Quantitative Methods in Finance

Step 2: Explain

Prof. Mike Aguilar

https:/www.linkedin.com/in/mike-aguilar-econ

Last updated: March 19, 2024

Audience and Prerequisites

Audience:

- These notes were prepared for Kenan Flagler Business School's Daytime MBA program
- The setting is a 14 session course
- These notes serve as reference materials to complement our in-class work using computational tools

Prerequisites:

- Some knowledge of probability, statistics are expected
- Knowledge of finance in general, and asset pricing, in particular is expected

Motivating Case Study

- The motivating case study is portfolio allocation
- However, the concepts and tools are widely applicable to a range of settings within Finance
- We group the concepts into the following functional steps within our case study
    1. Explore - Loading and cleaning data, EDA, etc..
    2. Explain - Factor modeling, etc..
    3. Forecast - Time series models, etc...
    4. Protect - Portfolio allocation, risk measurement, etc..

Housekeeping

- Throughout the slide deck you will see “Q”, which indicates a question to you, the reader
- You will also see “A”, which indicates the associated answer
- It is generally most efficient to learn this material through active participation. Whenever you encounter a “Q”, be sure to try and develop your answer before turning the page to the provided “A” answer
- I recommend reading through these slides before engaging in associated coding exercises

Outline

CAPM

$$E[r_t - r_f] = \alpha + \beta(E[r_{M,t} - r_f])$$

$$r_t - r_f = \alpha + \beta(r_{M,t} - r_f) + u_t$$

CAPM

$$E[r_t - r_f] = \alpha + \beta(E[r_{M,t} - r_f])$$

$$r_t - r_f = \alpha + \beta(r_{M,t} - r_f) + u_t$$

- How do we interpret α and β?
- How do we estimate α and β?
- How do we determine if $\beta = 1$?
- How do we use this model?

By the end of this section you should be able to

- Construct a population regression model
- Estimate such a model
- Assess the overall goodness of fit
- Empirically test hypotheses

Outline

1. Regression
   - Univariate Regression Models
   - Population Regression Model
   - Estimation
   - Interpretation
   - Inference
   - Goodness of Fit
   - Multiple Regression
   - Dummy Variables
   - Common Problems & Solutions
2. Modeling Events

Outline

1. Regression
   - Univariate Regression Models
   - Population Regression Model
   - Estimation
   - Interpretation
   - Inference
   - Goodness of Fit
   - Multiple Regression
   - Dummy Variables
   - Common Problems & Solutions

Outline

1. Regression
   - Univariate Regression Models
   - Population Regression Model
   - Estimation
   - Interpretation
   - Inference
   - Goodness of Fit
   - Multiple Regression
   - Dummy Variables
   - Common Problems & Solutions

$$Y = \beta_1 + \beta_2 X + u \tag{1}$$

$$Y = \beta_1 + \beta_2 X + u \tag{1}$$

Y	Dependent Variable or Regressand
X	Explanatory/Independent Variable or Regressor
β_k	Parameters
u	Disturbance Term

Why do the disturbance terms exist?

Why do the disturbance terms exist?

- Omission of explanatory variables
- Aggregation of variables
- Model misspecification
- Function misspecification
- Measurement error

Population Regression Function

Assumptions

Population Regression Function

Assumptions

- $E[u] = 0$

Population Regression Function

Assumptions

- $E[u] = 0$
- $E[u|X] = E[u]$

Population Regression Function

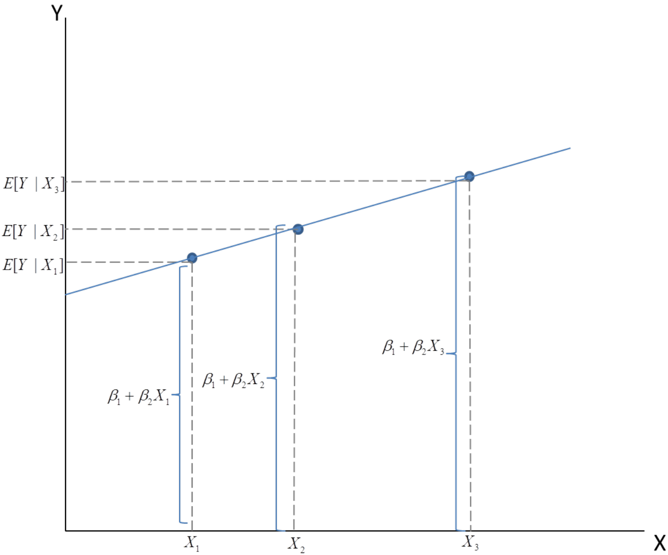
Assumptions

- $E[u] = 0$
- $E[u|X] = E[u]$

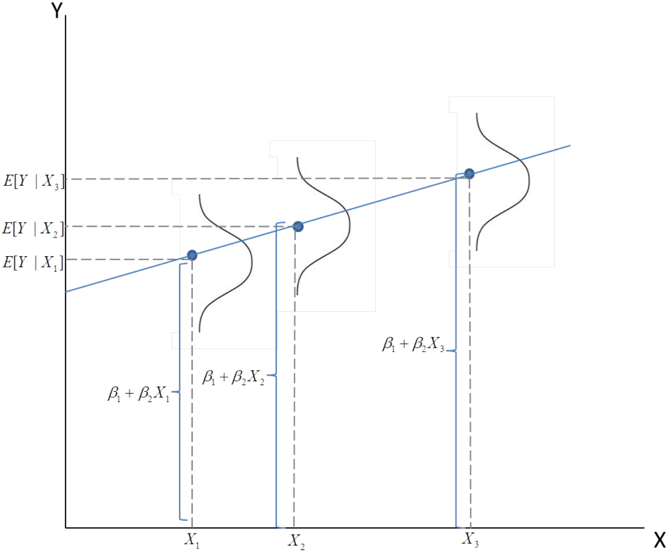
Population Regression Function

$$E[Y|X] = \beta_1 + \beta_2 X$$

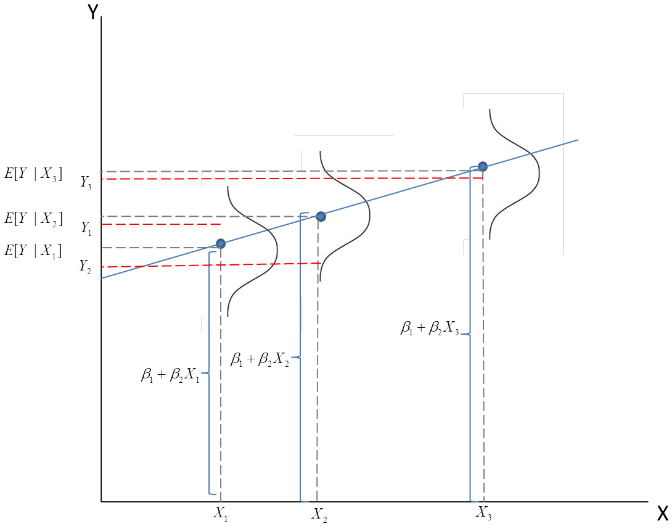
Disturbance Terms



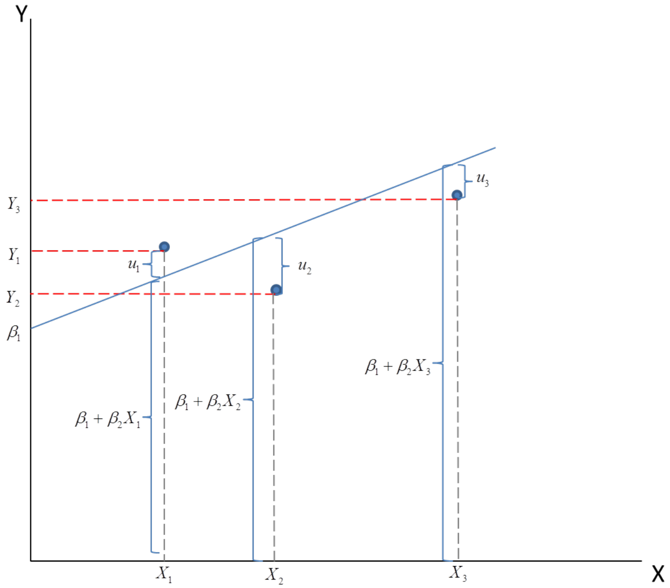
Disturbance Terms



Disturbance Terms



Disturbance Terms



Types of Data and Models

Data Types:

- Cross Sectional
- Time Series
- Panel

Types of Data and Models

Data Types:

- Cross Sectional
- Time Series
- Panel

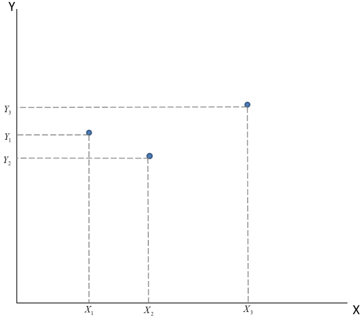
Model Types:

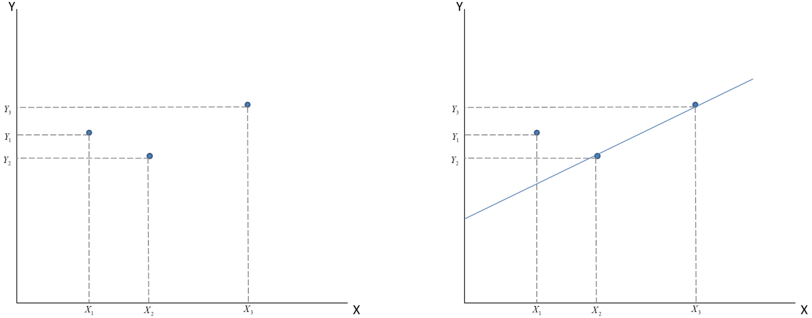
- X_i nonstochastic
- X_i stochastic; drawn randomly and independently from defined populations
- X_i stochastic; temporally persistent

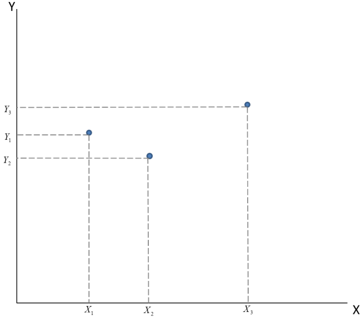
Standard OLS Assumptions - Simple Regression

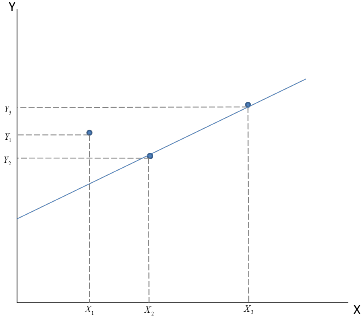
- Linear in Parameters
- Random Sampling
- Sample variation in the explanatory variable
- Zero Conditional Mean: $E[u|X] = 0$
- Homoscedasticity: $Var[u|X] = \sigma^2$

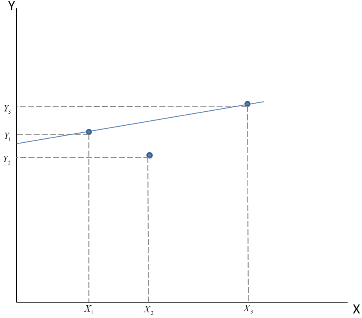
Outline

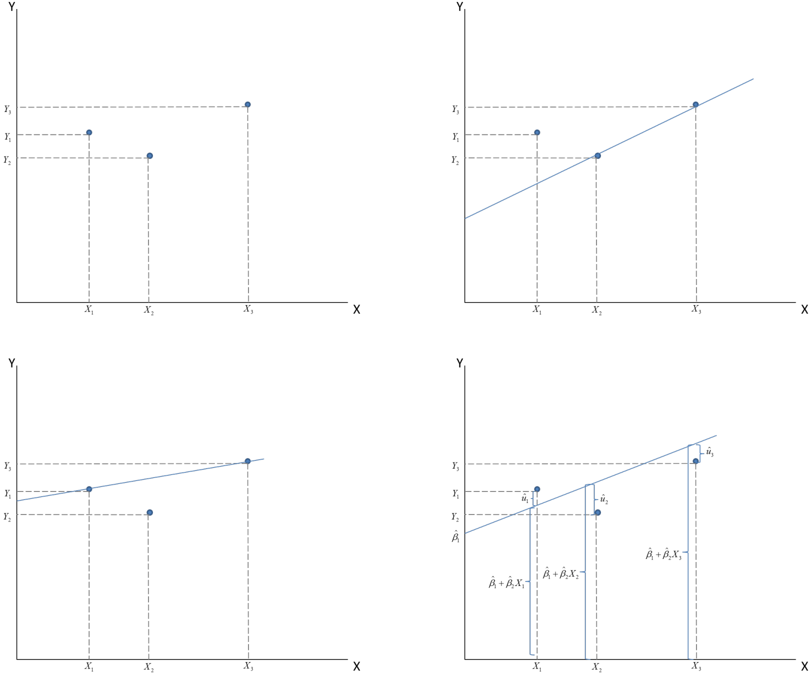

Y
Y_3
Y_1
Y_2
X_1
X_2
X_3
X


Y
Y_3
Y_1
Y_2
X_1
X_2
X_3
X
Y
Y_3
Y_1
Y_2
X_1
X_2
X_3
X


Y
Y_3
Y_1
Y_2
X_1
X_2
X_3
X


Y
Y_3
Y_1
Y_2
X_1
X_2
X_3
X


Y
Y_3
Y_1
Y_2
X_1
X_2
X_3
X


Y
X
Y_1
Y_2
Y_3
X_1
X_2
X_3
$\hat{\beta}_1$
$\hat{u}_1$
$\hat{u}_2$
$\hat{u}_3$
$\hat{\beta}_1 + \hat{\beta}_2 X_1$
$\hat{\beta}_1 + \hat{\beta}_2 X_2$
$\hat{\beta}_1 + \hat{\beta}_2 X_3$

Fitted Values

For ease of exposition, assume X and Y are mean zero.

$$\begin{aligned} Y_i &= \beta X_i + u_i \\ \hat{Y}_i &= \hat{\beta} X_i \\ \hat{u}_i &= Y_i - \hat{Y}_i \end{aligned}$$

Example (OLS Minimization)

1. Write the OLS objective function.
2. Construct the associated F.O.C.
3. Solve for the OLS estimator.

Example (OLS Minimization)

$$
\begin{aligned}
W &= \min_{\beta} \sum_{i=1}^{n} \hat{u}_i^2 && (2)\\
W &= \min_{\beta} \sum_{i=1}^{n} (Y_i - \beta X_i)^2 && (3)\\
\frac{\partial W}{\partial \beta} &= \sum_{i=1}^{n} -2X_i(Yi\hat{\beta}X_i) = 0 && (4)\\
\hat{\beta} &= \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} && (5)
\end{aligned}
$$

OLS Estimators

Definition (OLS Estimators)

Given the model

$$Y_i = \alpha + \beta X_i + u_i$$

with standard assumptions, the OLS estimators are

$$\begin{aligned}
\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X} \\
\hat{\beta} &= \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}
\end{aligned}$$

OLS Estimators

Definition (OLS Estimators)

Given the model

$$Y_i = \alpha + \beta X_i + u_i$$

with standard assumptions, the OLS estimators are

Coefficient	Estimator	Variance
α	$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$	$Var(\hat{\alpha}) = \frac{\sigma_u^2}{n} \frac{\sum X_i^2}{\sum (X_i - \bar{X})^2}$
β	$\hat{\beta} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$	$Var(\hat{\beta}) = \frac{\sigma_u^2}{\sum (X_i - \bar{X})^2}$

Standard Error of Regression Coefficient

Estimate variance of error:

$$\hat{\sigma}_u^2 = \frac{1}{n-2} \sum u^2$$

Coefficient	Estimator	Std Error
α	$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$	$s.e.(\hat{\alpha}) = \sqrt{\frac{\hat{\sigma}_u^2}{n} \frac{\sum X_i^2}{\sum (X_i - \bar{X})^2}}$
β	$\hat{\beta} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$	$s.e.(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}_u^2}{\sum (X_i - \bar{X})^2}}$

Under standard assumptions, the OLS estimators are

- Random Variables
- Normally Distributed
- Unbiased
- Efficient
- Consistent

Outline

1. Regression
   - Univariate Regression Models
   - Population Regression Model
   - Estimation
   - Interpretation
   - Inference
   - Goodness of Fit
   - Multiple Regression
   - Dummy Variables
   - Common Problems & Solutions

Model Interpretation

```
. reg ER MRP

      Source |       SS       df       MS              Number of obs =     150
-------------+------------------------------           F(  1,   148) = 1068.62
       Model |  .122240485     1  .122240485           Prob > F      =  0.0000
    Residual |  .016929808   148  .000114391           R-squared     =  0.8784
-------------+------------------------------           Adj R-squared =  0.8775
       Total |  .139170293   149  .000934029           Root MSE      =   .0107

------------------------------------------------------------------------------
          ER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         MRP |   .6354764   .0194396    32.69   0.000     .5970614    .6738914
       _cons |   .0201293   .0021349     9.43   0.000     .0159104    .0243481
------------------------------------------------------------------------------
```

Model Interpretation

Source	SS	df	MS
Model	.122240485	1	.122240485
Residual	.016929808	148	.000114391
Total	.139170293	149	.000934029

Model Interpretation

Source	SS	df	MS
Model	.122240485	1	.122240485
Residual	.016929808	148	.000114391
Total	.139170293	149	.000934029

ANOVA Table

- SS = sum of squares
- df = degrees of freedom
- MS = mean squares (SS/df)

- SS Model(ESS) $=\sum(\hat{Y}_i - \bar{Y})^2$
- SS Residual(RSS)$=\sum \hat{u}_i^2$
- SS Total(TSS)=ESS+RSS$=\sum(Y_i - \bar{Y})^2$
- Total df$=n-1$; Model df$=k-1$; Residual df$=n-1-(k-1)$

Model Interpretation

Number of obs	=	150
F(1, 148)	=	1068.62
Prob > F	=	0.0000
R-squared	=	0.8784
Adj R-squared	=	0.8775
Root MSE	=	.0107

Model Interpretation

```
Number of obs =      150
F(  1,    148) = 1068.62
Prob > F       =  0.0000
R-squared      =  0.8784
Adj R-squared  =  0.8775
Root MSE       =   .0107
```

Overall Model Fit

- Number of obs = # observations used in the regression
- F(1,148)=F-Statistic from test of overall goodness of fit
- Prob<F=p-value associated with F-Statistic
- R-squared=Proportion of variance in Y explained by X
- Adj.R-squared=R^2 with penalty for extra predictors
- Root MSE=$\sqrt{\text{MS Residual}}$.

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$\hat{\beta}_1$

$\hat{\beta}_2$

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$\hat{\beta}_1$	Fitted value of Y for X equal to 0
$\hat{\beta}_2$	

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$\hat{\beta}_1$	Fitted value of Y for X equal to 0
$\hat{\beta}_2$	One unit increase in X is associated with a $\hat{\beta}_2$ unit increase in Y

Interpreting Regression Coefficients

ER	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MRP	.6354764	.0194396	32.69	0.000	.5970614	.6738914
_cons	.0201293	.0021349	9.43	0.000	.0159104	.0243481

Interpreting Regression Coefficients

ER	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MRP	.6354764	.0194396	32.69	0.000	.5970614	.6738914
_cons	.0201293	.0021349	9.43	0.000	.0159104	.0243481

- ER: Dependent Variable
- MRP: Independent Variable
- _const: Constant term
- Coeff MRP(.6354)$=\hat{\beta}_2$
- Coeff _cons(.0201)$=\hat{\beta}_1$
- St.Err: Standard Errors for coefficients
- t: t-statistics for $H_0 : \beta_i = 0$
- $P > |t|$: 2 tailed p-vales for null above
- 95% Conf.Interval: 95% confidence intervals for coeff.

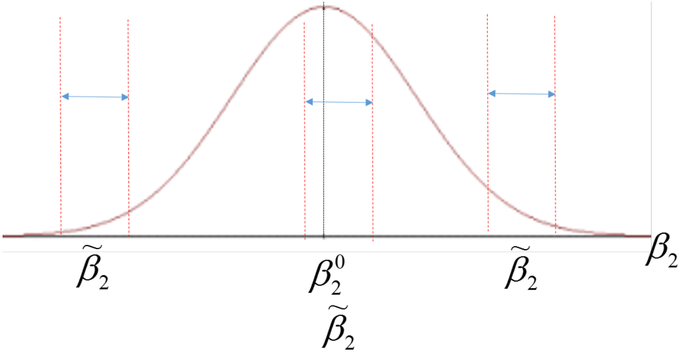
Outline

Inference

Consider the simple market model $r = \beta_0 + \beta_1 r^M + u$. Suppose you are proposing the inclusion of another "factor", such as Earnings Quality. You purport $r = \beta_0 + \beta_1 r^M + \beta_2 EQ + u$. How do we test this theory?

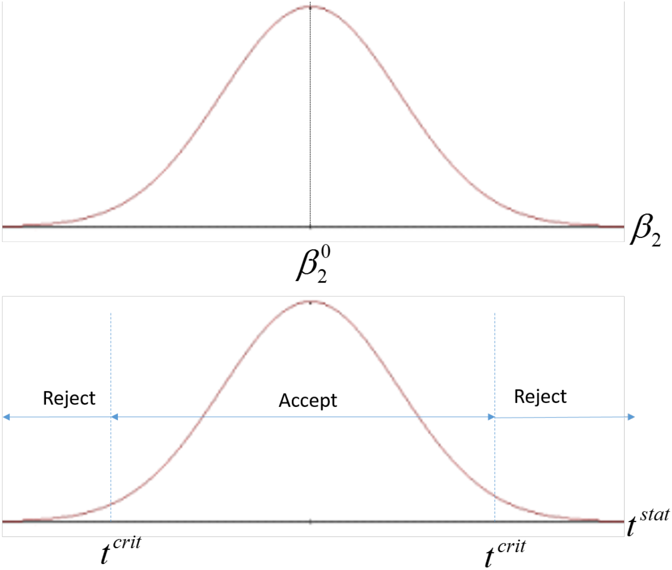
$$\begin{aligned} H_0 &: \quad \beta_2 = 0 \\ H_a &: \quad \beta_2 \neq 0 \end{aligned}$$

Inference



β_2^0: our hypothesized value of β_2
$\tilde{\beta}^2$: data-based estimates of β_2

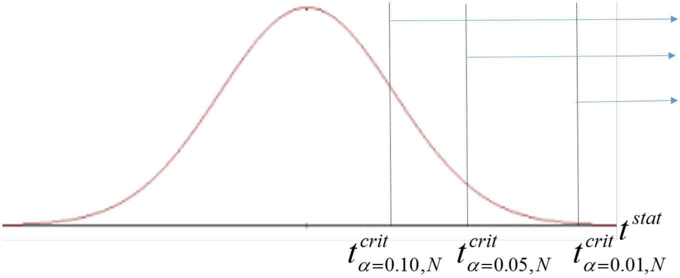
Inference



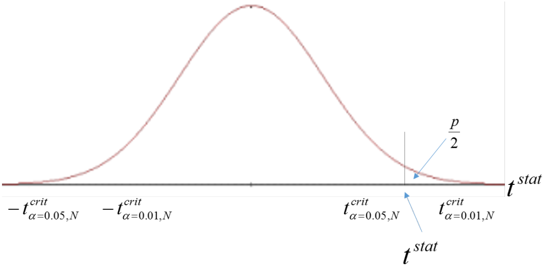
Inference

Where do we draw the t^{crit} values? Depends upon

1. Degrees of Freedom: $df = N - k$. As df increase the critical value decreases towards those for a standard normal distribution.
2. Mistakes:
    - $Pr[Reject\ H_0|H_0\ True] = \alpha$=Type 1 Error=Size
    - $Pr[Accept\ H_0|H_0\ False]$=Type 2 Error
    - $Pr[Reject\ H_0||H_0\ False]$ = Power


t^{stat}
$t^{crit}_{\alpha=0.10,N}$
$t^{crit}_{\alpha=0.05,N}$
$t^{crit}_{\alpha=0.01,N}$

Inference



- In this case, we reject at the 5% level, 4% level, etc..., but fail to reject at the 1% level.
- Rule of thumb: Reject if t-stat > 2.
- p-value is the smallest confidence level at which can reject the null.
- Reject null if $p < \alpha$

Testing Significance of a Regressor

Two Sided Test

$H_0 : \beta_2 = \beta_2^0$

$H_1 : \beta_2 \neq \beta_2^0$

$t^{stat} = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)}$

$t^{crit}_{\alpha,n-2} = T^{-1}(1 - \frac{\alpha}{2})$

Reject H_0 if

$|t^{stat}| > t^{crit}_{\alpha/2,n-2}$

Reject H_0 if $p < \alpha$

Testing Significance of a Regressor

<u>Two Sided Test</u>

$H_0 : \beta_2 = \beta_2^0$

$H_1 : \beta_2 \neq \beta_2^0$

$t^{stat} = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)}$

$t^{crit}_{\alpha,n-2} = T^{-1}(1 - \frac{\alpha}{2})$

Reject H_0 if

$|t^{stat}| > t^{crit}_{\alpha/2,n-2}$

Reject H_0 if $p < \alpha$

<u>One Sided Tests</u>

$H_0 : \beta_2 = \beta_2^0$

$H_1 : \beta_2 > \beta_2^0$

$t^{stat} = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\beta_2)}$

$t^{crit}_{\alpha,n-2} = T^{-1}(1 - \alpha)$

Reject H_0 if $t^{stat} > t^{crit}_{\alpha,n-2}$

Reject H_0 if $p/2 < \alpha$

Testing Significance of a Regressor

```
. reg ER MRP

      Source |       SS       df       MS              Number of obs =     150
-------------+------------------------------           F(  1,   148) = 1068.62
       Model |  .122240485     1  .122240485           Prob > F      =  0.0000
    Residual |  .016929808   148  .000114391           R-squared     =  0.8784
-------------+------------------------------           Adj R-squared =  0.8775
       Total |  .139170293   149  .000934029           Root MSE      =   .0107

------------------------------------------------------------------------------
          ER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         MRP |   .6354764   .0194396    32.69   0.000     .5970614    .6738914
       _cons |   .0201293   .0021349     9.43   0.000     .0159104    .0243481
------------------------------------------------------------------------------
```

The default null hypothesis in most statistical packages is that the parameter value in question is equal to zero. In this case $H_0 : \beta = 0$, against the two-tailed alternative $H_a : \beta \neq 0$. Given the low p-value for MRP, we can reject the null hypothesis at all conventional levels of significance (10%, 5%, 1%), suggesting that the MRP does indeed influence returns.

Outline

$R^2 \equiv$ Coefficient of Determination

$R^2 \equiv$ Coefficient of Determination

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \;=\; \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} \hat{u}_i^2$$

$R^2 \equiv$ Coefficient of Determination

$$
\begin{aligned}
\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} \hat{u}_i^2 \\
TSS &= ESS + RSS
\end{aligned}
$$

$R^2 \equiv$ Coefficient of Determination

$$\begin{aligned} \sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} \hat{u}_i^2 \\ TSS &= ESS + RSS \end{aligned}$$

Definition

The Coefficient of Determination is written as $R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$. Or equivalently, $R^2 = 1 - \frac{\sum_{i=1}^{n}\hat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$.

The F test of Goodness of Fit

$H_0 : \beta_2 = 0 \quad H_1 : \beta_2 \neq 0$
$F = \frac{(ESS/TSS)/(k-1)}{(RSS/TSS)/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$
where $k = \#$ of regressors (intercept and slope).

Testing Overall Goodness of Fit

```
. reg ER MRP

      Source |       SS       df       MS              Number of obs =     150
-------------+------------------------------           F(  1,   148) = 1068.62
       Model |  .122240485     1  .122240485           Prob > F      =  0.0000
    Residual |  .016929808   148  .000114391           R-squared     =  0.8784
-------------+------------------------------           Adj R-squared =  0.8775
       Total |  .139170293   149  .000934029           Root MSE      =   .0107

------------------------------------------------------------------------------
          ER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         MRP |   .6354764   .0194396    32.69   0.000     .5970614    .6738914
       _cons |   .0201293   .0021349     9.43   0.000     .0159104    .0243481
------------------------------------------------------------------------------
```

Testing Overall Goodness of Fit

```
. reg ER MRP

      Source |       SS       df       MS              Number of obs =     150
-------------+------------------------------           F(  1,   148) = 1068.62
       Model |  .122240485     1  .122240485           Prob > F      =  0.0000
    Residual |  .016929808   148  .000114391           R-squared     =  0.8784
-------------+------------------------------           Adj R-squared =  0.8775
       Total |  .139170293   149  .000934029           Root MSE      =   .0107

------------------------------------------------------------------------------
          ER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         MRP |   .6354764   .0194396    32.69   0.000     .5970614    .6738914
       _cons |   .0201293   .0021349     9.43   0.000     .0159104    .0243481
------------------------------------------------------------------------------
```

- $R^2 = 1 - RSS/TSS = 1 - .000114391/.000934029 \approx .8784$
- About 87% of the sample variation in ER is explained by MRP
- $F_{k-1,n-k} = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{R^2}{(1-R^2)/(n-2)} \approx 1068.62.$
- $Prob > F = 0.0000 \rightarrow$ Reject $H_0 : \beta_2 = 0.$

Outline

1. Regression
   - Univariate Regression Models
   - Population Regression Model
   - Estimation
   - Interpretation
   - Inference
   - Goodness of Fit
   - Multiple Regression
   - Dummy Variables
   - Common Problems & Solutions

Multiple Regression Model

Multiple Regression Model

$$Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \cdots + \beta_k X_{k,i} + u_i$$

Multiple Regression Model

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \cdots + \beta_k X_{k,i} + u_i \\
Y_i &= \beta_1 + \sum_{j=2}^{k} \beta_j X_{j,i} + u_i
\end{aligned}
$$

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2,i} + \hat{\beta}_3 X_{3,i}$$

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2,i} + \hat{\beta}_3 X_{3,i}$$

$\hat{\beta}_1$

$\hat{\beta}_2$

$\hat{\beta}_3$

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2,i} + \hat{\beta}_3 X_{3,i}$$

$\hat{\beta}_1$ Fitted value of Y for all X equal to 0

$\hat{\beta}_2$

$\hat{\beta}_3$

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2,i} + \hat{\beta}_3 X_{3,i}$$

$\hat{\beta}_1$	Fitted value of Y for all X equal to 0
$\hat{\beta}_2$	One unit increase in X_2 is associated with a $\hat{\beta}_2$ unit increase in Y, holding X_3 constant
$\hat{\beta}_3$	

Interpreting Regression Coefficients

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2,i} + \hat{\beta}_3 X_{3,i}$$

$\hat{\beta}_1$	Fitted value of Y for all X equal to 0
$\hat{\beta}_2$	One unit increase in X_2 is associated with a $\hat{\beta}_2$ unit increase in Y, holding X_3 constant
$\hat{\beta}_3$	One unit increase in X_3 is associated with a $\hat{\beta}_3$ unit increase in Y, holding X_2 constant

Inference on Multiple Regressors

$$
\begin{aligned}
\text{Unrestricted} \quad Y &= \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \\
\text{Restricted} \quad Y &= \beta_1 + \beta_4 X_4 + u
\end{aligned}
$$

$H_0 : \beta_2 = \beta_3 = 0$
H_1: at least one "$=$" doesn't hold
$F_{q,n-k} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(N-K)}$
Reject if $p < \alpha$
STATA: test X_2 X_3

RSS_{UR} :Residual Sum of Squares, unrestricted
RSS_R :Residual Sum of Squares, restricted
q :# of restrictions (e.g. number of variables set $=$ zero)
n :sample size
k :# of variables in unrestricted, including constant

Standard OLS Assumptions - Multiple Regression

- Linear in Parameters
- Random Sampling
- Sample variation in the explanatory variable
- Zero Conditional Mean: $E[u|X] = 0$
- Homoscedasticity: $Var[u|X] = \sigma^2$
- No Perfect Collinearity

Outline

Dummy Variables

$$r_t = \alpha + \beta(r_{M,t} - r_f) + \gamma_1 D_t + \gamma_2 (r_{M,t} - r_f) \times D_t + u_t$$

where $D_t = 1$ during recessions and 0 otherwise.

$D = 0 \quad E[r_t] = \hat{\alpha} + \hat{\beta}(E[r_{M,t}] - r_f)$

$D = 1 \quad E[r_t] = (\hat{\alpha} + \hat{\gamma}_1) + (\hat{\beta} + \hat{\gamma}_2)(E[r_{M,t}] - r_f)$

	$E[r_t \mid E[r_{M,t}] - r_f = 0]$	$\frac{\partial E[r_t]}{\partial E[r_{M,t}] - r_f}$
Non-Recession	$\hat{\alpha}$	$\hat{\beta}$
Recession	$\hat{\alpha} + \hat{\gamma}_1$	$\hat{\beta} + \hat{\gamma}_2$

$\hat{\gamma}_1$: difference in mean returns between recession and non-recession periods, when $E[r_{M,t}] - r_f = 0$.

$\hat{\gamma}_2$: difference in "beta" effect between recession and non-recession periods.

Outline

Multicollinearity

$$HPR_t^{IBM} = \beta_1 + \beta_2 HPR_t^{IVV} + \beta_3 HPY_t^{IVV} + u_t$$

- Problem: Regressors are (near) linear combinations of each other
- Consequence: OLS invalid (Unbiased but std errors too large)
- Detection: Correlation among regressors; VIF
- Remediation: Drop or combine troublesome regressors

Heteroscedasticity

$$\text{wage}_i = \beta_1 + \beta_2 \text{education}_i + u_i$$

- Problem: Variance of disturbance is not constant
- Consequence: Unbiased, but std errors are incorrect
- Detection: Plot residual (variance) on combo's of X; BP and White tests; STATA: estat hettest
- Remediation: Hetero. robust std errors

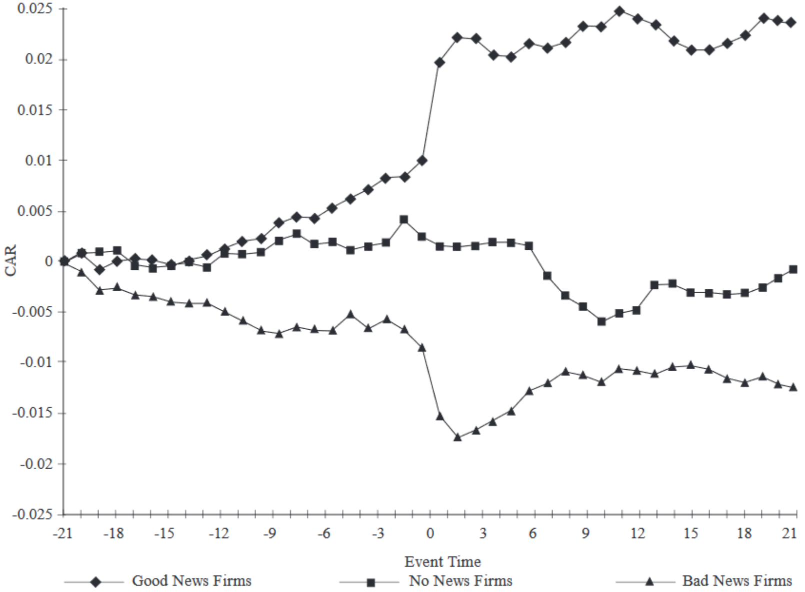
Serial Correlation

$$r_t = \beta_1 + \beta_2 F_{t-1} + u_t$$

- Problem: $Corr(u_t, u_{t-h}) \neq 0$
- Consequence: Unbiased but std errors are wrong
- Detection: Plot residual against each other over time; DW Test
- Remediation: Add lagged "Y", or use Serial Correlation robust std errors

Outline

Motivation ala Campbell, Lo, and MacKinlay



Steps for Conducting Event Study

1. Define the event window
2. Identify universe of assets
3. Define baseline return model
4. Define the estimation window
5. Calculate "abnormal" returns
6. Conduct tests and create visualizations

Define event and estimation windows

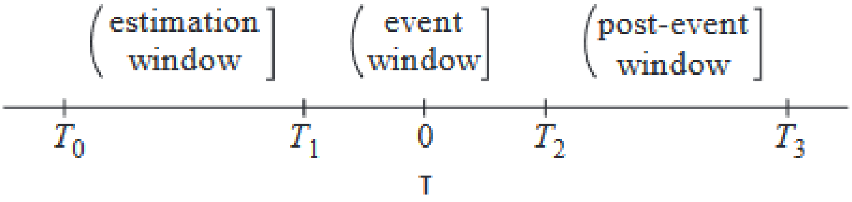


Figure 1. Time line for an event study.

Define baseline return model

Myriad models are possible. Two of the most common:

Constant Mean Return Model

$$R_{i,t} = \mu_i + \eta_{i,t} \quad \eta \sim N(0,1)$$

Market Model

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \eta_{i,t} \quad \eta \sim N(0,1)$$

Calculate and Report Abnormal Returns

Abnormal returns

$$AR_{i,\tau} = R_{i,\tau} - E[R_{i,\tau}|T_1] \;\; \forall \tau \in \{T_1, ..., T_2\}$$

where the $E[R_{i,\tau}|T_1]$ is the baseline model for returns using information up until the beginning of the event window T_1. Notice that $\hat{AR} = \hat{\eta}$

Cumulative Abnormal Returns

$$CAR_i(\tau_1, \tau_2) = \sum_{\tau=\tau_1}^{\tau_2} AR_{i,\tau}$$

where $T_1 < \tau_1 \leq \tau_2 \leq T_2$. These cumulative abnormal returns can then be aggregate across assets for each point in time, or aggregate across time for a given asset.