Life-Cycle Savings Hypothesis

Arden Miller

August 25, 1998

Executive Summary

The life-cycle savings hypothesis suggests that the savings rate for a country should be lower if non-members of the labour force constitute a large percentage of the population and should also depend on the rate of income growth. The level of income itself should not be an important factor according to this theory. Data collected on 50 countries over the period of 1960-1970 was analysed to see if it was in accord with this theory. The data is in general agreement with in that it indicates that variables associated with the percentage of the population not in the work force and the growth rate of disposable income are important predictors of savings whereas the level of disposable income is not an important predictor. However, it also indicates that much of the variability in savings from country to country is left unexplained by these predictors.

1 The Data

The data consists of observations for 50 countries taken over the period of 1960-1970. Values for the following five variables were recorded:

savings: the average percentage of aggregate personal savings over disposable income.

growth: the average percent growth rate of disposable income.

disp.inc: the average level of per-capita disposable income (measured in US dollars).

pop.75: the average percentage of the population over 75 years of age.

pop.15: the average percentage of population under 15 years of age.

The variables **pop.15** and **pop.75** are being used as surrogate variables for percentage of the population that are non-members of the labour force. Since the life-cycle savings hypothesis is not directly interested in the population profile of countries, these two variables were added together to create a new variable called **non.work**. Since in developing countries people may enter the work

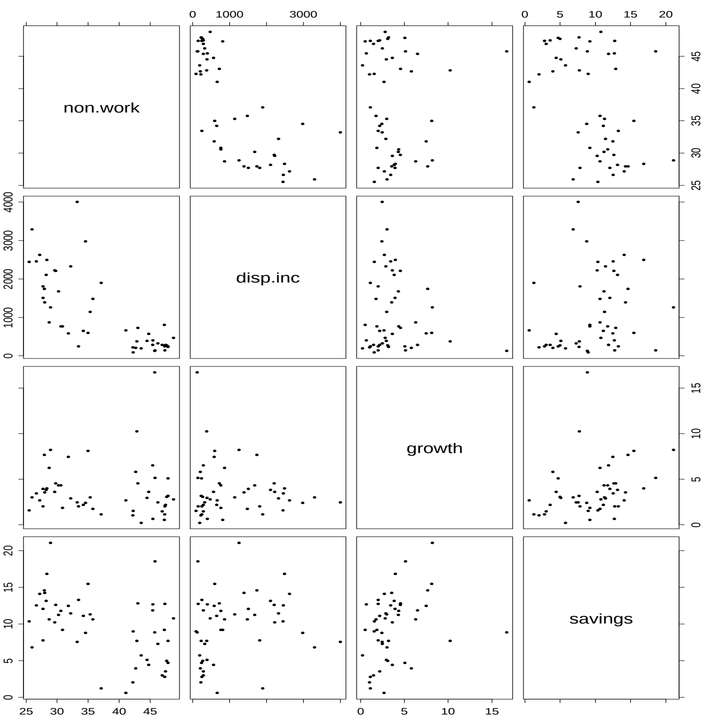

Figure 1: Pairwise scatterplots of data

force at an early age, it is not clear how well this new variable will perform as a surrogate for the percentage of the population that are non-members of the labour force.

Figure 1 contains pairwise scatterplots of the data. These plots suggest that the data may well be in accord with the life-cycle savings hypothesis. There appears to be a negative relationship between **non.work** and **savings**, a positive relationship between **growth** and **savings**, and no discernable relationship between **disp.inc** and **savings**. Another interesting feature of this data is that the countries appear to form two distinct clusters relative to the variable **non.work**. An inspection of the countries involved indicates that one of these groups consists of the industrialised countries and the other group consists of the developing countries.

2 A Model for Savings

The data was used to create the following model that relates **savings** to the other variables:

$$\textbf{savings} = 17.57 - 0.23\textbf{non.work} + 0.71\textbf{growth}.$$

The variable **disp.inc** is not in the model since the analysis indicated that including it does not appreciably increase the models ability to predict **savings**. This model indicates that if **growth** is held constant, the savings rate for a country tends to decrease by 0.23 for each percentage point increase in the percentage of the population outside the ages of 15 to 70. It also indicates that if the percentage of the population outside the ages of 15 to 70 is held constant, the savings rate for a country tends to increase at a rate of 0.71 per each percentage point increase in growth rate. These results agree well with the life-cycle savings hypothesis. However, the presented model only accounts for approximately 30 percent of the variability in **savings** so much of the differences in savings rates between countries is left not explained. This means there could well be important predictors of savings not considered in this analysis.

It was found that Libya was markedly different than the other countries in that it had an extremely high growth rate over this period and a much lower rate of savings than would be expected given it's rate of growth. As a result, the inclusion of Libya in the data set has a big impact on the coefficient for **growth** in the fitted model. For this reason, the model presented for **savings** was obtained ignoring the observation for Libya. If the observation for Libya is not ignored then the coefficient for **growth** would change from 0.71 to 0.44.

3 Industrialised and Developing Countries

As was noted, the data can be divided into two groups bases on the values of **non.work** that correspond to the industrialised and the developing nations. It is reasonable to investigate these two groups separately to determine if our conclusions are the same for both groups. If this is done we find that for the industrialised countries, there is evidence that **growth** and marginal evidence that **non.work** should be included in a model for **savings** but no evidence that **disp.inc** should be included. For the developing countries, there is no evidence that any of **growth**, **non.work**, or **disp.inc** should be included in a model for savings.

This suggests that the mechanism that determines savings may be different in industrialised countries and in developing countries. However, the data is inconclusive on this point. By splitting the data into two groups we have substantially reduced the precision with which we can estimate the coefficients in our models. The uncertainty for the estimated coefficients for the models obtained using the split data are large enough that we cannot rule out the possibility that the true coefficients are the same for both groups.

Statistical Appendix

Fitting the regression model consisting of the linear terms in all of the regressors yielded the following results:

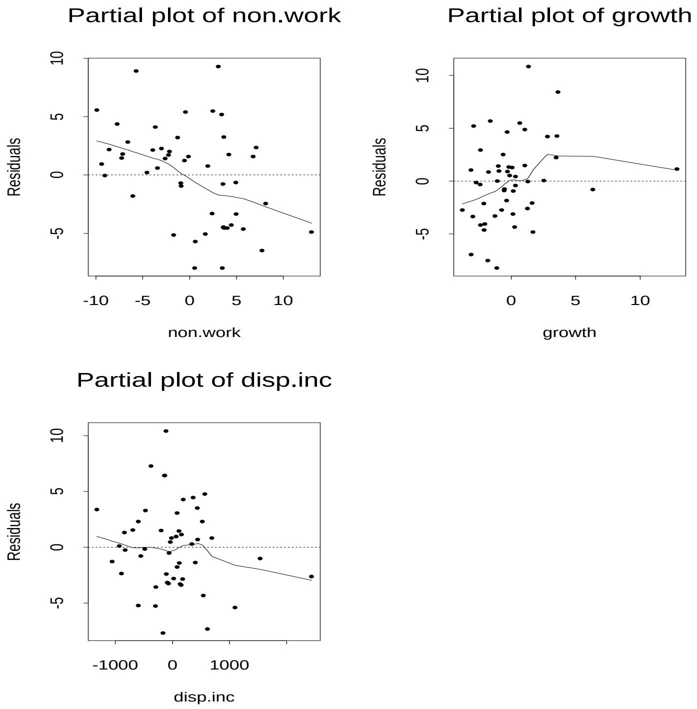

Figure 2: Partial regression plots

	value	std. err.	P-value
intercept	21.6096	4.8840	0.0001
non.work	-0.3336	0.1039	0.0024
growth	0.3910	0.1969	0.0530
disp.inc	-0.0008	0.0008	0.3224

The partial regression plots for this model are given in Figure 2. They do not indicate the need for any of the regressors to be transformed. Since it has such a large P-value, **disp.inc** was dropped from the model.

	Value	std. err.	P-value
intercept	17.5691	2.7461	0.0000
non-work	-0.2555	0.0685	0.0005
growth	0.4398	0.1907	0.0256

The residual plots for this new model are given in Figure 3. The plot of residuals versus fitted values and the normal probability plot of residuals do not indicate any serious problems with the model. The L-R plot indicates that observation 49 (Libya) is a high leverage point and observation 46 (Zambia) has

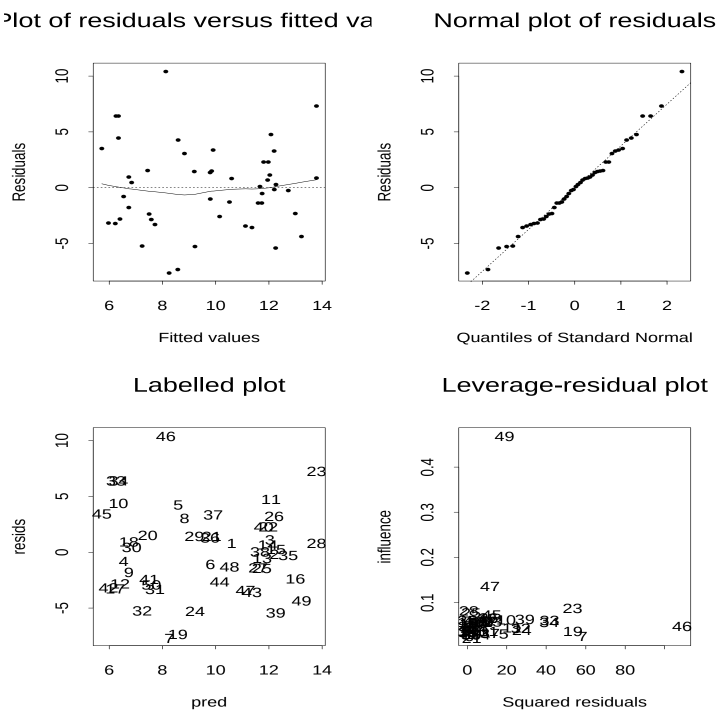

Figure 3: Residual plots

an usually large residual. Libya has an usually large Cook's distance (0.712) so the model was re-fitted with Libya deleted.

	Value	Std. Err.	P-value
Intercept	15.7363	2.9418	0.0000
non.work	-0.2288	0.0695	0.0019
growth	0.7057	0.2522	0.0075

There is a big change in the coefficient for **growth** but only a small change in the coefficient for **non.work**. This can be explained by noting that Libya is a high leverage point mainly due to its extremely high rate of growth. Residual plots for this model do not indicate any serious problems with this model.

The values for **non.work** form two distinct clusters that correspond to the industrialised countries and the developing countries. The data was divided into two groups: countries with **non.work** greater than 38 percent (developing countries) and those countries with **non.work** less than 38 percent (industrialised countries). Libya was not included in either group since it has already been found to be unusual. The two groups were analysed separately to see if the same conclusions would be reached for each group as were reached for the entire data set.

The fitted model for the developing countries indicated that none of the regressors were statistically significant.

	Value	Std. Err.	P-value
Intercept	-14.3062	20.9704	0.5038
non.work	0.4456	0.4489	0.3338
growth	0.4606	0.4345	0.3031
disp.inc	-0.0001	0.0052	0.9879

Partial regression plots do not clearly indicate the need to transform any of the regressors. The residual plots indicate one possible outlier (Zambia) but removing does not make any of the regressors significant.

The fitted model for the industrialised countries indicates that **growth** is significant and **non.work** is marginally significant.

	Value	Std. Err.	P-value
Intercept	18.4780	6.5176	0.0094
non.work	-0.2977	0.1803	0.1124
growth	0.9403	0.2931	0.0039
disp.inc	-0.0008	0.0007	0.2691

Again partial regression plots do not clearly indicate a need to transform any of the regressors and residual plots do not indicate any problems with this model.

The fact that **growth** is significant and **non.work** is marginally significant for the industrial countries but neither are significant for the developing countries seems to suggest that the life-cycle savings model may only apply to industrial countries. However, we cannot say for sure this is the case. If we compare the estimated coefficients for the two models taking into account their standard errors, it is possible that the true coefficients are actually the same for industrial and developing countries.

475.330 Assignment Two; Marking Guide

This assignment asks the students to use data collected on 50 countries to evaluate the life-cycle savings hypothesis. They were specifically asked to keep this in mind when doing their analysis. The assignment should be in two parts (same as assignment 1).

Report on what the data says about the life-cycle savings hypothesis (13 Marks)

This part of the assignment should discuss what the data indicates about the life-cycle savings hypothesis. The student's were specifically asked to produce a model and to explain what it means in non-statistical terms. They should then relate this model to the life-cycle savings hypothesis.

- 5 marks for a generally well laid out, coherent report. The reader should not have to search for the important parts among a lot of details. Give: (1) 5 marks for a clear, precise report, (2) 3 marks if the report would be difficult for a non-statistician to understand, and (3) 1 mark if it would be difficult for anyone to understand.

- 8 marks for a sensible discussion about the life-cycle savings hypothesis. Look for a sensible model and a correct interpretation of that model (3 marks for this part). In addition they need to relate their model(s) to the life-cycle savings hypothesis (3 marks). They should also discuss some (give 1 mark for each one they discuss intelligently up to a maximum of 2 marks) of the following issues that arise in this data set:

  - the suitability of using **pop.15** and **pop.75** as surrogate variables for the the percentage of non-members of the work force in a country (questionable especially for the developing countries).
  - combining **pop.15** and **pop.75** into a single variable since they are being used to represent the percentage of non-members of the work force (this is a good idea).
  - the data are clustered into two groups which correspond to industrialised and developing countries (this split is most evident in **pop.15** and **pop.75** or the variable formed by adding these together). It is a good idea to investigate whether the two groups are acting in a similar fashion (most people will divide the data into two sets but it could be done by defining new variables as well). They should find that none of the regressors are significant for the developing country group.
  - all reasonable models for this data have quite low R-square values indicating that much of the differences in savings rates between countries is not being explained by the regressors considered.

- Libya is a high leverage point that has a big impact on the fitted model. They should investigate what happens when Libya is removed from the data. A couple of other countries are also indicated as being either outliers or influential (Zambia and the USA, I believe).
- Any other issue that I may have missed.

Statistical Appendix (7 Marks)

Thus appendix should outline the reasons that they came to the conclusions they presented in the first part of the analysis. They are not required to give a detailed account of everything they did and they most certainly should not just have a whole bunch of computer output. Rather they should present a coherent account of the parts of their analysis that led them to their model for savings. For this assignment they need to discuss influence diagnostics and residual plots for their final model (including the output and plots is not absolutely necessary but it is helpful). Give: (1) 7 marks for a well documented account (2) 5-6 marks if they leave some important aspects unexplained or if they include lots of unimportant details (3) 3-4 marks if some of the conclusions they made are clearly not supported by their analysis (4) 2 or less marks if they just present output without reasons for their conclusions or if their reasons are not sensible.