Linear Regression for Time Series and Applications in Macro Forecasting

Laurent Ferrara[1]

[1]Banque de France

CEF, Nov. 2018

Plan

Basic specification

- Objective: Assess the explanatory power of k variables X_{it} for $i = 1, \ldots, k$, grouped in the vector X_t, for the variable of interest Y_t.
- From a statistical point of view, X_{it} and Y_t are stochastic processes, i.e; a sequence of random variables for which we observe realizations x_{it} and y_t, for $t = 1, \ldots, T$
- In the remaining, we assume that all the considered variables are stationary variables, adjusted from seasonal variations
- Some of the explanatory variables can be deterministic (intercept, trend, ...)

Basic specification

We assume the explanatory variables X_t have a linear impact on the dependent variable Y_t, such that for $t = 1, \ldots, T$:

$$Y_t = \beta_1 X_{1t} + \ldots + \beta_k X_{kt} + \varepsilon_t$$

with:

- the conditional expectation of Y_t knowing all the information on explanatory variables at t, denoted F_t, is given by:

$$E[Y_t | F_t] = \beta_1 X_{1t} + \ldots + \beta_k X_{kt}$$

- β_i measures the change in the expected value of Y_t when there is a marginal change in X_{it}, other variables constant
- ε_t: an error term capturing the part of Y_t not explained by X_t

Basic specification

The linear regression model can be expressed in a matrix form :

$$Y = X\beta + \varepsilon$$

by grouping:

- Y_t and ε_t into $(T \times 1)$ vectors Y and ε
- $\beta_1, \ldots, \beta_k$ into the $(k \times 1)$ vector β
- $X_{1t}, \ldots, X_{kt}$ into the $(T \times k)$ matrix X

Basic assumptions

We assume the following *Linear Regressions Assumptions*:

1. $E(\varepsilon) = 0$
2. $E(\varepsilon\varepsilon') = \sigma^2 I_T$
3. X is distributed independently of ε
4. $X'X$ is non-singular
5. X is weakly stationary

In addition we assume (i) no omitted variables, (ii) linearity in the relationship and (iii) stability of parameters overtime

Parameter estimation

- In the previous model the $k+1$ parameters $(\beta_1, \ldots, \beta_k, \sigma^2)$ are unknown and have to be estimated from the observed variables
- From the Gauss-Markov theorem, we know that the best linear unbiased estimator (BLUE) is the Ordinary Least-Squares (OLS) estimator given by:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

- The OLS estimator is unbiased as :

$$E(\hat{\beta}) = \beta$$

- The OLS estimator is the best unbiased as its variance is minimum and converges to 0 as $T \to \infty$:

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Parameter estimation

- The residuals are given by :

$$\hat{\varepsilon}_t = Y_t - X_t\hat{\beta}$$

- The residuals are related to the errors but are different:

$$\hat{\varepsilon} = (I - X(X'X)^{-1}X')\varepsilon$$

- An estimator for the variance of the errors is thus:

$$\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/(T - k)$$

- This estimator is unbiased as:

$$E(\hat{\sigma}^2) = \sigma^2$$

Parameter estimation

In addition to the 5 LR assumptions, if we assume that the errors are Normally distributed so that:

$$\varepsilon \sim \mathbb{N}(0, \sigma^2 I_T)$$

then the distribution of the OLS estimates is:

$$\sqrt{T}(\hat{\beta} - \beta) \sim \mathbb{N}(0, \sigma^2 (X'X/T)^{-1})$$

and

$$(T - k)\hat{\sigma}^2/\sigma^2 \sim \chi^2(T - k)$$

and $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

Similar results hold when $E(X\varepsilon) = 0$ and $T \to \infty$

Testing the parameters

- 2 types of tests: Parameters and Residuals.
- In a linear model, parameters β_i have to be different from 0, otherwise there is a specification issue (i.e.: X_{it} has no impact on Y_t).
- The Student test enables to test the null hypothesis: $H_0 : \beta_i = 0$ vs the alternative $H_1 : \beta_i \neq 0$
- There are 2 other possible alternative hypotheses: $H_1 : \beta_i > 0$ or $H_1 : \beta_i < 0$
- The Student t-stat is given by :

$$t = \frac{\hat{\beta}_i}{\sqrt{\hat{v}(\hat{\beta}_i)}}$$

where $\hat{v}(\hat{\beta}_i)$ is the i^{th} element on the diagonal of

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X/T)^{-1}$$

Testing the parameters

- It can be shown that under H_0, t follows a Student law with $T - k$ degrees of freedom, i.e. $t \sim t(T - k)$
- The rejection region depends on the chosen alternative hypothesis
- The decision rule boils to compare the t-stat realization t^* to the t_α the quantile of the Student law where α is the type I error
- Don't forget that α is chosen by the practitioner, thus the testing outcome can be different depending on the risk-appetite (risk-taker vs risk-adverse)

Testing the parameters

Alternatively, software report the probability, under the null hypothesis, that the absolute value of the test statistic t is larger than the realized value of the statistic, denoted t^*. This is the $p - value$ defined by :

$$p - value = \mathbb{P}_{H_0}(|t| > t^*)$$

or

$$p - value = \mathbb{P}_{H_0}(t > t^*) + \mathbb{P}_{H_0}(t < -t^*)$$

This $p - value$ can be directly compared to the type I error α: if $p - value > \alpha$ we accept the null H_0 with a type I risk α

Testing the parameters

Let's now consider the joint null hypothesis:

$$H_0 : \beta_1 = \ldots = \beta_k = 0$$

For testing H_0 we use the F-statistics defined as :

$$F = \frac{\hat{\beta}' X' X \hat{\beta}}{k\sigma^2} = \frac{\hat{\beta}' \hat{V}(\hat{\beta})^{-1} \hat{\beta}}{k}$$

It can be shown that when H_0 is true, $F \sim F(k, T-k)$, ie F follows a Fischer distribution.

If the $p-value$ associated to the test is such that $p-value > \alpha$, we accept the joint null H_0 with a type I risk α

Alternatively, there is a β_i such that $\beta_i \neq 0$

Testing the whiteness of residuals

Global significance of ACF using a "Portmanteau" statistics (Box-Pierce or Ljung-Box) given by:

$$Q_K = T(T+2)\sum_{k=1}^{K}\frac{\hat{\rho}_{\varepsilon}^{2}(k)}{T-k}. \tag{1}$$

Under the null hypothesis of non-correlation of the first K auto-correlations: $H_0 : \rho_\varepsilon(1) = \rho_\varepsilon(2) = \ldots = \rho_\varepsilon(K) = 0$, Q_K asymptotically follows a χ^2 distribution with $(K-p-q)$ degrees of freedom.

The null is rejected with a type I risk α if :

$$Q_K > \chi^2_{1-\alpha}(K-p-q)$$

or if: $p - value < \alpha$

The choice of the integer K has to be discussed.

Testing the Normality of residuals

Objectives

- Test if a given trajectory $(x_1, \ldots, x_T)$ comes from a **Gaussian** processes $(X_1, \ldots, X_T)$, based on higher moments of the distribution (3rd and 4th)
  Theoretical moments of order k:

$$m_k = \frac{E(X^k)}{E(X^2)^{k/2}}$$

Hypothesis

- H0: $(x_1, \ldots, x_T)$ comes from a Gaussian distribution $\mathbb{N}$
- H1: $(x_1, \ldots, x_T)$ doesn't come from a Gaussian distribution $\mathbb{N}$

Testing the Normality of residuals

Under H_0, we get the following asymptotical results:

$$\sqrt{T}\hat{m}_3 \xrightarrow{\mathscr{L}} \mathbb{N}(0,6) \quad (2)$$

$$\sqrt{T}\hat{m}_4 \xrightarrow{\mathscr{L}} \mathbb{N}(3,24) \quad (3)$$

$$\hat{m}_3 \text{ ind } \hat{m}_4 \quad (4)$$

The Jarque-Bera test statistics is:

$$JB = T\left(\frac{\hat{m}_3^2}{6} + \frac{(\hat{m}_4 - 3)^2}{24}\right)$$

Testing the Normality of residuals

Under H_0:

$$JB \sim \chi^2(2)$$

Usual quantiles are:

$$P(JB \geq 4,60) = 0,10$$

$$P(JB \geq 5,99) = 0,05$$

$$P(JB \geq 9,21) = 0,01$$

Exemples

If $JB^* > 9,21$, we reject H_0 with a type I risk $\alpha = 0,01$
If $JB^* \in [6;9]$, we reject H_0 with a type I risk $\alpha = 0,05$, but accept H_0 with a type I risk $\alpha = 0,01$

Measures of model fit

- A model that fits well in-sample, ie with small residuals, doesn't necessarily forecast well out-of-sample (*overfitting issue*)
- But a model with a poor fit generally leads to poor forecasts
- The coefficient of determination R^2 is a standard measure for the goodness-of-fit:

$$R^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'Y} = 1 - \frac{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\sum_{t=1}^{T} Y_t^2}$$

or

$$R^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y} = 1 - \frac{\sum_{t=1}^{T}(Y_t - \hat{\beta}X_t)^2}{\sum_{t=1}^{T} Y_t^2}$$

Thus: $0 \geq R^2 \geq 1$.

- It can be shown that : $R^2 = Corr(Y, \hat{Y})^2$

Measures of model fit

To account for the number of explanatory variables, the adjusted R^2 is defined as:

$$\bar{R}^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}/(T-k)}{Y'Y/(T-1)} = 1 - \frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2/(T-k)}{\sum_{t=1}^{T}Y_t^2/(T-1)}$$

Alternatively, information criteria can also be used to account for the number of explanatory variables:

$$\text{Akaike Information Criteria: } AIC = \log(\sigma^2) + 2k/T$$

$$\text{Schwartz Information Criteria: } SIC = \log(\sigma^2) + k\log(T)/T$$

$$\text{Hannan-Quinn Criteria: } HIC = \log(\sigma^2) + 2k\log\log(T)/T$$

Point forecasts

We are interested in forecasting the value Y_{T+h} based on the previously estimated model.

It can be shown that the best linear unbiased predictor of Y_{T+h} is given by:

$$\hat{Y}_T(h) = \hat{\beta} X_{T+h} \tag{5}$$

that is

$$\hat{Y}_T(h) = \sum_{i=1}^{k} \hat{\beta}_i X_{i,T+h}$$

in the sense of producing minimum forecast variance and zero

mean forecast error, where the forecast error is :

$$e_{t+h} = Y_{T+h} - \hat{Y}_T(h) \tag{6}$$

Point forecasts

The forecast error can be rewritten as:

$$e_{T+h} = (\beta - \hat{\beta})X_{T+h} + \varepsilon_{T+h} \tag{7}$$

The variance of the forecast error is thus given by:

$$V(e_{T+h}) = \sigma^2(1 + X_{T+h}(X'X)^{-1}X'_{T+h}) \tag{8}$$

Forecast uncertainty depends on:

1. on the variance of the error term σ^2
2. the variance of the estimated parameter $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$
3. the uncertainty around the forecasts of explanatory variables (generally unknown at $T + h$)

Criteria to assess forecast accuracy

- Mean Forecast Error:

$$MFE = E(e_{T+h})$$

- Mean Absolute Forecast Error:

$$MAFE = E(|e_{T+h}|)$$

- Mean Squared Forecast Error:

$$MSFE = E(e_{T+h}^2)$$

- Root Mean Squared Forecast Error:

$$RMSFE = \sqrt{E(e_{T+h}^2)}$$

Criteria to assess forecast accuracy

In practice, we compare over the horizon ex post observed values

$$(Y_{T+1}, \dots, Y_{T+h})$$

with forecasts

$$(\hat{Y}_T(1), \dots, \hat{Y}_T(h))$$

$$MSFE = \frac{1}{h} \sum_{k=1}^{h} (Y_{T+k} - \hat{Y}_{T+k})^2$$

Density and interval forecasts

Sometimes, forecasting the whole distribution may be of interest. Assume the LR1-LR5 are valid, T is large and the error term has a Normal distribution. In that case:

$$\left(\frac{Y_{T+h} - \hat{Y}_T(h)}{\sqrt{V(e_{T+h})}} \right) \sim \mathbb{N}(0,1)$$

implying

$$Y_{T+h} \sim \mathbb{N}(\hat{Y}_T(h), V(e_{T+h})) \tag{9}$$

Thus a $(1-\alpha)$ forecast interval is given by:

$$\left[\hat{Y}_T(h) - q_{1-\alpha/2}\sqrt{V(e_{T+h})}; \hat{Y}_T(h) + q_{1-\alpha/2}\sqrt{V(e_{T+h})} \right]$$

where $q_{1-\alpha/2} > 0$ is the $1-\alpha/2$ quantile of the Normal distribution.

Application 1: Forecasting Euro Area GDP

Objective:
Explain quarterly GDP growth (Y_t) by industrial production growth (ipr_t), stock prices growth (sr_t), European Sentiment Index growth (su_t) and inflation rate (pr_t).

The general model to estimate is the following:

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

Cf. Eviews exercise

Application 2: Taylor rule

Basic regression for central bank interest rates proposed by Taylor (1993):

$$r_t = \alpha + \beta g_t + \gamma(\pi_t - \pi*) + \varepsilon_t$$

where:
π_t is a measure of inflation (headline or core),
$\pi*$ is inflation targeting (assuming the CB is an inflation targeter)
g_t a measure of slack, usually output gap
ε_t is a white noise.

$\beta = 0.5$ in Taylor (1993), but $\beta = 1$ often used
$\gamma = 1.5$

Application 2: Taylor rule

In this framework, α can be seen as the neutral nominal interest rate.

Basic regression becomes:

$$r_t = (r_t^* + \pi^*) + \beta g_t + \gamma(\pi_t - \pi*) + \varepsilon_t$$

where: r_t^* is the real neutral interest rate that can be time-varying, as in Laubach and Williams (2003)

Application 2: Taylor rule

Extended regression for central bank interest by accounting for persistence in the interest rate (inertial Taylor rule):

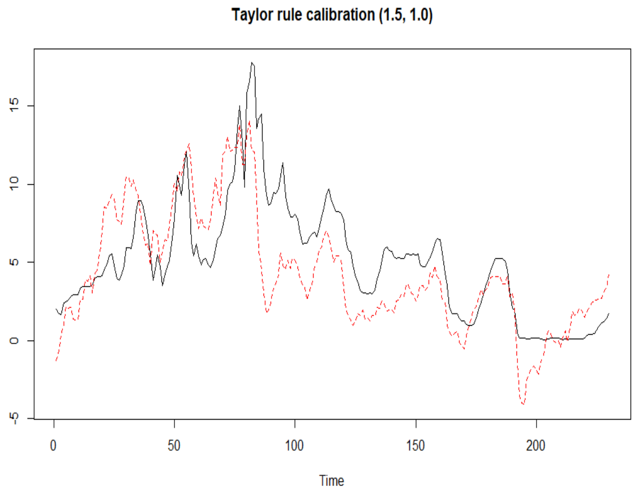
$$r_t = \rho r_{t-1} + \alpha + \beta g_t + \gamma(\pi_t - \pi*) + \varepsilon_t$$

where: ρ controls the persistence (generally estimated around 0.85).

Variant of the extended regression:

$$r_t = \rho r_{t-1} + (1-\rho)\{\alpha + \beta g_t + \gamma(\pi_t - \pi*)\} + \varepsilon_t$$

Calibration of the US Taylor rule



References

- Ghysels, E. and M. Marcellino (2018), *Applied Economic Forecasting using Time Series Methods*, Oxford University Press
- Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.