsymbol{B}_0} J_n(\beta). \end{aligned} \tag{37}$$

- Testing is based on the difference in the criterion function evaluated in both cases. The criterion-based (aka. *minimum-distance* or *likelihood-ratio-like*) test-statistic for $\mathbb{H}_0$ against $\mathbb{H}_1$ is proportional to

$$\begin{aligned} J &= \min_{\beta \in \boldsymbol{B}_0} J_n(\beta) - \min_{\beta \in \boldsymbol{B}} J_n(\beta) \\ &= J_n(\widetilde{\beta}) - J_n(\widehat{\beta}) > 0. \end{aligned} \tag{38}$$

Wald test as an F-test

- The linear null hypothesis $\mathbb{H}_0 : \boldsymbol{R}'\beta - \theta_0 = \mathbf{0}$ is often tested using an F-test:

$$F_n = \frac{(SSE_n(\widetilde{\beta}_{CLS}) - SSE_n(\widehat{\beta}))/q}{SSE_n(\widehat{\beta})/(n-k)} = \frac{n-k}{q}\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \tag{39}$$

  $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ are the restricted and unrestricted estimated residual variances, q is the number of restrictions.
- In the normal regression, F_n follows an exact F-distribution.
- For linear restrictions under homoscedastic errors, the Wald and F-statistics are related by

$$F_n = W_n^0/q. \tag{40}$$

- A special case of the F-statistic is when all coefficients except the intercept are set to zero, so the “constrained” model is the “intercept only” model and $\tilde{\sigma}^2 = \widehat{\mathrm{Var}}[Y]$ (provides the basis for figuring out the “significance” of R^2).

Hausman tests

- A famous, general idea about testing proposed by Hausman (1978) applies in situations where there are two estimators, $\widehat{\theta}_E$, $\widetilde{\theta}_I$ for which
  - under the null: both $\widehat{\theta}_E$ and $\widetilde{\theta}_I$ consistent, but $\widehat{\theta}_E$ is efficient
  - under the alternative: $\widehat{\theta}_E$ is inconsistent but $\widetilde{\theta}_I$ is not
- The test is based on the difference in the estimators $d = \widetilde{\theta}_I - \widehat{\theta}_E$ and is

$$H = d' \{\mathrm{Var}[d]\}^{-1} d. \tag{41}$$

  The problem is to find $\mathrm{Var}[d]$ and in particular, we need the covariance of $\widetilde{\theta}_I$ and $\widehat{\theta}_E$.

- Hausman showed that the *efficient estimator* $\widehat{\theta}_E$ and the *difference* between the efficient $\widehat{\theta}_E$ and the inefficient estimator $\widetilde{\theta}_I$ have a zero covariance:

$$\mathrm{Var}[\widehat{\theta}_E, \widehat{\theta}_E - \widetilde{\theta}_I] = \mathbf{0} \tag{42}$$

- It follows that

$$\mathrm{Var}[\widetilde{\theta}_I - \widehat{\theta}_E] = \mathrm{Var}[\widetilde{\theta}_I] - \mathrm{Var}[\widehat{\theta}_E]. \tag{43}$$

- The Hausman test statistic is

$$H = (\widetilde{\theta}_I - \widehat{\theta}_E)' \left(\mathrm{Var}[\widetilde{\theta}_I] - \mathrm{Var}[\widehat{\theta}_E]\right)^{-1} (\widetilde{\theta}_I - \widehat{\theta}_E) \tag{44}$$

- This is χ^2 distributed with q degrees of freedom (q depends on the details of the problem).
- The (estimated) variance matrix of the difference is often small and therefore difficult to invert so a generalized inverse (the generalized Moore-Penrose inverse) is used instead.

Multiple tests

- The typical empirical economics paper looks at a large number of coefficient estimates and their standard errors (as well as p-values etc).
- To examine the "significance" of a coefficient *after* having examined that of $k - 1$ others before it invokes the problem of *multiple testing* (alt. *multiple comparisons*).
- This is a complex topic which we here only scratch at the surface.

Multiple tests – example

- The following example is taken from Goldberger (1991, p. 262).
- Suppose there are k null hypotheses, all of which are true and which are tested using procedures with nominal size α.
- What is the likelihood of rejecting at least one of them?

$$\begin{aligned}\mathbb{P}[\text{at least one rejection}] &= 1 - \mathbb{P}[\text{all accepted}] \\ &= 1 - (1-\alpha)^k.\end{aligned} \tag{45}$$

- demo:
  Set $\alpha = \{0.100, 0.050, 0.010, 0.005\}$ and $k = \{2, 5, 10, 20, 50\}$. The resulting actual sizes/true significance levels are:

	k				
	2	5	10	20	50
α					
0.100	0.190	0.410	0.651	0.878	0.995
0.050	0.098	0.226	0.401	0.642	0.923
0.010	0.020	0.049	0.096	0.182	0.395
0.005	0.010	0.025	0.049	0.095	0.222

Bonferroni correction

- Suppose we examine k coefficients and the null hypotheses $\mathbb{H}_{j,0} : \beta_j = 0$ associated with them.
- Ignoring the repeated nature of testing, we would reject each of the nulls if the $z/t - statistic$ exceeds the $1 - \alpha$ critical value of normal distribution (≈ 1.96).
- Suppose at least one of the coefficients is observed to be "significant", i.e., has a p-value less than α.
- Now ask the question: under the *joint* hypothesis that the whole set ("family") of null hypotheses are true, what is the probability that the smallest p-value is less than α?
- This is hard to answer in general, but the Bonferroni correction bounds this by αk.
- Further: to bound the familywise error probability below α, reject only if the smallest. p-value is less than α/k (the familywise Bonferroni p-value is $k \min_{j \leq k} p_j$).

Bonferroni bounds

- Consider hypotheses $\mathbb{H}_{j,0}, j = 1, \ldots, k$ with associated tests and p-values p_j all such that when $\mathbb{H}_{j,0}$ is true, $\lim_{n\to\infty} \mathbb{P}(p_j < \alpha) = \alpha$.
- The event that (at least) one of the k is significant can be written as

$$\left\{\min_{j\leq k} p_j < \alpha\right\} = \cup_{j=1}^{k} \{p_j < \alpha\}. \tag{46}$$

- By Boole's inequality, we get

$$\mathbb{P}\left(\min_{j\leq k} p_j < \alpha\right) \leq \sum_{j=1}^{k} \mathbb{P}(p_j < \alpha) \to k\alpha \tag{47}$$

and

$$Pr\left(\min_{j\leq k} p_j < \alpha/k\right) \leq \sum_{j=1}^{k} \mathbb{P}(p_j < \alpha/k) \to \alpha. \tag{48}$$

Bonferroni bounds – illustration

- Suppose we have two coefficient estimates with p-values .04 ("significant") and .15 ("not significant"). A 5% Bonferroni test requires the smallest p-value be $< \alpha/2 = .025$ (so fails to reject) and a Bonferroni familywise p-value is $2 \times \min\{.04, .15\} = .08$ which fails at the .05-level.
- Contrast this with p-values .01, .15, with Bonferroni familywise p-value .02 ($< .05$, so rejects).

Case	p_1	p_2	Bonf test $\min p_j < \alpha/k$	Bonf familywise $k \min p_j < \alpha$
A	.04	.15	$.04 > .025$ (not reject)	$.08 > .05$ (not reject)
B	.01	.15	$.01 < .025$ (reject)	$.02 < .05$ (reject)

Power

- The power of a test is the likelihood $\pi(\theta)$ of rejecting $\mathbb{H}_0$ when $\mathbb{H}_1$.
- Consider e.g. $Y \sim N(\theta, \sigma^2)$ with known σ^2; the t-statistic $T(\theta) = \sqrt{n}(\bar{Y} - \theta)/\sigma$.
- The test-statistic for $\mathbb{H}_0$ is $T = T(0)$ which in turn is

$$T(0) = T(\theta) + \sqrt{n}\theta/\sigma. \tag{49}$$

- With the true θ, $T(\theta)$ is $N(0, 1) = Z$, so

$$\mathbb{P}(T > c|\theta) = \mathbb{P}(Z + \sqrt{n}\theta/\sigma > c) = 1 - \Phi(c - \sqrt{n}\theta/\sigma). \tag{50}$$

- This is a (monotonically) increasing function of θ, n and decreasing in σ, c.
- Finally, recall that a test is *consistent* if $\forall\theta \in \Theta_1$, $\mathbb{P}(\text{reject } \mathbb{H}_0|\theta) \longrightarrow 1$ as $n \longrightarrow \infty$.

Asymptotic local power – scalar case

- Suppose the restrictions are not true so rather than θ_0, we have some $\theta_n = r(\beta_n)$, not β. (Why we index by n becomes apparent below.)
- Then for a "localizing parameter" h we can write

$$\theta_n = \theta_0 + n^{-1/2}h. \tag{51}$$

  θ_n is "local to" θ_0 and as n grows becomes closer.
- We then have

$$\begin{aligned}\sqrt{n}(\widehat{\theta} - \theta_0) &= \sqrt{n}(\widehat{\theta} - \theta_n) + \sqrt{n}(\theta_n - \theta_0) \\ &= \sqrt{n}(\widehat{\theta} - \theta_n) + h.\end{aligned} \tag{52}$$

  We know that $\sqrt{n}(\widehat{\theta} - \theta_n) \xrightarrow{d} \sqrt{V_\theta} Z$ where $Z \sim N(0, 1)$ so

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \sqrt{V_\theta} Z + h \sim N(h, V_\theta). \tag{53}$$

- For instance, for the t-statistic we have

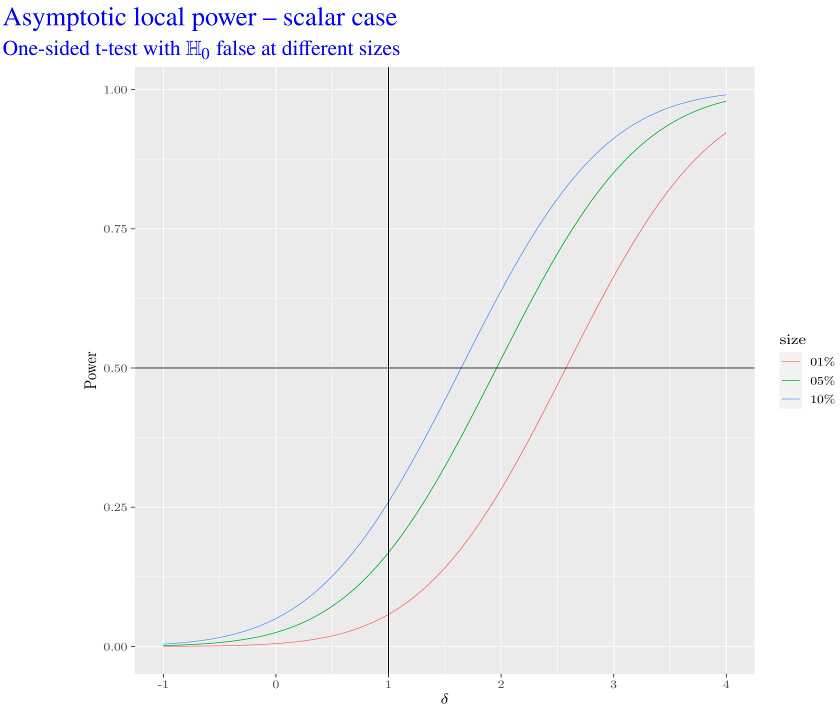
$$T = \frac{\widehat{\theta} - \theta_0}{s(\theta)} \xrightarrow{d} \frac{\sqrt{V_\theta} Z + h}{\sqrt{V_\theta}} \sim Z + \delta \tag{54}$$

  where $\delta = h/\sqrt{V_\theta}$.

Asymptotic local power – scalar case

- The *asymptotic local power* of a one-sided t-test is

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\text{reject } \mathbb{H}_0) &= \lim_{n\to\infty} \mathbb{P}(t > c) \\ &= \mathbb{P}(Z + \delta > c) = 1 - \Phi(c - \delta) = \pi(\delta) \end{aligned} \tag{55}$$


Asymptotic local power – scalar case
One-sided t-test with $\mathbb{H}_0$ false at different sizes
Power
1.00
0.75
0.50
0.25
0.00
-1
0
1
2
3
4
δ
size
01%
05%
10%

Asymptotic local power – vector case

- The vector values case is very similar; we now have $\theta_n = \boldsymbol{r}(\beta_n)$ and

$$\theta_n = \theta_0 + n^{-1/2}\boldsymbol{h} \tag{56}$$

and

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \sqrt{n}(\widehat{\theta} - \theta_n) + \boldsymbol{h} \xrightarrow{d} \boldsymbol{Z}_h \sim N(\boldsymbol{h}, \boldsymbol{V}_\theta). \tag{57}$$

- The Wald statistic is

$$\begin{aligned} W =& n(\widehat{\theta} - \theta_0)' \widehat{\boldsymbol{V}}_\theta^{-1} (\widehat{\theta} - \theta_0) \\ &\xrightarrow{d} \boldsymbol{Z}_h' \boldsymbol{V}_\theta^{-1} \boldsymbol{Z}_h \sim \chi_q^2(\lambda) \end{aligned} \tag{58}$$

where χ^2 is the non-central χ^2-distribution with q degrees of freedom and non-centrality parameter $\lambda = \boldsymbol{h}' \boldsymbol{V}_\theta^{-1} \boldsymbol{h}$.

- Goldberger, Arthur S (1991). *A Course in Econometrics*. Cambridge, Massachusettes: Harvard University Press.
- Gonick, Larry and Smith, Woollcott (1993). *The Cartoon Guide to Statistics*. New York: HarperPerennial.
- Hansen, Bruce E (2021). *Econometrics*. Madison, WI: University of Wisconsin.
- Hausman, J A (1978). "Specification tests in econometrics". *Econometrica* 46, pp. 1251–1271.