QUANTITATIVE METHODS

A: REGRESSION ANALYSIS

- Focus on **Linear Regression**
- Y: the **Dependant Variable**
- X: the **Independent Variable**
- Forms of Regressions:
  - *a.)* **Linear**: $Y = a + bX$
  - *b.)* **Loglinear**: $Y = ab^x \rightarrow \log Y = \log(a) + (\log(b))\, x$
  - *c.)* **Power Function**: $Y = aX^b \rightarrow \log Y = \log(a) + (b)\log(X)$
- Linear Regression focuses on the AVERAGE Relation between X & Y
- Look to the Line of Best Fit (sum of least squares method)
- Y → the Dependant Variable
- a → the Y intercept: in Modern Portfolio Theory, a.k.a. the Alpha of the Stock
- b → the Slope (REGRESSION Co-Efficient); in Modern Portfolio Theory, a.k.a. Beta of Stock
- *Example of a Simple Regression*

% Δ Market (Y)	% Δ Stock (X)
5	10
(10)	(15)
10	15
0	5
(10)	(5)

This Can Be Solved for on the Texas Instrument Calculator; 1st enter Data then do Statistical Program & Solve: $Y = 3.28 + 1.28X$

The Measure to which the line Does NOT fit the Data is the **Standard Error of the Estimate, S_{YX}**

- $S^2_{YX} = (\sum [Y_A - Y_E]^2) / (n-2)$
- $S_{YX} = (S^2_{YX})^{1/2}$
  - If on the Exam, Construct a Table to do the Following Calculations

X	Y_A	$Y_E = a + bX$	$Y_A - Y_E$	$(Y_A - Y_E)^2$
5	10	9.68	.32	.1024
(10)	(15)	(9.52)	(5.48)	30.0304
10	15	16.08	(1.08)	1.1664
0	5	3.28	1.72	2.9584
(10)	(5)	(9.52)	4.52	20.4304
				54.6808

$S^2_{YX} = (\sum [Y_A - Y_E]^2) / (n-2) \rightarrow 54.688 / (5-2) = 18.2293$

$S_{YX} = (S^2_{YX})^{1/2} \rightarrow (18.2293)^{1/2} = \pm$ **4.27**

Like OTHER estimates of standard error (like standard deviation), **68%** fall within 1 standard error of the line, and **95%** fall within 2 standard errors of the line. However, future data will not necessarily lie within 1 or 2 regression lines' standard error due to the fact that the sample data is NOT a complete population, but only an estimate of true values. Note, the confidence levels are Parabolic for future data.

B: CORRELATION & COVARIANCE

- **R^2 (the Coefficient of DETERMINATION)** measures the Goodness of Fit of a Regression Relation, it is derived from the Standard Error of Estimate. It is Defined as the Percentage of the Total Variation in Y that is Explained by the Regression Equation. It is obtained by dividing the Explained Variation in Y by its Total Variation.

  Explained Variation in Y = $S^2_Y - S^2_{YX}$

  $\mathbf{R^2_{YX}} = [S^2_Y - S^2_{YX}] / S^2_Y =$ **$\{1 - (S^2_{YX})/(S^2_Y)\}$**

  $\mathbf{r_{YX} = (R^2_{YX})^{1/2}}$

- $\mathbf{r_{YX}}$ is called the **correlation coefficient** and it measures the degree of linear association between X & Y. Since it is the Square Root of $\mathbf{R^2_{YX}}$, it may be either positive or negative. Although the Calculator gives it, may need to first adjust it.

  NOTE: When Calculating in Stat. Mode on the TI, will be given S_Y,. Square it to get S^2_Y. S^2_{YX} will have to be calculated as in the above calculation of determining the Standard Error of Estimate (and squaring it).

- **Covariance** is an Alternate way of calculating Correlation Coefficients.

  $\mathbf{COV_{YX} =([\sum (X - \underline{X}) (Y - \underline{Y})] / N)}$

  $\mathbf{r_{YX} = [(COV_{YX}) / (\sigma_Y \sigma_X)]}$

  If on the Exam, Might need to construct a table to perform the calculations

X	**Y**	**$X - X_{avg}$**	**$Y - Y_{avg}$**	**$(X - X_{avg})(Y - Y_{avg})$**
5	10	6	8	48
(10)	(15)	(9)	(17)	153
10	15	11	13	143
0	5	1	3	3
<u>(10)</u>	<u>(5)</u>	<u>(9)</u>	<u>(7)</u>	<u>63</u>
(5)	10	0	0	410

Note: TI will give σ_X and σ_Y. It will also give r_{YX}. However, to get the true R^2_{YX}, one must make an adjustment due to the small sample size.

Adjusted $R^2_{YX} = 1 - (1 - r^2_{YX})[(n-1)/(n-2)]$

Adjusted r_{YX} = (Adjusted R^2_{YX})

C: TESTING THE SIGNIFICANCE OF REGRESSION PARAMETERS

- It is Important for the Regression Coefficient to be SIGNIFICANT (a non-zero). Use a T-test, with the Null Hypothesis that the Regression Coefficient (b) = 0.
- $\mathbf{t_{calc} = |(b - b_0) / S_b|}$
- Since in the Null, b_0 = o, really, $\mathbf{t_{calc} = |b / S_b|}$
- If **$|t_{calc}| > t_{crit}$, ACCEPT the Null (ergo Insignificant)**
- If **$|t_{calc}|$ # t_{crit}, REJECT the Null (ergo Significant)**
- Critical Values are determined by Table… Depends on **Degrees of Freedom (n-2), & Level of Significance (.05)** → $t_{.05/2,\ 3\ d.f.}$
- The Exam will generally give you the S_b (standard error of the Regression Coefficient)
- Can Likewise Test for the Y Intercept (**a**) and Correlation Coefficient ($\mathbf{r_{YX}}$). Note, must first try to adjust it.
- May also test for Values Other Than 0, using b_0 = 1, or something as the Null Hypothesis.
- Will also Need to Determine Confidence Limits on the Regression Coefficient (b)

  A 95% Confidence Interval on the Regression Coefficient b would be:

  $b_{min} < b < b_{max}$

  $b_0 - t_{.05/2,\ 3\ d.f.} S_b < b < b_0 + t_{.05/2,\ 3\ d.f.} S_b$

D: MEASURING & FORECASTING GROWTH

- 2 Ways of Plotting Growth Rates. Simple Compound Calculation of Start & End (Average Growth). Or Conversion to Log Linear to Linear Regression. Not Likely to be on test.
- Interpretation of Difference: Calculation will Give the Actual Historic Growth Rate, where as Log Linear is a Good Predictor of the Future.

E: MULTIPLE LINEAR REGRESSION

- FORM: $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n$
- Often, will get the Following Multiple Regression Output from a Computer.

No. of Variables……………………3

Regression Coefficient	Value	Standard Error	t-Value
1	-.38	.23	-1.65
2	1.25	1.15	1.09
3	.05	.01	5.00

Intercept…………………………..0.157
(std. Error)………………………...0.083
No. of Observations……………100
Std. Error of Estimate…………….0.358
Coefficient of Mult. Correlation r...0.705
D.W………………………………2.18
F…………………………………31.62

Correlation Matrix

	X_1	X_2	X_3
X_1	1.00		
X_2	0.30	1.00	
X_3	0.15	0.20	1.00

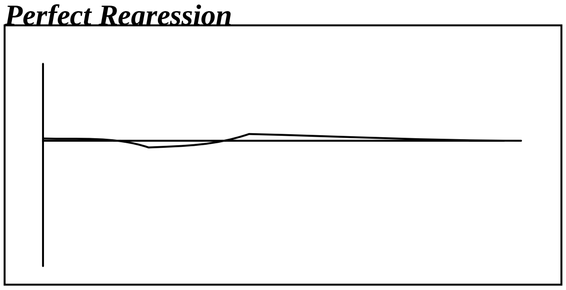
Partial Correlations

$r_{YX1X2X3}$=.35 $r_{YX3X2X1}$=.45 $r_{YX2X3X1}$=.46
$r_{YX2X1X3}$=.28 $r_{YX1X3X2}$=.58
$r_{YX3X1X2}$=.41

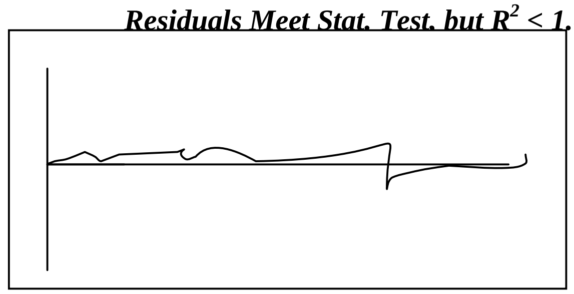
1. **Determine the Equation**
   - $Y = .157 - .38X_1 + 1.25X_2 + .05X_3$
     (.083) (.23) (1.15) (.01)
2. **Perform Significance Tests**
   - Note, when performing Multiple Linear Regression, **DF** = n – k – 1
   - If $t_{calc} < t_{crit}$ ACCEPT, & if $t_{calc} > t_{crit}$ REJECT the NULL
     - Here, Intercept is Insignificant, b_1 & b_2 are also Insignificant; but b_3 is significant
     - Since the *Standard Error of the Estimate* is .358 (S_{YX}), 68% of the Observations fall within +/- .358 of the Regression Line, 95% of the Observations fall within +/- (2)(.358) of the Regression Line.
     - Since the *Coefficient of Multiple Correlation* (r_{YX}) is .705, it is known that the *Coefficient of Multiple Determination* (R^2_{YX}) is .497. However, as variable increase, so will R^2, so it is not the best measure of goodness of fit. To see how significant R^2 is, do an **F-test** given the F value in the printout, and a critical F value, to determine significance of variables, taken together as a group.
3. **Look at the Correlation Matrix**
   - Ideally, want high correlation between X & Y, and Low correlation between X's themselves.
   - **Multicolinearity** occurs when the variables are highly correlated to each other. This will bias the R^2 upward and give a false impression of a good fitting equation. Try other equations, try to obtain more data. Symptoms: High R^2, but statistically insignificant variables.

4. **Look to the Partial Correlation Coefficients**
   - Measures the correlation between Y and some of the Independent Variables, while the other variables are held constant. I'm not really sure. Hopefully not on test.
5. **Perform Durbin Watson to see if there is AUTOCORRELATION**
   - Autocorrelation occurs when the residuals are systematically related. Rule of thumb, no Autocorrelation if DW is between 1.45 & 2.55. A/C may be caused by the omission of a relevant independent variable or by the cyclical behavior of the independent variables. Tends to reduce the standard errors of regression coefficients. Increases the probability that an insignificant variable will be viewed as significant. Can use Generalized Least Squares to Overcome.
6. **Plot the Residuals**

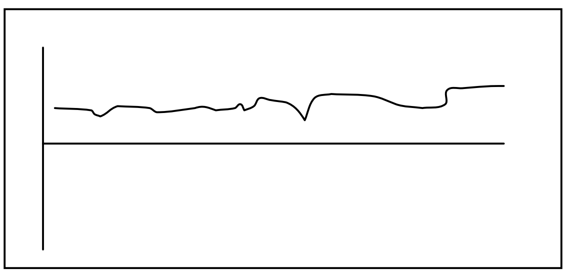
Perfect Regression



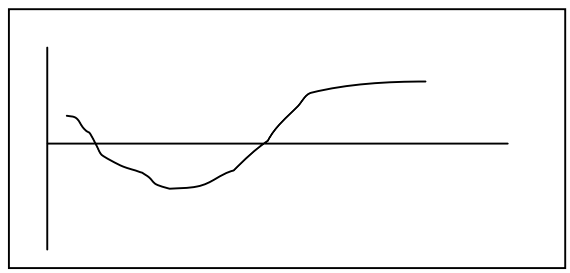
Residuals Meet Stat. Test, but $R^2 < 1.0$



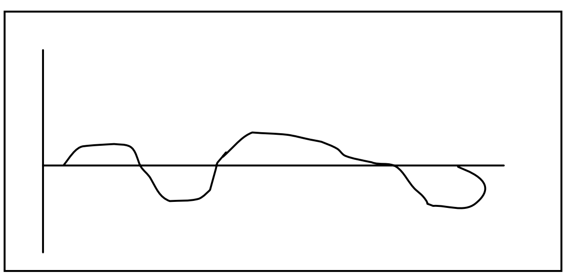
Bias



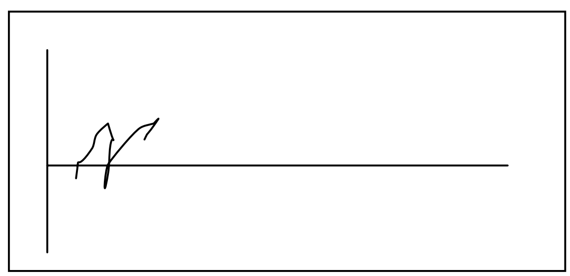
Non-linear, logarithmic relation



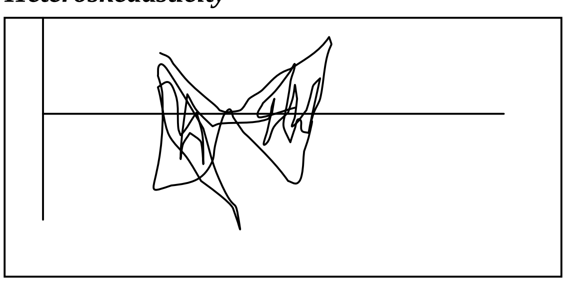
Autocorrelation



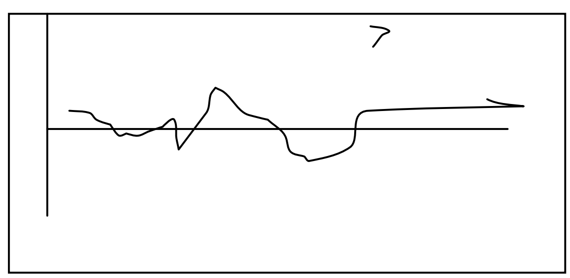
Omitted Trend Variable



Heteroskedasticity



Outliers



F: STATISTICAL PREDICTION & ECONOMETRICS

- Using Statistical Techniques to establish predictive relationships is called Statistical Prediction. Involves several steps.
  1. ***Use of Economic Theory to Identify which Variables should determine the Variable that is to be predicted as well as the Economic Form of the Relationship.***
     - Could Use dummy Variables in Model (a.k.a. Dichotomous or Binary Variables) as well as Interval (Continuous) Variables.
  2. ***Collect the Data for Analysis***
  3. ***Perform Regression using all Statistical Tests.***
     - Regression is Based on Several Assumptions
       a.) the Regression relationship is properly specified
       b.) the data is measured correctly
       c.) there is NO multicolinearity
       d.) Good Tests:
         - Good R^2
         - Pass F-test, t-test
         - No Multicolinearity
         - No Autocorrelation (pass DW test)
         - No biases
         - No Heteroskedasticity
         - Normal Distribution of Residuals.
  4. ***Sources of Error in Econometrics***
     a.) Measurement Error
     b.) Specification Error
     c.) Statistical Error
     d.) Human Indeterminacy