Multiple Regression Analysis: Inference

ECONOMETRICS (ECON 360)

BEN VAN KAMMEN, PHD

Introduction

When you perform statistical inference, you are primarily doing one of two things:
- Estimating the boundaries of an interval in which you suspect the population parameter to lie, or
- Testing the validity of some hypothesized value for the population parameter.

These are procedures you have practiced in introductory statistics courses. In ECON 360, we will apply these procedures to single regression coefficient estimates.
- Additionally we will apply them to combinations of estimates, e.g., we will ask questions like:
- Are *several* coefficients likely to be simultaneously equal to zero?
- What is the confidence interval for the *sum* of several coefficients?

Introduction (continued)

In order to answer questions like those above, one needs to know more about the distributions of the estimators he is using.

- Constructing a confidence interval requires a margin of error, that relies on a value from a statistical distribution, i.e., a "z" or "t" value.

- Testing a hypothesis about a parameter involves asking how likely one would be to get an estimate so far from the hypothesized value were the null hypothesis true. If that probability ("p value") is low, the null hypothesis is probably not true, but you need to know how the estimate's sampling distribution looks in order to assess the probability in the first place.

Outline

Sampling Distributions of the OLS Estimators.

Testing Hypotheses about a Single Parameter.

Confidence Intervals.

Testing Hypotheses about a Linear Combination of Parameters.

Testing Multiple Linear Restrictions, the F Test.
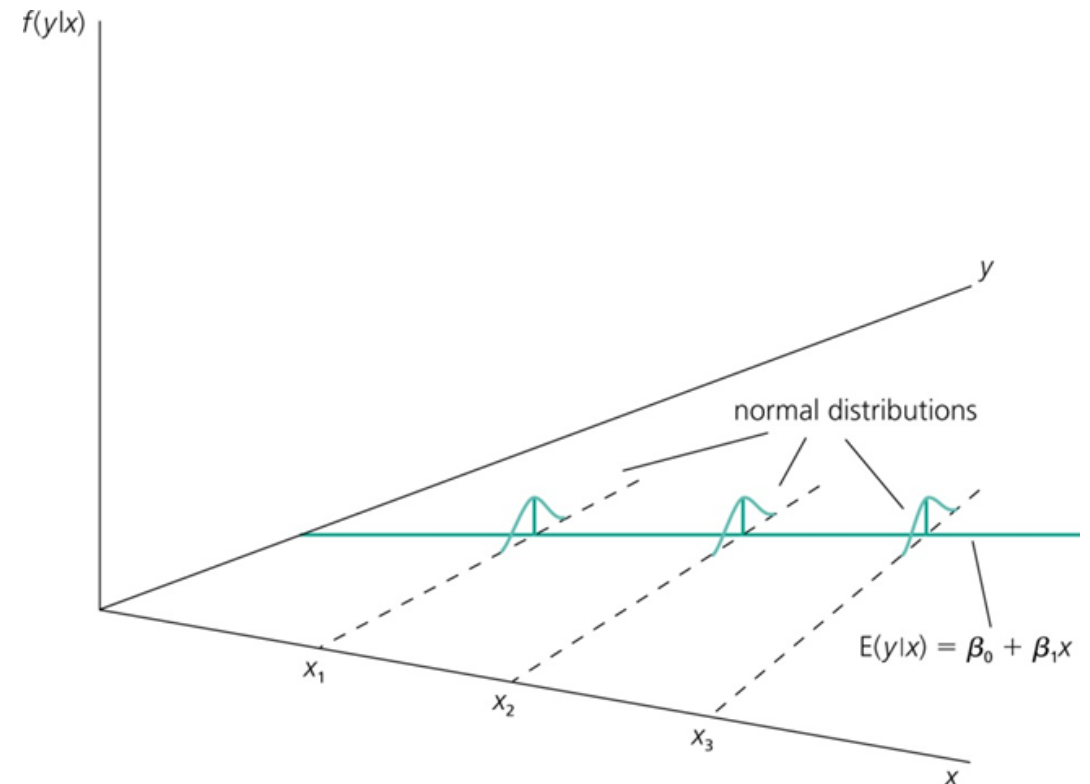
Reporting Regression Results.

Sampling distributions of the OLS estimators

Consider the consequences of assuming (Assumption MLR.6) that the error term is <u>normally distributed</u>,

◦ i.e., in addition to previous assumptions that it has zero expectation and constant variance.

All together these can be stated as follows:
$$u \sim Normal(0, \sigma^2).$$

The classical linear model (CLM)

We are still assuming <u>mean independence from the x variables</u>, as well.

A few consequences of this additional assumption:

◦ Combined with Assumptions MLR.1 through MLR. 5, the regression model is known as the <u>classical linear model</u> ("CLM").

◦ The CLM is now the best (least variance) estimator among estimators—not just linear ones, as before under the BLUE result.

◦ y, conditional on all the x variables, is also normally distributed; all the variance comes from the errors.

CLM (continued)

The last property is summarized on the figure(from the textbook) and in notation below:
$$(y|x_1 \ldots x_k) \sim Normal(\beta_0 + \beta_1 x_1 + \ldots \beta_k x_k, \sigma^2).$$

We will examine circumstances under which MLR.6 is likely to be a bad "unrealistic" assumption—and there are plenty of them—later.

◦ You can show that the regression coefficients are normally distributed as well.

CLM (continued)

Recall the "partialling out" expression for regression estimates:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}, \text{where}$$

- $\hat{r}_1$ is the residual from regressing x_1 on the other x variables in the model.
- $\hat{r}_{i1} \equiv x_{i1} - \hat{x}_{i1}; \hat{x}_{i1}$ is the fitted value.

$\widehat{\beta}_1$ can be decomposed into its expected value and a deviation from the expected value by substituting for y_i above:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + u_i)}{\sum_{i=1}^{n} \hat{r}_{i1}^2}.$$

CLM (continued)

Once again the $(\hat{r}_{i1})$ residuals are uncorrelated with the other x variables. Then,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1}(\beta_1 x_{i1} + u_i)}{\sum_{i=1}^n \hat{r}_{i1}^2} = \frac{\beta_1 \sum_{i=1}^n \hat{r}_{i1} x_{i1} + \sum_{i=1}^n \hat{r}_{i1} u_i}{\sum_{i=1}^n \hat{r}_{i1}^2}.$$

$$\Leftrightarrow \widehat{\beta}_1 = \frac{\beta_1 \sum_{i=1}^n \hat{r}_{i1}(\hat{x}_{i1} + \hat{r}_{i1}) + \sum_{i=1}^n \hat{r}_{i1} u_i}{\sum_{i=1}^n \hat{r}_{i1}^2} = \frac{\beta_1 \sum_{i=1}^n \hat{r}_{i1}^2 + \sum_{i=1}^n \hat{r}_{i1} u_i}{\sum_{i=1}^n \hat{r}_{i1}^2};$$

the last equality uses the fact, as in Chapter 3, that the residuals from regressing x_1 on $\forall\ x_{j \neq 1}$ are uncorrelated with $\hat{x}_1$, because it is a linear function of $x_{j \neq 1}$.

CLM (concluded)

Lastly a little simplification yields:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \hat{r}_{i1} u_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}.$$

And if the second term above is just a sum of normally distributed random variables, as we have assumed it is (because of u), the sum is also normally distributed

◦ (see Appendix B.5 in Wooldridge).

◦ Since they are functions of the independent variables, $\hat{r}$ are not treated as a random variables.

$$\widehat{\beta}_j \sim Normal\left(\beta_j, Var\left(\widehat{\beta}_j\right)\right), \text{where } Var\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{SST_j(1 - R_j^2)}.$$

Inference about the CLM

From introductory statistics, this means you can standardize $\widehat{\beta}_j$ by subtracting its mean and dividing by its standard deviation.

This turns it into a normally distributed random variable ("z") with mean zero and standard deviation 1:

$$z = \frac{\widehat{\beta}_j - \beta_j}{\left(Var\left(\widehat{\beta}_j\right)\right)^{\frac{1}{2}}} ; z \sim Normal(0,1).$$

Testing hypotheses about a single parameter

In practice when we perform tests on $\widehat{\beta}$ estimates, σ^2 is not known and is replaced with its sample counterpart.

This consumes a degree of freedom and requires the empiricist to use the "t" distribution (with the remaining $n - k - 1$ degrees of freedom) instead of "z".

◦ For testing whether the coefficient estimate is significantly different from zero, use a t test.

$$t_{n-k-1} = \frac{\widehat{\beta}_j - 0}{se(\widehat{\beta}_j)} = \left(\widehat{\beta}_j - 0\right)\left[\frac{\widehat{\sigma}^2}{SST_j\left(1 - R_j^2\right)}\right]^{-\frac{1}{2}}$$

Testing hypotheses about a single parameter (continued)

This test statistic is based on the common <u>null hypothesis</u> (H_0) that:

- "after controlling for the other explanatory variables, there is <u>no</u> relationship between x_j and y".
- Stated in formal notation, it is:

$$H_0: \ \beta_j = 0.$$

If "t" is large (in absolute value), we have a statistic "far away" from the null hypothesized value, which is unlikely to have been generated by an "unlucky" sample

- and therefore evidence in support of the <u>alternative hypothesis</u>:
- ("after controlling for the other explanatory variables, there <u>is</u> a relationship between x_j and y").
- Or,

$$H_1: \ \beta_j \neq 0.$$

Testing hypotheses about a single parameter (continued)

Naturally you can test hypotheses of the <u>one tail</u> variety, as well, e.g.,
$$H_0: \ \beta_j \leq 0 \text{ with } H_1: \ \beta_j > 0.$$

In either case "far away" (and the decision to reject the null hypothesis or not) depends on probability.

Specifically it depends on the probability density function (pdf) of the sampling distribution (t_{n-k-1}), as if the null hypothesis is true.

It may look like the t distribution on the next slide.

Example of a *t* distribution: with *p* value and critical value
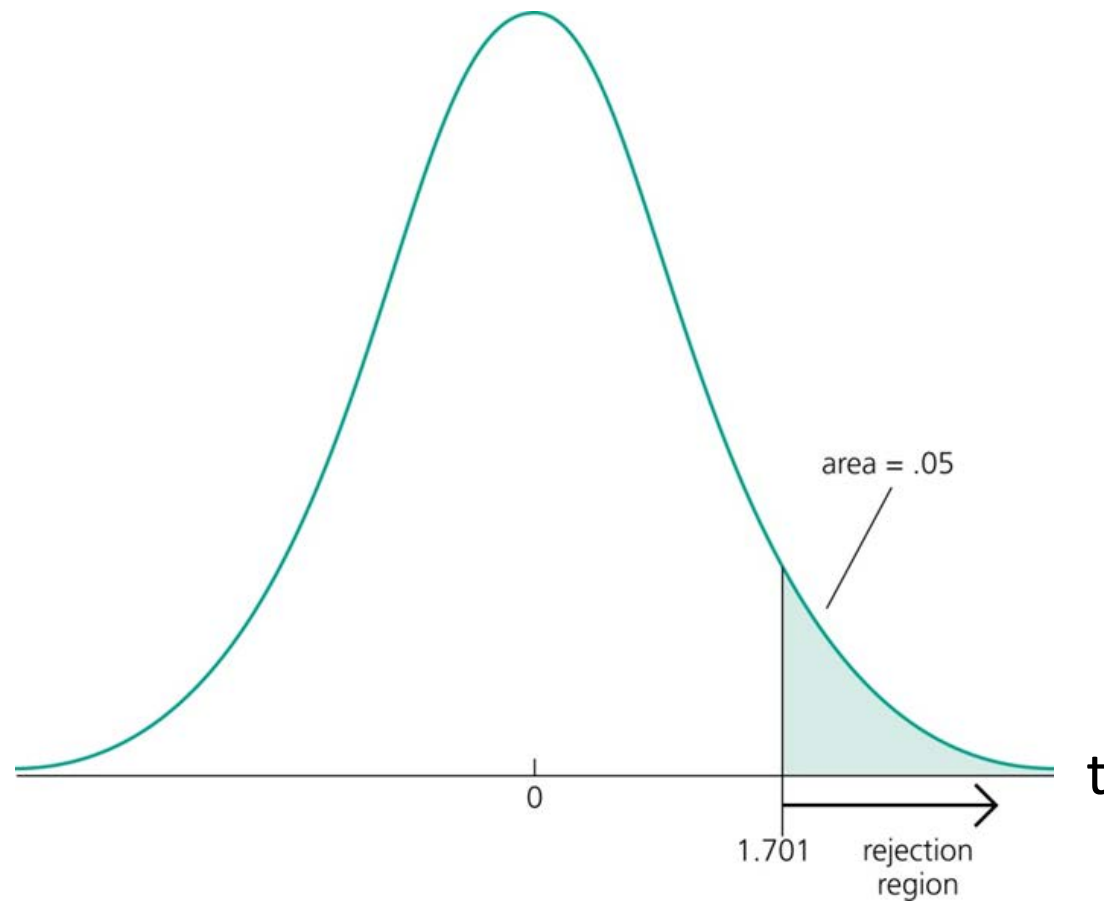
Testing hypotheses about a single parameter (continued)

The center (expected value) of the pdf is 0 if the null hypothesis is true, and values in the tails are increasingly improbable,

◦ reflecting the Law of Large Numbers and the rarity with which random samples deviate very far from the population mean.

◦ In the example above, there is only a 0.05 probability of sampling a t value greater than 1.701; this is reflected by calculating the area under the distribution in the upper tail of the curve.

◦ By now this should be ingrained, but it is worth mentioning one more time: the area in the left (of 1.701) tail is 0.95, i.e., the probability of sampling $t \leq 1.701$ is 0.95.

Testing hypotheses about a single parameter (continued)

The reason this example is used is that 0.05 is a commonly used standard for statistical significance in published empirical work.

In this case, a t statistic greater than 1.701 would compel the researcher to reject the null hypothesis that $\beta_j \leq 0$ and argue that there is a positive relationship between x_j and y.

◦ He would be confident that there is only a 0.05 (5%) likelihood that his finding is a Type I Error (rejecting H_0 when it is, in fact, true), i.e., 5% is his Confidence Level in the result.*

To satisfy an even greater standard for confidence in your results, an even larger t is necessary,

◦ e.g., to be 99% or 99.9%, t would have to be over 2.467 and 3.408, respectively.

*This is also sometimes called a "false positive".

Testing hypotheses about a single parameter (concluded)

Values can either be looked up on a "t chart" like the one in the Appendix of the textbook, or they can be calculated by software.

The upper tail probability can be calculated in STATA for a t distribution with "n" degrees of freedom using the function "ttail".

◦ Its syntax is: **ttail(n,t)**, where "t" is the cut-off for the upper tail.

You can also look up the value of "t" for which "p" is the upper tail area using "invttail".

◦ The syntax is: **invttail(n,p)**.

Using the estimated t value from a regression to look up the significance level (area in tail(s)) at which the null hypothesis can be rejected reveals the p-value for the result. It tells you the smallest probability of Type I Error with which the null hypothesis can be rejected.

Guidelines for discussing economic and statistical significance

If an estimate is statistically significant, discuss the magnitude of the coefficient to get an idea of its economic or practical importance.

The fact that a coefficient is statistically significant does not necessarily mean it is economically or practically significant!

If a variable is statistically and economically important but has the "wrong" sign, the regression model might be misspecified.

If a variable is statistically insignificant at the usual levels (10%, 5%, 1%), one may think of dropping it from the regression.

Confidence intervals

A <u>confidence interval</u> can be constructed <u>around the point estimate</u> of $\widehat{\beta}_j$ using the familiar procedure from introductory statistics:

$$\widehat{\beta}_j \pm se(\widehat{\beta}_j) * c,$$

where c comes from a t chart (or software) and reflects the desired confidence level and appropriate $(n - k - 1)$ degrees of freedom.

The two boundaries of the interval tell the reader a range in which the true parameter value, β_j, will lie—with the specified probability, e.g., 95%.

Confidence intervals (continued)

This result comes from taking the absolute value of the standardized test statistic. Since the t distribution is symmetric,

$$t = \frac{\widehat{\beta}_j - \beta_j}{se\left(\widehat{\beta}_j\right)} \rightarrow \Pr\left(\frac{\widehat{\beta}_j - \beta_j}{se\left(\widehat{\beta}_j\right)} \leq -c_\alpha\right) = \Pr\left(\frac{\widehat{\beta}_j - \beta_j}{se\left(\widehat{\beta}_j\right)} > c_\alpha\right) = \frac{\alpha}{2}\text{ , and}$$

$$\Leftrightarrow \Pr\left(\left|\frac{\widehat{\beta}_j - \beta_j}{se\left(\widehat{\beta}_j\right)}\right| > c_\alpha\right) = 2\frac{\alpha}{2} = \alpha, \text{where}$$

c_α is a constant that reflects the confidence level (specifically α is the probability the interval does not contain the parameter of interest).

◦ Lower alpha means a wider ("less precise") interval.

Confidence intervals (continued)

So:

$$\Pr\left(\left|\frac{\widehat{\beta}_j - \beta_j}{se\left(\widehat{\beta}_j\right)}\right| \leq c_\alpha\right) = (1 - \alpha), \text{ and } \Pr\left(\left|\widehat{\beta}_j - \beta_j\right| \leq c_\alpha se\left(\widehat{\beta}_j\right)\right) = (1 - \alpha).$$

Confidence intervals (concluded)

It follows from the properties of inequality expressions that:

$$\left|\widehat{\beta}_j - \beta_j\right| = \left|\beta_j - \widehat{\beta}_j\right| \leq c_\alpha se\left(\widehat{\beta}_j\right) \rightarrow -c_\alpha se\left(\widehat{\beta}_j\right) \leq \beta_j - \widehat{\beta}_j \leq c_\alpha se\left(\widehat{\beta}_j\right),$$

$$\text{and } \widehat{\beta}_j - c_\alpha se\left(\widehat{\beta}_j\right) \leq \beta_j \leq \widehat{\beta}_j + c_\alpha se\left(\widehat{\beta}_j\right)$$

The probability that β_j lies in the interval $\left[\widehat{\beta}_j \pm se\left(\widehat{\beta}_j\right)c_\alpha\right]$ is just $(1 - \alpha)$:

$$\Pr\left[\widehat{\beta}_j - c_\alpha se\left(\widehat{\beta}_j\right) \leq \beta_j \leq \widehat{\beta}_j + c_\alpha se\left(\widehat{\beta}_j\right)\right] = (1 - \alpha).$$

Testing hypotheses about a linear combination of parameters

The preceding methods are taught in all introductory courses in statistics, and applying them to the CLM should serve mostly as practice.

Not all introductory courses, however, teach students how to test hypotheses about how parameters relate <u>to each other</u>—<u>rather than</u> how they relate <u>to a constant</u>. These are the kind of hypotheses we consider now.

For example we will practice methods for testing hypotheses like the following:
$$H_0: \ \beta_1 = \beta_2 \Leftrightarrow \beta_1 - \beta_2 = 0.$$

Expected value of a linear combination of parameters

The second equality gives some insight about how a test statistic can beconstructed.

We're performing a test on a <u>linear combination of estimates</u>.

◦ If each of them is normally distributed, again, so is the linear combination.

The difficulty we now face is finding the standard error of the linear combination of estimates,

◦ i.e., what is $Var\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$?

◦ The point estimate of $(\beta_1 - \beta_2)$ is just the difference in coefficient estimates: $\widehat{\beta}_1 - \widehat{\beta}_2$.

◦ Once we ascertain the standard error we can perform a t test as before.

Variance of a linear combination of parameters

The expected value of the difference is:

$$E(\widehat{\beta}_1 - \widehat{\beta}_2) = E(\widehat{\beta}_1) - E(\widehat{\beta}_2) = \beta_1 - \beta_2, \text{variance is:}$$

$$Var(\widehat{\beta}_1 - \widehat{\beta}_2) = E\left(\widehat{\beta}_1 - \beta_1 - (\widehat{\beta}_2 - \beta_2)\right)^2$$

$$= E(\widehat{\beta}_1 - \beta_1)^2 + E(\widehat{\beta}_2 - \beta_2)^2 - 2E[(\widehat{\beta}_1 - \beta_1)(\widehat{\beta}_2 - \beta_2)], \text{or}$$

$$Var(\widehat{\beta}_1 - \widehat{\beta}_2) = Var(\widehat{\beta}_1) + Var(\widehat{\beta}_2) - 2Cov(\widehat{\beta}_1, \widehat{\beta}_2).$$

Inference about a linear combination of parameters

The (covariance) term is the only one that is not obvious among the regression output.

- ◦ This quantity can be derived formally; software can calculate it for you, too.*

But there is another clever way of performing the test—by transforming one of the variables in the regression to estimate the quantity $\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$ and its standard error directly.

*to have STATA merely perform the test, the syntax is:

lincom {x1 variable}-{x2 variable}.

Inference about a linear combination of parameters (continued)

It is quite easy to do using the "lincom" command followed by the expression you want to test in terms of variables' names.

Performing this test *without* relying on pre-programmed "black box" commands means defining a new parameter: "theta".

$$\theta_1 \equiv \beta_1 - \beta_2, \text{ and we can substitute it into the regression model.}$$

Inference about a linear combination of parameters (continued)

$$y = \beta_0 + (\theta_1 + \beta_2)x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u, \text{ which you can rewrite:}$$

$$y = \beta_0 + \theta_1 x_1 + \beta_2(x_1 + x_2) + \ldots + \beta_k x_k + u.$$

Transforming x_1 and x_2 into a sum, and regressing y on the sum instead of x_2 itself, enables you to estimate theta and its standard error directly from the regression.

The *t* test can be performed using the coefficient estimate and standard error from variable x_1 in this regression.

Testing multiple linear restrictions, the F test

Continuing with the theme of testing multiple parameter values, consider the possibility of testing multiple hypotheses simultaneously.

For example imagine a set of variables that you suspect does not affect y (once you control for the other x variables).

One could verify this suspicion by testing the null hypothesis:

$$H_0: \ \beta_3 = 0; \beta_4 = 0; \beta_5 = 0 \text{ against } H_1: \ at \ least \text{ one is } \neq 0.$$

Caveat: joint significance

This is **<u>not</u>** the same thing as testing 3 hypotheses about single parameters separately!

Those individual *t* tests would place no restrictions on the other parameters and do not constitute a test of the joint hypothesis.

Testing a joint hypothesis with the SSR

The test for this kind of hypothesis is based on the Sum of Squared Residuals (SSR).

Particularly it is based on how the SSR increases when the coefficients in question are <u>restricted</u> to being zero, i.e., we make <u>exclusion restrictions</u> by estimating the regression without them.

If the <u>unrestricted model</u>, (with x_3, x_4, x_5) fits significantly better than the <u>restricted model</u> (without them: $\beta_3 = \beta_4 = \beta_5 = 0$), then the null hypothesis should be rejected.

Testing a joint hypothesis with the SSR (continued)

To make this formal, consider the unrestricted model with $(k + 1)$ parameters.

In the restricted model q of them are restricted to zero.

◦ If the exclusions are put at the end of the series, it would look like:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k-q} x_{k-q} + u.$$

Testing a joint hypothesis with the SSR (continued)

Each residual is normally distributed, and the sum (n terms) of the standardized ($\hat{u}_i/\sigma$) squares of normally distributed random variables follows a Chi Squared (χ^2) distribution with ($n - k - 1$) degrees of freedom.

$$\sum_{i=1}^{n} \left(\frac{\hat{u}_i}{\sigma}\right)^2 = \frac{SSR}{\sigma^2} \sim \chi^2_{n-k-1}.$$

When Chi Squared variables are added (subtracted), their degrees of freedom are also added (subtracted), i.e.,

$$\frac{SSR_{restricted}}{\sigma^2} - \frac{SSR_{unrestricted}}{\sigma^2} \sim \chi^2_{n-k-1+q-(n-k-1)=q}.$$

Testing a joint hypothesis with the SSR (continued)

The ratio of two Chi Squared random variables (times the inverted ratio of their degrees of freedom) follows an *F* distribution with numerator and denominator degrees of freedom determined by the two Chi Squared statistics.

I.e.,

$$\frac{\dfrac{\chi^2_q}{q}}{\dfrac{\chi^2_{n-k-1}}{n-k-1}} \sim F(q, n-k-1).$$

Testing a joint hypothesis with the SSR (continued)

Since sums of residuals (and differences in sums) follow Chi Squared distributions, the ratio of change in SSR to unrestricted SSR can be the basis for an *F* Test.

$$\frac{\dfrac{SSR_r - SSR_{ur}}{q}}{\dfrac{SSR_{ur}}{n-k-1}} \sim F(q, n-k-1), \text{where}$$

q is called the numerator degrees of freedom and *n-k-1* is the denominator degrees of freedom.

One can calculate *F* from estimating the restricted and unrestricted models and obtaining the SSR figures.

Testing a joint hypothesis with the SSR (continued)

With a test statistic in hand, now
- What is the distribution of F? and
- how large the test statistic must be to reject H_0?

This can be found using a table like the one in the Appendix or by using software.
- If H_0 is true, the numerator will be approximately zero because the restrictions will not increase SSR very much.
- But if $(SSR_r - SSR_{ur})$ is large, it makes the F statistic large and contradicts the null hypothesis.

STATA can calculate the upper tail (p-value) probability using the **Ftail** command.
- The syntax is: **Ftail(n1,n2,F)**, where "n1" and "n2" are the numerator and denominator degrees of freedom, respectively, and "F" is the test statistic.

Testing a joint hypothesis with the SSR (concluded)

To perform the test, merely decide on a confidence level $(1 - \alpha)$ and calculate the value of F at which the upper tail contains α probability.

If the test statistic is greater than this <u>critical value</u>, reject the null and do not reject it otherwise.

As with t tests, the p-value indicates the smallest Type I Error probability with which you can reject the null hypothesis.

The R^2 version of the test

Because of the relationship between R^2 and SSR, the *F* test can be performed using R^2 as well.

$$R^2 = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST} \Leftrightarrow SSR = SST(1 - R^2).$$

F can be rewritten:

$$\frac{\dfrac{SST(1 - R_r^2) - SST(1 - R_{ur}^2)}{q}}{\dfrac{SST(1 - R_{ur}^2)}{n - k - 1}} \sim F(q, n - k - 1)$$

The R^2 version of the test (continued)

$$\text{simplifying yields, } F = \frac{\dfrac{(1-R_r^2)-(1-R_{ur}^2)}{q}}{\dfrac{(1-R_{ur}^2)}{n-k-1}} = \frac{\dfrac{(R_{ur}^2-R_r^2)}{q}}{\dfrac{(1-R_{ur}^2)}{n-k-1}}.$$

A special case of the *F* test ascertains whether the set of explanatory variables, collectively, explains y significantly.

◦ This is done by choosing $q = k$. Then *F* becomes:

$$F = \frac{\dfrac{(R_{ur}^2)}{q}}{\dfrac{(1-R_{ur}^2)}{n-k-1}}.$$

This tests the <u>overall significance of the regression</u>.

Even more complicated hypothesis tests

Finally consider testing <u>exclusion restrictions and an individual parameter value</u> simultaneously with an *F* test.

◦ The textbook uses an example in which all the coefficients except one are null hypothesized to be zero, and the remaining one is hypothesized to equal 1.

◦ This can be performed by transforming the dependent variable, subtracting the variable with the coefficient of 1. The restricted model is:

$$y = \beta_0 + 1x_1 + u, \text{and the restricted } SSR \text{ comes from,}$$
$$\pi = \beta_0 + u; \ \pi \equiv y - x_1.$$

Regressing only on a constant will just estimate the mean of the transformed LHS variable.

Reporting regression results

Always report the OLS coefficients.

◦ For the key variables, the author should offer interpretation, i.e., explain the estimated effects in the context of units of measurement, e.g., elasticity, semi-elasticity.

◦ And the practical significance of the results should be discussed.

Depending on the audience, report the standard errors (other economists who know how to calculate t statistics readily) or t statistics (other academics who don't use econometrics as fluently).

◦ Standard errors are more reliable generally, especially in instances in which the null hypothesized value is non-zero.

Reporting regression results (continued)

Report R^2 as a measure of goodness of fit.

◦ It also helps calculate *F* statistics for testing exclusion restrictions.

Although the results in the text are frequently reported in equation form for pedagogical purposes, most regression results in published papers are not.

Instead they are reported on one or several <u>tables</u>, in which the samples, sample sizes, R^2, and sets of control variables are noted in each column (for each specification).

Regression results tables

Tables should be "self-contained", i.e., with a title describing the model estimated, the dependent variable, intuitive (!) names for the explanatory variables in the first column, and a caption explaining any symbols, jargon, estimation methods, and anything else that could confuse the reader.

◦ Many veteran readers of empirical work will skim a paper merely by studying the tables, and it is desirable that anyone who does so be able to get the rudiments of your findings.

An example from a paper I have been working on (studying the effect of paid sick leave mandates on labor market outcomes) is shown on the next slide.

Regression results tables (continued)

Table 1: Mandates' Effects on Individual Industry Employment

	Effect on Log Industry Employment			
	1	**2**	**3**	**4**
Mandate Implemented (β1)	**0.2912**	**0.2926**	**-0.0069**	**-0.0070**
	(.1904)	(.1903)	(.0106)	(.0106)
Implied % Change Employment	-	-	-	-
Mandate*Unaffected (β2)	-0.2136	-0.2166	0.0306**	0.0305**
	(.1779)	(.1776)	(.0142)	(.0142)
Mandate, Unaffected Industry (β1+β2)	**0.0776**	**0.0760**	**0.0237***	**0.0235***
	(.0564)	(.056)	(.006)	(.006)
Implied % Change Employment	-	-	2.40%	2.37%
Observations	2,778,552	2,778,552	2,745,474	2,745,474
Panels	33,078	33,078	33,078	33,078
Year, Calendar Month Effects	Yes	Yes	Yes	Yes
Controls	Population	All	Population	All
Time Trends	Industry, Treatment & Control	Industry, Treatment & Control	Industry, County	Industry, County
Dynamic Treatment Effect	No	No	No	No

Estimates of model (2) using industry data. The panels in this model are (2 digit) industry, county combinations. Again standard errors are cluster robust around each panel (county-industry). Table 6a shows the employment effect estimates and 6b shows the wage effect estimates. Summary statistics for the industry characteristics are on a table in the appendix. The emphasis in these tables is on the source of employment and wage gains: the "unaffected" industries. This supports the hypothesis that the law's positive health externalities drive the county results and offset the costs to the firms who are bound by it.
* Significant at α=0.1
** Significant at α=0.05
*** Significant at α=0.01

Conclusion

As necessary, you should be able to test various forms of hypotheses about regression estimates:

◦ Value of a single parameter,

◦ Value of a sum of (difference between) multiple parameters (linear combination),

◦ Joint equality of multiple parameters,

◦ Significance of overall regression model.

Performing these tests with technical accuracy and being able to present the results of a regression is vital to an empiricist's credibility.

This lecture on inference is based on the classical linear model assumptions.

◦ Next we will consider what happens when the normality assumption is lifted.