Lecture 0 - Introduction to the Course

Pedro Valls EESP-FGV & CEQEF-FGV

July 5, 2024

FGV EESP

CENTRO DE ESTUDOS QUANTITATIVOS EM ECONOMIA E FINANÇAS

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts
- Interval density forecasts

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts
- Interval density forecasts
- Goodness of measure of forecast

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts
- Interval density forecasts
- Goodness of measure of forecast
- Parameter testing

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts
- Interval density forecasts
- Goodness of measure of forecast
- Parameter testing
- Variable Selection

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts
- Interval density forecasts
- Goodness of measure of forecast
- Parameter testing
- Variable Selection
- Multicollinearity

Outline

- Syllabus of the course Special Topics in Time Series Econometrics - Forecasting
- Overview of the course
- Applied paper and Seminar
- Referee Report
- Basic Specification of the Linear Regression Model
- Parameter Estimation
- Example using simulated data - EVIEWS or R commands
- Measures of model fit
- Constructing point forecasts
- Interval density forecasts
- Goodness of measure of forecast
- Parameter testing
- Variable Selection
- Multicollinearity
- Empirical Examples

Syllabus of Forecasting

Forecasting

Applied Paper and Seminar

- Any topic in Macro or Finance
- Need to use some topic presented in the course
- It is not allowed to use techniques not presented in the course.
- You have to submit a document similar to an article.
- The document must be printed and presented at a seminar in the last lecture.

Referee Report

- The report should include a summary of the paper's content and place in the relevant literature, a critical assessment of its contribution, and suggestions for improvement.
- The indicative length of the referee report is 1500 words (i.e., approximately 6 pages).
- Finally, students will have to carry out an empirical application (or simulations) intimately connected to the selected research paper.
- Examples NBER ; RBFin

Basic Specification of the Linear Regression Model I

- From the economics point of view
  - we want to model inflation considering cost push factors (unit labour costs and prices of intermediate inputs and energy) and inflation expectations.
- From statistical point of view
  - $Z = (y, X)$ is a gaussian stationary stochastic processes serially uncorrelated (Some of these hypotheses will be relaxed later);
  - we have a set of realizations $Z_t = (y_t, X_t)$ for $t = 1, \ldots, T$;
  - $$\begin{aligned} D(Z_1, \ldots, Z_T \mid \Theta) &= \prod_{t-1}^{T} D(Z_t \mid \Theta) \\ &= \prod_{t-1}^{T} D(y_t, X_t \mid \Theta) \\ &= \prod_{t-1}^{T} D(y_t \mid X_t, \Theta_1) D(X_t, \mid \Theta_2) \qquad (1) \end{aligned}$$

Basic Specification of the Linear Regression Model II

- (1) is the joint density of the data written as the product of the conditional density of y_t given X_t and the marginal density of X_t.
- If we want to recover the parameters of the joint distribuition Θ, we have to estimate both densities, i.e., th conditional and the marginal.
- If the parameters of interest are only the parameters of the conditional density, i.e., Θ_1, and if Θ_1 and Θ_2 are variation free (see [Engle et al., 1983] or [Valls Pereira, 2024][1]), , there is no lost in working only with the conditional model (this means that X_t is weakly exogenous for Θ_1 .
- Since the distribution of Z_t is gaussian, we have that the explanatory variables X have a linear impact on the dependent variable y we can write the model as:

$$y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + \cdots + X_{kt}\beta_k + \epsilon_t \quad t = 1, \cdots, T \tag{2}$$

[1]This notes can be downloaded here ADL Model

Basic Specification of the Linear Regression Model

- Model (2) implies that
  - $E(y_t|X_{1t}, X_{2t}, \cdots, X_{kt}) = X_{1t}\beta_1 + X_{2t}\beta_2 + \cdots + X_{kt}\beta_k$ - measure the change in the expected value of y_t when there is a marginal change is X_{it} and the other X_t are kept constant
  - $\epsilon_t = y_t - E(y_t|X_{1t}, X_{2t}, \cdots, X_{kt})$ is the non systematic part and by construction is a white noise

Basic Specification of the Linear Regression Model I

- In matrix form

$$\underset{(T\times 1)}{\boldsymbol{y}} = \underset{(T\times k)}{\boldsymbol{X}}\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\epsilon}} \tag{3}$$

- with the following assumptions
  - LR1: $E(\epsilon_t) = 0$
  - LR2: $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 \boldsymbol{I}_T$
  - LR3: $\boldsymbol{X}$ is distributed independently of $\boldsymbol{\epsilon}$
  - LR4: $\boldsymbol{X}'\boldsymbol{X}$ is non singular
  - LR5: X is weakly stationary
- the model is correctly specified - no omitted variables
- model is linear in variables and parameters
- the parameters, β and σ_ϵ^2 are time invariants.

- Can not use (3) for forecasting since the parameter are unknown
- Several estimators for the parameters but if for the moment X is assumed to be deterministic from Gauss Markov theorem the Best Linear Unbiased Estimator (BLUE) for β is

$$\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{4}$$

- it is called ordinary least squared (OLS) estimator
- Can be derived as the minimised with respect to β of $\boldsymbol{\epsilon}\boldsymbol{\epsilon}' = \sum_{t=1}^{T} \epsilon_t^2 = \sum_{t=1}^{T} (y_t - \boldsymbol{X}_t\boldsymbol{\beta})$ where $\boldsymbol{X}_t = (X_{1t}, X_{2t}, \cdots, X_{kt})$
- The OLS estimator is unbiased

$$E(\widehat{\beta}) = \beta \tag{5}$$

Parameter estimation II

- $\widehat{\beta}$ has a minimal variance in the class of estimators that are linear and unbiased and has variance given by:

$$Var(\widehat{\beta}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{6}$$

- Can rewrite (6) as:

$$Var(\widehat{\beta}) = \frac{\sigma^2}{T}\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{T}\right)^{-1} \tag{7}$$

- Since

$$\frac{\boldsymbol{X}'\boldsymbol{X}}{T} \underset{T\to\infty}{\longrightarrow} Q_X \text{ positive definite matrix} \tag{8}$$

- and

$$\frac{\sigma^2}{T} \underset{T\to\infty}{\longrightarrow} 0 \tag{9}$$

Parameter estimation III

- using (8) and (9) into (7) we have

$$\lim_{T\to\infty} Var(\widehat{\beta}) = 0 \tag{10}$$

- By (10) and (5) we have that the OLS estimator of β is consistent.
- The residuals are given by:

$$\widehat{\epsilon}_t = y_t - X_t\widehat{\beta} \tag{11}$$

- and in matrix form

$$\begin{aligned} \widehat{\boldsymbol{\epsilon}} &= \boldsymbol{y} - \boldsymbol{X}\widehat{\beta} \\ &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\epsilon \end{aligned} \tag{12}$$

Parameter estimation IV

- can use the residuals to construct an estimator of the error variance σ^2

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}}{T-k} \tag{13}$$

- It can be proven that $\widehat{\sigma}^2$ is an unbiased estimator of σ^2
- By the normality of the data it follows that the errors are also normally distributed, that is

$$\boldsymbol{\epsilon} \sim NI(\mathbf{0}, \sigma^2 I_T) \tag{14}$$

- then the distribution of the OLS estimator is

$$\sqrt{T}(\widehat{\beta} - \beta) \sim \boldsymbol{N}\left(\mathbf{0}, \sigma^2 \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{T}\right)^{-1}\right) \tag{15}$$

Parameter estimation V

- and
$$\frac{\widehat{\sigma}^2(T-k)}{\sigma^2} \sim \aleph^2(T-k) \tag{16}$$
- $\widehat{\beta}$ and $\widehat{\sigma}^2$ are independent

Example using simulated data I

- Consider the following DGP

$$y_t = 1 + D_t + D_t x_t + x_t + 0.5x_{t-1} + 0.5y_{t-1} + \epsilon_t \qquad (17)$$

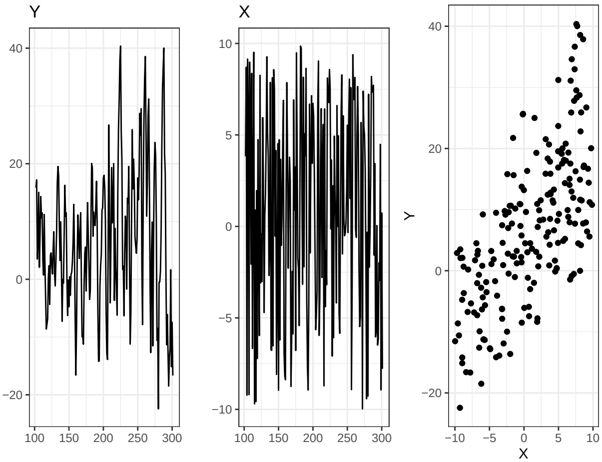
  - where $y_t = \ln(Y_t)$ is the dependent variable,
  - x_t is the independent variable,
  - lags of $\ln(Y_t)$ and x_t appear also as explanatory variables,
  - D_t is a dummy variable equal to 1 for the observations from 202 to 501 that determines a change in the intercept and in the coefficient of x_t
  - $\epsilon_t \sim NI(0,1)$
  - $t = 1, \cdots, 501$ but we discard the first 101 observations and
  - Estimation sample: observations $102 - 301$
  - Forecasting sample: observations $302 - 501$

Example using simulated data - R commands I

simulated data

Example using simulated data - EVIEWS commands

Figure: Y, X and Scatted Plot of X and Y



Measures of model fit I

- A common measure of model fit is the coefficient of determination R^2, that compares the variability in the dependent variable y with that in the model (OLS) residual $\widehat{\epsilon}$

$$\begin{aligned} R^2 &= 1 - \frac{\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}}{\boldsymbol{y}'\boldsymbol{y}} = \frac{\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}}}{\boldsymbol{y}'\boldsymbol{y}} \\ &= 1 - \frac{\sum_{t=1}^{T} \widehat{\epsilon}_t^2}{\sum_{t=1}^{T} y_t^2} = \frac{\sum_{t=1}^{T} (y_t - X_t\widehat{\beta})^2}{\sum_{t=1}^{T} y_t^2} \end{aligned} \tag{18}$$

- where

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{19}$$

  - $0 \leq R^2 \leq 1$
  - $R^2 = 0$ when $\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = \boldsymbol{y}'\boldsymbol{y}$
  - $R^2 = 1$ if the fit is perfect
  - $R^2 = Corr(y, \widehat{y})^2$
  - R^2 is monotonically increasing

Measures of model fit II

- Modified version of R^2 is the adjusted R^2 defined as

$$\overline{R}^2 \quad = \quad 1 - \frac{\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}/(T-k)}{\mathbf{y}'\mathbf{y}/(T-1)} \quad = \quad 1 - \frac{\frac{\sum_{t=1}^{T}\widehat{\epsilon}_t^2}{T-k}}{\frac{\sum_{t=1}^{T} y_t^2}{T-1}} \tag{20}$$

- Information criteria provide an alternative method for model evaluation
- They combine model fit with a penalty related to the number of model parameters

Example using simulated data - R commands

- `# correlation matrix between y and x`
- `cor(sim.data$est$x,sim.data$est$y)`
- 0.6721621
- `# ordinary least square analysis`
- `ols.fit <- lm(y` $\sim x, data = sim.data\$est$`)`
- print(xtable(ols.fit), floating = F) # LaTeX output

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.5053	0.6616	8.32	0.0000
x	1.4781	0.1157	12.77	0.0000
$R^2 = 0.4518$		$\overline{R}^2 = 0.449$		$AIC = 7.3034$
$\widehat{\sigma} = 9.233$				$BIC = 7.3530$
$F(1,198) = 16.2$		$p - value = 0.000$		

Constructing point forecasts I

- Forecasting y_{T+h} using the linear model (3) with parameter estimated as (4) assuming the future value of the regressors, i.e. X_{T+h} are known
- The best linear unbiased forecast of y_{T+h} is given by

$$\widehat{y}_{T+h} = \underset{1\times k}{X_{T+h}} \underset{k\times 1}{\widehat{\beta}} \tag{21}$$

- Forecast error given by

$$e_{T+h} = y_{T+h} - \widehat{y}_{T+h} \tag{22}$$

- Linear prediction of y is given by

$$y_{T+h} - l'y = (X_{T+h} - l'X)\beta + \epsilon_{T+h} - l'\epsilon \tag{23}$$

Constructing point forecasts II

- Taking expectations

$$E(y_{T+h} - l'y) = (X_{T+h} - l'X)\beta \tag{24}$$

- and the forecast is unbiased if $l'X = X_{T+h}$
- the variance of the forecast error is given by

$$E(y_{T+h} - l'y)^2 = E(\epsilon_{T+h} - l'\epsilon)^2 = \sigma^2(1 + l'l)$$

- We want to solve the following problem

$$\begin{array}{cc} \min\limits_{l} & \sigma^2(1 + l'l) \\ \text{subject} & l'X = X_{T+h} \end{array} \Leftrightarrow \begin{array}{cc} \min\limits_{l} & l'l \\ \text{subject} & l'X = X_{T+h} \end{array} \tag{25}$$

Constructing point forecasts III

- Construct the Lagrangian

$$= l'l - \underset{1\times k}{\lambda'}(X'l - \underset{k\times 1}{X_{T+h}}) \quad (26)$$

- First order conditions

$$\begin{aligned} \underset{(T\times 1)}{\partial/\partial l} &= 2l - X\lambda \\ \underset{(k\times 1)}{\partial/\partial \lambda} &= X'l - X'_{T+h} \end{aligned}$$

- In matrix form

$$\begin{bmatrix} \underset{T\times T}{2I} & \underset{T\times k}{-X} \\ \underset{k\times T}{X'} & \underset{k\times k}{0} \end{bmatrix} \begin{bmatrix} \underset{T\times k}{l} \\ \underset{k\times 1}{\lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ X'_{T+h} \end{bmatrix}$$

Constructing point forecasts IV

- The optimal solution is

$$\left[\begin{array}{c} l^* \\ \lambda^* \end{array}\right] = \left[\begin{array}{cc} 1/2\ I - 1/2\ X(X'X)^{-1}X' & X(X'X)^{-1} \\ -(X'X)^{-1}X' & 2(X'X)^{-1} \end{array}\right] \left[\begin{array}{c} 0 \\ X'_{T+h} \end{array}\right]$$

- and the optimal forecast is

$$\widehat{y}_{T+h} = l^{*\prime} y = X_{T+h}(X'X)^{-1}X'y = X_{T+h}\widehat{\beta}$$

- The associated forecast error is

$$e_{T+h} = X_{T+h}(\beta - \widehat{\beta}) + \epsilon_{T+h} \tag{27}$$

- and its variance is given by

$$\begin{aligned} E(y_{T+h} - \widehat{y}_{T+h})^2 &= E(X_{T+h}(\beta - \widehat{\beta}) + \epsilon_{T+h})^2 \\ &= \sigma^2 \left(1 + \underset{1\times k}{X_{T+h}} \underset{k\times k}{(X'X)^{-1}} \underset{k\times 1}{X'_{T+h}}\right) \end{aligned} \tag{28}$$

Interval density forecasts

- Assumptions LR1-LR5 are valid
- The sample size T is large
- The error has a normal distribution
- From the definition of the forecast error $e_{T+h} = y_{T+h} - \widehat{y}_{T+h}$ it follows

$$\left(\frac{y_{T+h} - \widehat{y}_{T+h}}{\sqrt{Var(e_{T+h})}}\right) \sim N(0,1)$$

- implies

$$y_{T+h} \sim N(\widehat{y}_{T+h}, Var(e_{T+h}))$$

- If using parameter estimators for β and σ^2 and the sample size is finite then the standardized forecast error has a Student t distribution with $T-k$ degrees of freedom.
- Forecast density can be used to construct interval forecasts for y_{T+h}

$$[\widehat{y}_{T+h} - c_{\alpha/2}\sqrt{Var(e_{T+h})} : \widehat{y}_{T+h} + c_{\alpha/2}\sqrt{Var(e_{T+h})}] \tag{29}$$

- where $c_{\alpha/2}$ is the $(\alpha/2)\%$ critical value for the standard normal

Goodness of measure of forecast

- Two goodness of measure of forecast

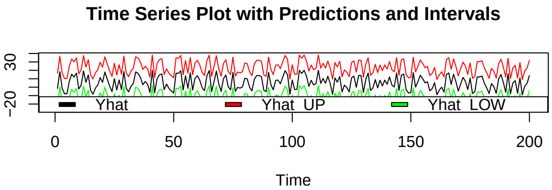
$$RMSE = \sqrt{\frac{1}{h}\sum_{j=1}^{h}(y_{T+j} - \widehat{y}_{T+j})^2}$$

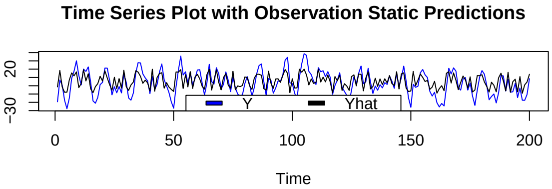
$$MAE = \frac{1}{h}\sum_{j=1}^{h}|y_{T+j} - \widehat{y}_{T+j}| \tag{30}$$

Example using simulated data - R commands I

simulated data

Example using simulated data - R commands - static forecasting

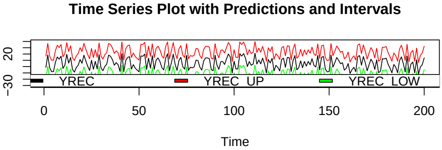


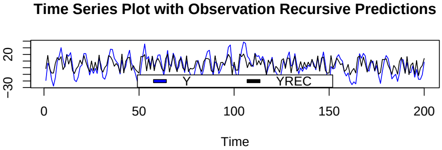


Example using simulated data - R commands - dynamic forecasts I

simulated data

Example using simulated data - R commands - dynamic forecasts





Example using simulated data - R commands - Goodness of measure of forecast I

simulated data

Example using simulated data - EVIEWS commands - Goodness of measure of forecasts

	Static	Recursive
RMSE	9.835744	**9.641340**
MAE	8.141826	**8.002104**

Parameter testing I

- Consider the linear regression model

$$y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + \cdots + X_{kt}\beta_k + \epsilon_t \quad t = 1, \cdots, T \tag{31}$$

- in matrix form

$$\boldsymbol{y} = \boldsymbol{X}\ \boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{32}$$

- assume Assumptions LR1-LR5 and normality of the errors
- We want to test

$$H_0 : \beta_i = c \ (\text{ or } \boldsymbol{\beta} = \boldsymbol{c})$$

- against

$$H_1 : \beta_i \neq c \ (\text{ or } H_1 : \text{ at least one } \beta_i \neq \boldsymbol{c}\ )$$

Parameter testing II

- Test statistic is given by

$$t - stat = \frac{\widehat{\beta}_i - c}{\sqrt{\widehat{Var}(\widehat{\beta}_i)}} \overset{H_0}{\sim} t(T-k) \tag{33}$$

- where $\widehat{Var}(\widehat{\beta}_i)$ is the $i-th$ element on the diagonal of

$$\widehat{Var}(\widehat{\beta}) = \widehat{\sigma}^2 \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{T} \right)^{-1}$$

- When the null is

$$H_0 : \beta_1 = c_1, \cdots, \beta_k = c_k \tag{34}$$

- with the alternative

$$H_1 : \text{ at least one } \beta_i \neq c_i \tag{35}$$

- The test statistic in this case is

$$F - stat = \frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{c})' \boldsymbol{X}' \boldsymbol{X} (\widehat{\boldsymbol{\beta}} - \boldsymbol{c})}{k\widehat{\sigma}^2} = \frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{c})' \widehat{Var}(\widehat{\beta})^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{c})}{k\widehat{\sigma}^2} \overset{H_0}{\sim} F(k, T-k) \tag{36}$$

Variable Selection I

- GET methodology starting from a general statistical model, exploits standard testing procedures to reduce its complexity by eliminating statistically insignificant variables and checking that the resulting model satisfies the underlying assumptions by means of diagnostic test - [Hoover and Perez, 1999] and [Hendry and Krolzig, 2005].
- The software Autometric, [Doornik, 2009] permits a practical implementation and the forecasts from the resulting models seem to be rather good
- Alternative can be classified into hard-and-soft-thresholding rules - [Bai and Ng, 2008].
- Under the hard-thresholding a regressor is selected according to the significance of its correlation with the target
  - shortcoming is that it only takes into account the bivariate relationship between the target variable and each regressor
  - tends to select highly collinear targeted predictors

Variable Selection II

- problems of multicollinearity

- Under the soft-thresholding the selection is basis on the following minimization problem:

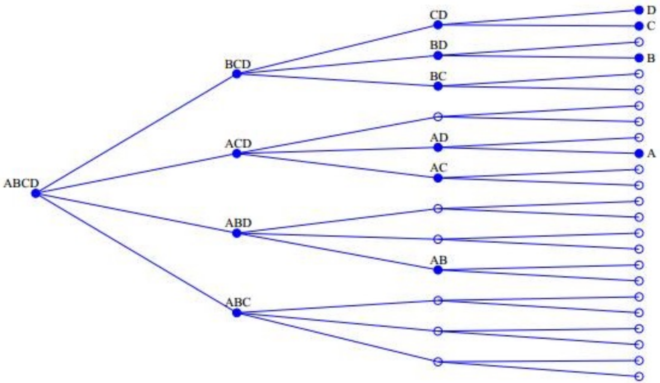
$$\min_{\beta} \Phi(RSS) + \lambda \Psi(\beta_1, \cdots, \beta_j, \cdots, \beta_N)$$

- where RSS indicates the Residual Sum of Squared
- The parameter λ governs the shrinkage.

Autometrics I

- **Dummy Saturation**

1. [Doornik, 2009]: *Autometrics*



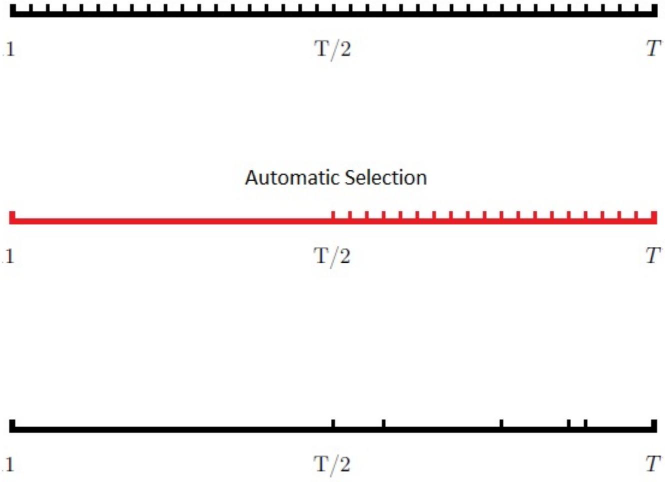
Autometrics II

2. The general idea of *Autometrics* is to estimate all the nodes from the above Figure, making joint significance and specification tests, and assessing information criteria.

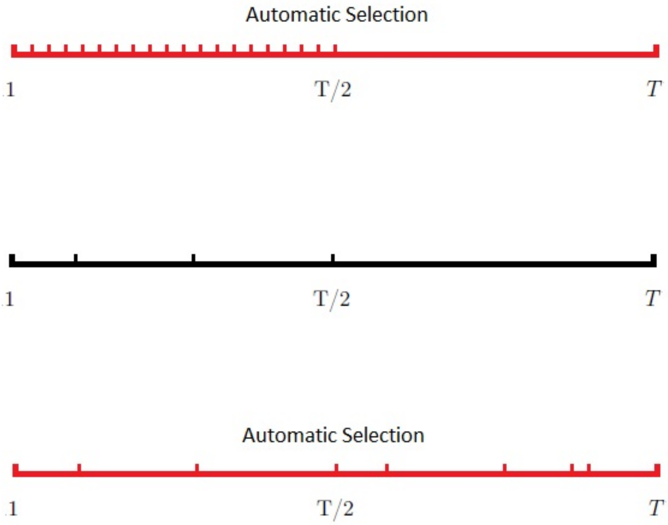
- In a Dummy Saturation, we create one or more dummies to each observation.

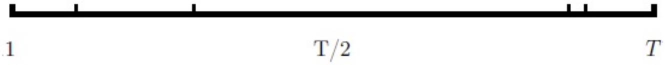
- As there are more variables than observations, estimating the GUM is inviable.
- Autometrics, however, applies a reduction technique for these models, initially proposed by [Hendry and Krolzig, 2005] and applied in the context of IIS by [Santos et al., 2008].

Autometrics III



Autometrics IV




1
T/2
T

Forward selection (FWD)

- FWD consists of regressing y on x_1 storing the residual $(\widehat{\epsilon}_1)$
- Next look for the covariate in the X information set with the highest correlation with the residual, say x_2
- The residual $\widehat{\epsilon}_1$ is projected onto x_2 and a new residual $\widehat{\epsilon}_2$ is stored and the covariate mostly correlated with $\widehat{\epsilon}_2$ is identified.
- The procedure continues until all the variables in the information set have been ranked or when a given criteria is satisfied, e.g. the adjusted R^2.
- FWD is exactly the opposite of that behind hard-thresholding, while the latter can select a large number of regressors very correlated with each other, FWD tends to keep fewer variables as orthogonal as possible to each other.
- Other variable selection metods are surveyed in [Kapetanios et al., 2014]

Least angle regressions (LARS)

- Starts as FWD by identifying the covariate that has the highest correlation with the target
- The largest step in he direction of this covariate x_1 is taken until a new predictor x_2 has as much correlation with the current residual.
- After this step LARS proceeds equiangularly between x_1 and x_2 rather than orthogonal as in FWD.
- After k steps there are k variables in the active set.
- The algorithm is stopped here and the coefficients of the remaining $N - k$ regressors are all set to zero.

RIDGE, LASSO and elastic net estimator (NET) I

- LASSO can be obtained in the LARS algorithm by imposing at each step a restriction on the sign of the correlation between the new candidate regressor and the projection along the equiangular direction in the previous step.
- LASSO can be related to RIDGE estimator which is a constrained OLS estimator that penalizes overfitting. Given M regressors, RIDGE coefficients are obtained by solving the following minimization problem

$$\min_{\beta} RSS + \lambda \sum_{j=1}^{M} \beta_j^2$$

- The Lagrange multiplier λ governs the shrinkage: the higher λ, the higher the penalty for having extra regressors.

RIDGE, LASSO and elastic net estimator (NET) II

- LASSO modifies the penalty function of RIDGE in the following minimization problem

$$\min_{\beta} RSS + \lambda \sum_{j=1}^{M} |\beta_j|$$

- Unlike RIDGE, in LASSO some regression coefficients are set exactly to zero.
- The Elastic Net (NET) estimator is a refinement of LASSO and it is the solution of the following minimization problem:

$$\min_{\beta} RSS + \lambda_1 \sum_{j=1}^{M} |\beta_j| + \lambda_2 \sum_{j=1}^{M} \beta_j^2$$

- The parameters λ, λ_1 and λ_2 are often selected via cross-validation.

RIDGE, LASSO and elastic net estimator (NET) III

- In a first step, a training sample is used to estimate de model for various values of λ.
  - Compute a loss function of interest, e.g. MSFE
  - Choose the value of λ which minimizes the loss
- In a second step a new sample is used to compute the loss function for the selected value of λ
  - Check if this choices also produce good results outside the training sample.

- Cross validation is valid for cross-section data by **not for time series data**.
- the choose of the hyperparameters , in the time series context, is done using AIC or BIC.
- could use part of the sample as training sample and the testing sample should be increasing sample, in order not to lose the data dependency.

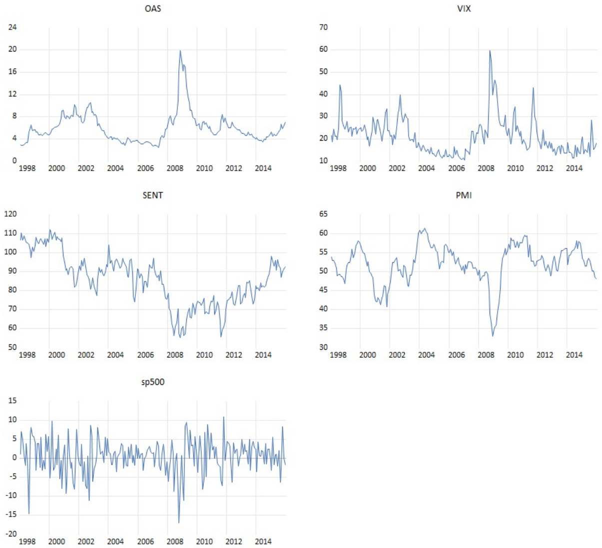
Multicollinearity

- It can happen that some of the explanatory variables are highly, though not perfectly, correlated
- In this case it is difficult to identify all the β parameters
- Large standard errors are associated to the OLS estimators
- Hypothesis testing can be affected by multicollinearity.
- To solve extend the sample can alleviate the problem
- Re-parameterize the model or use principal component
- This is not a problem in Autometrics .

Empirical example - Forecasting default risk

simulated data

Empirical example - Forecasting default risk



Empirical example - Forecasting default risk

- `# OLS for` $OAS_t = \alpha + \beta_1 VIX_{t-1} + \epsilon_t.$
- α `# coefficients for the constant`
- β_1 `# coefficients for the regressor`

	Estimate	Std. Error	t value	Pr(>|t|)
α	0.05436	0.31010	0.175	0.861
β_1	0.28027	0.01364	20.549	0.0000
$R^2 = 0.6647$	$\overline{R}^2 = 0.6631$			$AIC = 3.822267$
$\widehat{\sigma} = 1.621$				$BIC = 3.869299$
$F(1, 213) = 422.2$	$p - value = 0.000$			$LogLik = -407.8937$

Empirical example - Forecasting default risk

- `# OLS for` $OAS_t = \alpha + \beta_2 SENT_{t-1} + \epsilon_t$
- α `# coefficients for the constant`
- β_2 `# coefficients for the regressor`

	Estimate	Std. Error	t value	Pr(>|t|)
α	13.75283	1.09252	12.588	0.0000
β_2	-0.09037	0.01259	-7.178	0.0000
$R^2 = 0.1948$		$\overline{R}^2 = 0.191$		$AIC = 4.698361$
$\widehat{\sigma} = 2.512$				$BIC = 4.745393$
$F(1, 213) = 51.52$		$p - value = 0.000$		$LogLik = -502.0738$

Empirical example - Forecasting default risk

- `# OLS for` $OAS_t = \alpha + \beta_3 PMI_{t-1} + \epsilon_t$
- α `# coefficients for the constant`
- β_3 `# coefficients for the regressor`

	Estimate	Std. Error	t value	Pr(>|t|)
α	13.75283	1.09252	12.588	0.0000
β_3	-0.09037	0.01259	-7.178	0.0000
$R^2 = 0.1948$		$\overline{R}^2 = 0.191$		$AIC = 4.698361$
$\widehat{\sigma} = 2.512$				$BIC = 4.745393$
$F(1, 213) = 51.52$		$p - value = 0.000$		$LogLik = -502.0738$

Empirical example - Forecasting default risk

- `# OLS for` $OAS_t = \alpha + \beta_4 SP500_{t-1} + \epsilon_t$
- α `# coefficients for the constant`
- β_4 `# coefficients for the regressor`

	Estimate	Std. Error	t value	Pr(>|t|)
α	6.10840	0.17961	34.009	0.0000
β_4	-0.21997	0.04008	-5.488	0.0000
$R^2 = 0.1239$		$\overline{R}^2 = 0.1198$		$AIC = 4.782746$
$\widehat{\sigma} = 2.62$				$BIC = 4.829778$
$F(1,213) = 30,11$		$p - value = 0.000$		$LogLik = -511.1452$

Empirical example - Forecasting default risk

- # OLS for $OAS_t = \alpha + \beta_1 VIX_{t-1} + \beta_2 SENT_{t-1} + \beta_3 PMI_{t-1} + \beta_4 SP500_{t-1} + \epsilon_t$
- α # coefficients for the constant
- $\beta_1, \ldots, \beta_4$ # coefficients for the regressorz

	Estimate	Std. Error	t value	Pr(>|t|)
α	15.899229	1.413175	11.251	0.0000
β_1	0.183322	0.014438	12.697	0.0000
β_2	-0.047540	0.006598	-7.205	0.0000
β_3	-0.186019	0.021207	-8.772	0.0000
β_4	-0.043921	0.021376	-2.055	0.0411
$R^2 = 0.7979$		$\overline{R}^2 = 0.794$		$AIC = 3.344027$
$\widehat{\sigma} = 1.267$				$BIC = 3.438091$
$F(1, 210) = 207.2$		$p - value = 0.000$		$LogLik = -353.4829$

Empirical example - Forecasting default risk

Table: Table 1.13.6

Equation	BIC
VIX	3.869299
$SENT$	4.745393
PMI	4.282438
$SP500$	4.829778
ALL	3.4438091

Empirical example - Forecasting default risk

Table: Table 1.13.14 - RMSE and MAE for Stattic and Dynamic Foreacasts

Equation	Forecasting	Method
	Static	Recursive
RMSE_VIX	2.038440	1.723875
MAE_VIX	1.350642	1.213021
RMSE_SENT	3.358557	2.851563
MAE_SENT	1.892727	2.128676
RMSE_PMI	2.614534	2.235882
MAE_PMI	1.744641	1.597488
*RMSE_SP*500	3.362887	3.141333
*MAE_SP*500	1.963813	2.036924
RMSE_ALL	1.451455	1.158489
MAE_ALL	1.002002	0.804399
RMSE_ALL_WSENT	1.819475	1.460782
MAE_ALL_WSENT	1.371211	1.093121

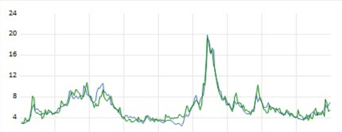
Empirical example - Forecasting default risk - Autometrics

- `rgets(tpval=0.001, pet=0.001, eq=eq_rget) oas c vix(-1) sent(-1) pmi(-1) sp500(-1) @`

Dependent Variable: OAS
Method: Least Squares
Date: 10/01/18 Time: 17:25
Sample (adjusted): 1998M02 2015M12
Included observations: 215 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.812584	1.522511	4.474570	0.0000
VIX(-1)	0.195337	0.013878	14.07545	0.0000
SENT(-1)	-0.002472	0.009713	-0.254535	0.7993
PMI(-1)	-0.140621	0.020135	-6.984018	0.0000
SP500(-1)	0.009567	0.017660	0.541754	0.5886
@AFTER("1999M06")	1.700113	0.380887	4.463563	0.0000
@AFTER("2000M05")	1.618159	0.394374	4.103106	0.0001
@AFTER("2002M07")	-1.061685	0.272922	-3.890072	0.0001
@AFTER("2008M08")	4.319213	0.462234	9.344213	0.0000
@AFTER("2009M03")	-3.248173	0.458850	-7.078935	0.0000
@AFTER("2011M09")	-0.289693	0.974479	-0.297280	0.7666
@AFTER("2011M10")	-0.370935	0.991391	-0.374156	0.7087

R-squared	0.891636	Mean dependent var	6.008000
Adjusted R-squared	0.885764	S.D. dependent var	2.792485
S.E. of regression	0.943827	Akaike info criterion	2.776448
Sum squared resid	180.8342	Schwarz criterion	2.964576
Log likelihood	-286.4681	Hannan-Quinn criter.	2.852460
F-statistic	151.8468	Durbin-Watson stat	1.192668
Prob(F-statistic)	0.000000		



Referências I

Bai, J. and Ng, S. (2008).
Forecasting economic time series using targeted predictors.
Journal of Econometrics, 146(2):304–317.

Doornik, J. A. (2009).
Autometrics.
In Castle, J. and Shephard, N., editors, *The Methodology and Practice of Econometrics: A Frestschrift in Honour of David F. Hendry*, chapter 4, pages 88–121. Oxford University Press.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983).
Exogeneity.
51(2):277.

Hendry, D. F. and Krolzig, H.-M. (2005).
The properties of AutomaticGetsModelling.
The Economic Journal, 115(502):C32–C61.

Referências II

Hoover, K. D. and Perez, S. J. (1999).
Data mining reconsidered: encompassing and the general-to-specific approach to specification search.
The Econometrics Journal, 2(2):167–191.

Kapetanios, G., Marcellino, M. G., and Papailias, F. (2014).
Variable selection for large unbalanced datasets using non-standard optimisation of information criteria and variable reduction methods.
SSRN Electronic Journal.

Santos, C., Hendry, D. F., and Johansen, S. (2008).
Automatic selection of indicators in a fully saturated regression.
Computational Statistics, 23(2):317–335.

Valls Pereira, P. L. (2024).
Modelos autorregressivos-defasagensdistribuídas (adl).