2026 Level II Fact Sheet

Key Facts & Formulas for the CFA®

Quantitative Methods

Multiple linear regression model

$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \cdots + b_kX_{ki} + \varepsilon_i$, i = 1, 2,….n

Assumptions of multiple linear regression

1. Linearity: The relationship between the dependent variable and the independent variables is linear.
2. Homoskedasticity: The variance of the regression residuals is the same for all observations.
3. Independence of errors: The observations are independent of one another. This implies the regression residuals are uncorrelated across observations.
4. Normality: The regression residuals are normally distributed.
5. Independence of independent variables:
   - 5a. Independent variables are not random.
   - 5b. There is no exact linear relation between two or more of the independent variables or combinations of the independent variables.

Commonly used diagnostic plots

Scatterplots of dependent and independent variables are used to check if the assumptions of 'linearity' and 'independence of independent variables' have been violated.
Scatterplots of residuals are used to check if the assumptions of 'homoskedasticity' and 'independence of errors' have been violated.
A 'Q-Q' plot of residuals is used to check if the assumption of 'normality' has been violated.

Goodness of Fit

R-squared (coefficient of determination) measures the fraction of the total variation in the dependent variable that is explained by the independent variable.

$$R^2 = \frac{\text{explained variation}}{\text{total variation}}$$

Total variation = unexplained variation + explained variation
AIC and BIC: lower values of both measures are better.
AIC is preferred if the model is used for prediction purposes.
BIC is preferred when goodness of fit is the goal.

Joint F-test

$$\text{F stat} = \frac{(\text{Sum of squares error restricted model} - \text{Sum of squares error unrestricted})/q}{\text{Sum of squares error unrestricted model}/(n-k-1)}$$

where: q is the number of restrictions, i.e. the number of variables omitted
If the calculated F-stat exceeds the critical F-value, we can conclude that at least one of the omitted variables is statically significant.

General linear F-test

$$\text{F stat} = \frac{\text{mean regression sum of squares}}{\text{mean squared error}} = \frac{MSR}{MSE}$$

If the calculated F-stat exceeds the critical F-value, we can conclude that at least one of the slope coefficients is different from 0.

Functional Form

Failures in Regression Functional Form	Explanation	Consequence
Omitted variables	One or more important variables are omitted from the regression	May lead to heteroskedasticity or serial correlation
Inappropriate form of variables	Ignoring a nonlinear relationship between the dependent and independent variable	May lead to heteroskedasticity
Inappropriate variable scaling	One or more regression variables may need to be transformed before estimating the regression.	May lead to heteroskedasticity or multicollinearity
Inappropriate data pooling	Regression model pools data from different samples that should not be pooled	May lead to heteroskedasticity or serial correlation

Violation of Regression Assumptions

Assumption	Violation	Issue	Detection	Correction
Homoskedastic error terms	Heteroskedastic error terms	Biased estimates of coefficients' standard errors	Visual inspection of residuals; Breusch–Pagan test	Revise model; use robust standard errors
Independence of observations	Serial correlation	Inconsistent estimates of coefficients and biased standard errors	Breusch–Godfrey test	Revise model; use serial-correlation consistent standard errors
Independence of independent variables	Multicollinearity	Inflated standard errors	Variance inflation factor	Revise model; increase sample size

High leverage point: a data point having an extreme value of an independent variable (X).
Outlier: a data point having an extreme value of the dependent variable (Y).

Detecting Influential Points

- Leverage (detects high leverage points): if leverage > $3\left(\frac{k+1}{n}\right)$: observation is potentially influential
  where k = number of independent variables
- Studentized residual (detects outliers):

If	Then
$\lvert t_{i^*} \rvert > 3$	Flag observation as being an outlier
$\lvert t_{i^*} \rvert$ > critical value of the t-statistic with n-k-2 degrees of freedom	Flag outlier observation as being potentially influential

Qualitative Independent Variables

- Dummy variables take on a value of 1 if a particular condition is true and 0 if that condition is false. To distinguish among *n* categories, the model must include *n – 1* dummy variables.
- An intercept dummy adds to or reduces the original intercept if a specific condition is met. When the intercept dummy is 1, the regression line shifts up or down parallel to the base regression line.

$Y_i = b_0 + d_0D_i + b_1X_i + \varepsilon_i$.

- A slope dummy allows for the slope of the regression line to change if a specific condition is met.

$Y_i = b_0 + b_1X_i + d_1D_iX_i + \varepsilon i$.

- It is also possible for a regression model to use both intercept and slope dummy variables.

$Y_i = b_0 + d_0D_i + b_1X_i + d_1D_iX_i + \varepsilon_i$.