Mark Meldrum
BY CERIFI

2026
Level 2 - Formula Sheet

Quantitative Methods

LM1: Basics of Multiple Regression and Underlying Assumptions
LOS: Formulate a MR equation

General Equation: $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n$

k	Slope Coefficients
K + 1	Coefficients
df	Degrees of Freedom = n – (k +1) = n – k -1
$\boldsymbol{X_1, X_2 \ldots X_n}$	Partial Regression Coefficients

Dummy Variables: Take the value of 0 or 1. Use n – 1 dummy variables. If all dummy variables =0, then the regression represents the i^{th} condition (i.e. the intercept).

Log-log Regression: $\ln(Y) = b_0 + b_1\ln(X_1) + b_2\ln(X_2) + \ldots + b_n\ln(X_n)$

LM2: Evaluating Regression Model Fit and Interpreting Model Results
LOS: Measures of goodness of fit

R-Squared: $R^2 = \frac{SSR}{SST}$

Adjusted R-Squared: $\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2)$

or: $\bar{R}^2 = 1 - \left[\frac{SSE/(n-k-1)}{SST/(n-1)}\right]$

Akaike's information criterion: $n\ln\left(\frac{SSE}{n}\right) + 2(k+1)$

Schwarz's Bayesian information criterion: $n\ln\left(\frac{SSE}{n}\right) + \ln(n)(k+1)$

Testing Joint Hypotheses for Coefficients: $F = \frac{(SSE\ restricted\ model - SSE\ unrestricted)/q}{SSE\ unrestricted\ model/(n-k-1)}$

LM3: Model Misspecification
LOS: Heteroskedasticity and Serial Correlation

Breusch-Pagan test for conditional heteroskedasticity:

$$\varepsilon^2 = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

- If no conditional heteroskedasticity, R^2 will be very low
- H_0 : no conditional heteroskedasticity
- $nR^2 \sim \chi^2$, df = number of independent variables, then compare with χ_c^2 (one-tailed test)
- Do not want to reject H_0

Durbin-Watson test for SC:

$$DW \approx 2(1 - r)$$

$$r = \text{corr}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})$$

- H_0: DW = 2
- DW ≈ 2(1 – r) where r = corr($\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}$)
- Reject H_0 if DW < d_L
- Do not reject H_0 if DW > d_U

Breusch-Godfrey test for SC:
Original regression:

$$y_t = b_0 + b_1X_{1t} + b_2X_{2t} + \boldsymbol{u}_\mathbf{t}$$

$$\widehat{\mathbf{u}}_\mathbf{t} = a_0 + a_1X_{1t} + a_2X_{2t} + p_1u_{t-1} + e_t$$

- H_0 : No serial correlation
- Chi-square test statistic for lag = 1
- F stat with n-p-k-1 degrees of freedom for p>1

LOS: Explain multicollinearity

Variance Inflation Factor (VIF) test for MC:

$$VIF_j = \frac{1}{1 - R_j^2}$$

- R_j^2 is calculated by regressing each independent variable against the k-1 other independent variables
- VIF_j >5 warrants investigation
- VIFj >10 indicates serious multicollinearity

LM4: Extensions of Multiple Regression
LOS: Formulate and interpret a logistic regression model

Logistic Regression:

$$\ln\frac{P}{(1-P)} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon$$

$$P = \frac{1}{1 + exp\left[-\left(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n\right)\right]}$$

LM5: Time Series Analysis

LOS: Calculate and evaluate trend models

Linear: $y_t = b_0 + b_1 t + \varepsilon_t$

If SC, DW ≠ 2, SE biased downwards

Log-linear: $\ln(y_t) = b_0 + b_1 t + \varepsilon_t$

b_1	Trend Coefficient
t	Time
ε_t	Error Term

LOS: Describe the structure of AR(p)

AR(1) $x_t = b_0 + b_1 x_{t-1} + \varepsilon_t$

AR(2) $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t$

AR(p) $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + b_p x_{t-p} + \varepsilon_t$

Covariance Stationary

- Constant mean, variance and Covariance(x_t,x_{t-s}) in all periods. If $b_1 < 1$, then the series is mean reverting and is Covariance stationary. All Cov. St. T have a finite mean reverting level. If not, then move on to first-differencing.
- Cannot use the DW statistic anymore to test for SC (once the model has as IV that is a lagged value of the DV). Use the t-statistics of the autocorrelations of the error terms.

LOS: Calculate mean-reverting level

$$x_t = \frac{b0}{(1 - b1)}$$

LOS: Calculate the root mean squared error

The square root of the average squared error:

$$RMSE = \sqrt{\frac{\sum \varepsilon^2}{n}}$$

The smaller the RMSE, the more accurate the model.

LOS: Unit roots, random walks

Random Walk: $b_0 = 0$ and $b_1 = 1$

After first differencing:
$b_0 = 0$ and $b_1 = 0$
(Cov. St., but no forecasting power)

Random Walk with drift: $b_0 \neq 0$

- Cannot estimate an AR(1) on a random walk
- Test for a RW: Dickey-Fuller
- H_0: $g = 0$, H_a: $g < 0$, where $g = (b_1 - 1)$

LOS: ARCH models

$$\hat{\varepsilon}_t^2 = a_0 + a_1\hat{\varepsilon}_{t-1}^2 + \mu$$

Test: H_0: $a_1 = 0$ versus H_a: $a_1 \neq 0$
Rejecting H_0 implies ARCH(1)
Thus, if $a_1 = 0$, we have constant variance

LOS: Two time-series

Engle-Granger DF test: $(\varepsilon_t - \varepsilon_{t-1}) = b_0 + g\varepsilon_{t-1} + \mu$

- H_0: $g = 0$, H_a: $g = 1$
- Reject H_0, regression output is OK.

DV and IV are both TS

Neither TS has a unit root Both have a unit root but are cointegrated	**No issues**	**All else, not OK**

LM6: Machine Learning

LOS: supervised versus unsupervised