Heteroskedacity	**1. Definition** : variance of residuals is not the same across all observations in the sample - Unconditional heteroskedacity : heteroskedacity is not related to the level of independent variables → cause no major problems with the regression - Conditional heteroskedacity : heteroskedacity is related to the level of independent variables → significant problems for statistical inference **2. Effect on Regression Analysis** - Unreliable estimates of standard errors - No impact on coefficient estimates - If underestimate standard of errors → Overestimate t-statistics → null hypothesis is rejected too often - Unreliable f-test **3. Detection** - Method #1 : Examine scatter plot of residuals - Method #2 : apply Breusch-Pagan chi-square test $BP\ chi-square\ test = n \times R_{residual}{}^2$ $df = k = number\ of\ variables\ to\ be\ tested$ **4. Correction** - Use robust standard errors (White-corrected standard errors, or heteroskedacity-consistent standard errors) instead of normanl standard errors (*) - Use generalised least squares → eliminate heteroskedacity by modifying original equation (*) Recommended by CFA. White-corrected standard errors are used when only heteroskedacity appears
Serial correlation	**1. Defintion** : Residual terms are correlated with on another - Positive serial correlation : positive regression error in a period → higher probability of positive regression error for the next period - Negative serial correlation : positive regression error in a period → higher probability of negative regression error for the next period **2. Effect on Regression Analysis** : - Underestimate coefficient standard errors → Overestimate t-statistic → Type 1 error : rejection of null hypothesis when it is actually true - Underestimate MSE → unreliable F-test → Type 1 error **3. Detection :** - Method #1 : Examine scatter plot of residuals - Method #2 : Apply Durbin-Watson statistic $DW = \frac{\sum_{t=2}^{T}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T}\hat{\varepsilon}_t{}^2}$ or $DW = 2 \times (1 - r)$ $DW < d_l \rightarrow error\ terms\ are\ positively\ serially\ correlated$ $d_l < DW < d_U \rightarrow inconclusive$ $d_U < DW \rightarrow fail\ to\ reject\ the\ null\ of\ no\ positive\ serial\ correlation$ **4. Correcting serial correlation** - Adjust the coefficient standard error using Hansen method (recommended by CFA, use Hansen method whenever there is serial correlation problem) - Improve the specification of the model by incorporating the time series nature of the data
Multicollinearity	**1. Definition** : 2 or more independent variables are highly correlated **2. Effect on Regression Analysis :** - Unreliable slope coefficients - Artificially inflated standard errors → greater probability for incorrect conclusion that variable is significant **3. Detection** : t-test indicate of each individual coefficient is not significantly different than zero, while F-test is significant and Coefficient of Determination is high **4. Correction** : omit one or more of correlated independent variables
Model misspecification	**Categories of model mispecification :** 1. Misspecified functional form - Model misspecification #1 : Omitted important variables - Model misspecification #2 : Variables should be transformed - Model misspecification #3 : Data is improperly pooled 2. Independent variables are correlated with error term in time series model - Model misspecification #4 : Using lagged dependent variables as independent variable - Model misspecification #5 : Forecasting the past - Model misspecification #6 : Measuring independent variables with error 3. Other time-series misspecifications that result in nonstationarity **Effect of model misspecification :** Model misspecification → Biased and inconsistent regression coefficients → Unreliable hypothesis testing and inaccurate predictions
Misspecification #1 : Omitting a Variable	**1. Defintion** : fail to include an important variable in the regression **2. Effect :** - Biased and inconsistent regression coefficients → unreliable hypothesis tests and predictions
Misspecification #2 : Variable should be transformed	Dependent variable is not linearly related to independent variables → should transform the independent variable Fail to transform the independent variable → Misspecify the model
Misspecification #3 : Incorrectly pooling data	Relationship between returns and independent variables during Y1 to Y3 is different than Y4 to Y6 If pooling the data and estimating regression of Y1 to Y3 over the entire period , rather than estimating separate regression → misspecify the model