# Prediction

Econometrics. BAM.

# 1. Introduction

– The information that estimated models provide it can be used for many purposes.
– One of them, but not the unique, it is to provide information about the future values of the dependent variable.
– These predictions can be useful in order to adopt economic policies or to take decisions in order to get the objective of sales of a firm.
– A problematic point is whether the future values of the explanatory variables are known. We will assume that they are known, but we will probably have to predict them in practice.

# 2. Predictor

Let us assume that we have the following model:

$$y_t = x_t{}' \beta + u_t,\ t = 1, 2, ..., T$$

Where all the assumptions of the GLM hold.

Now, we are interesting in predicting the values of the dependent variable for a prediction horizon of m periods. This new model can be stated as follows:

$$y_{T+\ell} = x_{T+\ell}{}' \beta + u_{T+\ell},\ \ell = 1, 2, ..., m$$

# 2. Predictor

Then, the best (minimum variance) linear predictor is defined as follows:

$$\hat{y}_T(\ell) = E\left(y_{T+\ell} / \Im_T\right) =$$

$$= E\left[\left(x_t' \beta + u_t\right) / \Im_T\right] =$$

$$= x_t' \beta$$

# 2. Predictor

We should substitute the vector of parameters for a vector of consistent estimators, as the OLS are. Then, we have that the predictor is defined:

$$\hat{y}_T\left(\ell\right) = x_t{}'\,\hat{\beta}$$

# 2. Predictor

The forecast error is defined as follows:

$$\hat{u}_T(\ell) = y_{T+\ell} - \hat{y}_T(\ell) =$$

$$= x_{T+\ell}{}' \beta + u_{T+\ell} - x_{T+\ell}{}' \hat{\beta} =$$

$$= x_{T+\ell}{}' \left( \beta - \hat{\beta} \right) + u_{T+\ell}$$

# 2. Predictor

Then, the variance of the forecast error is defined as follows:

$$Var\left[\hat{u}_T(\ell)\right] = Var\left[x_{T+\ell}'\left(\beta - \hat{\beta}\right) + u_{T+\ell}\right] =$$

$$= x_{T+\ell}' Var\left(\hat{\beta}\right) x_{T+\ell} + Var\left(u_{T+\ell}\right) + 2Cov\left(\hat{\beta}, u_{T+\ell}\right) =$$

$$= \sigma^2 x_{T+\ell}' \left(X'X\right)^{-1} x_{T+\ell} + \sigma^2 =$$

$$= \sigma^2 \left[1 + x_{T+\ell}' \left(X'X\right)^{-1} x_{T+\ell}\right]$$

# 2. Predictor

Again, we can substitute s2 by a consistent estimation and, consequently, we have that:

$$V\hat{a}r\left[\hat{u}_T(\ell)\right] = \hat{\sigma}^2\left[1 + x_{T+\ell}'\left(X'X\right)^{-1}x_{T+\ell}\right]$$

# 2. Predictor

This allows us to obtain the interval prediction

$$\hat{y}_{T+\ell} \pm t_{T-k}^{\alpha/2} \sqrt{\hat{\sigma}^2 \left[ 1 + x_{T+\ell}{}' \left( X'X \right)^{-1} x_{T+\ell} \right]}$$

# 2. Predictor

Similarly, we can test for structural permanence by using the following statistic:

$$t_{PE} = \frac{y_{T+\ell} - \hat{y}_{T+\ell}}{\sqrt{\hat{\sigma}^2 \left[ 1 + x_{T+\ell}{}' \left( X ' X \right)^{-1} x_{T+\ell} \right]}} \sim t_{T-k}$$

$H_o$: (post-sample) structural permanence

$H_A$: (post-sample) structural change

# 3. Goodness of prediction fit

1. Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{m}\sum_{\ell=1}^{m}\left[y_{T+\ell} - \hat{y}_T(\ell)\right]^2}$$

# 3. Goodness of prediction fit

2. Mean Absolute Error

$$MAE = \frac{1}{m} \sum_{\ell=1}^{m} \left| y_{T+i} - \hat{y}_T(\ell) \right|$$

# 3. Goodness of prediction fit

3. Theil's U coefficient

$$U = \sqrt{\frac{\sum_{\ell=1}^{m}\left[y_{T+\ell} - \hat{y}_T(\ell)\right]^2}{\sum_{\ell=1}^{m} y_{T+\ell}^2}}$$

This coefficient is related to the R2 but it is not bounded by 0 and 1. Large values indicate a poor forecasting performance