

ECON 4115/5115 Outline of Lecture Notes

Chapter 7. Time Series Regression Models

➤ The linear model

- Econometric regression model: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$
- y_t is the forecasting, dependent or endogenous variable
- x_t is the predictor, independent or exogenous variable
- ε_t is the error term
- β_0 is the intercept; β_1 is the slope
- *TSLM()* function will estimate the model in-sample
- Multivariate regression model: $y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_t$
- *ggpairs()* can help figure out which predictors to include
- Error term assumptions: no autocorrelation, heteroscedasticity or correlation w/ x_t
- R application: WY Okun's Law

➤ Least squares estimation

- The β 's are estimated using a least squares criterion: $\min \sum e_t^2$
- *TSLM()* function in R calculates least squares estimates in-sample
- Plotting fitted values vs. actual values
- Goodness of fit: $R^2 = \frac{\sum(y_t - \hat{y}_t)^2}{\sum(y_t - \bar{y})^2}$ and \bar{R}^2
- Be cautious of overfitting model in-sample

➤ Evaluating the regression model

- ACF plot of residuals, $e_t = y_t - \hat{y}_t$

- Histogram of residuals
 - *ggtsresidual()* function in R
 - Residual plot versus regressors and fitted values
 - Outliers and influential observations
 - Non-stationary series and spurious regressions
- Some useful predictors: linear trends and dummy variables
- A linear trend can be a regression predictor: $y_t = \beta_0 + \beta_1 t + \varepsilon_t$
 - A trend can be added in R using the *trend()* function
 - A dummy or binary variable (“yes” = 1; “no” = 0) can be a predictor
 - Can be used to remove outliers or extreme observations
 - Can be used to model “seasonality” using the *season()* function in R
 - Example: $y_t = \beta_0 + \beta_1 t + \beta_2 Mon_t + \beta_3 Tue_t + \dots + \beta_6 Fri_t + \beta_7 Sat_t + \varepsilon_t$
 - Make sure to avoid “dummy variable trap” by omitting one category
 - Dummy variables can be used to model transitory or permanent “interventions”
- Selecting predictors
- Scatter plots and p values are not a great method for selecting predictors
 - The *glance()* function in R will calculate five model selection measures:
 - Adjusted R^2 , $\bar{R}^2 = 1 - (1 - R^2) \left(\frac{T-1}{T-k-1} \right)$
 - Cross validation (CV): Omit one observation, fit model on remaining data, predict omitted observation, calculate MSE of out-of-sample forecast errors
 - $AIC = T \cdot \log \left(\frac{SSE}{T} \right) + 2(k + 2)$

- Small-sample corrected AIC: $AIC_c = AIC + \frac{2(k+2)(k+3)}{T-k-3}$
- $BIC = T \cdot \log\left(\frac{SSE}{T}\right) + \log(T)(k + 2)$
- Another possibility is a backward or forward stepwise regression

➤ Forecasting with regression

- In-sample predictions: $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t}$
- Ex-ante forecasts: Genuine forecasts using only information available in advance
- Ex-post forecasts: Forecasts using predictor information available with hindsight
- Comparisons of ex-ante and ex-post forecasts can help identify forecasting uncertainty
- Scenario-based forecasting (e.g., CBO forecasts)
- One option is to use lagged values of predictors: $\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t}$
- Forecast confidence intervals are difficult to calculate without matrices
- Example: 95% confidence interval for $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$: $\hat{y} \pm 1.96 \hat{\sigma}_e \sqrt{1 + \frac{1}{T} + \frac{(x-\bar{x})^2}{(T-1)s_x^2}}$

➤ Matrix formulation (advanced graduate material)

- The multiple regression model: $y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_t$
- Can be written as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{Y} = (y_1, y_2, \dots, y_T)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)'$, and $\mathbf{X}_{T \times k}$ is the explanatory variable matrix.
- The ordinary least squares formula is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

➤ Nonlinear regression

- The regression model needs to be linear in the parameters, but not the variables.
- Nonlinear regression examples:

- Log-log model: $\log(y_t) = \beta_0 + \beta_1 \log(x_t) + \varepsilon_t$
- Semi-log model: $\log(y_t) = \beta_0 + \beta_1 x_t + \varepsilon_t$, where often $x_t = t$
- Spline or piecewise linear trend
- Quadratic or higher order trends are not recommended

➤ Correlation, causation and forecasting

- Correlation does not imply causation
- It's possible X causes Y, but Y also causes X (two-way causality)
- It's also possible that a confounding variable Z, causes changes in both X and Y
- When forecasting, it's okay if two predictors X1 and X2 are highly correlated
- Hard to isolate the effects of X1 and X2 on Y, but together X1 and X2 can predict Y