Some Useful Econometric Techniques

Outline

- Descriptive Statistics
- Ordinary Least Squares
- Regression Tests and Statistics
- Violation of Assumptions in OLS Estimation
  - Multicollinearity
  - Heteroscedasticity
  - Autocorrelation
- Specification Errors
- Forecasting
- Unit Roots, Spurious Regressions, and cointegration

Descriptive Statistics

- **Useful estimators summarizing the probability distribution of a variable:**
- **Mean**

$$\mu = \frac{\sum_{i=1}^{T} X_i}{T}$$

- **Standard Deviation**

$$\sigma = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(X_i - \mu\right)^2}$$

Descriptive Statistics (Cont.)

- **Skewness (symmetry)**

$$S = \frac{1}{T}\sum_{i=1}^{T}\frac{\left(X_i-\mu\right)^3}{\sigma^3}$$

- **Kurtosis (thickness)**

$$K = \frac{1}{T}\sum_{i=1}^{T}\frac{\left(X_i-\mu\right)^4}{\sigma^4}$$

Ordinary Least Squares (OLS)

- **Estimation**
  - **Model**

$$Y_t = \beta_0 + \beta_1 X_{1t} + e_t$$

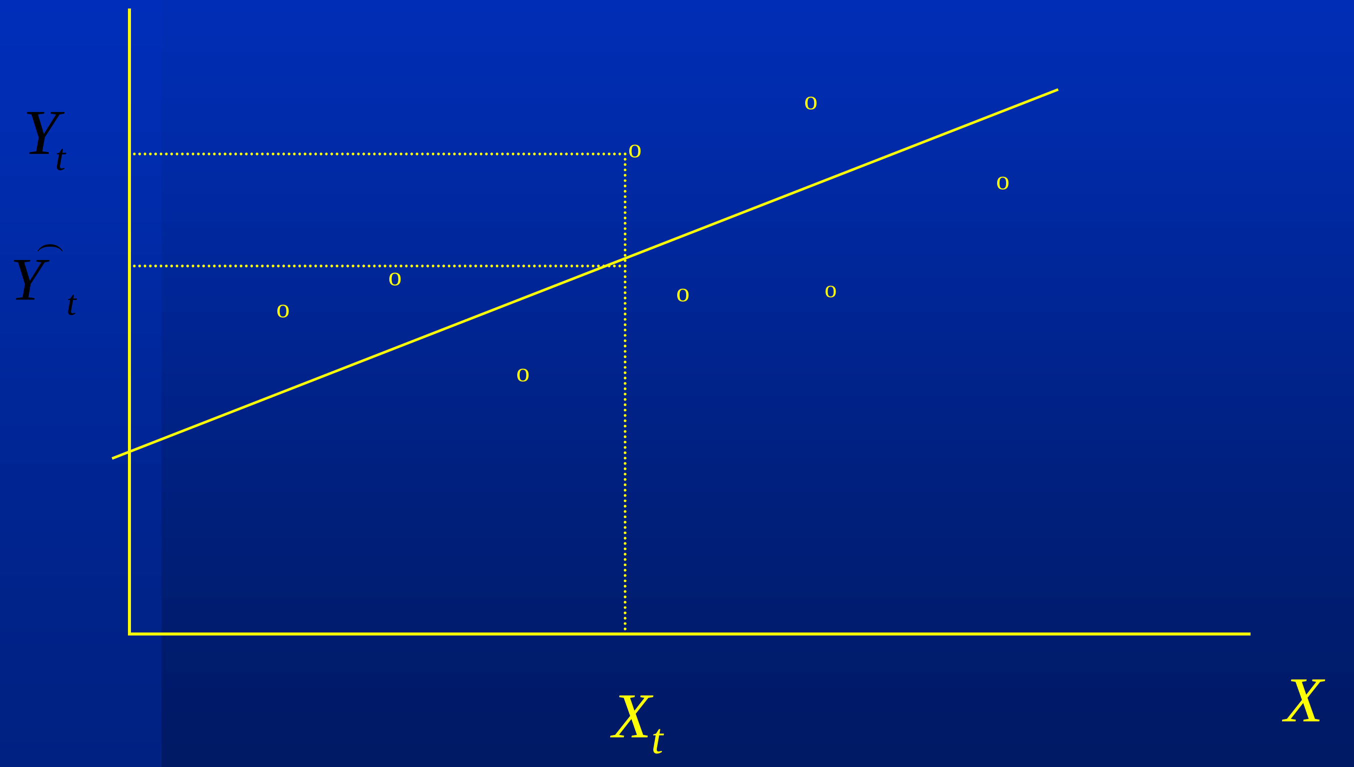
  - **The OLS requires:**
    - **Linear relationship between *Y* and *X*,**
    - ***X* is nonstochastic,**
    - **$E(e_t) = 0$, $Var(e_t) = s^2$ and $Cov(e_t, e_s)=0$ for *t* not equal to *s*.**

Ordinary Least Squares (OLS) (Cont.)

- **The OLS estimator for β_0 and β_1 are found by minimizing the sum of squared errors (SSE):**

$$\sum_{i=1}^{T} e_t^2 = \sum_{i=1}^{T} \left(Y_t - \widehat{Y}_t\right)^2 = \sum_{i=1}^{T} \left(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_t\right)^2$$

Ordinary Least Squares (OLS) (Cont.)



Ordinary Least Squares (OLS) (Cont.)

- **Minimizing the SSE is equivalent to:**

$$\frac{\partial\left(\sum_{i=1}^{T}\hat{e}_t^2\right)}{\partial\beta_0}=0\,,\quad\frac{\partial\left(\sum_{i=1}^{T}\hat{e}_t^2\right)}{\partial\beta_1}=0$$

- **Estimators are:**

$$\hat{\beta}_0=\vec{Y}-\hat{\beta}_1\overline{X}$$

$$\hat{\beta}_1=\frac{Cov(X,Y)}{Var(X)}=\frac{\sum_{i=1}^{T}\left(X_t-\overline{X}\right)\left(Y_t-\overline{Y}\right)}{\sum_{i=1}^{T}\left(X_t-\overline{X}\right)^2}$$

Ordinary Least Squares (OLS) (Cont.)

- **Properties of OLS estimators:**

$$\hat{\beta}_0 \text{ and } \hat{\beta}_1 \text{ are unbiased estimators}$$

$$E(\hat{\beta}_0) = \beta_0, \quad E(\hat{\beta}_1) = \beta_1$$

$$\hat{\beta}_0 = \mathrm{N}(\beta_{0,}, \sigma_{\mathrm{b}_0}{}^2), \quad \hat{\beta}_1 = \mathrm{N}(\beta_{1,}, \sigma_{\beta_1}{}^2)$$

  - **They are normally distributed**
  - **Minimum variance and unbiased estimators**

Example: Private Investment

- $FIR_t = b_0 + b_1RINT_{t-1} + b_2INFL_{t-1} + b_3RGDP_{t-1} + b_4NKFLOW_{t-1} + e_t$
- One can run this regression to estimate private fixed investment
  - A negative function of real interest rates (RINT)
  - A negative function of inflation (INFL)
  - A positive function of real GDP (RGDP)
  - A positive function of net capital flows (NKFLOW)

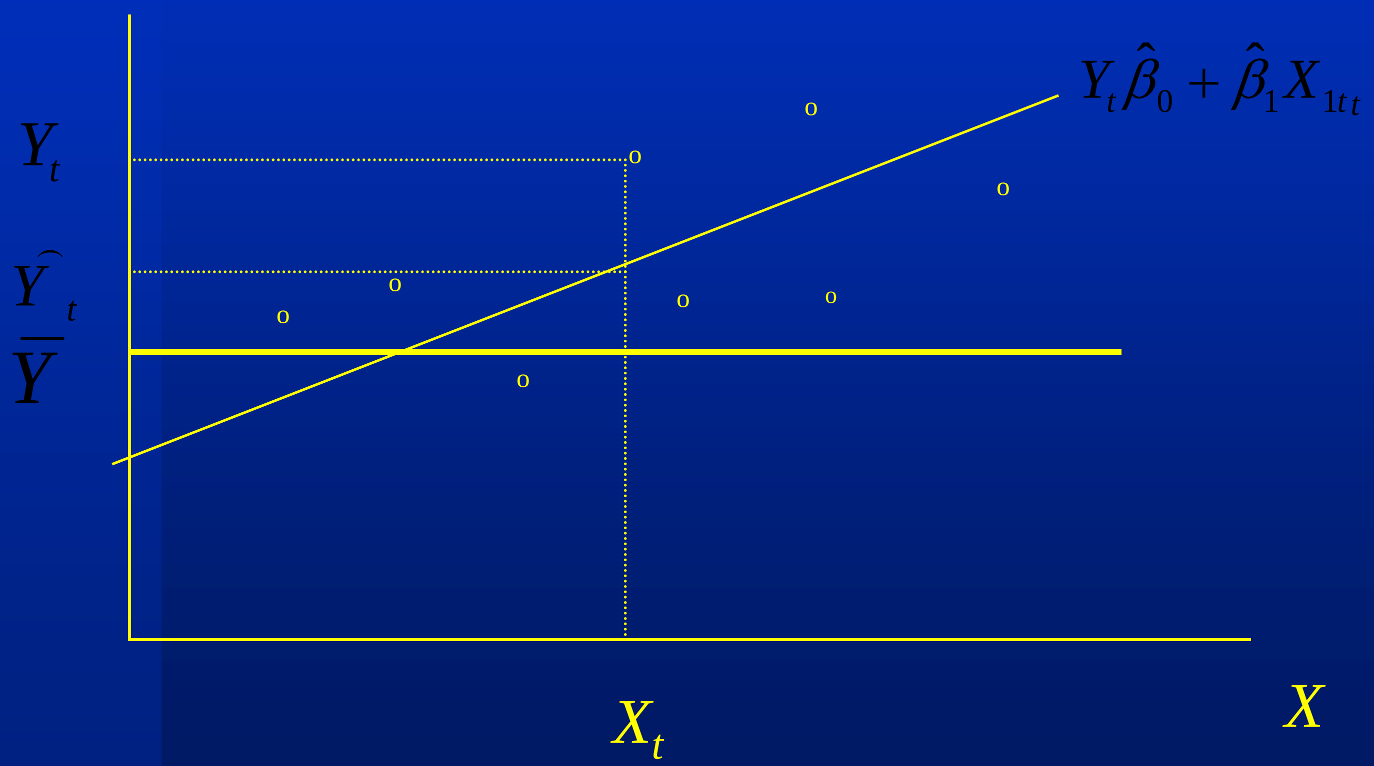
Regression Statistics and Tests

- R^2 is the measure if goodness of fit:

$$R^2 = \frac{SSR}{TSS} = \frac{\sum_{i=1}^{T}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{T}\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum_{i=1}^{T}\left(Y_i - \hat{Y}\right)^2}{\sum_{i=1}^{T}\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{SSE}{TSS}$$

- Limitations:
  - Depends on the assumption that the model is correctly specified
  - R^2 is sensitive to the number of independent variables
  - If intercept is constrained to be equal to zero, then R^2 may be negative.

Meaning of R^2



Regression Statistics and Tests

- **Adjusted R^2 to overcome limitations of**
- **R^2 = 1-SSE/(T- K)/TSS/(T-1)**
- **Is β_i statistically different from zero?**
- **When e_t is normally distributed, use *t-statistic* to test the null hypothesis $\beta_i = 0$.**
  - **A simple rule: if $t_{(T-k)} > 2$ then β_i is significant.**

$$t_{(T-k)} = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$$

Regression Statistics and Tests

- **Testing the model:**
  - **F-test: F-statistics with k-1 and T-k degrees of freedom is used to test for the null hypothesis:**
    - **$\beta_1=\beta_2=\beta_3=\ldots=\beta_k=0$**
    - **The f-statistics is:**
    - $$F_{(k-1,T-k)} = \frac{(T-k)R^2}{(k-1)(1-R^2)}$$
  - **The F test may allow the null hypothesis $\beta_1=\beta_2=\beta_3=\ldots=\beta_k=0$ to be rejected even when none of the coefficients are statistically significant by individual t-tests.**

Violations of OLS Assumptions

- **Multicollinearity**
  - **When 2 or more variables are correlated (in the multi variable case) with each other. E.g.,**

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t$$

  - **Result: high standard errors for the parameters and statistically insignificant coefficients.**
  - **Indications:**
    - **Relatively high correlations between one or more explanatory variables.**
    - **High R^2 with few significant t-statistics. Why?**

Violations of OLS Assumptions (Cont.)

$$\sigma^2 (X'X)^{-1} \rightarrow \infty$$

and

$$\frac{\hat{\beta}_i}{\hat{\sigma}_{\beta_i}} \rightarrow 0$$

Violations of OLS Assumptions (Cont.)

- **Heteroscedasticity: when error terms do not have constant variances σ^2.**
  - **Consequences for the OLS estimators:**
  - **They are unbiased [E(β)=β] but not efficient. Their variances are not the minimum variance.**
  - **Test: White's heteroscedasticty test.**

Violations of OLS Assumptions (Cont.)

- **<u>Autocorrelation</u>: when the error terms from different time periods are correlated [$e_t = f(e_{t-1}, e_{t-2}, \ldots)$]:**
  - **Consequences for the OLS estimators:**
    - **They are <u>unbiased</u> [$E(\beta)=\beta$] but <u>not efficient</u>.**
  - **Test for serial correlation: Durbin-Watson for first order serial correlation:**

$$DW = \frac{\sum_{t=2}^{T} \left(\hat{e}_t - \hat{e}_{t-1}\right)^2}{\sum_{t=1}^{T} \left(\hat{e}_t\right)^2}$$

Violations of OLS Assumptions (Cont.)

- **<u>Autocorrelation (cont.)</u>:**
- **Test for serial correlation (cont.)**
- **Durbin-Watson statistic (cont.)**
- **The DW statistic is approximately equal to:**

$$DW \approx 2(1-\rho_1) = 2\left(1-\frac{Cov(\hat{e}_t,\hat{e}_{t-1})}{Var(\hat{e}_t)}\right)$$

where

$$e_t = \rho_1 e_{t-1} + u_t$$

- **Note, if ρ_1=1 then DW =0. If ρ_1=-1 then DW =4. For ρ_1=0, DW =2.**
- **Ljung-Box Q test statistic for higher order correlation.**

Specification Errors

- **Omitted variables:**
  - **True model:**

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t$$

  - **Regression model:**

$$Y_t = \beta_0 + \beta_1 X_{1t} + e_t$$

  - **Then, the estimator for β_1 is biased.**

$$E(\beta_1^*) = \beta_1 + \beta_2 \frac{Cov(X_1, X_2)}{Var(X_2)}$$

Specification Errors (Cont.)

- **Irrelevant variables:**
  - **True model:**

$$Y_t = \beta_0 + \beta_1 X_{1t} + e_t$$

  - **Regression model:**

$$Y_t = \beta_0 + \beta^*_1 X_{1t} + \beta_2^* X_{2t} + e_t$$

  - **Then, the estimator for β_1 is still unbiased. Only efficiency declines, since the variance of β_1^* will be larger than the variance of β_1.**

Forecasting

- **A forecast is:**
  - **A quantitative estimate about the likelihood of future events which is developed on the basis of current and past information.**
  - **Some useful definitions:**
  - **<u>Point forecast</u>: predicts a single number for Y in each forecast period**
  - **<u>Interval forecast</u>: indicates an interval in which the realized value of Y will lie.**

Unconditional Forecasting

- **First estimate the econometric model**

$$Y_t = \beta_0 + \beta_1 X_{1t} + e_t$$

$$e_t \sim N\left(0, \sigma^2\right)$$

- **Then, compute:**

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{1T+1}$$

assuming X_{T+1} is known. This is the point forecast.

Unconditional Forecasting (Cont.)

- **The forecast error is:**

$$\hat{e}_{T+1} = \hat{Y}_{T+1} - Y_{T+1} = \left(\hat{\beta}_0 - \beta_0\right) + \left(\hat{\beta}_1 - \beta_1\right) X_{T+1} - e_{T+1}$$

- **The 95% confidence interval for Y_{T+1} is:**

$$\hat{Y}_{T+1} - t_{0.5} s_f \le Y_{T+1} \le \hat{Y}_{T+1} + t_{0.5} s_f$$

- **where**

$$s_f^2 = \hat{\sigma}^2 \left[1 + \frac{1}{2} + \frac{\left(X_{T+1} - \overline{X}\right)^2}{\sum_{t=1}^{T} \left(X_t - \overline{X}\right)^2} \right]$$

- **Which provides a good measure of the precision of the forecast.**