Session 6: Model evaluation

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

Last time...

- Inference

Today:

- Model evaluation
- R^2, RMSE, F statistic

Readings:

- Stock and Watson (2011), ch 5, 7
- Verbeek (2018), ch 2.4-2.6, 4

Model fit

- Can we say anything about how well our estimates fit the data? Three approaches
- R^2 ("R-squared")
- Standard error of the regression (SER) aka mean squared error (MSE)
- F statistic
- Note: none of these tells us whether our estimates are unbiased/consistent/efficient. These only tell us how our model (i.e., regression equation as we specified it) performs.

R^2

- Idea behind R^2: what share of the variance of Y is explained by our regression model?
- The higher the share, the better our model fits observed outcome Y

$$
\begin{aligned}
\text{Residual Sum of Squares (RSS)} &= \sum u_i^2 \\
\text{Explained Sum of Squares (ESS)} &= \sum (\hat{y}_i - \bar{y})^2 \\
\text{Total Sum of Squares (TSS)} &= \sum (y_i - \bar{y})^2 \\
\text{TSS} &= \text{ESS} + \text{RSS} \\
R^2 &= \frac{\text{ESS}}{\text{TSS}} \in [0, 1] \\
&= 1 - \frac{RSS}{TSS}
\end{aligned}
$$

- R^2 is easy to interpret (fraction)
- It is weakly increasing in independent variables (numerator doesn't change but residuals decrease)
- Adjusted R^2: penalizing for number of independent variables

$$R^2_{adj} = 1 - \frac{\frac{RSS}{n-k-1}}{\frac{TSS}{n-1}}$$

- Less interpretable; adjusted R^2 can be negative
- Sample-dependent: holding everything constant, increasing variation in X increases R^2 for the same model

SER or MSE

- Standard error of the regression (aka mean squared error or MSE)

$$\widehat{\sigma^2} = \frac{\sum u_i^2}{n-k-1}$$

- Measured in units of the dependent variable
  - Imprecise def: on average how off you are
  - Eg regressing income on age, with $\widehat{\sigma^2} = 3$: you are off by $3
  - Smaller = better
- Estimate of a population parameter and thus not directly affected by sample

Illustration

$$\text{Growth}_c = \alpha + \beta \text{Mean years of schooling}_c + \varepsilon_c$$

```
> summary(lm_robust(data=growth, growth ~ yearsschool))

Call:
lm_robust(formula = growth ~ yearsschool, data = growth)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
(Intercept)   0.9583    0.44514   2.153 0.035173  0.06875    1.848 63
yearsschool   0.2470    0.08204   3.011 0.003744  0.08308    0.411 63

Multiple R-squared:  0.1096 ,   Adjusted R-squared:  0.09543
F-statistic: 9.066 on 1 and 63 DF,  p-value: 0.003744
```

```
> estimatr::extract.lm_robust(model)


                 coef.       s.e.           p
(Intercept) 0.9582918 0.44514240 0.035172783
yearsschool 0.2470275 0.08204111 0.003743522

                    GOF dec. places
R$^2$       0.10956008         TRUE
Adj. R$^2$  0.09542611         TRUE
Num. obs.  65.00000000        FALSE
RMSE        1.80433333         TRUE
```

F test

- F test: test the idea that all covariates have **jointly** no effect on Y
- Concretely:
  - $H_0 : \beta_1 = \beta_2 = ... = \beta_k = c$ (generally $c = 0$)
  - H_a : at least one parameter $\neq c$

- To do so: capitalize on the following variable that has an F distribution:

$$F_{k-1,n-k} = \frac{SSE/(k-1)}{SSR/(n-k)}$$

- Decreases with poor model (large SSR), increases with better explanation (large SSE)
- Large F: model explains Y better than without any of the variables
- Note: F distribution has two inputs (k-1, n-k) that change its shape a lot

In theory

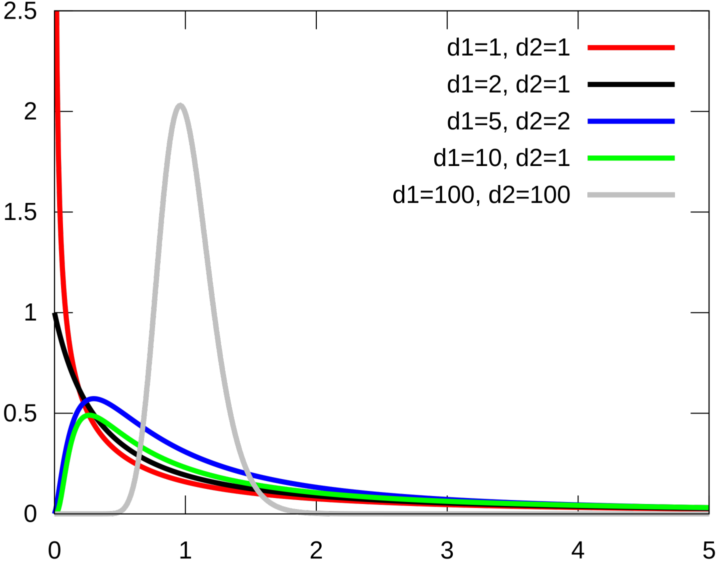

Figure 1: Source: wikipedia.

In practice

```
> model2 = lm_robust(data=growth, growth ~ yearsschool + tradeshare)
> summary(model2)

Call:
lm_robust(formula = growth ~ yearsschool + tradeshare, data = growth)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper DF
(Intercept)  -0.3702    0.59735 -0.6197 0.5377536 -1.56423   0.8239 62
yearsschool   0.2500    0.07621  3.2808 0.0017015  0.09769   0.4024 62
tradeshare    2.3313    0.63215  3.6879 0.0004781  1.06765   3.5949 62

Multiple R-squared:  0.2359 ,   Adjusted R-squared:  0.2113
F-statistic:  10.7 on 2 and 62 DF,  p-value: 0.000102
```

F statistic: 10.7, with $n-k=62$ and $k-1=2$ degrees of freedom.

- F statistic will be useful in instrumental variables
- Can be used to test **nested models**
- $Y = a + bX + cW + fZ$ vs. $Y = a + bX$

Conclusion

- Several methods to evaluate models
- R^2 is probably the most common, easy to interpret, but considerable limitations
- SER is more robust and informative
- F statistic is less intuitive but can be applied to a range of situations

Questions?

References

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd Edition*. Pearson.

Verbeek, Marno. 2018. *A Guide to Modern Econometrics 5th Edition*. Wiley.