FORECASTING FOR BUSINESS & ECONOMICS EBC2089

Prof. dr. Alain Hecq
A3.14
a.hecq@maastrichtuniversity.nl

February 2023

Aims

- **Prediction** in cross sectional data:
  - OLS estimator and fitted values for some values of the regressors.
  - Application to house evaluation, crime rate
- **Forecasting** in time series:
  - Estimation of a model, an equation (OLS, ARMA models,...) in data with many features such as trends, seasonality, cycles, ...
  - On your data.

Econometrics

- Econometrics is largely concerned with quantifying the relationship between one or several response variable (dependent variable) y_t and a one or several explanatory variables x_t, the regressors.
- For instance (cross section)

$$wage_i = \alpha + \beta.educ_i + e_i, \; i = 1...N$$

- Might be used to propose a salary (prediction of wage given the level of education, other variables like skills are important)
- New challenges with big data

Time Series

- In time series, the explanatory variables will be some
  - deterministic functions of the time (trend, seasonality,..),
  - the past of the series: dynamics/memory
  - as well as additional potential regressors and their past.
- For instance

$$prices_t = \alpha + \beta.trend + \rho.prices_{t-1} + e_t, \ t = 1...T$$

Brush-up basic econometric notations

- Consider the population equation (i.e. theoretical behavior of the process), with n individuals ($i = 1...n$), linking a dependent variable y to an explanatory variable x, an intercept (constant term) and the disturbance (error term) ε

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- For instance for a time series sample

$$consumption_t = \beta_0 + \beta_1 income_t + \varepsilon_t$$

- Those population coefficients are unknown $=>$ must be estimated from a sample (data).

Brush-up basic econometrics: numerical properties

- OLS (ordinary least squares, i.e. the estimator that minimizes the sum of squared residuals) in the bivariate case

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

- Or in the general multivariate case, instead of writing

$$y_i = \beta_0 + \beta_1 x_{1i}.. + \beta_k x_{ki} + \varepsilon_i$$

we use

$$y_i = X_i' \beta + \varepsilon_i$$

with $X_i' = (1 : x_{1i} : ... x_{ki})$

Brush-up basic econometrics: numerical properties

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\hat{\beta} = (\sum x_i x_i^{'})^{-1} \sum x_i y_i$$

or $\hat{\beta} = (X'X)^{-1} X'y$ with $X = (X_1', X_2', ... X_n')'$

Brush-up basic econometrics: statistical properties

- OLS estimator $\hat{\beta}$ is the best linear unbiased estimator (BLUES), i.e.
  - unbiased
  - and there does not exist another linear estimator with a smaller variance.
- This is true
  - under a set of assumptions about the error term ε
  - under the hypothesis about the correlation between the explanatory variables and the disturbance term.

More formally those important assumptions

- Strict exogeneity:

$$E(\varepsilon_i|X) = 0, \qquad \forall i \qquad\qquad (E(\varepsilon_i|X_1, \ldots X_i, ...X_n) = 0 \qquad \forall i$$

- That means that conditional on all observations (current, future and past, other individuals) the mean of $\varepsilon_i = 0$ and consequently no observations on X convey information about the expected value of the disturbance.
- Usually not met in time series because X contains the lagged dependent variables $=>$ consistency principle (in large sample) will replace unbiased (in small sample): a necessary "second best "
- In cross section, this assumes no omitted variables, no endogeneity, etc $=>$ no consistency here if not met.

Important assumptions

- Spherical disturbances: **homoscedasticity** and **no autocorrelation**
  - Constant variance: homoscedasticity (mainly for cross-section or in financial econometrics, volatility)

$$Var(\varepsilon_i|X) = \sigma^2, \qquad \forall i = 1 \ldots n$$

  - No autocorrelation (for time series)

$$Cov(\varepsilon_i, \varepsilon_j|X) = 0, \qquad \forall i \neq j$$

What is it? (example:

$$\begin{aligned} consumption_t &= \alpha + \beta.income_t + \varepsilon_t \\ \varepsilon_t &= \rho\varepsilon_{t-1} + v_t \quad \text{for } \rho \neq 0 \end{aligned}$$

Important assumptions

- Rank condition: the number of observations is larger than the number of regressors, otherwise cannot apply OLS.
- i.e. in a regression with k explanatory variables and n observation

$$y_i = \alpha_0 + \alpha_1 x_{1i} ... + \alpha_k x_{ki} + \varepsilon_i, \quad i = 1...n$$

- we must have

$$n > k$$

- Big data issues/solutions when k is too large compared to n:
  - **Dense models** such as factor models (PCA), imposes restrictions by e.g. combining regressors.
  - **Sparse models** such as Lasso, delete non relevant variables.

General consequences if no spherical disturbances

- Invalid inference: we cannot use t-tests (and F-tests) anymore but coefficients "still valid" (unbiased or consistent).
- Indeed if

$$Var(\varepsilon|X) = \sigma^2\Omega \text{ instead of } Var(\varepsilon|X) = \sigma^2 I$$

- OLS is still

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

- But the variance (and hence s.e. in the second column of EViews) of the parameters is not anymore

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

- But a larger one

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

General consequences if no spherical disturbances

- How to know that there are some dynamics and/or heretoskedasticity left: **test statistics, graphs**.
- What to do in time series if misspecification pops up: my preference in time series (not easy to improve model in cross section) goes to considering a dynamic model ($\neq$ from basic econometric textbook where correction is used)

$$\begin{aligned} consumption_t &= \alpha + \beta.income_t + \varepsilon_t \\ shift \text{ in } t-1 &=> \\ consumption_{t-1} &= \alpha + \beta.income_{t-1} + \varepsilon_{t-1} \end{aligned}$$

$$\begin{aligned} consumption_t &= \alpha + \beta.income_t + \rho\varepsilon_{t-1} + v_t \\ &= \alpha + \beta.income_t + \rho consumption_{t-1} \\ &\quad -\rho\alpha - \rho\beta income_{t-1} + v_t \end{aligned}$$

Alternative strategies

In cases the model cannot (or you don't want) be improved because forecasting is not the issue for instance:

- Strategy I: GLS/FGLS estimators
  - Weighted LS (for heteroskedasticity)
  - Cochrane Orcutt (quasi differences) for autocorrelation
- Strategy II: robust estimators
  - White robust standard errors (HCSE) if heteroskedasticity only
  - Newey-West robust standard errors (HAC) if heteroskedasticity and autocorrelation

Robust standard errors White HCSE estimator

- Use OLS because OLS is unbiased and consistent even when there is heteroskedasticity.
- Save residual sequence e_i and compute robustify standard errors

$$V^{HCSE}(\hat{\beta}|X) = \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 X_i X_i'\right)\left(\frac{X'X}{n}\right)^{-1}$$

- Instead of

$$V(\hat{\beta}|X) = \left(\frac{1}{n-k}\sum_{i=1}^{n} e_i^2\right)\left(X'X\right)^{-1} = s^2\left(X'X\right)^{-1}$$

  See option in EViews

- VERY IMPORTANT RESULT: without knowing the form of Ω we have a consistent estimator for $V(b)$ (ROBUST ESTIMATOR, **HCSE**).

Hypothesis testing

- Inference: using Student's t-test, i.e. if the null hypothesis is $H_0 : \beta = 0$

$$t - test = \frac{\hat{\beta}}{se_{(\hat{\beta})}}$$

  (this is what you have in the output of statistical packages)

- If another hypothesis is considered, for instance $H_0 : \beta = 1$,

$$t - test = \frac{(\hat{\beta} - 1)}{se_{(\hat{\beta})}}$$

  (this can easily be done using some simple calculus)

- Also using robustified HCSE $se_{(\hat{\beta})}$ when heteroskedasticity.

Hypothesis testing

- Standard error for a bth variable

$$se_{(\hat{\beta})} = s(X'X)^{-1}_{b\_th} = \sqrt{(\frac{1}{n-k}\sum e_t^2)(\sum x_i x_i')^{-1}_{b\_th}}$$

- You should compare with critical values on correct tables depending on your significance level (5%: critical value$=\pm$1.96 for a two sided test) and the type of hypothesis ($t-test$ vs. $F-test$, one versus two sided tests).
- $F-test$

$$\frac{(RSSR-USSR)/df_1}{USSR/(T-k)} \sim F(df_1, T-k)$$

- Other important concepts:$p-value$, confidence interval, R-squared (adjusted)....

Goodness of fit

- R-squared (WITH THE SAME DEPENDENT VARIABLE)

$$R^2 = 1 - \frac{\Sigma_{t=1}^{T} e_t^2}{\Sigma_{t=1}^{T} (y_t - \bar{y}_t)^2}$$

- Adjusted R-squared (better for comparing models WITH THE SAME DEPENDENT VARIABLE)

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-k} \Sigma_{t=1}^{T} e_t^2}{\frac{1}{T-1} \Sigma_{t=1}^{T} (y_t - \bar{y}_t)^2}$$

Information criteria

- Minimize the sum of squared residuals instead of maximizing the goodness of fit.
- However: Variance of the residuals doesn't enough take into account the number of explanatory variables k

$$s^2 = \frac{1}{T-k} \sum_{t=1}^{T} e_t^2$$

- Information criteria are better. They are also independent from the LINEAR transformation
- => i.e. models in level and in first difference have the same ICs but $\neq R^2$)

Information Criteria to be minimized

Akaike information criterion: (AIC)

$$AIC = \ln s^2 + \frac{2}{T}k$$

Schwarz criterion: (SC of SBC or BIC)

$$SC = \ln s^2 + \frac{\ln T}{T}k$$

HQ

$$HQ = \ln s^2 + \frac{\ln \ln T}{T}2k$$

Many others (with a different penalty). AIC overestimate, SC and HQ are consistent for the CORRECT model.

Mispecification 1: Normality Test

- Jarque-Bera is a test statistic for testing whether the series is normally distributed.
- The test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution, respectively 0 and 3.
- The statistic is computed as:

$$BJ = \frac{T-k}{6}\left(\widehat{SK}^2 + \frac{1}{4}\widehat{EK}^2\right) \sim \chi^2_{(2)}$$

  where k represents the number of estimated coefficients used to create the series.
- Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as with 2 degrees of freedom.
- The reported Probability in EViews is the probability that a Jarque-Bera statistic exceeds (in absolute value) the observed value under the null. A small probability value leads to the rejection of the null hypothesis of a normal distribution.

Mispecification 2: Heteroskedasticity and White's General Heteroskedasticity Test

- General because it doesn't precise the form of heteroskedasticity $\neq$ previously developed tests

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \varepsilon_i$$

- Heteroskedasticity if $\sigma_i^2 = V(\varepsilon_i)$ is not constant
- White's test looks at dependence between σ_i^2 and
  - (1) the explanatory variables,
  - (2) their squares and
  - (3) cross-product.

Mispecification 2: Heteroskedasticity and White's General Heteroskedasticity Test

- Question of the test: Can the variance be explained by those explanatory variables?
- If yes, the variance is not constant $=>$ heteroskedasticity.
- Steps
  - Estimate $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \varepsilon_i$ by OLS and store the (unbiased) residuals e_i
  - Approximate σ_i^2 by e_i^2 because OLS is still unbiased in the presence of heteroskedasticity.

White's General Heteroskedasticity Test

- Regress

$$
\begin{aligned}
e_i^2 &= \alpha_1 + \alpha_2 x_{1i} + \alpha_3 x_{2i} \\
&\quad + \alpha_4 x_{1i}^2 + \alpha_5 x_{2i}^2 \\
&\quad + \alpha_6 x_{1i} x_{2i} + u_i
\end{aligned}
$$

- No heteroskedasticity if

$$H_0 : \alpha_2 = \ldots = \alpha_6 = 0$$

$\Rightarrow$ use $F-$test reported in EViews

Mispecification 3: No Linearity (RESET test)

- We test the null of linearity against an unknown presence of non linear features
- If we reject we have to improve the model but we do not know exactly how (breaks, quadratic,...)
- Steps:
    - Run your model.
    - Take the fitted value.
    - Run an auxiliary model in which you add the fitted value (squared or cube) as an additional variable.