Boston University Questrom School of Business

MF793 – Fall 2021

Eric Jacquier

Testing and Forecasting with the OLS regression

A Needed results on quadratic forms

B Testing and Confidence Intervals for a single coefficient

C Significance of the regression as a whole

R^2, ANOVA table and the "F test".

D. Forecasting

Readings Hansen Chapter 2, 3, 4

Greene Ch. 3, 4 (only relevant sections, ignore asymptotics)

A Quadratic forms and the χ^2 distribution

All you wanted to know and never dared to ask about χ^2 distributions

A1 LN 8: $\chi2(N)$ can be written as a sum of N squared independent standard Normals.

N x 1 vector z ~ N(0,1) => $z'z \sim \chi^2(N)$ **[A1]**

A2 $x \sim N(0, V)$ then $x'x$ is **not** $\chi^2(N)$. Can we make a χ^2 from this x vector? Yes we can!

Result: for any covariance matrix V, we can find P_N so that $V = P P'$. Then $P^{-1} V P'^{-1} = I$

Consider $z = P^{-1} x$ $E(z z') = E(P^{-1} x x' P'^{-1}) = P^{-1} E(xx') P'^{-1} = P^{-1} V P'^{-1} = I$

So we have: $z'z \sim \chi^2(N)$ by [A1]

What is z'z? $z' z = x' P'^{-1} P^{-1} x = x' V^{-1} x$

For the Nx1 vector $x \sim N(0,V)$, the quadratic form $x' V^{-1} x \sim \chi2(N)$ **[A2]**

$\frac{e'e}{\sigma^2} = \frac{\varepsilon'}{\sigma} M \frac{\varepsilon}{\sigma}$

A3 What if x is **not** full rank, i.e, there are linear combinations within x? There is no V^{-1}!

E.g.; $\Sigma (y_i - \bar{y})^2$ is only T-1 independent normals, the SSR $e'e = \Sigma (y_i - \hat{\alpha} - \hat{\beta}' x_i)^2$, is not full rank.

- Take $z \sim N(0, I_N)$

  Take Q an NxN **idempotent** symmetric matrix (QQ' = Q) of rank ν=N-q,

  Then: tr(Q) = N-q , det(Q) = 0

  What can we say of z'Qz ?

- Take x = Q z a **projection** of z on a subspace of dimension N-q (like Mε in the regression)

  E(xx') = Q E(z z') Q' = Q The covariance matrix of x is Q, it is of rank N-q and non-invertible. *q* of the x's can be written as linear combinations of the other *N-q*

- Now look at x' x = z'Q' Q z = z' Q z (like ε'Mε in the regression)

  Let us use this result (no proof):

  Any symmetric idempotent matrix Q_N of rank N-q can be written as Q = PP', where P is a N x (N-q) matrix and $P'P = I_{N-q}$

  Bingo! x'x = z' Q z = z' P P' z = u' u Dimension of (u): ________ E(uu') = ________

For an i.i.d. vector $z \sim N(0, I_N)$, an idempotent matrix Q_N, the quadratic form z' Q z with rank(Q)=ν<N is a χ2(ν) **[A3]**

$\frac{e'e}{\sigma^2} = \frac{e'}{\sigma} M \frac{\varepsilon}{\sigma}$

M: rank T-K

$\frac{e'e}{\sigma^2} = \frac{(T-k)s^2}{\sigma^2}$

B Testing , confidence intervals for individual coefficient estimates

- Asymptotically or exactly, we have $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$
  - If $\hat{\beta}$ is multivariate normal, any element of $\hat{\beta}$ is univariate normal: no proof

$$\hat{\beta}_i \sim N(\beta_i\,,\ \sigma^2\,\{(X'X)^{-1}\}_{ii}) \qquad [1]$$

$\tilde{R} \sim MVN(\mu, \Sigma)$ $\qquad$ $R_p = w'\tilde{R} \sim N(w'\mu, w'\Sigma w)$

  - Any linear combination of $\hat{\beta}_i s$ is also normal. (portfolio return is a lin. comb. of stock returns)

  Consider δ a linear combination of β: $\qquad \delta \ = \ q' \ \beta \qquad$ Then:

$$\hat{\delta} = q'\,\hat{\beta} \ \sim N(q'\beta,\ \sigma^2\, q'(X'X)^{-1}q) \qquad \text{<- as in a portfolio variance} \qquad [2]$$

  We do need the off-diagonal elements of $(X'X)^{-1}$: $Cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2\,\{(X'X)^{-1}\}_{ij}$

- To make [1] **feasible**, we estimate σ by s.

  Recall: $\qquad \nu s^2 / \sigma^2 \sim \chi^2(\nu)$, where ν = T-k

Then $\qquad (\hat{\beta}_i - \beta_i) / \sigma_{\hat{\beta}_i} \sim N(0,1)$ becomes $(\hat{\beta}_i - \beta_i) \ / \ s_{\hat{\beta}_i} \sim t(\nu = T\text{-}k) \qquad$ [3]

Proof: same as for Student-t proof in LN8

B Significance of the Regression as a whole: R^2, ANOVA, F test

B1 R^2

- What if some elements $\widehat{\beta}_i{}'s$ are significant and others are not?

- R^2 gives us a unified answer for the effectiveness of the regression as a whole:
  What fraction of the LHS variation is jointly explained by the RHS variables.

- Is there a statistical significance test connected to the R^2 ? Doh Yes!

  The sample R^2 is a random estimate (for this sample) of the true unknown R^2

  We test the null H_0: the regression explains nothing, zip, zilch. i.e.: True $R^2 = 0$.

  We can show that the distribution of the sample R^2 under H_0 with normal errors is:

  $$R^2 \sim \text{Beta}((K-1)/2\,,\,(T-k)/2) \qquad \text{No proof !}$$

  Few people know or use this. But it is pretty trivial, get the Beta 95% cutoff and conclude!

- The **F-test** is more common, it comes out of the **analysis of variance** (**ANOVA**).

- There is a one-to-one correspondence between significance of the F-statistic and of the sample R^2.
  Given the F quantity in table B2, it is easy to prove that:

$$\frac{R^2/(k-1)}{(1-R^2)/(T-k)} \sim F(k-1, T-k) \qquad \text{Prove it}$$

B2 Analysis of Variance (ANOVA) and the F-statistic

Model: $Y = X\beta + \varepsilon$, where X is T x k Estimates: $\hat{Y} = X\hat{\beta} + e$

Pythagoras, LN10, p. 13: $(Y - \mathbf{1}\bar{Y})'(Y - \mathbf{1}\bar{Y}) = (\hat{Y} - \mathbf{1}\bar{Y})'(\hat{Y} - \mathbf{1}\bar{Y}) + e'e$

ANOVA table

Source of variation	Sum of Squares	df	Mean Square	F-statistic
Total SST	$\sum (y_i - \bar{y})^2$	T-1		
Regression SSR	$\sum (\hat{y}_i - \bar{y})^2$	k-1	MSR = SSR/k-1	
Error SSE	$\sum (y_i - \hat{y}_i)^2$	T-k	MSE = SSE/T-k	**MSR/MSE**

k-1: number of slopes

- SST = (T-1) $s^2 \sim \sigma^2 \chi^2(T-1)$

- **SSE = σ^2 χ^2(T-k)**

We have: $e'e / \sigma^2 = (\varepsilon'/\sigma) M (\varepsilon/\sigma) \sim \chi^2(\nu = \text{rank}(M)) =$ $\boldsymbol{\chi^2(T\text{-}k)}$ by [A3]

- **SSR $\sim \sigma^2 \chi^2$(k-1) under H_0** What happens to SSR if H_0 is exactly true for the data?

$$(\hat{Y} - \mathbf{1}\bar{Y})'(\hat{Y} - \mathbf{1}\bar{Y}) = (X\hat{\beta} - \bar{X}\hat{\beta})'(X\hat{\beta} - \bar{X}\hat{\beta}) = \hat{\beta}'(X - \bar{X})'(X - \bar{X})\hat{\beta}$$
$$= \hat{\beta}_d' X_d' X_d \hat{\beta}_d$$

Prove it, see addendum

$_d$ indicates regression in deviations, β_d is (k-1) x 1 (β without the intercept), X_d is T x (k-1)

Then $\hat{\beta}_d \sim N(\beta_d\,,\ \sigma^2(X_d{}'X_d)^{-1})$ if $\varepsilon \sim N(0, \sigma^2 I)$.

=> $(\hat{\beta}_d - \beta_d)'X_d{}'X_d(\hat{\beta}_d - \beta_d) \sim \boldsymbol{\sigma^2\,\chi^2(K\text{-}1)}$ by [A2]

Then, under H_0: $\beta_d = 0$, $\hat{\beta}_d{}'X_d{}'X_d\hat{\beta}_d \sim \boldsymbol{\sigma^2\,\chi^2(K\text{-}1)}$

- Then, with normal errors, the ratio **MSR / MSE ~ F(k-1 , T-k)**

  σ^2 drops out in the ratio
  Why are the numerator and denominator independent?

  This is the basis for a test of significance of the regression as a whole

- Warning: Significance tests are often not very useful by themselves. Could have a significant F statistic with a small R^2.... Or vice-versa!

- Generalization to a subset of the coefficients

  Let $X = (X_1 \mid X_2)$, with $k = k_1 + k_2$. Consider the **unrestricted** regressions of Y on X, and the **restricted** regression **of** Y on X_1. There are k_2 restrictions. One can show that:

$$((SSR_u - SSR_r)/k_2) \,/\, (SSE_u/(T\text{-}k)) \sim F(k\text{-}1\ ,\ T\text{-}k) \qquad \text{No proof}$$

k_2

Exercise: easy to rewrite this in term of the restricted and unrestricted SSEs or R^2s.

C Forecasting

C1 Assumptions:

- Regression: $y = x'\beta + \varepsilon$,

  $E(\varepsilon \mid x) = Cov(\varepsilon, x) = 0$

  $E(\varepsilon\varepsilon') = \sigma^2 I$

- Forecast is always conditional on x: we <u>know</u> x_f', the vector of **out-of-sample** Xs

- out-of-sample X?
  - Obvious for a *time-series* regression, where observations are in calendar time:

    y_f is a future data point
  - The regression can be *cross-sectional*, then y_f refers to any additional data not used in the estimation

- What **point forecast** to choose?

  Conditional mean $E(y_f \mid x_f)$ minimizes MSE of prediction

$$\widehat{y_f} = x_f'\hat{\beta}$$

- True future value: $y_f = x_f'\beta + \varepsilon_f$

- Forecast error: $e_f = \hat{y}_f - y_f = \underbrace{x_f'(\hat{\beta} - \beta)}_{\text{Fitting part}} + \underbrace{\varepsilon_f}_{\text{prediction part}}$

- Common sense assumptions:

  Future error ε_f independent from past x, from future x_f, and from past errors ε.

  => **ε_f is independent from $\hat{\beta}$,** **Why?**

Results:

- Forecast is unbiased:

$$\boldsymbol{E(e_f|X, x_f)} = E(x_f'(\hat{\beta} - \beta)) + E(\varepsilon_f)$$
$$=$$
$$\boldsymbol{= 0}$$

- Forecast error variance: $\boldsymbol{V(e_f|X, x_f)} = V\left(x_f'(\hat{\beta} - \beta)\right) + \sigma^2 + 0$ prove it

$$\boldsymbol{V(e_f|X, x_f)} = E(x_f'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'x_f) + \sigma^2$$

$$= x_f' E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) x_f + \sigma^2$$

$$= x_f' \sigma^2 (X'X)^{-1} x_f + \sigma^2$$

$$= \boldsymbol{\sigma^2 \left[x_f' (X'X)^{-1} x_f + 1 \right]} \qquad \textbf{[4]}$$

  - The variance of forecast error is a function of x_f.

    Maybe the forecast is more precise for some x_f than others?

  - The σ^2 term accounts for the *prediction part* of the error

- Summary: If the assumptions of the linear model are met out-of-sample as well as in-sample

    - Forecast error is unbiased and has variance [4] above
    - Forecast error is normally distributed if the model has normal errors
    - Confidence interval for the forecast is easily computed using [4] and normality
    - Forecast error becomes Student-t (T-k) as one must estimate σ by s   Prove it

- But the formula in [4] misses **a fundamental intuition in forecasting** !

The variance of forecast is smaller the closer x_f is to $\bar{x}$

Intercept and one variable:

$$y_f = \alpha + \beta x_f + \varepsilon_f$$
$$\hat{y}_f = \hat{\alpha} + \hat{\beta} x_f$$

$$V(e_f) = V(\hat{\alpha}) + V(\hat{\beta}) x_f^2 + 2Cov(\hat{\alpha}, \hat{\beta}) x_f + \sigma^2$$

You will write in your homework $V(\hat{\alpha})$, $V(\hat{\beta})$, $Cov(\hat{\alpha}, \hat{\beta})$

... so you can easily show that:

$$V(e_f) = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_f - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right] \qquad [5]$$

prove it

$V(e_f) = \sigma^2 \left[\mathbf{1} + \frac{1}{T} + \frac{(x_f - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right]$ is the basis for a confidence interval of forecast

With σ^2 :	prediction	interval	variance of one prediction
Without σ^2:	fit	interval	variance of a large number of repeated prediction