1: Regression Analysis

GECO 6281 Advanced Econometrics 1 (Lab)

Patrick Mokre

Fall 2020

Econometrics

- Extracting Relationships from Data (Correlation Analysis).
- Using assumptions about the data generating process to reveal data structure.
- Use Abstraction to find patterns not visible to the open eye.
- Frequentist econometrics is built around hypothesis testing
- Is there a “data generating process”?
- How would you describe the result of a hypothesis test?
- Is hypothesis testing necessary for theoretical advancement?

Econometrics: Terminology

- probability and probability distribution
- linear relationship
- regression
- ordinary least squares
- generalized least squares
- homoskedasticity
- multicollinearity
- endogeneity
- maximum likelihood
- hypothesis
- degrees of freedom
- fat tail

Advanced Econometrics 1

Learning Goals:

- Understanding of panel data analysis
- Ability to apply panel data analysis to research questions
- Confidence to read and comment on quantitative work
- Possibly ideas for a PhD chapter?

Your expectations and learning goals?

- What can you not do, but know it exists, and want to know after this term?
- How important is this class for your research projects?
- Do you think Econometrics should be an obligatory course?
- How would you want to teach this class?

Linear Relationships

Consider a linear approximation for some observations $y_1, ..., y_n$ with $x_1, ..., x_k$ vectors of length n.

$$\beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

Then the difference between observations and approximation can be written as follows.

$$y_i - x_i' \beta$$

Then a loss function can be used to minimize different aspects of the difference between observations and approximations. In OLS, the function used is squared deviation.

$$S(\hat{\beta}) = \sum_i^n (y_i - x_i' \hat{\beta})^2$$

Solving the oridnary least squares problem

To find a vector $\hat{\beta}_{OLS}$ that minimizes the quadratic loss function, one can derive K First-Order-Conditions by deriving $S(\hat{\beta})$ with respect to $\hat{\beta}$

$$
\begin{aligned}
\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} &= -2 \sum_{i}^{N} x_i (y_i - x_i' \hat{\beta}) = 0 \\
\sum_{i}^{N} [x_i' x_i] \hat{\beta} &= \sum_{i}^{N} x_i y_i \\
\hat{\beta}_{OLS} &= [\sum_{i}^{N} x_i' x_i]^{-1} \sum_{i}^{N} x_i y_i \\
&= (X'X)^{-1} X'Y
\end{aligned}
$$

Best linear unbiased estimator (BLUE)

$\hat{\beta}_{OLS}$ is the best linear unbiased estimator in a linear relationship (and $\hat{y} = x_i'\hat{\beta}_{OLS}$ the best linear approximation) under certain assumptions or conditions about the structure of the covariates $b_1, ..., b_k$ and the error term vector $e = y - x'\hat{\beta}_{OLS}$

- **How are these conditions called**?
- **What are these conditions**?

Gauss-Markov conditions

The standard Gauss-markov conditions are:

- $E\epsilon_i = 0 \quad \forall i \in N$
- $\epsilon_1, ..., \epsilon_n$ and $x_1, ..., x_n$ are *independent*
- $Var(\epsilon_i) = \sigma^2 \quad \forall i \in N$
- $cov(\epsilon_i, \epsilon_j) = 0 \quad \forall i, j \in N, \forall j \neq i$

From assumptions 1, 3 and 4 follows that the error term is independent of the observed covariates, $E(\epsilon \mid X) = E(\epsilon) = 0$ and $Var\epsilon \mid X = V(\epsilon) = \sigma^2 I_N$ (where I_N is an N-dimensional diagonal identity matrix).

Gauss Markov conditions 2

$\hat{\beta}_{OLS}$ is **unbiased** because $E(\hat{\beta}_{OLS}) = E((X'X)^{-1}X'Y) = E(\beta + (X'X)^{-1}X'\epsilon) = \beta + E((X'X)^{-1}X'\epsilon) = \beta + 0 = \beta$. Note that the estimator is unbiased even if **heteroskedasticity** or **autocorrelation** are present.

Clearly, $Var(\hat{\beta}_{OLS})$ can be read as a measure of how likely it is that the estimator is a bad guess. The variance is *deterministic in X*, and under assumptions 2, 3 and 4, $\hat{\beta}_{OLS}$ is **efficient**, i.e. the **best** estimator.

$$Var(\hat{\beta}_{OLS}) = E(\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta) = E((X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}) = E((X'X)^{-1}X'(\sigma^2 I_N)X(X'X)^{-1}) = \sigma^2(X'X)^{-1}$$

Distribution of the error terms

Assumptions 1, 3 and 4 can be replaced by assuming that errors are independently Gaussian Normal distributed: - $\epsilon \sim N(0, \sigma^2)$.

This implies that $y_i \mid x_i$ is also Gaussian Normal distributed.

In summary, one can say that $\hat{\beta}_{OLS} \sim N(\beta, \sigma^2(X'X)^{-1})$.

How does one estimate how well a model fits the data?

Goodness of Fit

The most popular measure of the goodness of fit in an OLS model is R^2, the proportion of explained variance to observed variance in the model.

$$R^2 = \frac{\hat{V}(\hat{y})}{\hat{V}(y)} = \frac{1/(N-1)\sum_i^N (\hat{y} - \bar{y})^2}{1/(N-1)\sum_i^N (y - \bar{y})^2}$$

Since $y_i = \hat{y}_i + \epsilon_i$ and $\sum_i^N y_i \epsilon_i = 0$, it follows that $\hat{V}(y_i) = \hat{V}(\hat{y}_i) + \hat{V}(\epsilon_i)$. Then R^2 can be calculated from observed values only.

$$R^2 = 1 - \frac{\hat{V}(\epsilon_i)}{\hat{V}(y_i)} = 1 - \frac{1/(N-1)\sum_i^N \epsilon_i^2}{1/(N-1)\sum_i^N (y_i - \bar{y})^2}$$

Hypothesis Testing

A hypothesis is a sentence. A “supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.”. *Most econometric results are hypotheses*.

In hypothesis testing, one measures how likely it is for data from the same time and place to look very different. Necessary assumptions for this are the existing of a **true data generating process** and **possible infinite repition data generation**.

Hypothesis Testing 2: Significant Coefficients

The OLS results about β are hypotheses. In the simplest case, one wants to test if β_k is different from zero. Let c_{kk} be the k-th entry in the diagonal of $(X'X)^{-1}$ and $\hat{\beta}_{OLS,k} \sim N(\beta_k, \sigma^2 c_{kk})$.

Construct a variable z, which follows a Gaussian Normal distribution.

$$z = \frac{\hat{\beta} - \beta}{\sigma\sqrt{c_{kk}}}$$

However, σ is unknown, but there exists an unbiased estimator $s^2 = \frac{1}{N-K}\sum_i^N e_i^2$ which is independent of $\hat{\beta}$. Furthermore, $\frac{(N-K)s^2}{\sigma^2} \sim \chi^2_{N-K}$.

Hypothesis Testing 3: Significant Coefficients

$$t = \frac{\hat{\beta} - \beta}{s\sqrt{c_{kk}}}$$

t is the ratio of a Gaussian Normal and a χ^2 variable, and thus follows a Student-t distribution with (N - K) degrees of freedom. The Student-t distribution looks a lot like the Gaussian Normal for large *degrees of freedom*, but has fatter tails.

Note that hypothesis testing using the t-distribution is testing under the assumption that the null hypothesis is correct.

Why are we interested in the distribution of t? What are degrees of freedom?

Degrees of Freedom

The degrees of freedom (DF) indicate the number of independent values that can vary in an analysis without breaking any constraints. it increases in independent information you can use for parameter estimation, and decreases in parameters you have to estimate due to your modeling choices.

In frequentist statistics, hypothesis testing is based in the assumption that coefficient estimates (such as $\hat{\beta}$ follow some distribution, where the shape is co-determined by the degrees of freedom (Student T, χ^2, ...).

For low degrees of freedom, these distributions become very narrow, making hypothesis testing difficult. Coefficient estimates become unreliable, and hypothesis tests lose testing power.

Joint Hypothesis Testing

A more general Null hypothesis is that J out of K coefficients are equal to zero. The standard application is to test if all coefficients apart from the intercept are zero.

$$H_0 : \beta_{K-J+1} = ... = \beta_K = 0$$

The test is implemented by comparing the sum of residuals of the model S_1 and a restricted model, in which the covariates we assume to be zero are not even implemented, S_0. Under the Null, $S_0 - S_1 = 0$, and $\frac{S_0 - S_1}{\sigma^2} \sim \chi^2_J$.

$$\begin{aligned} f &= \frac{(S_0 - S_1)/J}{S_1/(N-K)} \\ &= \frac{(R_0^2 - R_1^2)/J}{(1 - R_1^2)(N-K)} \end{aligned}$$

f is F-distributed with $N - K$ degrees of freedom and for a given significance level.

Generalized Least Squares

When the 5 Gauss-markov Conditions do not hold, $\hat{\beta}_{OLS}$ is no longer BLUE.

GLS gives an estimator that reduces to $\hat{\beta}_{OLS}$ when the Gauss-Markov conditions are satisfied but still holds as BLUE otherwise.

The relevant conditions are summarized as:

$$E(\epsilon \mid X) = E(\epsilon) = 0$$
$$V(\epsilon \mid X) = V(\epsilon) = \sigma^2 I$$

Generalized Least Squares 2

Both Autocorrelation and Heteroskedasticity imply that the variance condition no longer holds and can be re-written as an error-correlation matrix Ψ.

$$\begin{aligned} V(\epsilon \mid X) &= \sigma^2 \Psi \\ V(\hat{\beta}) &= \sigma^2 (X'X)^{-1} X' \Psi X (X'X)^{-1} \end{aligned}$$

If $\Psi = I$, this reduces to $V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

Generalized Least Squares 3

The intuition behind the generalized least squares (GLS) procedure is have a model which does **not satisfy Gauss Markov 1-4** and transform it such that it does.

One can define a square and non-singular matrix P for which:

$$\begin{aligned}\Psi^{-1} &= P'P \\ \Psi &= (P'P)^{-1} = P^{-1}(P')^{-1} \\ P\Psi P' &= PP^{-1}(P')^{-1}P' = I\end{aligned}$$

For this (**not necessarily unique**) matrix P for which $P\epsilon$ satisfies the Gauss-Markov conditions.

$$\begin{aligned}E(P\epsilon \mid X) &= PE(\epsilon \mid X) = 0 \\ V(P\epsilon \mid X) &= PV(\epsilon \mid X)P' = \sigma^2 P\Psi P' = \sigma^2 I\end{aligned}$$

Generalized Least Squares 4

One can use the GLS transformation on the model to retrieve $\hat{\beta}_{GLS}$.

$$
\begin{aligned}
Py &= PX\beta + P\epsilon \\
y^* &= X^*\beta + \epsilon^* \\
\hat{\beta}_{GLS} &= ({X^*}'X)^{-1}{X^*}'y^* = (X'\Psi X)^{-1}X'\Psi y
\end{aligned}
$$

In this form, the exact matrix P is irrelevant, however the choice of Ψ^{-1} is crucial. When Ψ is known, the modle is deterministic, otherwise it has to be **estimated**.

Summary

- What is Econometrics?
- what is a probability distribution?
- What are the learning goals in this class?
- What does BLUE stand for?
- Define homoskedasticity, consistency and efficiency.
- What is the intuition behind an R^2 statistic?
- When are Student-t and F-distributions of use?