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 ones 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 regression coefficients

- Estimate the coefficients in a regression, specify a null hypothesis of interest (for example, in CAPM regression the constant is zero).
- Derive a statistic (i.e. a quantity function of the regression coefficients) whose distribution is known under the null hypothesis, compute the observed value of the statistics
- Compute p as be the probability (under the null) of getting the value you have observed for the statistics
- p is called the p-value. Adopt a decision rule about p, call it p* and reject the null if the observed value of your statistic is smaller than p*. For example, if you take p=0.05 you reject the null everytime your observed statistics is smaller than 0.05. In this case you make the call that the observation of an event that has very low probability under the null is an indication that the null is rejected.

Statistical Significance of regression coefficients

- Of course by using the criterion adopted you run the risk of rejecting an hypothesis when that hypothesis is true. This is called the Probability of Type I error or the size of your test.
- There is another risk that you run: the probability of type II error, that is the probability of not rejecting a null when it is false. Think about an alternative hypothesis on the coefficients, you can compute the probability with which your statistics will be smaller than the cutoff point to which you associate a probability p*. That is the probability of type II error. The power of the test is 1- Pr (type II error).
- Note that the p-value can be computed in two ways i) by deriving the relevant distribution under the null ii) by simulating via Monte-Carlo or bootstrap the relevant distribution under the null. Using simulation makes easy to calculate the power of your test against given alternatives.

Relevance of regression coefficients

- Estimate the coefficients in a regression (for example,a CAPM regression) and keep them fixed at their point estimate
- Run an experiment by changing the conditional mean of the dependent variable via a shock to the regressors
- Assess how relevant is the shock to the regressor(s) (say, the excess market returns) to determine the dependent variables (say,excess returns on asset i)

- Inference in the Linear Regression Model is about design the appropriate statistics to test the hypothesis of interest on the coefficients in a linear model. We shall address this process in two steps
  - how to formalize the relevant hypothesis
  - on how to build the statistics.

How to formalize the relevant hypothesis

Given the genral representation of the linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

Our general case of interest is that of r restrictions on the vector of parameters with $r < k$. If we limit our interest to the class of linear restrictions on coefficients, 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.

How to formalize the relevant hypothesis

To illustrate how **R** and **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:

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

How to build the statistics

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

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

or, equivalently

$$y \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{X}\beta, \sigma^2 I\right), \tag{3}$$

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

$$\left(\widehat{\beta} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\beta, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right). \tag{4}$$

How to build the statistics

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

$$\left(\mathbf{R}\widehat{\beta} - \mathbf{r} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\mathbf{R}\beta - \mathbf{r}, \sigma^2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right). \tag{5}$$

And the relevant test can be constructed by deriving the distribution of (**??**) under the null $\mathbf{R}\beta - \mathbf{r} = \mathbf{0}$.
Unfortunately, using the normal distribution would require the knowledge of σ^2, which in general is not known. Fortunately, a statistics can be built based on the OLS estimate for σ^2.

How to build the statistics

Fortunately, a statistics can be built based on the OLS estimate for σ^2. In fact, it can be shown that

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

that can be used to test the relevant hypothesis.
Notice that, as we know that in the case r=1, $t_{t-k}=\sqrt{F\left(1,T-k\right)}$,if we are interested in testing hypothesis on a single coefficients (say β_1) we can use the following statistic:

$$\frac{\hat{\beta}_1-\beta_1}{\left(Var\left(\hat{\beta}_1\right)\right)^{1/2}} \sim t\left(T-k\right) \text{ under } H_0.$$

Therefore, an immediate test of significance of the coefficient can be performed, by taking the ratio of each estimated coefficient and the associated standard error.

The Partitioned Regression Model

Given the linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

Partition $\mathbf{X}$ in two blocks two blocks of dimension (Txr) and $(Tx\,(k-r))$ and $\boldsymbol{\beta}$ in a corresponding way into $\left[\begin{array}{cc} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 \end{array}\right]$. The partitioned regression model can then be written as follows

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon},$$

The Partitioned Regression Model

It is useful to derive the formula for the OLS estimator in the partitioned regression model. To obtain such results we partition the 'normal equations' $\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}^{'}\mathbf{y}$ as:

$$\begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \mathbf{y},$$

or, equivalently,

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix}. \tag{6}$$

The Partitioned Regression Model

System (6) can be resolved in two stages by first deriving an expression $\widehat{\boldsymbol{\beta}}_2$ as:

$$\widehat{\boldsymbol{\beta}}_2 = \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\left(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1\right),$$

and then by substituting it in the first equation of (6) to obtain

$$\mathbf{X}_1'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_1'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\left(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1\right) = \mathbf{X}_1'\mathbf{y},$$

from which:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \\
\mathbf{M}_2 &= \left(\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\right).
\end{aligned}$$

The Partitioned Regression Model

Note that, as $\mathbf{M}_2$ is idempotent, we can also write:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{M}_2'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2'\mathbf{M}_2\mathbf{y},$$

and $\widehat{\boldsymbol{\beta}}_1$can be interpreted as the vector of OLS coefficients of the regression of $\mathbf{y}$ on the matrix of residuals of the regression of $\mathbf{X}_1$ on $\mathbf{X}_2$. Thus, an OLS regression on two regressors is equivalent to two OLS regressions on a single regressor (Frisch-Waugh theorem).

The Partitioned Regression Model

Finally, consider the residuals of the partitioned model:

$$\begin{aligned}
\widehat{\epsilon} &= \mathbf{y} - \mathbf{X}_1\widehat{\beta}_1 - \mathbf{X}_2\widehat{\beta}_2, \\
\widehat{\epsilon} &= \mathbf{y} - \mathbf{X}_1\widehat{\beta} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\left(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\widehat{\beta}_1\right), \\
\widehat{\epsilon} &= \mathbf{M}_2\mathbf{y} - \mathbf{M_2}\mathbf{X}_1\widehat{\beta}_1 \\
&= \mathbf{M}_2\mathbf{y} - \mathbf{M_2}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \\
&= \left(\mathbf{M}_2 - \mathbf{M_2}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\right)\mathbf{y},
\end{aligned}$$

however, we already know that $\widehat{\epsilon} = \mathbf{My}$, therefore,

$$\mathbf{M} = \left(\mathbf{M}_2 - \mathbf{M_2}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\right). \tag{7}$$

Testing restrictions on a subset of coefficients

In the general framework to test linear restrictions we set $\mathbf{r} = \mathbf{0}$, $\mathbf{R} = \begin{bmatrix} I_r & 0 \end{bmatrix}$, and partition $\boldsymbol{\beta}$ in a corresponding way into $\begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 \end{bmatrix}$. In this case the restriction $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ is equivalent to $\boldsymbol{\beta}_1 = 0$ in the partitioned regression model.
Under H_0, $\mathbf{X}_1$ has no additional explicatory power for $\mathbf{y}$ with respect to $\mathbf{X}_2$, therefore:

$$H_0\text{: } \mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \qquad (\boldsymbol{\epsilon} \mid \mathbf{X}_1, \mathbf{X}_2) \sim N\left(0, \sigma^2 I\right).$$

Note that the statement

$$\mathbf{y} = \mathbf{X}_2\gamma_2 + \boldsymbol{\epsilon}, \qquad (\boldsymbol{\epsilon} \mid \mathbf{X}_2) \sim N\left(0, \sigma^2 I\right),$$

is always true under our maintained hypotheses. However, in general $\gamma_2 \neq \boldsymbol{\beta}_2$.

Testing restrictions on a subset of coefficients

To derive a statistic to test H_0 remember that the general matrix $\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'$ becomes now the upper left block of $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, which we can now write as $\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}$. The statistic then takes the form

$$\frac{\widehat{\beta}_1'\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)\widehat{\beta}_1}{rs^2} = \frac{\mathbf{y}'\mathbf{M_2}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}}\frac{T-k}{r} \sim F\left(T-k, r\right).$$

Given (7), (6) can be re-written as:

$$\frac{\mathbf{y}'\mathbf{M_2}\mathbf{y} - \mathbf{y}'\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}}\frac{T-k}{r} \sim F\left(T-k, r\right), \tag{8}$$

where the denominator is the sum of the squared residuals in the unconstrained model, while the numerator is the difference between the sum of residuals in the constrained model and the sum of residuals in the unconstrained model.

Testing restrictions on a subset of coefficients

Consider the limit case $r = 1$ and β_1 is a scalar. The F-statistic takes the form

$$\frac{\widehat{\beta}_1^2}{s^2\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)} \sim F\left(T-k, r\right), \text{ under } H_0,$$

where $\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}$ is element $(1,1)$ of the matrix $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.
Using the result on the relation between the F and the Student's t-distribution:

$$\frac{\widehat{\beta}_1}{s\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{1/2}} \sim t\left(T-k\right) \text{ under } H_0.$$

Therefore, an immediate test of significance of the coefficient can be performed, by taking the ratio of each estimated coefficient and the associated standard error.

The Relevance of a Regression

- 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.
- a variable can be very significant in explaining the variance of $E(\mathbf{y} \mid \mathbf{X})$, but little of the unconditional variance of $\mathbf{y}$ can be explained by the variance of $E(\mathbf{y} \mid \mathbf{X})$.
- statistical significance does not imply relevance

The R-squared as a measure of relevance

To illustrate the point let us consider two specific cases of applications of the CAPM:

$$\begin{aligned}
\left(r_t^i - r_t^{rf}\right) &= 0.8\sigma_m u_{m,t} + \sigma_i u_{i,t} \\
\left(r_t^m - r_t^{rf}\right) &= \mu_m + \sigma_m u_{m,t} \\
\begin{pmatrix} u_{i,t} \\ u_{m,t} \end{pmatrix} &\sim n.i.d.\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] \\
\mu_m &= 0.0065, \sigma_m = 0.054, \sigma_1 = 0.09, \sigma_2 = 0.005
\end{aligned}$$

We simulate an artificial sample of 1056 obs (same length with the sample July 1926-June2014) observations. μ_m and σ_m are calibrated to match the first two moments of the market portfolio excess returns over the sample 1926:7-2014:7. The standard errors of the two excess returns are calibrated to deliver R^2 of respectively about .22 and .98.

2

To calibrate the R^2 consider that

$$\begin{aligned} R^2 &= \frac{0.8^2\sigma_m^2}{0.8^2\sigma_m^2 + \sigma_i^2} \\ \sigma_i^2 &= \frac{0.8^2\sigma_m^2}{R^2} - 0.8^2\sigma_m^2 \\ \sigma_i^2 &= 0.8^2\sigma_m^2\left(\frac{1}{R^2} - 1\right) \end{aligned}$$

The R-squared as a measure of relevance

By running the two CAPM regressions on the artificial sample:

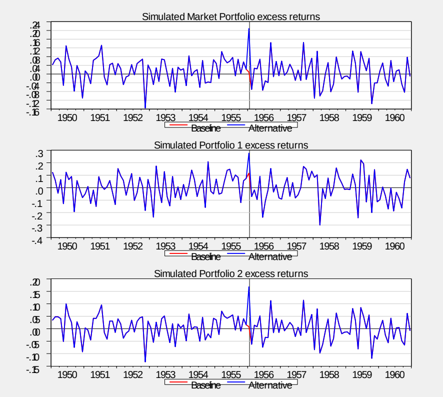
TABLE 3.1: The estimation of the CAPM on artificial data

Dependent Variable $\left(r_t^1 - r_t^{rf}\right)$				
Regressor	Coefficient	Std. Error	t-ratio	Prob.
$\left(r_t^m - r_t^{rf}\right)$	0.875		17.48	0.000
R^2 0.22	S.E. of regression 0.0076			
Dependent Variable $\left(r_t^2 - r_t^{rf}\right)$				
Regressor	Coefficient	Std. Error	t-ratio	Prob.
$\left(r_t^m - r_t^{rf}\right)$	0.793		201.86	0.000
R^2 0.972	S.E. of regression 0.0000			

In both cases the estimated beta are statistically significant and very close to their true value of 0.8.

The R-squared as a measure of relevance

Simulate again the processes but introduce at some point a temporary shift of two per cent in the excess returns in the market portfolio.



In both experiments the conditional expectation changes of the same amount but the share of the unconditional variance of **y** explained by the regression function is very different, as different are the R^2s.

The partial regression theorem

The Frisch-Waugh Theorem described above is worth more consideration.
The theorem tells us than any given regression coefficient in the model $E(y \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ can be computed in two different but exactly equivalent ways:
1) by regressing y on all the columns of $\mathbf{X}$,
2) by first regressing the j-th column of $\mathbf{X}$ on all the other columns of $\mathbf{X}$, computing the residuals of this regression and then by regressing y on these residuals.
This result is relevant in that it clarifies that the relationships pinned down by the estimated parameters in a linear model do not describe the connections between the regressand and each regressor but the connection between the part of each regressor that is not explained by the other ones and the regressand.

What if analysis

- The relevant question in this case becomes “how much shall y change if I change $\mathbf{X}_i$?”
- The estimation of a single equation linear model does not allow to answer that question, for a number of reasons.
- First, estimated parameters in a linear model can only answer the question how much shall $E(y \mid \mathbf{X})$ if I change $\mathbf{X}$? We have seen that the two questions are very different if the R^2 of the regression is low, in this case a change in $E(y \mid \mathbf{X})$ may not effect any visible and relevant effect on y.
- Second, a regression model is a conditional expected value GIVEN $\mathbf{X}$. In this sense there is no space for “changing” the value of any element in $\mathbf{X}$.

What if analysis

- Any statement involving such a change requires some assumption on how the conditional expectation of y changes if $\mathbf{X}$ changes and a correct analysis of this requires an assumption on the joint distribution of y and $\mathbf{X}$.
- Simulation might require the use of the multivariate joint model even when valid estimation can be performed concentrating only on the conditional model.
- Strong exogeneity is stronger than weak exogeneity for the estimation of the parameters of interest.

What if analysis

Think of a linear model with know parameters

$$y = \beta_1 x_1 + \beta_2 x_2$$

What is in this model the effect of on y of changing x_1 by one unit while keeping x_2 constant ? Easy β_1.
Now think of the estimated linear model:

$$y = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{u}$$

Now y is different from $E(y \mid \mathbf{X})$ and the question "what is in this model the effect of on $E(y \mid \mathbf{X})$ of changing x_1 by one unit while keeping x_2 constant ?" does not in general make sense.

What if analysis

- Changing x_1 keeping x_2 unaltered implies that there is zero correlation among this variables.
- But the estimates $\hat{\beta}_1$and $\hat{\beta}_2$ are obtained by using data in which in general there is some correlation between x_1 and x_2.
- Data in which fluctuations in x_1 do not have any effect on x_2 would have most likely generated different estimates from those obtained in the estimation sample.
- The only valid question that can be answered using the coefficients in linear regression is "What is the effect on $E(y \mid \mathbf{X})$ of changing the part of each regressors that is orthogonal to the other ones".
- "What if" analysis requires simulation and in most cases a low level of reduction than that used for regression analysis.

The semi-partial R-squared

- When the columns of **X** are orthogonal to each other the total R^2 can be exactly decomposed in the sum of the partial R^2 due to each regressor x_i (the partial R^2 of a regressor i is defined as the R^2 of the regression of y on x_i).
- This is in general not the case in applications with non experimental data: columns of **X** are correlated and a (often large) part of the overall R^2 does depend on the joint behaviour of the columns of **X**.
- However, it is always possible to compute the marginal contribution to the overall R^2 due to each regressor x_i, defined as the difference between the overall R^2 and the R^2 ot the regression that includes all columns **X** except x_i. This is called the semi-partial R^2.

The semi-partial R-squared

Interestingly, the the semi-partial R^2 is a simple tranformation of the t-ratio:

$$spR_i^2 = \frac{t_{\beta_i}^2 \left(1 - R^2\right)}{(T - k)}$$

This result has two interesting implications.

- First, a quantity which we considered as just a measure of statistical reliability, can lead to a measure of relevance when combined with the overall R^2 of the regression.
- Second, we can re-iterate the difference between statistical significance and relevance. Suppose you have a sample size of 10000 and you have 10 columns in **X** and the t-ratio on a coefficient β_i is of about 4 with an associate P-value of the order .01: "very" statistical significant! The derivation of the semi-partial R^2 tells us that the contribution of this variable to the overall R2 is at most approximately 16/(10000-10) that is: less than two thousands.