Formula Sheet

Jeff Nguyen

01/03/2020

Regression

Three important concepts of regressions: causality, can use several variables (multiple regression) and forecasting.

When run regression always include intercept, otherwise the estimate would be biased, inefficient and inconsistent. However, intercept can be dropped when forecasting.

ANOVA

ANOVA is a statistical test of whether two or more population means are equal. Up to two population tests of equality of means is Z or t. More than two, the test is an F test.

H_0: All slope coefficients jointly equals 0. H_a: At least one slop coefficient does not equal 0.

One population:

H_0: $\mu = c$
H_a: $\mu \neq c$
The test for Z or t depends on the size of the sameple with Z or $t = \frac{\overline{X} - c}{SE}$
Two population:
H_0: $\mu_1 = \mu_2$
H_a: $\mu_1 \neq \mu_2$
The test for Z or t depends on the size of the sameple with Z or $t = \frac{\overline{X_1} - \overline{X_2}}{SE}$

Three or more population:

H_0: $\mu_1 = \mu_2 = \mu_3 = ...$
H_a: $\mu_1 \neq \mu_2 \neq \mu_3 \neq ...$
The test is an F test with $F = \frac{MSB}{MSW}$ where MSB is the mean sum square variations between the groups and MSW is the mean sum square variations within the groups.

Sum Squares Variations Between the groups $SSB = n_1(\overline{X_1} - X)^2 + n_2(\overline{X_2} - X)^2 + n_3(\overline{X_3} - X)^2$
$MSB = \frac{SSB}{k-1}$ if there are k groups.

Sum Squares Variations Within the groups:

$$\begin{aligned} SSW &= \sum(X_{1i} - \overline{X_1})^2 + \sum(X_{2i} - \overline{X_2})^2 + \sum(X_{3i} - \overline{X_3})^2 \\ MSW &= \frac{SSW}{n_1 + n_2 + n_3 - 3} \\ F &= \frac{MSB}{MSW} \end{aligned} \tag{1}$$

Step-by-step ANOVA

Step 1: Write data in one column

y
S&P 500
S&P 500
S&P 500
Dow
Dow
Dow
FTSE
FTSE
FTSE
CAC
CAC
CAC

Step 2: Since there are N group, there should be $N-1$ dummy variables instead of N because columns of the matrix must not be the same ("Dummy Trap").

y	D_{SP500}	D_{Dow}	D_{FTSE}
S&P 500	1	0	0
S&P 500	1	0	0
S&P 500	1	0	0
Dow	0	1	0
Dow	0	1	0
Dow	0	1	0
FTSE	0	0	1
FTSE	0	0	1
FTSE	0	0	1
CAC	0	0	0
CAC	0	0	0
CAC	0	0	0

Dummy variable selection does not matters, it is selected randomly. If one dummy is insignificant, drop it and add the other one.

Step 3: Run ANNOVA in Excel.

Step 4: Look at F-statistics. Accept if $p_{value} > .05$ and reject if $p_{value} < .05$.

Step 5: Conclusion: if accept, insiginificant, there's no difference on returns.

Scenario 2: Test for effect of COVID. Add COVID dummy variable:

y	D_{SP500}	D_{Dow}	D_{FTSE}	D_{COVID}
S&P 500	1	0	0	0
S&P 500	1	0	0	1 (affected period)
S&P 500	1	0	0	0
Dow	0	1	0	0
Dow	0	1	0	1 (affected period)
Dow	0	1	0	0
FTSE	0	0	1	0
FTSE	0	0	1	1 (affected period)
FTSE	0	0	1	0
CAC	0	0	0	0
CAC	0	0	0	1 (affected period)
CAC	0	0	0	0

Additional Steps: Check the significant level for each variable.
If significant: $intercept + slope + covid$.
If not significant: $intercept + covid$.

Fama-French Three Factor Model

The Fama-French three factor model is expressed as:

$R_a - R_f = \alpha + \beta_1(R_m - R_f) + \beta_2 SMB + \beta_3 HML + \epsilon$

where, $R_a - R_f$ is the risk premium of asset or portfolio a, $R_m - R_f$ is the market risk premium, and R_f is the risk-free return rate.

The three factors, β_1 is analogous to the original CAPM β but not equal to it since there are now two additional factors. SMB is the returns of "Small market capitalization Minus Big cap firm" and HML stands for the return of "High book-to-value ratio firm Minus Low book to value firm".

Hypothesis Testing

Run an F-Test
H_0: No relationship.
H_a: At least 1 has a relationship.

Event Study

Step 1:

Choose one of the following models, one factor or two factor:

$R_{at} = \alpha + \beta R_{mt} + \epsilon_t$
$R_{at} = \alpha + \beta_1 R_{mt} + \beta_2 R_{indt} + \epsilon_t$

and run the regression of the daily return to asset a on the daily returns to market or the market and the industry, say for 252 days before T_1.

Step 2:

Find $E[R_a] = \hat{\alpha} + \hat{\beta}_1 R_m$ for the time period T_1 to T_2 including T_0.

Step 3:

Calculate AR, abnormal returns, which is $AR = R_a - E[R_a]$, and CAR, cumulative abnormal return for T_1 to T_2.

Step 4:

Estimate the t-statistics as $t = \frac{AR}{SE}$, where SE is the standard error of regression. For any $t > 1.96$, the abnormal return is significant at 95% level of confidence.

Step 5:

CAR looks at the cumulative effect of the event. CAR is the cumulative sum of AR. Plotting AR shows the impact of the event and the efficiency of the market.

Step 6:

If market overacts the effect of the event on the value of the firm, the CAR will increase, then it will decrease when the event is announced.

If the market underreacts or underestimates the effect of the event on the value of the firm, the CAR will increase beyond the announcement date and will level off at the price that reflects all the available information in the market.

Ex-post Forecasting

Ex-post forecasting is splitting the data set and forecast the last part of the data based on the first part of the data.

Ex-ante Forecasting

Ex-ante forecasting is using the entire data set and X value of future period to forecast Y-value of future period.

Ratios:

Coefficient of variations:

$CV = \frac{\sigma}{\mu}$

Sharpe Ratio:

$Sharpe = \frac{E[r_p]-r_f}{\sigma_p}$

Treynor Ratio:

$Treynor = \frac{r_p-r_f}{\beta_p}$

Sortino Ratio:

Assuming minimum acceptable returns $MAR = r_f$.
Calculate deviation from MAR: $r_p - MAR$.
Get the negative value only from above $subset < 0$.
Calculate the Lower Partial Moment: var.
Calculate downside deviateion: $sqrt(var)$.
Sortino Ratio: $Sortino = \frac{r_p-r_f}{\sigma_{downside}}$.

Linear Regression: Manual

Need to find:

$$\begin{aligned} \overline{X} =& \text{Average of all X} \\ \overline{Y} =& \text{Average of all Y} \\ x =& X - \overline{X} \\ y =& Y - \overline{Y} \\ \sum x.y \\ \sum x^2 \\ \sum y^2 \\ \sum X^2 \end{aligned} \tag{2}$$

Regression Output:

$$\begin{aligned} \hat{\beta}_1 =& \frac{\sum xy}{\sum x^2} \\ \hat{\beta}_0 =& \overline{Y} - \hat{\beta}_1 \overline{X} \end{aligned} \tag{3}$$

Statistical Analysis:

R^2 implies % of variation of Y is explained by X

$$\begin{aligned} R^2 =& \frac{\hat{\beta_1}^2 \sum y^2}{\sum y^2} \\ SE =& \sqrt{\frac{\sum y^2 - \hat{\beta_1}^2 \sum x^2}{n-k-1}} \\ ForecastingEfficiency =& \frac{SE}{\overline{Y}} \end{aligned} \tag{4}$$

Hypothesis Testing Manual

Standard Form of Regression

$\hat{Y} = \hat{B_0} + \hat{B_1} X_1$

Hypothesis for Statistical Significance B_1

Hypothesis

H_0: $B_1 = 0$
H_a: $B_1 \neq 0$

Statistical Analysis

Find $S\hat{B_1} = \frac{SE}{sqrt \sum x^2}$

Find $t - stat = \frac{\hat{B_1}}{S\hat{B_1}}$

Hypothesis for Statistical Significance B_0

Hypothesis

H_0: $B_0 = 0$
H_a: $B_0 \neq 0$

Statistical Analysis

Find $S\hat{B_0} = S\hat{\beta_1} \sqrt{\frac{\sum X^2}{n}})$

Find $t - stat = \frac{\hat{B_0}}{S\hat{B_0}}$

If $-1.96 < t - stat < 1.96$ failed to reject H_0, otherwise reject.

Hypothesis if $B_1 = \alpha$

Hypothesis

H_0: $B_1 = \alpha$
H_a: $B_0 \neq \alpha$

Statistical Analysis

Find $z - stat = \frac{\hat{\beta_1} - X}{S\hat{B_1}}$

Correlation $\frac{\overline{xy}}{\sqrt{x^2}\sqrt{y^2}}$

Gauss-Markov Theorem

OLS (Ordinary Leasts Squares): minimized sum of squared residuals.
MLE (Maximum Likelihood Equation): if all 6 assumptions of Gauss-Markov Theorem holds then OLS becomes MLE. Errors become normally distributed. This means you have the best estimation.

Three property of Regression:

Unbiased:

$E[\hat{\beta}] = \beta$. If biased: $E[\hat{\beta}] = \beta + \epsilon$.

Efficient:

$Var[\hat{\beta}] = minVariance$

Consistent:

$E[\hat{\beta}] = \beta + \epsilon_n$
$plim_{n \to \infty} E[n] = 0$
Error becomes 0 as the number of observation increases.

First Assumption: $E[\epsilon] = 0$

If violated: biased, inefficient and inconsistent.
Violate when don't include intercept (omitted variable biased).
To fix: include intercept in the regression.

Hypothesis:

H_0: $E[\hat{\epsilon}] = 0$ Estimate is unbiased, efficient and consistent.
H_a: $E[\hat{\epsilon}] \neq 0$ Estimate is biased, inefficient and inconsistent.

Hypothesis Testing:

This assumption cannot be tested for.

Second Assumption: $E[x\epsilon] = 0$ Endogeneity Problem

Implies there is no correlation between the error of regression and the independent variable. Violates if $E[x\epsilon] \neq 0$, i.e. either 2-way causality or omitted variable. If violated: biased, inefficient and inconsistent. Fixed by using Instrumental Variable (IV) on the problematic variable. The IV must have high correlation with X and no correlation with ϵ. Most macroeconomic model have 2-way causality.

Regression analysis assumes 1-way causality $X \to Y$.

Hypothesis:

H_0: $E[x\epsilon] = 0$
H_a: $E[x\epsilon] \neq 0$

Hypothesis Testing:

This assumption cannot be tested for.

Possible problems:

Two-way Causality: $X \to Y$ X affects Y
$Y \to X$ Y affects X

Example: Macro Model
Regression: $Q = \beta_0 + \beta_1 P + \epsilon$.
Q and P has two way causality.

If there is a two-way causality, can do a Granger Causality Test.

Omitted Variable Bias: $Q = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \beta_3 P_3 + \beta_4 Y + \epsilon$

If you drop a variable that is statistically significant $\to$ biased estimate.

Solutions: Solving 2-way Causality

The Instrumental Variable regression is unbiased, inefficient and inconsistent. You need large sample to fix efficiency.

$Y = \beta_0 + \beta_1 X + \epsilon$

Change of variable to:
$y = \beta_1 x + \epsilon$, where
$y = Y - \overline{Y}$ and $x = X - \overline{X}$

If there is a 2-way causality, i.e. $E[x\epsilon] \neq 0$, pick a variable $z = Z - \overline{Z}$ where 1) z is highly correlated with x and 2) z has no correlation with ϵ.

Multiply both side of the equation by $\sum z$:
$\sum z.y = \beta_1. \sum z.x + \sum z.\epsilon$. Thus:

$\beta_1 = \frac{\sum z.y}{\sum z.x}$ and $\beta_{OLS} = \frac{\sum xy}{\sum x^2}$

This is called Instrumental Estimation where $OLS : lm(Y \sim X, data = data)$ and $ivreg(Y \sim X|z, data = data)$

Example: Regression Equation: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

With $E[X_2 \epsilon] \neq 0$. This is a known problematic error.

We use 2 staged Leasts Squares 2SLS:

Step 1: Run regression of X_2, the problematic variable on X_1 and X_3:

$X_2 = \alpha_0 + \alpha_1 X_! + \alpha_2 X_3 + \epsilon$

Find the estimate value of X_2 which is $\hat{X}_2$ then run a regular regression with $\hat{X}_2$: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.

The R code is $ivreg(Y \sim X_2|X_1 + X_3, data = data.$

Granger Causality Test

First Regression: If X cause Y, past values of X will have forecasting power over Y.

$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \epsilon$

Hypothesis: H_0: $\alpha_1 = \alpha_2 = \alpha_3 = ... = 0$
H_a: At least one α is different from 0.

Do an F test. Failed to reject null meaning X does not cause Y. Reject null meaning X does cause Y.

Second Regression: $X_t = \lambda_0 + \lambda_1 X_{t-1} + \lambda_2 X_{t-2} + ... + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + ... + \epsilon$

Hypothesis: H_0: $\gamma_1 = \gamma_2 = \gamma_3 = ... = 0$
H_a: At least one γ is different from 0.

Do an F test. Failed to reject null meaning Y does not cause X. Reject null meaning Y does caue X.

Combination of results from 2 regression: If X cause Y without Y cause X: X Granger-cause Y.
If X cause Y AND Y cause X: either there is 2-way causality or there is no causality. Then we have to do a qualitative assessment:
1) If the variables are economic, there is likely 2-way causality.
2) If the variables are financials, there is likely no causality.

Third Assumptions: $E[\epsilon_t \epsilon_{t-1}] = 0$ Serial Correlation or Autocorrelation

ϵ_t is the error of regression at time t. ϵ_{t-1} is the lagged. They should not be correlated. If they are correlated, then $E[\epsilon_t \epsilon_{t-1}] \neq 0$, and we have serial correlation/ autocorrellation problem.

If violated, then estimate is unbiased and inefficient.

Hypothesis

H_0: $E[\epsilon_t \epsilon_{t-1}] = 0$
H_a: $E[\epsilon_t \epsilon_{t-1}] \neq 0$

Hypothesis Testing

$t - stat = \frac{\hat{\beta}}{SE_{\hat{\beta}}}$ is not reliable so one shouldn't use it because SE would be large and thus, bad for forecasting.

To fix: Durbin-Watson Test or add lagging variable.

Add lagging variable Y_{t-1} Original Equation:
$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$

Add lagging variable:
$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + \epsilon_{t-1}$

Multiply by ρ:

$\rho Y_{t-1} = \beta_0 \rho + \beta_1 \rho X_{t-1} + \rho \epsilon_{t-1}$

Subtract from original equation:
$Y_t - \rho Y_{t-1} = (\beta_0 - \beta_0 \rho) + \beta_1 (X_t - \rho X_{t-1}) + (\epsilon_t - \epsilon_{t-1})$

$Y_t^* = \beta_0^* + \beta_1 X_{t-1}^* + \epsilon_t^*$

- Run the regression and find $\hat{\epsilon_t}$.
- Run regression of $\hat{\epsilon_t}$ on ϵ_{t-1} and find $\hat{\beta}$.
- We have $\hat{y_t^*} = y_t - \hat{\rho} y_{t-1}$ and $x^* = x - \hat{\rho} x_{t-1}$, $Y_t = \beta_0 + \beta_1 X_{t-1}^* + \epsilon_t$

Do constant recursive substitution until population = sample so the estimate becomes unbiased and efficient.

First order:
$\epsilon_t = \rho \epsilon_{t-1} + \epsilon_t$
$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho \epsilon_{t-2} + \epsilon_t$

Durbin-Watson Test

Hypothesis: H_0: no first order autocorrelation
H_a: first order autocorrelation exists

Assumptions are: the error are normally distributed with a mean of 0 and the errors are stationary.

Test-statistics: $DW = \frac{\sum_{t=2}^{T} (\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^{T} \epsilon_t^2}$
where ϵ_t are residuals from an OLS.

The DW test reports a test statistic, with value from 0 to 4, where:
- 2 is no autocorrelation.
- 0 to 2 is a positive autocorrelation (common in time series data)
- 2 to 4 is negative autocorrelation (less common in time series data)

To correct autocorrelation/ serial correlation, add the first order autoregressive term AR(1) into the model. If AR(1) doesn't fix, we can use AR(2) as well.

Fourth Assumptions: Homoskedacity $E[\epsilon^2] = \sigma_\epsilon^2 I$

This assumption is specific to cross-sectional data.

Violates if $E[\epsilon^2] = \sigma_\epsilon^2 V$ (hetereo-scedacity)– determined by high variation. V is the variance-covariance matrix. If violates, estimate is unbiased and inefficient.

Hypothesis

H_0: $E[\epsilon^2] = \sigma_\epsilon^2 I$
H_a: $E[\epsilon^2] = \sigma_\epsilon^2 V$

Hypothesis Testing

To test: Goldfield-Quandt Test, White Test, Breusch-Pagan Test.

Regression Equation:
$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$

Goldfield-Quandt Test: 1- Run regression, estimate ϵ_i.
2- Sort $\hat{\epsilon}$ in ascending order based on X_2 data.
3- Divide ordered $\hat{\epsilon}$ to 3 parts: 40%, 20% and 40%.

H_0: $\epsilon_1^2 = \epsilon_3^2$
H_a: $\epsilon_1^2 \neq \epsilon_3^2$

If X_2 is the cause of hetereoscedacity, multiply all variables by $\frac{1}{\sqrt{X_2}}$, so the variable being $\frac{Y}{\sqrt{X_2}}$, $\frac{X_1}{\sqrt{X_2}}$.

Regress R Test + Breusch-Pagan Test Testing for heteroscedacity using error^2 of regression
reg <- lm(y~x1+x2, data=data6)
regresid <- reg$residuals
regresid2 <- regresid^2
residtest <- lm(regresid2~x1*x2+I(x1$^{2)+I(x2}$2), data=data6)
summary(residtest)

If x2 is significant then it could be the variable that cause hetereoscedacity.

Run the Breusch-Pagan Test for Hetereoscedacity:
bptest(reg, ~x1*x2+I(x1$^{2)+I(x2}$2), data=data6)
If p-value > .05 then it's homoscedatic.

Fifth Assumptions: Multi-Collinearity

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

Checks

If $cor(X_1, X_2) \geqslant .9 \rightarrow$, then Xs are highly correlated.

Second check: regress X_1 on independent variables and check R^2. If high R^2 then multicollinearity.

Check relationships:
- inconsistent signs of β
- inconsistent regression statistics: R^2, $t-value$, $F-stat$. Either all have to be significant or all insignificant.

For example:
Coefficient of Income and Advertising should be positive, otherwise there could be multicollinearity.

Other example:
$\beta_1 = (2.8)$
$\beta_2 = (4.5)$
$\beta_3 = (3.1)$
The absolute value of all these coefficient have to be bigger than 1.96.

$R^2 = .26$ meaning only 26% explained, i.e. inconsistent.
$F_{stat} < 4$ is insignificant.

Test for Multi-Collinearity:

Test 1: Using Correlation Matrix

In R, run cor.test(x1, x2, x3)

Test 2: Variance Inflation Factor (VIF)

Run 3 separate regressions:
$X_1 = \alpha_0 + \alpha_1 X_2 + \alpha_2 X_3$
$X_2 = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_3$
$X_3 = \lambda_0 + \lambda_1 X_1 + \lambda_2 X_2$

Find the R^2 of each regression. And calculate $VIF = \frac{1}{1-R^2}$. If there is at least 1 $VIF > 10$ then we have multi-collinearity.

Once we know there is multi-collinearity, we need to find out which 2 variables that are highly correlated and delete the less important one. But this will cause omitted variable problem. Therefore the result is unbiased, inefficient, and inconsistent.

BLUE OLS

Best (minimum variance), Linear, Unbiased, Estimation.

Demand Model

Linear model:

$\beta_0 + \beta_1 P + \beta_2 P_s + \beta_3 P_c + \beta_4 Y + \beta_5 AD + \epsilon$

With linear model:
Price elasticity: $\epsilon_P = \hat{\beta}_1 \frac{\overline{P}}{\overline{Q}}$
Income elasticity: $\epsilon_Y = \hat{\beta}_4 \frac{\overline{Y}}{\overline{Q}}$
Advertising elasticity: $\epsilon_{AD} = \hat{\beta}_5 \frac{\overline{AD}}{\overline{Q}}$

Log Linear Model:

$\ln Q = \ln \beta_0 + \beta_1 \ln P + \beta_2 \ln P_s + \beta_3 \ln P_c + \beta_4 \ln Y + \epsilon$

If the model is log linear then:
Price elasticity: $\epsilon_P = \beta_1$
Income elasticity: $\epsilon_Y = \beta_4$
Advertising elasticity: $\epsilon_{AD} = \beta_5$

Cobb-Douglas Model:

$Q = \beta P^{\beta_1} P_s^{\beta_2} P_c^{\beta_3} Y^{\beta_4} AD^{\beta_5} \epsilon$

Price elasticity: $\epsilon_P = \beta_1$
Income elasticity: $\epsilon_Y = \beta_4$
Advertising elasticity: $\epsilon_{AD} = \beta_5$

Change in demand YoY:

$\Delta Q = \epsilon_P \Delta P + \epsilon_Y \Delta Y + \epsilon_{AD} \Delta AD$

Note that intercept isn't included in ΔQ because it's a constant.

Calculation of Optimum Advertising Amount:

$AD^* = P.Q.\frac{\epsilon_{AD}}{|\epsilon_P|}$

Normal, Luxury, Necessity Goods

Recall:
$\epsilon > 1$ Elastic (luxury good): P up, TR down; P down, TR up
$\epsilon < 1$ Inelastic (necessity): P up, TR up. P down, TR down
$\epsilon = 1$ Unit elastic: P up down, TR unchanged.

How do you know which model is better

Do a Ramsey Reset Test.

H_0: Specification is correct.
H_a: Specification is incorrect.

Run a χ^2 test. If p-value $< .05$ then reject H_0.

Chow Test

Is the test of joint restriction on regression coefficient.

1. Unrestricted Regression Equation:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Or $Q = \beta_0 N^{\beta_1} K^{\beta_2} \epsilon$

Recall:
Economies of Scale $\beta_1 + \beta_2 > 1$
Decreasing return to scale $\beta_1 + \beta_2 < 1$

Hypothesis

H_0: $\beta_1 + \beta_2 = 1$
H_a: $\beta_1 + \beta_2 \neq 1$

Hypothesis Testing:

Objective: Test if there is constant return to scale.

Restricted Regression Equation:

$Y = \beta_0 + \beta_1 X + (1 - \beta_1) X_2 + \epsilon$
Algebraic transform:

$Y - X_2 = \beta_0 + \beta_1 (X - X_2) + \epsilon$
where $Z = Y - X_2$ and $W = X - X_2$

Chow Test: $F = \frac{R^2_{unrestricted} - R^2_{restricted}}{1 - R^2_{unrestricted}} . \frac{T-K}{m}$
where T is the number of slope betas, K is the number of variable in the unrestricted regression and m is the number of restriction

Rule of thumb:

t-test: > 2 , reject null, significant
F-test: > 4, reject null, significant

Permanent Income Hypothesis

The permanent income hypothesis is a theory of consumer spending stating that people will spend money at a level consistent with their expected long-term average income. The level of expected long-term income then becomes thought of as the level of "permanent" income that can be safely spent. A worker will save only if their current income is higher than the anticipated level of permanent income, in order to guard against future declines in income.

Spending Equation:

$Y_t = \beta_0 + \beta_1 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2}$
where Y_t is spending and X_t is income.

Problems with this model:
- Multi-collinearity (multiple X variables).

- Serial correlation.
- Loss of degree of freedom (unknown how many lag variables to include).

Solution: Distributed Lag Model:

$Y_t = \beta_0 + \beta X_t + \lambda Y_{t-1} + \epsilon_t$

Good:
- There is no multi-collinearity because of separate X and Y variables.
- There is no losing degree of freedom because there is no lag variable.

Run the regression.

Result: Marginal Propensity to Consumption

Short Run: $MPC = \beta$ is estimated from the regression.
Long Run: $MPC = \frac{\beta}{1-\lambda}$

These metrics can be used to derive the multiplier of spending vs income: $Multiplier = \frac{1}{1-MPC}$