Ecn 102 - Analysis of Economic Data
University of California - Davis
John Parman

February 7, 2011

Problem Set 4 - Solutions

This problem set will be graded and is due by 5pm on Friday, February 18th. It may be turned in earlier either in class or to your TA's mailbox in the economics department mailroom. You may work in groups but everyone in the group must write up their own solutions including creating their own graphs and tables.

The Marginal Propensity to Consume

At this point in the course, you have had a fair amount of practice working with data in Excel. Now the problem sets will shift their focus from walking through the steps in Excel to thinking about how to use the tools you now have to analyze economic questions. Consequently, an important part of the problem sets from this point on will be figuring out the best approach to take to answer the question of interest. There will be many correct ways of going about solving these problems (and many incorrect ways). This problem set will help you start making these decisions of how to do analysis on your own rather than simply following steps given to you. It will be up to you to decide when a variable should be transformed, when a new variable needs to be created, when observations should be dropped, and so on.

(a) The goal of this problem is to gain some insight into the marginal propensity to consume (MPC). The MPC is a concept economists use to describe how much of each additional dollar in income goes toward consumption. So if the value of a consumer's MPC is 0.4, when the consumer gets an extra dollar in income she will spend 40 cents of it and save the remaining 60 cents. Given this definition of the MPC, write down an equation that gives money spent on consumption goods as a function of income and the MPC. (Note that the MPC will rarely tell us exactly how much a person consumes given their income. Consumption will often times be a little higher or a little lower for many reasons other than just differences in income. Your equation should account for this additional variation in consumption with an error term.)

By the definition given above, consumption is just equal to income multiplied by the MPC. So our equation would look like:

$$C_i = MPC \cdot I_i \quad (1)$$

where C_i is person i's consumption and I_i is person i's income. However, consumption is also going to vary across individuals for reasons other than just changes in income. For example, a person may have unexpected medical bills or house repairs that they need to pay for, increasing his or her spending even though income hasn't gone up. To account for this, we should include an error term in our model:

$$C_i = MPC \cdot I_i + \varepsilon_i \tag{2}$$

(b) Consumers have to spend a certain amount of money even if they have no income. They still need to pay for necessities like rent and food. Let's assume that this is a fixed amount of spending and that all additional spending as income starts to go up from zero is determined by the MPC. Write down an equation that gives money spent on consumption goods as a function of income, the MPC and the fixed amount of spending on basic necessities (we'll call it autonomous consumption, C_a).

Now, our model should have individuals consuming a fixed amount C_a and then an additional amount that is proportional to their income:

$$C_i = C_a + MPC \cdot I_i + \varepsilon_i$$

We still include an error term for the same reasons as above.

(c) We are going to use data from the *Economic Report of the President* to try and learn about the marginal propensity to consume. The following website contains downloadable versions of various statistical tables from the 2010 *Economic Report of the President*:

http://www.gpoaccess.gov/eop/tables10.html

Suppose we wanted to get an estimate of the MPC for the United States. What would your ideal data look like?

The way we have been talking about the MPC is in terms of individual consumers. If we are trying to model the behavior of individual consumers, we would like to have data on individual consumers. In this specific case, cross-sectional data giving consumption and income for a large number of individuals would be particularly useful. We would like this sample of individuals to be a truly random sample of the population and therefore representative of the population. If we thought that the MPC may have changed over time, we may actually prefer to have panel data tracking consumption and income across individuals and over time. If we thought that the MPC may differ across different groups of people we would like our dataset to contain information on the characteristics that define these groups (variables like wealth, geographical location, family size, etc.). This would allow us to control for these characteristics when we estimate the MPC.

(d) Look through the data available from the *Economic Report of the President*. Most likely, your ideal data is not there. Determine which available data series would be best for estimating the MPC. How do these data differ from your ideal data? What implications does this have about how your estimate of the MPC may differ from the true value (or the value you would estimate with your ideal data)?

There are several different series that would make sense. For the solutions I have chosen the disposable per capita income in chained 2005 dollars from Table B31 as the measure of disposable income and the personal consumption expenditures per capita in chained 2005 dollars from the same table. It is very important that the data series you choose are measured in the same units. In other words, if you use per capita income you should be using per capita consumption. If you use nominal income you should be using nominal consumption. In this particular case, it makes sense to focus on per capita variables measured in real (rather than nominal) dollars. We have phrased our discussion in terms of an individual consumer which suggests we should focus on per capita figures. Given that we will be using observations from different years, it is important to use real values instead of nominal values to make certain that the units are the same from one observation to the next.

There are several problems with these data. First, they are aggregate data. Changes in income may be driven by different people in the population than changes in consumption. Additionally, the MPC and the value of autonomous consumption may be very different for different people in the population. Suppose that for most people, the MPC is fairly high but for the very wealthy, the MPC is very low (they have too much money to know what to do with). If changes in income are driven mostly by changes in income for the wealthy, we would estimate an MPC that is much lower than the average MPC among the population. We would prefer to have individual level data. This would let us see how person i's consumption changes when person i's income goes up, not when average income for the population goes up. A second problem is that we are limited to one observation per year. If the MPC is changing over time this is going to be a problem. Multiple observations at each point in time would help us deal with this problem.

(e) Choose the best data to estimate the MPC, download the data and combine it into a single Excel spreadsheet so that you can analyze it. Use your data to estimate the MPC assuming that autonomous consumption is zero (note that one of the regression options in Excel is 'Constant is zero'.) Based on your results, calculate a 90% confidence interval for the MPC.

See ps4-solutions.xlsx. Note that when we calculate the 90% confidence interval, we use $t_{\frac{\alpha}{2},n-1}$ instead of $t_{\frac{\alpha}{2},n-2}$. The reason for this is that we have forced the intercept to be zero so we are actually only estimating a single parameter (the slope coefficient). This means we need a minimum of one data point to get the estimate and then all of the data points after that are helping us refine that estimate. This is just like when we did univariate statistical inference before, we needed a minimum of one data point to estimate the mean and the data points after that first one were helping us get a better estimate. Therefore, we have $n-1$ degrees of freedom. In the standard bivariate regression, we are trying to estimate an intercept and a slope coefficient. To do this, we need a minimum of two data points and we therefore have $n-2$ degrees of freedom.

(f) Drop the assumption that autonomous consumption is zero. Use your data to get estimates of both autonomous consumption and the MPC. Give a 90% confidence interval for each of these estimates.

See ps4-solutions.xlsx.

(g) Use the results of your regression and a scatterplot of your data to argue whether the assumption in part (e) that the autonomous consumption is zero seems reasonable.

See ps4-solutions.xlsx for the scatterplot. If you looked just at the scatterplot, it looks like the intercept could very well be something very close to zero. However, from the regression results in part (f) we got a highly statistically significant negative intercept, suggesting that at any reasonable significance level we would reject the null hypothesis that C_a is equal to zero. What may seem odd is that the autonomous consumption estimate turned out to be negative. If our model is correct, this would suggest that when people earn zero income, they actually spend a negative amount which does not make much sense. As we will see in the next part, what is really going on is that our model is doing a poor job of fitting the data.

(h) Create a scatterplot of the residuals from your regression in part (f). The residuals $(y_i - \hat{y}_i)$ should be measured on the y-axis and the value of income should be measured on the x-axis. Given this scatterplot, does it appear that any of the assumptions we made about the error terms are violated? (These assumptions are listed on page 109 of your textbook.)

See ps4-solutions.xlsx for the calculation of the residuals and the scatterplot of the residuals (note that you can also have Excel automatically create a scatterplot of the residuals as one of the regression options). From the scatterplot, it is clear that we have violated the assumption that the value of the error is uncorrelated with the value of the independent variable. There is a very clear pattern between the residuals and the values of income. The pattern is actually a common result of a situation where the relationship between Y and X is a convex curve rather than a straight line. If you try to fit a line through a convex curve, points on either end of the range of X values will tend to be above the line and points in the middle of the range of X values will tend to be below the line. You can try to fix this by using a polynomial in X rather than just X when doing the regression. In ps4-solutions.xlsx I have included a regression of consumption on income and income-squared along with a plot of the resulting residuals. You can see that the scatter of residuals looks much more random than when we just used income. One other thing to note is that this has given us a very different result for the autonomous consumption. In the quadratic model, the estimated intercept was 1929. This suggests that if disposable income is zero, a person still spends \$1,929. This is a much more reasonable estimate of the autonomous consumption than what we got from the linear model.

On the last page of these solutions I have included several graphs of residuals for other situations where our assumptions are violated. In graph (A), similar to the situation in this problem, the value of the residuals is clearly not unrelated to the value of X. In graph (B), the mean of the residuals is clearly not zero. In graph (C), the variance of the residuals is not constant and independent of X. You can see that the variance decreases as X increases.

(i) Use your data to test the claim that people save more than 50% of each dollar they earn.

Let's use the results from part (f). We are testing the following set of hypotheses:

$$H_0\text{: } MPC \geq 0.50$$

$$H_a\text{: } MPC < 0.50$$

Given our regression results, our test statistic is:

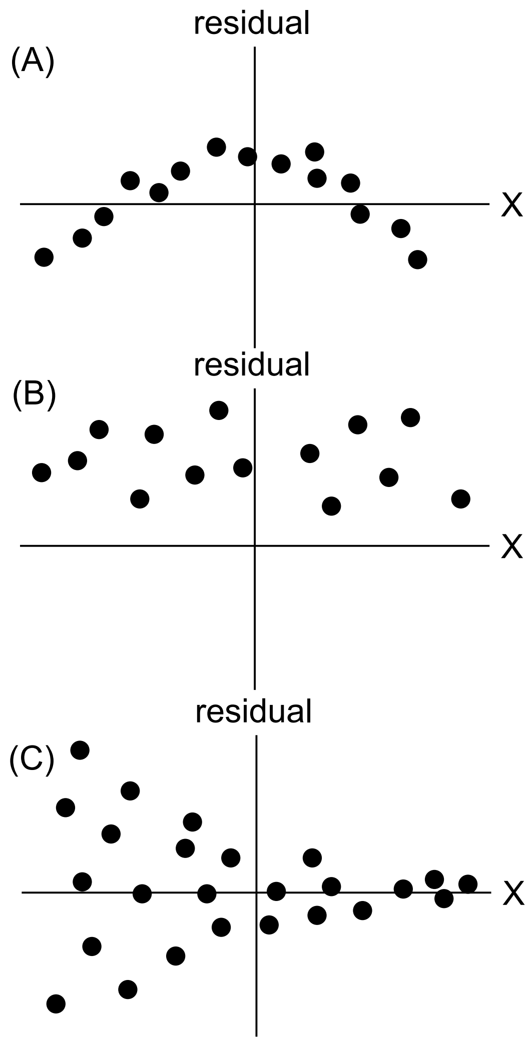
$$t^* = \frac{0.97 - 0.50}{0.008}$$

$$t^* = 58.75$$

The p-value for this test statistic is:

$$p = 1 - TDIST(58.75, 50 - 2, 2) = .9999...$$

This is an incredibly large p-value so there is no way that we would reject the null hypothesis that people have an MPC that is more than 0.5 (meaning they save less than half of their income). If you had set this up as an upper one-tailed test, you would have gotten an incredibly small p-value and rejected the null hypothesis that people save more than half of their income at any reasonable significance level.


residual
(A)
X
residual
(B)
X
residual
(C)
X