|  |
| --- |
|  |
| Nominal Personal Savings as a function of Disposable Income, 1964 to 2010 |
| Linear Regression Analysis |
|  |
| **Caleb** |
| **2/28/2014** |

|  |
| --- |
|  |

Introduction

The objective of this paper is to perform a linear regression analysis on a simple Keynesian savings function dependent on disposable personal income. Developed and covered in full detail in *The General Theory of Employment, Intrest, and Money*, Keynesian consumption and savings functions are used to calculate the total amount of consumption/savings in the economy. Due to the scope of this paper a simple linear regression is used.

The Model

Personal savings as a function of disposable personal income utilizes the Keynes model of consumption. Understanding the savings function used in this model requires an understanding of the following linear consumption function shown in Table 1,

|  |
| --- |
| **Table 1: The Consumption Model**  [1]  C = Consumption  Autonomous consumption  c = Marginal propensity to consume  Y = Disposable personal income |

Autonomous consumption () is defined as consumption independent of personal income. The independent nature of autonomous consumption implies that as personal income varies, will remain a constant value. Expressed graphically, autonomous consumption represents the point where the function intercepts the vertical axis. When working with savings functions, we use to denote our autonomous savings term. If disposable income were equal to zero we would expect to be negative reflective of dissaving occurring to maintain positive.

Marginal propensity to consume (), abbreviated MPC, is the derivative of consumption with respect to disposable income, . The MPC is approximately equal to the fractional expression, by which consumption increases given a marginal increase in disposable income, . Graphically the MPC represents the instantaneous slope of the consumption function. Typically, the value of the MPC falls between 0 and 1. In the context of working with the savings function we are concerned with the marginal propensity to save, abbreviated MPS which is, . We suspect the sign of the MPS to be positive reflecting some proportional increase in savings as income increases.

The term Y represents disposable personal income or personal income minus personal taxes paid. To derive the savings function begin by making the assumption that disposable income () is equal to the sum of consumption () and savings (),

[2]

By substituting equation [1] for and solving for an equation that models savings as a function of income is shown in Table 2.

|  |
| --- |
| **Table 2: The Savings Model**  [3]    = Personal savings  = Autonomous savings (vertical intercept)  = marginal propensity to save (slope of the function)  = disposable personal income |

Hypothesis

The results produced by a regression estimating an economic model such as the savings function are notably less valuable unless their statistical significance confirmed. With the use of statistical software, the significance of the results being produced by a given regression can be tested. The statistical significance of a result is the likelihood of whether the result was the outcome of chance given a prescribed probability threshold. A statistical software package known as SPSS is used to calculate statistics for the estimated regression. The statistical analysis performed contains two tests to verify the statistical significance of various parts of the function. The two tests performed consist of an F-test designed to test the significance of the of the savings function as a whole and t-tests designed to test the significance of the individual coefficients.

The desired outcome of the tests is either to reject the null hypothesis and affirm the alternative hypothesis. Comparing the value of the resulting test statistic generated by SPSS to a value from a table of a given F or t distribution, the null hypothesis will be affirmed or rejected. The value from a table of given F or t distributions is known as a critical value and is dependent on the degrees of freedom and confidence interval. The confidence interval is used as an indication to the reliability that a range of specific values contain an unknown population parameter. Over the course of this analysis all tests are interpreted at α=.05, this implies that there exists a 5% chance of wrongly rejecting the null hypothesis, a type II error. The amount 5% is referred to as the significance level and is typically represented by the Greek letter alpha (α).

Using the F-test we assess the significance of the estimated regression. The null hypothesis was that the regression is not statistically significant whereas the alternative hypothesis states that the regression is statistically significant as shown in Table 2. By comparing the value that SPSS gave for the F-test statistic to the value taken from the F-table distribution, we then decide whether to reject or affirm our null hypothesis at a .05 significance level.

The t-test is used to measure the significance of each individual coefficient. The null hypothesis and alternative hypothesis differ from the F-test in that we no longer are looking for significance of the whole function; instead, we assess the statistical significance that each coefficient is different from zero. The null hypothesis states that the value of coefficients are not significantly different from zero, statistically speaking, thus = 0 and = 0 as shown in Table 3. The alternative hypothesis states that the coefficients are statistically significant difference from zero, ≠ 0 and ≠ 0 as shown in Table 2. As with the F-test, the results are evaluated at a 5% significance level.

|  |
| --- |
| **Table 3: The Hypotheses**  F-test (statistical significance of function)  : The regression is not statistically significant  The regression is statistically significant  t-test (statistical significance of coefficients)  : |

Methodology

Regression is used in statistical analysis as a means to make predictions based upon statistical significance of the testing. The general form of the regression equation is shown in equation [4],

[4]

Equivalently the regression equation can be expressed in matrix notation shown below[[1]](#footnote-1),

[5]

where the matrices , X, , and have n rows depending on the n amount of observations in the data set. The term y is the dependent variable, is a constant equivalent to the autonomous savings term, is the coefficient in front of the independent variable x equivalent to MPS () and disposable income (), respectively and is an error vector corresponding to values determined by the SPSS output. Due to the scope of the project, a simple version of the regression is implemented. The simple regression fits a straight line to a set of n points in a manner to minimize the vertical distance between the regression line and the actual data points. This implies that the simple regression is operation as a minimization function of the sum of the squared residuals,, where Y is the observed Y data point and Y’ is the value predicted by the regression line. The least squares method produces the line of best fit to a set of points.

R2 represents a measurement of how well the regression line fits the data set. Defined on the interval [0,1], the R2 is known as the coefficient of determination and is equal to the explained variation divided by the total variation as shown in Table 3. A R2 value of one implies that the regression line is a perfect fit to the actual data points. We interpret R2 as a percentage of variation in savings (or consumption) explained by the regression equation.

SPSS reports the F-ratio, which represents the variation in Y attributed to the regression divided by the unexplained variation as shown in Table 3. We find F-critical by comparing the value of the F-ratio to the value found using the degrees of freedom and appropriate significance level in a table for an F distribution we can accept or reject the null hypothesis. The F-table is unique in that it requires two degrees of freedom; one for the numerator (the number of coefficients including the intercept – 1); and one for the denominator (the number of observations – the number of coefficients including the intercept). If the value of the F-ratio is greater than or equal to the critical value in the table, then we reject the null hypothesis and confirm the regression is statistically significant, thus implying there is a relationship between the independent variables and the dependent variable.

The t statistic represents the regression coefficient divided by the standard error of the regression coefficient as shown in Table 4. Confirming whether or not to reject the null hypothesis for a t-test is handled in the same manner as with the F-test, that is selecting the appropriate degrees of freedom and appropriate significance level. There is a slight variation to the t statistic due to the two-tailed characteristic of the distribution; the t statistic is expressed as an absolute value. If the absolute value of the t-statistic is greater than or equal to the absolute value of critical t we reject the null hypothesis and confirm the alternative hypothesis. The confirmation of the alternative hypothesis implies that the coefficient is significantly different from zero

|  |
| --- |
| **Table 4: Statistical Formulas**  R2 :The coefficient of determination ≝  F:The F-statistic ≝ where K is the # of coefficients (including the intercept)  t:The t statistic ≝ where is the standard error of the regression coefficient |

Data

The linear regression was performed using personal savings and income data from the *2013 Economic Report of the President*. Specifically, the data was gathered over the time interval [1964-2010] from the Disposable Personal Income and Personal Savings columns in Table B-30 and is measured in billions of dollars. An important distinction to make is that the time period for the data is cyclically consistent, that is, the data begins in an expansionary year and ends in an expansionary year. (Research)

The data values for disposable income exhibit a steady positive trend from the year 1964 until the year 2009 where a negative dip occurs suddenly, this is shown in Plot 1.

|  |
| --- |
| **Plot 1**    Notice the dip in disposable personal income occurring around the year 2009. |

The behavior of the personal savings data erratic and fluctuates often. From the years 1964-1979 savings follow an increasing trend followed by a period of oscillatory behavior. The oscillatory behavior is dominant over the years 1979-2003 when a massive spike occurs in personal savings, this is shown in Plot 2.

|  |
| --- |
| **Plot 2**    Notice the erratic behavior of person savings in particular the sudden increase in personal savings during the 2008 economic crisis. |

Important statistical characteristics of the variables are shown in Table 5.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 5: Descriptive Statistics**  Disposable Personal Income (Measured in billions of dollars)   |  |  |  |  | | --- | --- | --- | --- | | Maximum Value | Minimum Value | Mean Value | Standard Deviation | | 11,127.1 | 514.3 | 4,290.7 | 3,368.7 |   Personal Savings (Measured in billions of 2013 dollars)   |  |  |  |  | | --- | --- | --- | --- | | Maximum Value | Minimum Value | Mean Value | Standard Deviation | | 592.3 | 40.5 | 220.95 | 127.4 | |

Using SPSS a scatter plot of personal savings as a function of disposable personal income was generated and is shown in Figure 1.

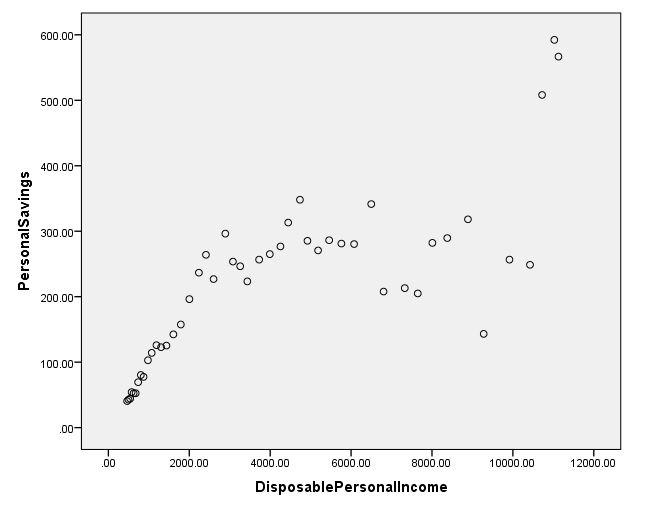


Figure 1

**Disposable Personal Income**

As seen in figure 1 as the points increase vertically they fan out horizontally. Initially we see a positive relationship. With the exception of the points in the top right quadrant, the points on the scatter plot resemble a quadratic function. Overall there is a positive relationship between disposable personal income and personal savings.

Results

The results of the regression estimation of the coefficients are shown in Table 6.

|  |
| --- |
| **Table 6:Nominal Personal Savings as function of Disposable Personal Income**  S = 95.043 + 0.29Y  (4.917)  (8.248)\*\*  R2 = .776 F = 68.031  (t-statistic) \*t is significant at .05 |

The R2 value is .776 which implies that 77.6% of the variation in savings is explained by the regression equation. The value of the F statistic is equal to 68.031 compared to the F critical value of 4.05 for a given significance interval of 95% with 1 degree of freedom in the numerator and 45 degrees of freedom in the denominator (Health). In this case F statistic ≥ F critical, so we reject the null hypothesis. Because we reject the null hypothesis, and confirming the alternative hypothesis yielding the conclusion, affirmation of the alternative hypothesis signifies there is a relationship between the disposable personal income variables and the savings variable. The t statistic is equal to 4.917 for the constant (autonomous savings) and 8.248 for the slope (MPS); both of these are greater than the t critical value of 2.0141 given significance level of 5% and 45 degrees of freedom (Washington). Therefore, we reject the null hypothesis and affirm that the coefficients are significantly different from zero. The constant term implies that if disposable personal income were zero, personal savings would be equal to 95.043 billion dollars. The slope term represents the change in savings given a marginal increase in disposable personal income and is equal to .029. While the MPS is non-zero the impact it has on savings overall is surprisingly low.

Conclusion

The linear regression analysis performed over the years 1964-2010 confirms a relationship between the dependent and independent variable.  SPSS estimated the marginal propensity to save to be .029; this is small because the marginal propensity to consume is high.  The positive sign of the constant was unexpected.  The constant being positive can be explained by looking at the residual and scatter plot and analyzing the general concavity the points constitute.  The regression is fitting a line to the data points; this forces the constant term to be positive.  This suggests a specification bias in linear Keynes model.

     Future research of Keynesian savings functions could be improved a number of ways.  By changing the functional form of the regression, a better fitting line is achieved.  In addition to changing the functional form of the regression, the admission of additional variables to the equation would provide a better fit to the data.

# Works Cited

Advisers, Chairmen of the Council of Economic. "Economic Report of the President." 2013. Document.

Health, Boston University School of Public. "CRITICAL VALUES for the "F" Dist." 2013. *http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704\_HypothesisTesting-ANOVA/ftable.pdf.* PDF. 28 February 2014.

Research, The National Bureau of Economic. *http://www.nber.org/cycles/cyclesmain.html*. 2013. PDF. 28 February 2014.

Sharabati, Walid. *http://www.stat.purdue.edu/~wsharaba/stat512/topic3.pdf*. 9 August 2013. Handout. 28 February 2013.

Washington, University of. "Crititical valuues for t distribution." 2012. *http://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/.* Webpage. 28 February 2014.

1. Terms in bold imply the respective matrix ie.= [↑](#footnote-ref-1)