Intermediate Macroeconomics Recitation 5

Topics: Consumption over the life-cycle, regressions and causality, graphing growing variables

1 Consumption over the life-cycle

Recall the consumption problem from last week, which can be summarized in a utility function and an intertemporal budget constraint:

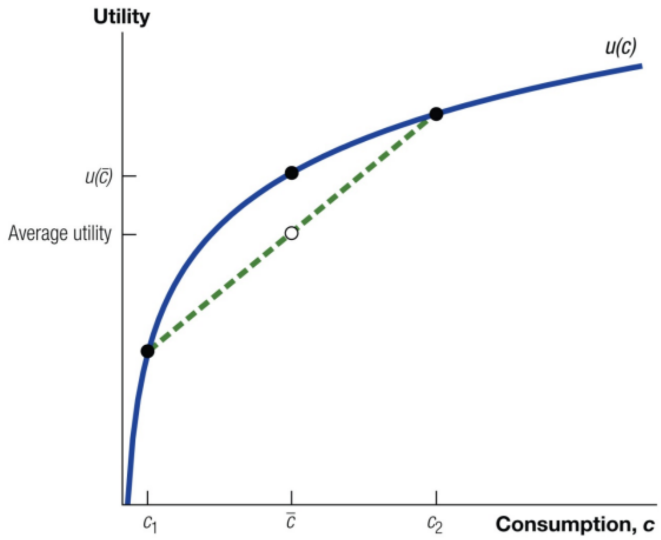
$$\max_{C_1,C_2} u(C_1) + \beta u(C_2) \qquad \text{subject to:} \quad C_1 + \frac{C_2}{1+r} = Y_1 + \frac{Y_2}{1+r}$$

Last recitation, we made the assumption that the utility function was $u(C) = \log(C)$, so that $u'(C) = 1/C$. Then, the consumption euler equation becomes:

$$\frac{C_2}{C_1} = \beta(1+r)$$

C_2/C_1 is the growth rate of consumption. Hence, we fine that consumption growth is independent of income growth. This is counterintuitive since we might expect that if an agent receives more income when they are young than when they are old relative to another agent that they will consume relatively more when they are young and thus their consumption growth will be low. Why do we not get this result? Because agents want to consumption smooth. They want to ensure that they consume similar amounts when they are young and when they are old.

Let's consider why consumers want to consumption smooth. Imagine that the consumer consumes c_1 at time 1 and c_2 at time 2 where $c_1 \neq c_2$. Then, due to the concavity of their utility function u, agents will get higher utility from consuming the average of c_1, c_2 i.e. $0.5c_1 + 0.5c_2$ compared to before (we've ignored discounting here for simplicity). Here is a graphical representation of this example:



We can also solve for the level of consumption in each period:

$$\begin{aligned} C_1 &= \frac{1}{1+\beta}\left(Y_1 + \frac{Y_2}{1+r}\right) \\ C_2 &= \beta\,(1+r)\,C_1 \end{aligned}$$

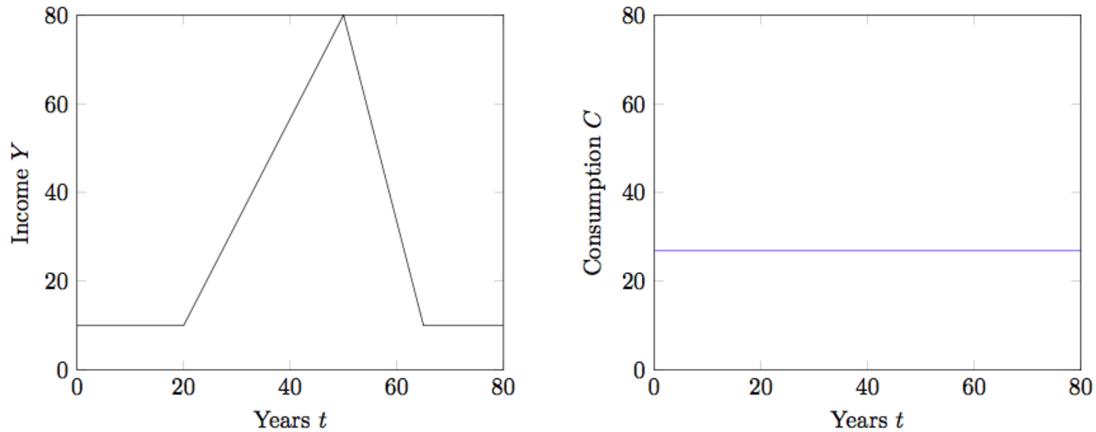
This shows that the level of consumption does not depend upon the level of period specific income. Instead, it only depends upon the present value of the consumer's remaining lifetime income, and that the level of consumption is just some fraction of consumer's remaining lifetime income. The idea that the only part of income that determines the level of consumption today is the present value of remaining lifetime income (and so income today conditional on lifetime income does not matter) is called the Permanent Income Hypothesis.

We can generalise these consumptions savings results to many periods. In particular, we can find:

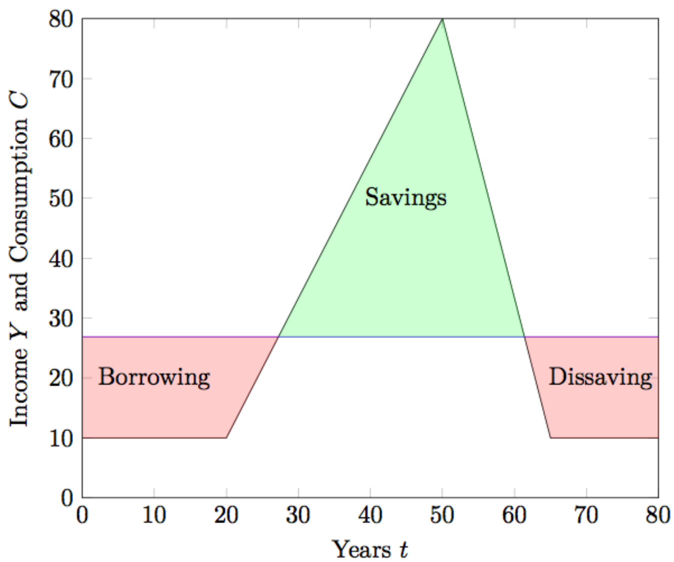
$$C_{t+1} = \beta(1+r)C_t$$

Thus, consumption should grow at a constant rate $\beta(1+r)$ throughout an agent's lifetime. If we make the simplifying assumption that $\beta(1+r) = 1$ then we get that result that consumers will consume the same in every period.

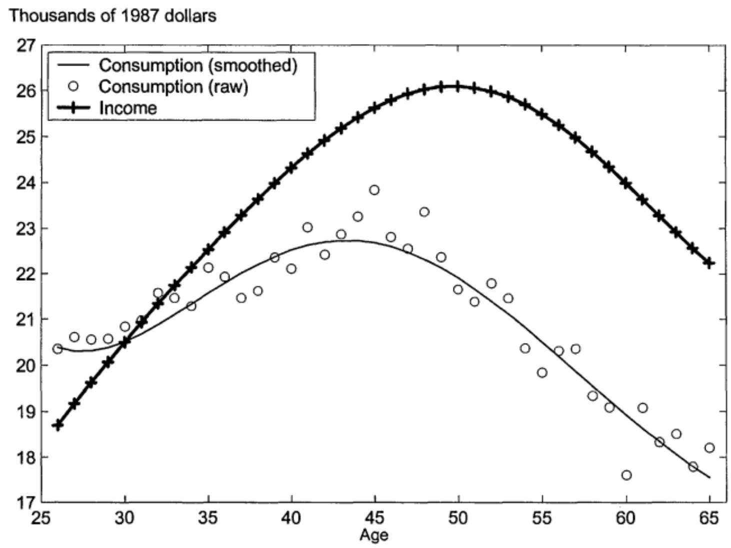
In the simplified case where $\beta(1+r) = 1$ then we know that even if income varies across a consumer's lifetime, they will consume the same in every period. This is represented in the following graph:



This would require the following profile of borrowing and savings:



We saw the figure in by Gourinchas and Parker (2002) in lecture.



Does the model fit the data? No, the model implies a constant growth rate of consumption (Euler equation) which is independent of the income profile. The data shows a hump shaped pattern of consumption over the life cycle. The consumption profile seems to mimic the income profile.

What are potential reasons for the differences between the model and the data?

- Borrowing constraints: Agents might want to borrow a large amount of money when they are young but banks probably will not give it to them due to the fact that young people hold little collateral and banks don't like to give large loans without collateral since they can't force people to work. Borrowing constraints explain why consumers never borrow large amounts. They do not explain why consumers do not save large amounts. This could just be because consumers naturally have a low β and tend to prefer to spend whenever they can.

- Precautionary savings: Most people will receive higher incomes as they get older. However, not everyone will. Some will get unlucky in their careers or other aspects of their lives. Agents may wish to save when they are young in case they are unlucky (since they will get extremely low utility if they get unlucky and run out of money).

- Changing preferences over time: When you are young, you only have to look after yourself. When you are middle-aged, many people choose to have children which are costly. When you are old, you no longer need to work so you can more easily spend less i.e. take longer but cheaper transport, make food rather than eat out, do your chores rather than hire help.

- Myopia: Agents may not fully take into account their preferences so, even though, they may want to save for the long-term, they may always put it off for another day. This is often used to explain why retirement savings are so low.

2 Statistical Relationships and Causality

2.1 Correlation and Regressions

We need to understand what correlations and regressions mean. However, you don't need to be able to compute these. See the appendix if you are interested in how these are computed.

Correlation: Correlation gives us a measure of whether two variables have a positive relationship. For example, we might expect to find a positive correlation between height and weight since people who are taller tend to be heavier.

Regressions: Regressions are method of estimating how one variable (the dependent variable) will change when we adjust other variables (the explanatory variables). For example, we could regress weight on height in order to see how weight changes when we change height. We might get the following relationship:

$$Weight(pounds) = 10 + 2 * Height(inches)$$

This would suggest that if all we know about someone is their height (60 inches) then they will weigh 130 pounds.

We can add multiple variables into a regression. For example, we can introduce the variable Age:

$$Weight(pounds) = 6 + 2 * Height(inches) + 0.1 * Age(years)$$

This would suggest that if all we know about someone is their height (60 inches) and age (20 years old) then they will weigh 128 pounds.

This implies that if we have two different people, A and B, with the same explanatory variables except for $Height$ and A has a height that is one inch taller than B then A will have a weight that is 2 pounds heavier. It does NOT suggest that if we raise person B's height by one inch that we will raise their weight by 2 pounds. This is the key difference between a regression relationship and causality. A regression is just a statistical relationship that does not imply causality.

2.2 Misinterpreting Statistical Relationships as Causality

Here are some examples where people have mistaken a statistical relationship for causality:

- From the New York Times on February 22nd 2010 (`http://www.nytimes.com/2010/02/23/health/23mind.html?mtrref=www.google.com&gwh=EF9F1767766F53D7963BD3871A4731CF&gwt=pay`), a study analysed the performance of a basketball player and the degree to which he touched other players. Mr. Kraus and his co-authors, Cassy Huang and Dr. Keltner, report that with a few exceptions, good teams tended to be touchier than bad ones. The most touch-bonded teams were the Boston Celtics and the Los Angeles Lakers, currently two of the league's top teams; at the bottom were the mediocre Sacramento Kings and Charlotte Bobcats. The same was true, more or less, for players. The touchiest player was Kevin Garnett, the Celtics star big man, followed by star forwards Chris Bosh of the Toronto Raptors and Carlos Boozer of the Utah Jazz. 'Within 600 milliseconds of shooting a free throw, Garnett has reached out and touched four guys,' Dr. Keltner said.

- OkCupid note that it is a myth that if a man posts a picture of their abs, it reduces your chances of getting matched with a date (`https://theblog.okcupid.com/the-4-big-myths-of-profile-pictures-41bedf26e4d`). They point out that males who post pictures of their abs actually get more dates and argue that this counters the myth that posting a picture of your abs is only for 'bozos'.

There are two main problems with inferring a causal relationship from a statistical relationship:

- Reverse causation: In the basketball example, it might appear that touching improves a basketball team's results. However, there may be reverse causality. It could be that the causality goes in the other direction. Teams touch to celebrate good performance. Therefore, it is not clear that if teams start to touch more that it will improve their performance.
- Omitted variable bias: In the OkCupid example, if a guy is more attractive it would appear that he is both more likely to get dates and more likely to post pictures of his abs. Thus, the fact that guys who post pictures of their abs get more dates may just be because these guys are more attractive. We cannot draw a causal relationship. Indeed, it seems entirely possible that a guy who is less attractive would not get more dates from posting a picture of their abs.

2.3 Assessing Causality

There are two standard ways of assessing causality in Economics:

- Field experiments: Economists deliberately introduce random variation in treatment and compare the performance of different treatments. These are the equivalent of clinical trials.
- Natural experiments: Economists look for natural random variation in treatment and compare the performance of different treatments.

Examples of field experiments:

- Mosquito nets: We want to investigate how mosquito nets affect health outcomes. We can give mosquito nets to some people and not to others where we randomly select who receives the nets. We can then compare health outcomes across these villages.

- Assessing the impact of school vouchers on agents. We can give school vouchers to some children and not to other children where we randomly select who receives the vouchers. We can then compare how the children perform.

It is easy to see how field experiments are an effective method of assessing causality. However, they are not always practical because:

- Expensive: They can be very expensive to run and manage.
- Impractical: It's difficult to run field experiments with Macroeconomic policy since citizens might not be happy if the government deliberately tries a policy that it knows is unlikely to work i.e. setting the nominal interest rate too high to see how it affects the economy.

Natural experiments are a good alternative. Here economists have to search for a source of natural variation. Here are some examples:

- Impact of fiscal stimulus: A standard question in Economics is how does government spending affect output. We can't just look at the relationship directly because if output changes it may affect government spending (reverse causality). Nakamura and Steinsson (2014) analyse this question book looking at how output changes within states when the central governments adjusts military procurement from different states. The idea is that central government procurement of military equipment from states will not depend upon the state's output so will not be affected by the reverse causality issue.
- Returns of education: A standard question in Economics is how much schooling impacts earnings. We can't easily assess this since people who get more education have different characteristics to those who get less (omitted variables bias). Angrist and Krueger (1991) attempted to solve this problem by arguing that agents who are born in the first quarter of the school year are less likely to finish school than those born in the last quarter of the school year (since they can drop out sooner). Thus, they can compare returns to schooling on people born in the earlier quarters of the school year to later quarters of the school year (via a little econometrics) to get an estimate of returns to education.

3 Graphing Growing Variables

Consider a time variable Y_t which could represent something like GDP. We can rewrite Y_t as follows:

$$\begin{aligned} Y_t &= Y_{t-1} + (Y_t - Y_{t-1}) \\ &= Y_{t-1}(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}} \end{aligned}$$

We define the percentage change of a variable Y_t from $t-1$ to t as:

$$g_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

Therefore:

$$Y_t = Y_{t-1}(1 + g_t)$$

This is called exponential growth because if $g_t > 0$ normally then at time t Y_t will increase by $g_t Y_{t-1}$ whereas at time $t-1$ Y_{t-1} will increase by $g_{t-1}Y_{t-2}$. We know that $Y_{t-1} > Y_{t-2}$ so Y_t probably increases by more in later periods than earlier periods. Indeed, if you plot this variable (with $g_t > 0$ normally) in levels, it will look like all the growth appears in recent years. Therefore, it's not terribly helpful to plot this graph in levels.

We can see this in per capita GDP in levels:

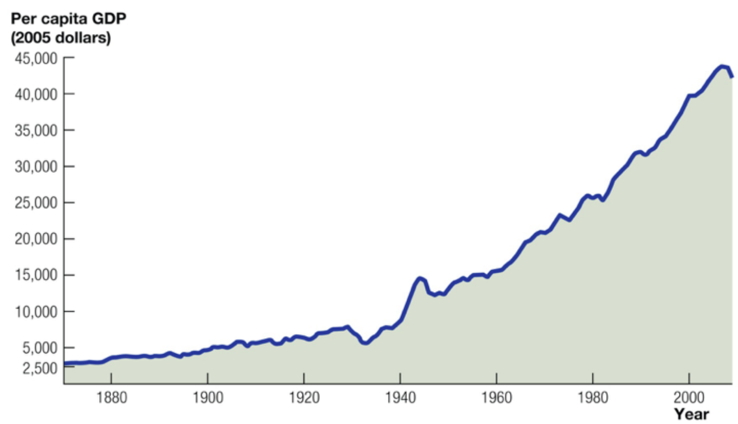


FIGURE 3.2 Per Capita GDP in the United States, 1870–2009

It looks like all the growth has happened in recent years.

Let's take logs (it doesn't matter which base) and see if this helps us draw a more easily interpretable graph. We get:

$$\log(Y_t) = \log(Y_{t-1}) + \log(1 + g_t)$$

Then we have that $\log(Y_t)$ will be roughly a straight line with gradient $\log(1 + g_t)$ if g_t is roughly constant. Often, we don't actually plot $\log(Y_t)$ but we plot in a log scale. This is where the scale corresponds to the log of the variable i.e. each tick on the y-axis corresponds to multiplying Y_t by some base number.

We can observe GDP per capita with a log scale (of base 2):

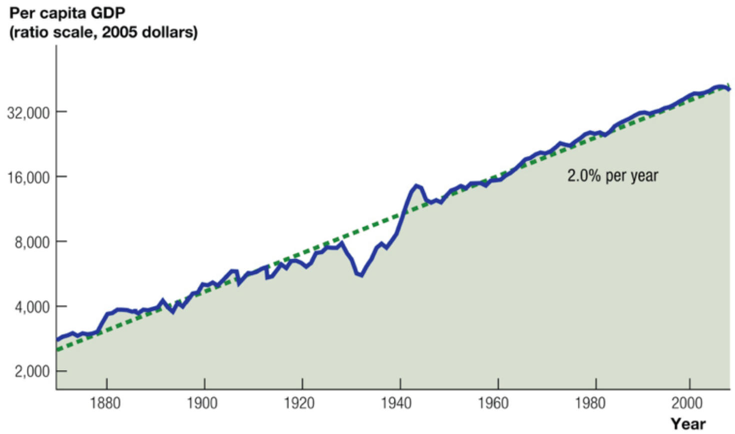


FIGURE 3.5 Per Capita GDP in the United States, 1870–2009: Ratio Scale

We see that GDP per capita with a log scale is approximately a straight line since growth is roughly constant in the long-term.

A Basic Statistics

A.1 Summary Statistics

Assume that we are given joint observations (x,y). (income,test score) for a group of individuals is an example of such obs. **Cross-sectional** data refers to data that comes at a single time, from multiple individuals. On the contrary, we might observe a **time series** data for one individual (or more likely, a country) that comes at a series of time. And lastly, we may have joint observations from multiple individuals across multiple period of time. This is called a **panel** data. Tools to analyze a given dataset are different depending on whether the data is cross-sectional, time series, or panel. To avoid complexity, we assume that the data we observe is cross-sectional.

Given our joint observations, we first concentrate on a single set of observations, say income. You want to understand the properties of your given income data, and summarize it into a few numbers. The first statistic that comes into mind is be the average income. This is known to be the mean of the data and is given by the following.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + ... + x_n}{n}$$

However, this summary may not be enough to fully understand the property of your income data. To illustrate, say that you are given 5 observations of income by two different datasets $(5, 5, 5, 5, 5)$ and $(1, 2, 2, 10, 10)$. Both have the same mean which is 5. However, both datasets are quite different since for the first case, all people have equal income while for the second case, there are large dispersions observed. Hence to capture this difference, we may also want to summarize the dispersion of the dataset into one number. Variance measures the dispersion of the data and is given by the following.

$$var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Continuing with our example, we see that the variance of your first income set is 0, while the variance of our second set is 16.8. Since the variance is higher in the second sample, we may conclude that the second income set is more disperse, while the first dataset is more concentrated. In practice, we more often use the standard deviation (or standard error) as a measure of dispersion. This is simply taking the square root of the variance. Here, the standard deviation is 0 for the first sample, while it is 4.1 for the second dataset.

We can go on and on to define more summary statistics that capture all the properties of the data. However, if we come up with too many summary statistics, this may not really be a summary of your dataset. For the above example, if you define more than 5 summary statistics to represent your dataset, you might as well express the raw data itself, rather than coming up with more summaries than what is there.

Now, suppose that we have the joint observation (x,y). Besides the single observation properties, we may also be interested in their relations and summarize it into one number. Covariance captures this concept and is given by the following.

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

From the definition, it is immediate to see that a positive covariance exists between the joint observation if both x and y move in the same direction, while a negative covariance exists vice versa. However, we may want a measure of the relation between the two datasets that is scale invariant. Correlation between the two variables captures this and is given as follows.

$$corr(x,y) = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$$

It can be shown that this correlation is bounded in the interval [-1,1]. This summarizes the degree of **linear** relation between two datasets. It has been emphasized in class that a causal inference is not immediate from observing a strong correlation between the two datasets, but the opposite deserves some attention as well. That is, even if we observe 0 correlation, this does not necessarily imply that there is no causal relation between the two variables. For example, suppose that we extract 4 observations $\{(-2,4),(-1,1),(1,1),(2,4)\}$ from the function $y = x^2$. It is clear that these joint observations are perfectly related in a nonlinear fashion. However, covariance is 0 and hence correlation turns out to be 0. Again, correlation only summarizes the degree of linear relation between the two datasets.

A.2 Regression Analysis

Suppose that we want to go more deep, and study not only the correlation of two variables but also the marginal impact of one variable on the other. In this case we conduct a regression between the two variables. Regression analysis does not verify causality, but is nevertheless useful in that it tests the first condition for causality exercise, and also gives a handy way to forecast. We will only touch the surface of this topic, and econometrics deals with this topic in more detail. Back to the joint observations (x,y), we set up a simple linear regression model where x is defined to be an independent variable while y is a dependent variable.

$$y_i = a + bx_i + e_i \qquad i = 1, 2, ..., n$$

Here, e_i is an error term that we assume to be mean zero. Our interest is in estimating b (and a), and we take the least squares approach. That is, our objective is to minimize the discrepancy of y and $a + bx$, which is summarized by the squared sum of the difference. Mathematically we choose a and b such that the following is minimized.

$$\min_{a,b} \sum_{i=1}^{n} e_i^2 = \min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

The first order condition to choose a and b is by taking the derivatives with respect to each parameter, and equating them to 0.

$$\begin{aligned} [a] \qquad & -2 \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i) = 0 \\ [b] \qquad & -2 \sum_{i=1}^{n} x_i (y_i - \hat{a} - \hat{b}x_i) = 0 \end{aligned}$$

Solving these two so-called 'normal equations', we reach the following estimate for a and b.

$$\begin{aligned} \hat{a} &= \overline{y} - \hat{b}\overline{x} \\ \hat{b} &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{cov(x,y)}{var(x)} \end{aligned}$$

(Details on the derivation of this result is in any econometrics textbook.) Increasing x by 1 unit will 'predict' an increase of y by $\hat{b}$ units. In most cases, we have more than 2 joint observations. Suppose that we have 3 joint observations (x_1, x_2, y). We may suspect that x_1 and x_2 (linearly) explain y and want to formulate this idea. Then the proper regression model is the following.

$$y_i = a + bx_{1i} + cx_{2i} + e_i \qquad i = 1, 2, ..., n$$

By similar process as above, we can obtain an estimate of a, b and c that gives the following result.

$$y_i = \hat{a} + \hat{b}x_{1i} + \hat{c}x_{2i} + \hat{e}_i$$

Here, the interpretation of $\hat{b}$ is the increase of y when x_1 increases by 1 unit, holding x_2 fixed. That is, $\hat{b}$ measures the marginal effect of x_1 on y.