Econometrics Cheat Sheet 0.8

Basic concepts

Definitions

Econometrics - is a social science discipline with the objective of quantify the relationships between economic agents, contrast economic theories and evaluate and implement government and business policies.

Econometric model - is a simplified representation of the reality to explain economic phenomena.

Data types

Cross section - data taken at a given moment in time, an static "photo". Order does not matter.

Temporal series - observation of one/many variable/s across time. Order does matter.

Panel data - consist of a temporal series for each observation of a cross section.

Pooled cross sections - combines cross sections from different temporal periods.

Phases of an econometric model

1. Specification
2. Estimation
3. Validation
4. Utilization

Econometric model assumptions

Under this assumptions the estimators of the parameters will present "good properties". **Extended Gauss Markov assumptions**:

1. Parameters linearity. y must be a linear function of the β's.
2. The sample (data) taken from the population is random. Characteristics:
   - Independence: all the co-variances between the x's are zero.
   - Identical distribution: all the expected values and variances of the observations are the same.
3. $E(u|x_1, ..., x_k) = 0$, the estimations are unbiased. Implications:
   - $E(u) = 0$, there are no systematic errors.
   - $Cov(u, x_1) = ... = Cov(u, x_k) = 0$, there are no relevant variables left out the model.
   - $E(y|x_1, ..., x_k) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$, the linear relation between y and the x's is fulfilled, at least in average.
4. Homocedasticity: $Var(u_i|x_{1i}, ..., x_{ki}) = \sigma^2$, the variability of the residual is the same for all levels of x.
5. No auto-correlation: $Cov(u_i, u_j|x) = 0$, the residuals do not contain information about other residuals.
6. The residuals are normally distributed.
7. No multicolineality: none of the independent variables is constant nor exist an exact (or approximate) linear relation between them, they are linearly independents.
8. The number of available data is greater than $k + 1$ (parameters to estimate).

Interpretation of coefficients

Model	Dependent	Independent	β_1 interpretation
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$log(x)$	$\Delta y = (\beta_1/100)[1\%\Delta x]$
Log-level	$log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$log(y)$	$log(x)$	$\%\Delta y = \beta_1 \%\Delta x$
Quadratic	y	$x + x^2$	$\Delta y = (\beta_1 + 2\beta_2 x)\Delta x$

Ordinary Least Squares

Objective of OLS - minimize the square sum of residuals:

$Min \sum_{i=1}^{n} \hat{u}_i^2 = Min \sum_{i=1}^{n} [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)]^2$

Simple regression model

Equation: $y_i = \beta_0 + \beta_1 x_{1i} + u_i; i = 1, ..., n$

Estimation: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ $\hat{u}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

With

$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$

$\hat{\beta}_1 = \frac{Cov(Y,X)}{Var(X)}$

Multiple regression model

$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i, i = 1, ..., n$

$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i + ... + \hat{\beta}_k X_{ki})$

Objective:

$Min \sum_{i=1}^{n} \hat{u}_i^2$

Then

$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - ... - \hat{\beta}_k \overline{X}_k$

$\hat{\beta}_j = \frac{Cov(Y, resid(X_j))}{Var(resid(X_j))}$

Properties of OLS

- Linearity in Y.
- Normality: Y/X $N(\beta_0 + \beta_1 X, \sigma^2)$
- Expected value of the estimator: $E(\hat{\beta}_1/X_i) = \beta_1$, then $\hat{\beta}_1$ is an unbiased estimator of β_1
- Variance of the estimator: $Var(\hat{\beta}_1/X_i) = \frac{\sigma^2}{nVar(X_i)}$

Efficiency of OLS estimators, Gauss-Markov Theorem. In the context of the simple or multiple linear regression model, the OLS estimators of the parameters are those with the lowest variance between the lineal and unbiased estimators

Central Limit Theorem

Under the CLT, $\hat{\beta}_j$ is a consistent estimator of the population parameter β_i.

$plim \hat{\beta}_i = \beta_i$

The Central Limit Theorem allow us to obtain (asymptotically):

$\frac{\hat{\beta}_i - \beta_i}{s(\hat{\beta}_i)} \sim N(0, 1)$

Goodness of the fit, R-Squared

The R2 is a measure of the goodness of the fit, how the OLS fit to the data.

Is the proportion of variability of the dependent variable explained by the regression line:

$R^2 = \frac{\sum_i (\hat{Y}_i - \overline{Y}_i)^2}{\sum_i (Y_i - \overline{Y}_i)^2} = 1 - \frac{\sum_i \hat{\epsilon}_i^2}{nS_y^2}$

The R^2 takes values between 0 (no lineal explanation of the variations of Y) and one (total explanation of the variations of Y)

Is a descriptive measure of the global fit of the model.

The R^2 measures the percentage of variation of Y that is linearly explained by the variations of X.

The R^2 increments it's value when increments the number of regressors, whatever they are relevant or not.

For eliminate the above phenomena, there is a R^2 corrected by degrees of freedom ($\overline{R}^2$).

$\overline{R}^2 = 1 - \frac{n-1}{n-k-1} \frac{\sum_i \hat{\epsilon}_i^2}{\sum_i (Y_i - \overline{Y}_i)^2} = 1 - \frac{n-1}{n-k-1}(1 - R^2)$

For big sample sizes:

$\overline{R}^2 \approx R^2$

Errors

Standard error of the regression is a measure of the goodness of the fit.

$\hat{\sigma} = \sqrt{\frac{\sum_i \hat{\epsilon}_i^2}{n-k-1}}$

It's value decreases as the number of regressors increase, so it have the same problem as the R^2

Hypothesis testing

An hypothesis test is a rule designed to explain from a sample, if exist evidence or not to reject an hypothesis that is made on one or more population parameters.
Elements of an hypothesis contrast:

- Null hypothesis (H_0): is the hypothesis that you want to contrast.
- Alternative hypothesis (H_1): is the hypothesis that cannot be rejected when the null hypothesis is rejected.
- Statistic of contrast: is a random variable with a known distribution that allow us to see if we reject (or not) the null hypothesis.
- Significance level (α): is the probability of rejecting the null hypothesis being true (Error type I). Is chosen by who conduct the contrast. Commonly is 0,10, 0.05, 0.01 or 0,001
- Critic value: is the value that, for a determined value of α, determines the reject (or not) of the null hypothesis.
- p-value: is the highest level of significance for what we do not reject (accept) the null hypothesis (H_0).

The rule is: if p-value is lower than α, there is evidence at that given α to reject the null hypothesis (accept the alternative instead).

Individual significance contrasts

- $H_0 : \beta_j = 0$
- $H_1 : \beta_j \neq 0$

Supposing that the model's errors are distributed as a normal distribution.
$Y_i/X_i, ..., X_k \sim N(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k, \sigma^2)$
Then, under H_0:
$t = \frac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \sim t_{n-k-1,\alpha/2}$
If $\mid t \mid > t_{n-k,\alpha/2}$ there is evidence to reject the null hypothesis.

Confidence intervals

Supposing normality of the residuals:
$Y_i/X_i, ..., X_k \sim N(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k, \sigma^2)$
Then, $t = \frac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \sim t_{n-k-1}$
The confidence interval:
$P[\hat{\beta}_j - t_{n-k-1,\alpha/2} s(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t_{n-k-1,\alpha/2} s(\hat{\beta}_j)] = 1 - \alpha$

The F contrast

It uses a non restricted model and a restricted model to do assumptions about the parameters.

- Non restricted model: is the model on which we want to make the hypothesis contrast.
- Restricted model: is the model on which the hypothesis that we want to contrast have been imposed.

Then, looking at the errors, there is:

- $\sum_{i=1}^{n} \hat{\epsilon}_N^2$: is the sum of the OLS residuals of the non restricted model (SRN).
- $\sum_{i=1}^{n} \hat{\epsilon}_N^2$: is the sum of the OLS residuals of the restricted model (SRR).

Then:
$F = \frac{SRR - SRN}{SRN} \frac{(n-K-1)}{q} \sim F_{q,n-K-1}$
Where K is the number of parameters in the non restricted model and q is the number of linear hypothesis.
When $F_{tables} > F_{q,n-K-1}$

Dummy variables and structural change

Dummy (or binary) variables are used for qualitative information: sex, married or not, etc.
The dummy variables get the value of 1 in a category, and 0 on the rest.
The dummy variables are used to analyze and modeling structural changes in the models parameters.
We denominate structural changes to the modifications in the value of the parameters of the models for different sub-populations.
In the constant, their associate parameter represents the difference in mean between the values.
In the parameters that determines the slope of the regression line, the associate parameter represents the difference in the effect between the values.

The Chow's structural contrast

When we want to analyze the existence of structural changes in all the model parameters, is more common to use a particular expression of the F contrast known as the Chow's contrast.
It defines two non restricted models (with structural change):
$Y_i = \beta_0^A + \beta_1^A X_{1i} + ... + \beta_k^A X_{ki} + \epsilon_i$ sub-sample A
$Y_i = \beta_0^B + \beta_1^B X_{1i} + ... + \beta_k^B X_{ki} + \epsilon_i$ sub-sample B
Restricted model (without structural change):
$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \epsilon_i$
With the restriction:
$H_0 : \beta_j^A = \beta_j^B; j = 0, 1, ..., k$
Be SRN the sum of the OLS square residuals of the non restricted model: $SRN = SR_A + SR_B$
Be SRR the sum of the OLS square residuals of the restricted model.
Then:
$F = \frac{SRR - SRN}{SRN} \frac{n-2(k+1)}{k+1} \sim F_{k+1,n-2(k+1)}$
If $F_{table} < F$, is evidence to reject the null hypothesis.

Multicolineality

If there is exact multicolineality, the equation system of OLS cannot be solved due to infinite solutions.

- Approximate multicolineality: when one or more variables are almost a constant or there is a linear relation between them. In this context, there is not a problem, given the classic requirements of OLS, and the inference is valid. But, there are some empiric consequences of this:
  - Small sample variations can induce to big variations in the OLS estimations.
  - The variance of the OLS estimator of the X variables that are collinear $Var(\hat{\beta}_j)$ increments, then, the inference of the parameter is affected → The estimation of the parameter is very imprecise (big confidence interval)

Calculating the Variance Inflation Factor to analyze multicolineality problems:
$VIF(\hat{\beta}_j) = \frac{1}{1-R_j^2}$

Indicates the increment of $V(\hat{\beta}_j)$ because of the multicolineality.

- If it is bigger than 10, indicates that there are multicolineality problems.
- From 4 onwards, it is advisable to analyze in more detail if there may be multicolineality

One typical characteristic of multicolineality is that the regression coefficients of the model are not individually different from zero (because the high variances), but jointly different from zero.

Heterocedasticity

The residuals ϵ_i of the population regression function don't have the same variance σ^2:
$Var(\varepsilon \mid X) = Var(Y \mid X) \neq \sigma^2 \rightarrow$ HETEROCEDASTICITY

Consequences

The estimators still are unbiased. The estimators still are consistent. The variance of the estimators is biased: the construction of confidence intervals and the

hypothesis contrast are not valid. In this context, OLS is not an unbiased lineal estimator of minimum variance. There is an alternate unbiased lineal estimator of minimum variance denominated estimator of minimum pondered squares (MPS) or minimum generalized squares (MGS).
Under the Gauss-Markov Theorem assumptions, OLS estimators are not efficient. The estimations of the variance of the estimators are biased. The hyphotesis contrast and the confidence intervals are not reliable.

Detection

Plots (look for structures in plots with the square residuals) and contrasts: Park test, Goldfield-Quandt, Bartlett, Breush-Pagan, CUSUMQ, Spearman, White. White's null hypothesis:
$H_0 = HOMOCEDASTICITY$

Correction

- When the variance structure is known, use weighted least squares.
- When the variance structure is not known: make assumptions of the possible structure and apply weighted least squares (factible weighted least squares)
- Supposing that σ_i^2 is proportional to x_i^2, divide by x_i

Autocorrelation

The "natural" context of this phenomena is in temporal series.
The residual of any observation, u_i is correlated with the residual of any other observation. The observations are not independent $E(u_i, u_j) \neq 0; i \neq j$
Causes of autocorrelation:

- By the existence of tendencies and/or cycles in the data.
- By the aggregation of the data.
- By the existence of an especification error (omission of relevant variables, bad functional form, etc.)
- Model specification error
-

Regression Analysis

Study and predict the mean value of a variable regarding the base of fixed values of other variables. We usually use Ordinary Least Squares (OLS).

Correlation Analysis

The correlation analysis not distinguish between dependent and independent variables. **Simple Correlation** Measure the grade of lineal association between two variables.

Utilization

Interpretation of the model