# Interpreting Regression Results

Carlo Favero

# Interpreting Regression Results

Interpreting regression results is not a simple exercise. We propose to split these procedure in three steps.

- First, introduce a measure of sampling variability and evaluate again what you know taking into account that parameters are estimated and there is uncertainty surrounding your point estimates.
- Second, understand the relevance of our regression independently from inference on the parameters. There is an easy way to do this: suppose all parameters in the model are known and identical to the estimated values and learn how to read these.
- Third, remember that each regression is run after a reduction process has been, explicitly or implicitly implemented. The relevant question is what happens if something went wrong in the reduction process? What are the consequences of omitting relevant information or of including irrelevant one in your specification?

# Statistical Significance and Relevance

- Relevance of a regression is different form statistical significance of the estimated parameters.
- In fact, confusing statistical significance of the estimated parameter describing the effect of a regressor on the dependent variable with practical relevance of that effect is a rather common mistake in the use of the linear model.
- Statistical inference is a tool for estimating parameters in a probability model and assessing the amount of sampling variability. Statistics gives us indication on what we can say about the values of the parameters in the model on the basis of our sample.
- The relevance of a regression is determined by the share of the unconditional variance of $\mathbf{y}$ that is explained by the variance of $E(\mathbf{y} \mid \mathbf{X})$. Measuring how large is the share of the unconditional variance of $\mathbf{y}$ explained by the regression function is the fundamental role of $R^2$.

# Statistical Significance of CAPM coefficients

- Estimate the coefficients in a CAPM regression
- Suppose the null hypothesis is true and compute $p$ as be the probability (under that hypothesis) of getting results as extreme as those observed
- $p$ is called the p-value. If it is very small the results are statistically significance in the sense that the null is rejected as the probability of observing what you have observed under the null is small.
- Note that the p-value can be computed in two ways i) by deriving the relevant distribution aunder the null ii) by simulating via Monte-Carlo or bootstrap the relevant distribution under the null.

# Relevance of CAPM coefficients

- Estimate the coefficients in a CAPM regression and keep them fixed at their point estimate
- Run an experiment by changing the conditional mean via a shock to the regressor
- Assess how relevant is the shock in excess market returns to determine excess returns on asset i

# Inference in the Linear Regression Model

- Users of econometric models in finance attributes high priority to the concept of "statistical significance" of their estimates. In the standard statistical jargon an estimate of a parameter is “statistical significant” if its estimated value, compared with its sampling standard deviation makes it unlikely that in other samples the estimate may change of sign.
- In the linear regression model the statistical index mostly used is the t-ratio and an estimated parameter has a significance which is usually measured in terms of its P-value, the probability with which that coefficient is equal to zero.
- In the previous section we have discussed the common confusion between “statistical significance” and “relevance”
- In this section we illustrate the basic principles that allow us to evaluate statistical significance and to perform test of relevant hypothesis on the estimated coefficient in a linear model.

# Elements of Distribution theory

We consider the distribution of a generic $n$-dimensional vector $\mathbf{z}$, together with the derived distribution of the vector $\mathbf{x} = g(\mathbf{z})$ which admits the inverse $\mathbf{z} = h(\mathbf{x})$, with $h = g^{-1}$. If $prob(z_1 < z < z_2) = \int_{z_1}^{z_2} f(z)\,dz$, and $prob(x_1 < x < x_2) = \int_{x_1}^{x_2} f^*(x)\,dx$, then:

$$f^*(x) = f(h(x))\,J,$$

where $J = \begin{vmatrix} \frac{\partial h_1}{\partial x_1} & \ldots & \frac{\partial h_n}{\partial x_1} \\ \ldots & \ldots & \ldots \\ \frac{\partial h_1}{\partial x_n} & \ldots & \frac{\partial h_n}{\partial x_n} \end{vmatrix} = \left| \frac{\partial \mathbf{h}}{\partial \mathbf{x}'} \right|$.

# The normal distribution

The standardized normal univariate has the following distribution:

$$\begin{aligned} f(z) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), \\ E(z) &= 0, \qquad var(z) = 1. \end{aligned}$$

By considering the transformation $x = \sigma z + \mu$, we derive the distribution of the univariate normal as:

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \\ E(x) &= \mu, \qquad var(x) = \sigma^2. \end{aligned}$$

# The normal multivariate distribution

Consider now the vector $\mathbf{z} = (z_1, z_{2,} ..., z_n)$ , such that

$$f(\mathbf{z}) = \prod_{i=1}^{n} f(z_i) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right).$$

$\mathbf{z}$ is, by construction, a vector of normal independent variables with zero mean and identity variance covariance matrix. The conventional notation is $\mathbf{z} \sim \mathbf{N}(0, I_n)$.

# The normal multivariate distribution

Consider now the linear transformation,

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu},$$

where $\mathbf{A}$ is an $(n \times n)$ invertible matrix. We consider the following transformation $\mathbf{z} = \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ with Jacobian $J = \left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}$. By applying the formula for the transformation of variables, we have:

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \left|\mathbf{A}^{-1}\right| \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \mathbf{A}^{-1\prime}\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),$$

which, by defining the positive definite matrix $\sum = \mathbf{A}\mathbf{A}'$, equals

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \left|\sum^{-\frac{1}{2}}\right| \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \sum^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

The conventional notation for the multivariate normal is $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \sum)$.

# The transformation of normal multivariate

**Theorem**

*For any* $\mathbf{x} \sim \mathbf{N}\left(\mu, \sum\right)$*, given any* $(m \times n)$ $\mathbf{B}$ *matrix and any* $(m \times 1)$ *vector,* $\mathbf{d}$*, if* $\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{d}$*, this implies* $\mathbf{y} \sim \mathbf{N}\left(\mathbf{B}\mu + \mathbf{d}, \mathbf{B}\sum \mathbf{B}'\right)$*.*

Consider a partitioning of an $n$-variate normal vector in two sub-vectors of dimensions $n_1$ and $n - n_1$:

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right).$$

1. $\mathbf{x}_1 \sim \mathbf{N}\left(\mu_1, \sum_{11}\right)$, which follows from applying the general formula in the case $\mathbf{d} = \mathbf{0}$, $\mathbf{B} = \left(I_{n_1} \;\; \mathbf{0}\right)$;
2. $(\mathbf{x}_1 \mid \mathbf{x}_2) \sim \mathbf{N}\left(\mu_1 + \sum_{12} \sum_{22}^{-1} (\mathbf{x}_2 - \mu_2), \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}\right)$, which is obtained by applying the general formula to the case $\mathbf{d} = \sum_{12} \sum_{22}^{-1} \mathbf{x}_2, \mathbf{B} = \left(I_{n_1} \;\; -\sum_{12} \sum_{22}^{-1}\right)$.

# Distributions derived from the normal

Consider $\mathbf{z} \sim \mathbf{N}(0, I_n)$, an $n$-variate standard normal. The distribution of $\omega = \mathbf{z}'\mathbf{z}$ is defined as a $\chi^2(n)$ distribution with $n$ degrees of freedom. Consider two vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ of dimensions $n_1$ and $n_2$ respectively, with the following distribution:

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \sim \mathbf{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} I_{n_1} & 0 \\ 0 & I_{n_2} \end{pmatrix} \right).$$

We have $\omega_1 = \mathbf{z}_1'\mathbf{z}_1 \sim \chi^2(n_1)$, $\omega_2 = \mathbf{z}_2'\mathbf{z}_2 \sim \chi^2(n_2)$, and $\omega_1 + \omega_2 = \mathbf{z}_1'\mathbf{z}_1 + \mathbf{z}_2'\mathbf{z}_2 \sim \chi^2(n_1 + n_2)$. In general, the sum of two independent $\chi^2(n)$ distributions is in itself distributed as $\chi^2$ with a number of degrees of freedom equal to the sum of the degrees of freedom of the two $\chi^2$.

# Distributions derived from the normal

Our discussion of the multivariate normal concludes that if $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \sum)$, then $(\mathbf{x} - \boldsymbol{\mu})' \sum^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2(n)$.
A related result establishes that if $\mathbf{z} \sim \mathbf{N}(0, I_n)$ and $\mathbf{M}$ is a symmetric idempotent $(n \times n)$ matrix of rank $r$, then $\mathbf{z}'\mathbf{M}\mathbf{z} \sim \chi^2(r)$.
Another distribution related to the normal is the $F$-distribution. The $F$-distribution is obtained as the ratio of two independent $\chi^2$ divided by the respective degrees of freedom. Given $\boldsymbol{\omega}_1 \sim \chi^2(n_1)$, and $\boldsymbol{\omega}_2 \sim \chi^2(n_2)$, we have:

$$\frac{\boldsymbol{\omega}_1 / n_1}{\boldsymbol{\omega}_2 / n_2} \sim F(n_1, n_2).$$

# Distributions derived from the normal

The Student's $t-$distribution is then defined as:

$$t_n = \sqrt{F(1,n)}.$$

Another useful result establishes that two quadratic forms in the standard multivariate normal, $\mathbf{z}'\mathbf{M}\mathbf{z}$ and $\mathbf{z}'\mathbf{Q}\mathbf{z}$, are independent if $\mathbf{MQ} = \mathbf{0}$. We can finally state the following theorem, which is fundamental to the statistical inference in the linear model:

**Theorem**

*If* $\mathbf{z} \sim \mathbf{N}(0, I_n)$, $\mathbf{M}$ *and* $\mathbf{Q}$ *are symmetric and idempotent matrices of ranks* $r$ *and* $s$ *respectively and* $\mathbf{MQ} = \mathbf{0}$, *then* $\frac{\mathbf{z}'\mathbf{Q}\mathbf{z}}{\mathbf{z}'\mathbf{M}\mathbf{z}}\frac{r}{s} \sim \mathbf{F}(s, r)$.

# The conditional distribution

$y \mid X$

To perform inference in the linear regression model, we need a further hypothesis to specify the distribution of **y** conditional upon **X:**

$$y \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I\right), \tag{1}$$

or, equivalently

$$\epsilon \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{0}, \sigma^2 I\right). \tag{2}$$

Given $(1)$ we can immediately derive the distribution of $\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)$ which, being a linear combination of a normal distribution, is also normal:

$$\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right). \tag{3}$$

# The conditional distribution

$y \mid X$
Equation (3) constitutes the basis to construct confidence intervals and to perform hypothesis testing in the linear regression model. Consider the following expression:

$$\begin{aligned} \frac{\left(\widehat{\beta}-\beta\right)' \mathbf{X}'\mathbf{X}\left(\widehat{\beta}-\beta\right)}{\sigma^2} &= \frac{\boldsymbol{\epsilon}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\epsilon}}{\sigma^2} \\ &= \frac{\boldsymbol{\epsilon}'\mathbf{Q}\boldsymbol{\epsilon}}{\sigma^2}, \\ \mathbf{Q} &= \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \end{aligned}$$

and, applying the results derived in the previous section, we know that

$$\frac{\boldsymbol{\epsilon}'\mathbf{Q}\boldsymbol{\epsilon}}{\sigma^2} \mid \mathbf{X} \sim \chi^2(k). \tag{4}$$

# The conditional distribution

$y \mid X$
Equation $(4)$ is not useful in practice, as we do not know $\sigma^2$. However, we know that

$$\frac{S\left(\widehat{\beta}\right) \mid \mathbf{X}}{\sigma^2} = \frac{\epsilon' \mathbf{M} \epsilon}{\sigma^2} \mid \mathbf{X} \sim \chi^2 \left(T-k\right). \quad (5)$$

$$\mathbf{M} = \mathbf{I} - \mathbf{Q} \quad (6)$$

Since $\mathbf{MQ} = \mathbf{0}$, we know the distribution of the ratio of $(4)$ and $(5)$; moreover, taking the ratio, we get rid of the unknown term $\sigma^2$:

$$\frac{\left(\widehat{\beta}-\beta\right)' \mathbf{X}'\mathbf{X} \left(\widehat{\beta}-\beta\right) / \sigma^2}{s^2 / \sigma^2} = \frac{\epsilon' \mathbf{Q} \epsilon}{\epsilon' \mathbf{M} \epsilon} \left(T-k\right) \sim kF\left(k, T-k\right). \quad (7)$$

# Confidence Intervals for beta

We use result $(7)$ to obtain from the tables of the $F$-distribution the critical value $F_{\alpha}^{*}(k, T-k)$ such that

$$prob\left[F(k, T-k) > F_{\alpha}^{*}(k, T-k)\right] = \alpha, \qquad 0 < \alpha < 1,$$

for different values of $\alpha$ we are in the position of evaluating exactly an inequality of the following form:

$$prob\left\{\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^{\prime}\mathbf{X}^{\prime}\mathbf{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \leq ks^{2}F_{\alpha}^{*}(k, T-k)\right\} = 1-\alpha,$$

which defines confidence intervals for $\widehat{\boldsymbol{\beta}}$ centred upon $\boldsymbol{\beta}$.

# Hypothesis Testing

- Hypothesis testing is strictly linked to the derivation of confidence intervals.
- When testing the hypothesis, we aim at rejecting the validity of restrictions imposed on the model on the basis of the sample evidence. Within this framework, (**??**) $-$ (3) are the maintained hypothesis and the restricted version of the model is identified with the null hypothesis $H_0$. Following the Neyman$-$Pearson approach to hypothesis testing, one derives a statistic with known distribution under the null. Then the probability of the first-type error (rejecting $H_0$ when it is true) is fixed at $\alpha$.
- For example, we use a test at the level $\alpha$ of the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, based on the $F$-statistic, when we do not reject the null $H_0$ if $\boldsymbol{\beta}_0$ lies within the confidence interval associated with the probability $1 - \alpha$.
- However, in practice, this is not a useful way of proceeding, as the economic hypotheses of interest rarely involve a number of restrictions equal to the number of estimated parameters.

# Hypothesis Testing

The general case of interest is therefore the one when we have $r$ restrictions on the vector of parameters with $r < k$. If we limit our interest to the class of linear restrictions, we can express them as

$$H_0 = \mathbf{R}\boldsymbol{\beta} = \mathbf{r},$$

where $\mathbf{R}$ is an $(r \times k)$ matrix of parameters with rank $k$ and $\mathbf{r}$ is an $(r \times 1)$ vector of parameters. To illustrate how $\mathbf{R}$ and $\mathbf{r}$ are constructed, we consider the baseline case of the CAPM model; we want to impose the restriction $\beta_{0,i} = 0$ on the following specification:

$$\begin{aligned} \left(r_t^i - r_t^{rf}\right) &= \beta_{0,i} + \beta_{1,i}\left(r_t^m - r_t^{rf}\right) + u_{i,t}, \qquad (8) \\ \mathbf{R}\boldsymbol{\beta} &= \mathbf{r}, \\ \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_{0,i} \\ \beta_{1,i} \end{pmatrix} &= (0)\,. \end{aligned}$$

The distribution of a known statistic under the null is derived by applying known results.

# Hypothesis Testing

If $\left(\widehat{\beta} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\beta, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$, then:

$$\left(\mathbf{R}\widehat{\beta} - \mathbf{r} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\mathbf{R}\beta - \mathbf{r}, \sigma^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right). \quad (9)$$

The test is constructed by deriving the distribution of (9) under the null $\mathbf{R}\beta - \mathbf{r} = \mathbf{0}$.
Given that

$$\left(\mathbf{R}\widehat{\beta} - \mathbf{r} \mid \mathbf{X}\right) = \mathbf{R}\beta - \mathbf{r} + \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\epsilon,$$

under $H_0$, we have:

$$\begin{aligned}
& \left(\mathbf{R}\widehat{\beta} - \mathbf{r}\right)'\left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\widehat{\beta} - \mathbf{r}\right) \\
= \; & \epsilon'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\epsilon \\
= \; & \epsilon'\mathbf{P}\epsilon.
\end{aligned}$$

where $\mathbf{P}$ is a symmetric idempotent matrix of rank $r$, orthogonal to $\mathbf{M}$.

# Hypothesis Testing

Then

$$\frac{\left(\mathbf{R}\widehat{\beta}-\mathbf{r}\right)'\left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\widehat{\beta}-\mathbf{r}\right)}{s^2}\sim \mathbf{rF}\left(r,T-k\right), \qquad \text{under } H_0,$$

which can be used to test the relevant hypothesis.