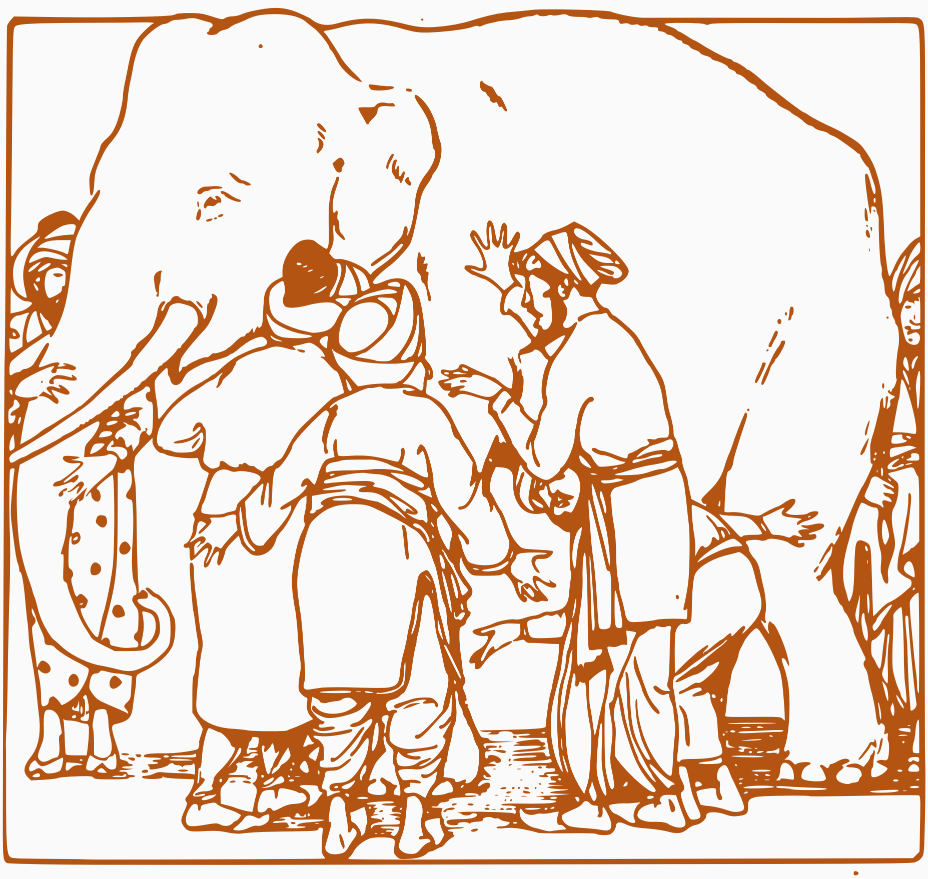
macroeconomics

Jyotirmoy Bhattacharya

v4.3.1



You can find the LaTeX source of this book at https://github.com/jmoy/jmoy-macroeconomics

Email: jyotirmoy@jyotirmoy.net

Contents

CHAPTER 1

Introduction

Macroeconomics tries to understand the dynamics of economic aggregates like national income, the price level or the rate of unemployment. Our interest in macroeconomics arises from the following questions about modern capitalistic economies:

(1) Economies and groups of economies go through periods of general underutilisation of resources. Unemployment coexists with unutilised plant and machinery. This is evidently inefficient since putting the unemployed workers to work on the unused machines would produce additional output that can make at least some people better off without making anyone worse off. Yet, in a slump the market mechanism does not seem to work towards eliminating this inefficiency. At least not fast enough. The Great Depression was the most dramatic of such episodes but smaller slumps occur quite regularly. Is this periodic inefficiency intrinsic to a capitalistic economy or can it be eliminated without any major changes in the structure of the economy?

(2) Economies differ dramatically in their average standard of living and there is no systematic tendency for this gap to close. What are the economic forces that make some countries rich and others poor? Why haven't the poor countries been able to close this gap by accumulating capital and adapting the technology available to rich countries?

(3) Governments think, and citizens often agree, that the two problems above can be at least mitigated through the choice of appropriate public policies. Is this really so? And if so, which policies should be adopted in which circumstances?

There exist diverse ways of approaching these questions. The present book follows the 'neoclassical' approach which is currently the most popular.

The first feature of this approach, as it applies to macroeconomics, is that we start with households, firms and governments as our basic units of analysis. Each household, firm and government is composed of diverse individuals who interact with each other in complex ways. In macroeconomics we usually ignore this interaction and consider each of these units as a black box.

Second, we assume that each household and firm maximises a well-defined objective function subject to the constraints imposed on it by the institutional framework of the economy. This is a major assumption. First, contrary to soap operas and corporate thrillers, we assume that conflicts of interests between the different individuals constituting a household or firm work themselves out in a way that the unit has a whole appears to be pursuing a coherent goal. Second, we assume that regardless of the complexity of the environment facing the unit it can rank all the alternatives available to it and choose the best. Thus there is no limit to the information processing and decision making sophistication of the economic units.

This assumption certainly captures important aspects of reality. Economic decisions are certainly goal-oriented and often when the stakes are high we spend considerable effort in trying to determine which choice is the best. Yet, both introspection and systematic research shows that we are not really the superoptimizers of the last paragraph. Faced with complex situations we fall back to using simple rules of thumb rather than carrying out the impossibly complex task of finding the best alternatives. Our decisions are often subject to unconscious biases. Incorporating these departures from full optimization into economic modes is among the most active areas of current research. However, this research is yet to reach a consensus. Therefore, in this text we limit ourselves for the most part to models based on full optimization.

Third, in the neoclassical approach we look at *equilibrium* states—states where the desired actions of different agents are all consistent with each other. The exact form of this equilibrium condition depends on the particular institutional structure being studied. In competitive markets it takes the form of the equality of demand and supply. Where strategic interactions are important we use equilibrium concepts from game theory, the most important of which is that of Nash equilibrium.

The justification for limiting our attention to equilibrium states is that in any other state some agents will find that they cannot carry out their plans or that their plans do not have the expected outcomes. This will make them change their behaviour. Thus a non-equilibrium state cannot persist. This does not by itself imply that an equilibrium will ultimately come about. The system may keep moving from one non-equilibrium state to another forever. Only if we think that this is unlikely and that a system away from equilibrium will move close to equilibrium rapidly enough are we justified in studying only equilibrium states.

References

Other graduate-level treatments of macroeconomics in the neoclassical paradigm, in order or increasing difficulty, are: [**Rom11**], [**LS12**], [**SLP89**]. Introductions to other approaches are [**Dav11**] and [**Tay04**]. [**Kah13**] is a popular account of the psychology of decision-making.

CHAPTER 2

The AS-AD model

In this chapter we begin our study of short-run fluctuations by reviewing the AS-AD model that you must have already encountered as an undergraduate. We assume that the economy is closed.

1. Background

Throughout we assume that there is a single produced good in terms of which we measure real output and expenditure and a single labour market. We also assume that there are only two assets—money and bonds—and a single nominal interest rate which measures the return from bonds.

The AS-AD model analyses the economy in a single time period during which we assume that the stock of capital and the state of expectations remain unchanged. We refer to this by saying that AS-AD is a model of the "short-run".

2. Aggregate Demand

The demand for goods (Z) is made up of consumption (C), investment (I) and government expenditure (G). All these variables are measured in real terms.

We take G to be given exogenously.

Households decide how much to consume based on their current disposable income and wealth, expectations of future disposable income and current and future needs. Of these, all variables other than current disposable income are held constant in the short-run. So we can write $C = C(Y - T)$ where Y is current income and T is net taxes. We assume that T is given exogenously.

Firms decide how much to invest based on the current level of the capital stock, the current and future levels of output and the real rate of interest.

The economic story for investment is that the desired level of capital stock depends positively on present and future expected profitability and negatively on the real rate of interest (since a higher real rate of interest implies that a unit of real output in the future is worth relatively less in terms of present output). The higher the gap between desired and actual capital stock, the higher is the rate of investment. Once again suppressing the variables fixed in the short-run we have $I = I(Y, i - \pi^e)$ where $i - \pi^e$, the difference between the nominal rate of interest and expected inflation, equals the real rate of interest.

In equilibrium, the total output of goods must equal the total demand for goods, that is it must be the case that

$$Y = C(Y - T) + I(Y, i - \pi^e) + G \tag{1}$$

The satisfaction of this equation in necessary for goods market equilibrium, but it is not sufficient. We have not yet discussed the supply decision of firms and therefore it is not yet clear that firms would want to supply a quantity of output that would satisfy this equation.

Given our assumptions the (Y, i) combinations which satisfy (1) form a downward-sloping curve in the (Y, i) space. It is known as the IS curve.

Looking at asset markets we assume that the demand for money takes the form

$$M^d = PL(Y, i)$$

where P is the price level. The demand for money is increasing as a function of Y since a higher level of output also implies a higher volume of transactions and hence a higher demand for money to finance those transactions. It is decreasing as a function of i since by holding money consumers must forego the interest earnings on bonds and the higher this opportunity cost the more consumers would economise on the holding of money.

The strict proportionality between the demand for money and the price level needs comment. Imagine a doubling of all current prices and wages. Assuming a fixed expected inflation rate this also implies a doubling of all expected future prices and wages. As a result the real opportunities available to economic agents remain unchanged. We therefore believe that agents would carry out the same real transactions. This assumption that real demands and supplies depend only on real opportunities and not on nominal quantities is known as the lack of money illusion. But given that prices and wages have doubled the old level of real transactions would now require exactly double the amount of money as before to carry out.

The supply of money (M) is assumed to be exogenous. Equality of the supply and demand for money give

$$M/P = L(Y, i) \tag{2}$$

For a given value of P this is an upward-sloping curve in the (Y, i) space called the LM curve.

For each possible value of P the intersection of the IS and LM curves (or what is the same thing, the simultaneous solution of (1) and (2)) gives us unique values of Y and i. We can express them as $Y(P)$ and $i(P)$. $Y(P)$ is decreasing and $i(P)$ is increasing since an increase in P moves the LM curve leftward.

The locus $Y(P)$ in the (Y, P) space is called the aggregate demand (AD) curve. Though the names are similar, this curve is very different from the demand curve for a single good that we study in microeconomics.

First, the AD curve *does not* show the quantity demanded of the single good in the economy for a given price while everything else is being held constant. If you recall, both consumption and investment demand depend on the level of current income. But the level of income *is not* held constant while deriving the IS curve. Rather the IS curve is the locus of points where income is chosen such that demand equals output. Thus the AD curve is seen as a locus of (Y, P) pairs that are consistent with equilibrium in the goods and asset markets.

Second, the demand curve for a single normal good slopes downwards because an increase in price decreases the real income of consumers (whose money income is assumed to be given) and gives them an incentive to substitute away from the good whose price has increased. The reason for the AD being downward sloping is entirely different. The AD curve slopes downward because at a higher price the demand for money is lower, which causes the LM curve to shift leftward, decreasing the level of Y where the IS and the LM curves intersect.

3. The AS curve

In deriving the IS curve (1) we mentioned that goods market equilibrium additionally requires consideration of the supply decision of firms.

One important aspect of supply decisions in the real world is that money wages and prices are "sticky", i.e. they don't immediately adjust fully to changes in economic condition.

Starting from this observation we may make the simple assumption that in the short-run wages and prices remain fixed at whichever levels they were set at in the past and that firms supply the level of output demanded at the given prices. Under this assumption the (Y, P) combinations consistent with the supply decisions of the firms is a horizontal line in the (Y, P) space—the so-called "Keynesian AS curve"—and the intersection of this curve with the AD curve gives us the equilibrium level of Y and P. Since the level of P is exogenously given, we may have as well derived the equilibrium level of Y from the IS-LM apparatus by drawing the LM curve corresponding to this P.

Even if the exogeniety of wages and prices were to be accepted, the "Keynesian AS curve" is not consistent with perfect competition in the goods and labour markets. Perfectly competitive firms choose the amount of labour that they employ and hence the output that they produce by equating the marginal product of labour to real wage. Similarly in perfectly competitive labour markets the supply of labour is determined by workers equating their marginal disutility from labour to the real wage. For a given level of prices and wage, the output and employment implied by the profit maximisation decision of firms, the utility maximisation decisions of workers and by the intersection of the AS-AD curve may all three be different.

In his *General Theory* ([**Key36**]) Keynes actually allowed the level of prices to be flexible and required that it adjust in order to ensure that the profit maximising output of firms equal the output at which IS and LM intersected. However, he argued that in general the economy might be in equilibrium even when the real wage is not equal to the marginal disutility of labour.

Keynes's framework runs into problems. First prices are in fact sticky. They do not change immediately when aggregate demand and output changes. Second, if we assume diminishing marginal productivity of labour and a real wage equal to the marginal productivity of labour then we would expect that real wages would decrease when output increases and vice-versa, i.e. real wages would be countercyclical. However, empirical studies of real wages do not find this countercyclical pattern. Finally, it is not made clear why money wages would not fall in a situation of excess supply of labour, even if we were to accept, as Keynes seems to argue, that a fall in money wages would not restore equilibrium.

As a result, nowadays instead of perfect competition and flexible prices we work with models of sticky prices and wages and use the existence of imperfect competition in the goods and labour markets to explain why firms and workers may be willing to accommodate changes in demand at unchanged prices and wages. The "Keynesian AS curve" can be thought of as a simplified expression of these theories.

However this is too drastic a simplification since it makes P and W entirely exogenous and therefore does not allow us to discuss how these variables respond to economic changes. So, for example, we cannot discuss as important an issue as inflation in this framework.

Yet, even wages and prices which are "sticky" are not fixed for ever. They do adjust with time. Firms and workers do reset prices and wages from time to time and when they do so they take into account the current economic conditions as well as their expectations of future economic condition. One observation common to many models is that higher levels of economic activity are, other things being constant, associated with higher prices. This is the basis of the upward-sloping AS curves that you can find in undergraduate textbooks. We shall not derive such a curve here (though see the exercises) but postpone the discussion to later chapters where we can approach it with better tools.

4. The Way Forward

The IS-LM and AS-AD models are still the models many macroeconomists reach for when first trying to understand questions related to aggregate fluctuations. But they have shortcomings.

First, being limited to a single short-run does not allow us to discuss the evolution of macroeconomic variables and the effects of policies over time. For example, if we want to understand the effects of a permanent increase in the tax rate it is not enough to know what happens in the period in which the increase is imposed. We would also like to know the impact as the continued increase affects expectations and asset choices in the economy, something which we cannot do within the AS-AD model.

Limiting our study to a single short-period also creates an artificial separation between the study of growth and fluctuations. The long-run over which growth happens is stitched from a sequence of short-runs and at least in principle our short-run theories must be consistent with our long-run ones.

Second, even in a short-run the assumption that expectations are exogenously fixed is often not plausible. While capital stocks can only adjust slowly over time because of the finite speed of the physical processes involved, there is no such friction holding back changes in human beliefs. Participants in the economy continuously revise their beliefs in the light of new information. The kind of policy interventions—such as changes in government expenditure or the money stock—that we study using the AS-AD framework also convey new information to agents and very likely change their beliefs. Therefore there is always the likelihood of error in studying the effects of policy changes like these while holding beliefs constant.

In the rest of this book we study the behaviour of firms and households in greater detail than we have done in this chapter. We will take up issues like the role of credit market imperfections in determining consumption and investment or how job market search and asymmetries of information make the labor market so different from competitive commodity markets.

In developing our models we shall also make the role of stocks and expectations explicit. Combining these models with assumptions regarding the evolution of stocks and expectations will then also enable us to go beyond a single-short run and address the limitations of the AS-AD model discussed above.

Exercises

Exercise 2.1. From your favourite undergraduate texts find at least three different derivations of an upward-sloping AS curve. State precisely the assumptions regarding firm and worker behaviour used in each.

Exercise 2.2. Explain the difference between the real and the nominal rate of interest. In the text we claimed that the real rate of interest is the nominal rate of interest minus the expected rate of inflation. Why is this so?

EXERCISE 2.3. We derived the IS-LM model holding the stock of money as fixed. However, most central banks nowadays set the nominal interest rate and allow the stock of money to adjust so as to equal the demand for money. How would you modify the IS-LM model to incorporate this fact? (For more on this see [**Rom00**].)

References

The AS-AD model is covered in most undergraduate texts.

It turns out that setting up an AS-AD model in a way that is logically and economically consistent is somewhat tricky and there are different modelling choices that can be made. See [**Dut02**] for a history of the model and alternative ways of setting it up. For critiques, see [**Bar94**] and [**Col95**].

CHAPTER 3

Consumption: Certainty

Savings is not an end in itself. Rather savings is the means a household adopts in order to provide for future needs. Therefore we can hope to be able to better analyse and predict consumption-saving behaviour if instead of thinking of the household as choosing between consumption and savings we think of the household as choosing between fulfilling present needs and fulfilling future ones. This is the starting point of the intertemporal theory of consumption. To more clearly see this central tradeoff between the present and the future we begin by assuming in this chapter that the household faces no uncertainity regarding future needs and opportunities. Of course, in reality uncertainty has an important influence on intertemporal choices and we extend our discussion to take it into account in Chapter 7

1. Two-period case

1.1. Budget constraint. Consider a consumer who lives for two periods, has an endowment of y_1 and y_2 units of goods in the two periods respectively and can borrow and lend any amount that they like at the real rate of interest r.

Suppose the consumer consumes c_1 in the first period. Then she will have to take a loan of $c_1 - y_1$ to finance her consumption. (This number can be negative, in which case the consumer is lending rather than borrowing.) In the next period the consumer will therefore have to make loan repayments of $(1 + r)(c_1 - y_1)$. Assume that the consumer does not want to make any bequests and cannot die with any outstanding loans, consumption in the second period must be,

$$c_2 = y_2 - (1 + r)(c_1 - y_1)$$

Simplifying and rearranging we have

$$c_1 + \frac{c_2}{1 + r} = y_1 + \frac{y_2}{1 + r} \tag{3}$$

This is the budget constraint faced by the consumer. We can interpret this to mean that the present value of the consumer's consumption stream must equal the present value of their incomes.

1.2. Utility maximization. Suppose the consumer maximises a quasiconcave utility function $U(c_1, c_2)$ subject to this budget constraint. Then the consumer's first-order conditions are

$$U_1(c_1, c_2) = \lambda \tag{4}$$

$$U_2(c_1, c_2) = \lambda/(1 + r) \tag{5}$$

where λ is the Lagrange multiplier corresponding to the budget constraint and $U_i(c_1, c_2)$ denotes the partial derivative $\partial U/\partial c_i$. We have explicitly shown the dependence of the partial derivatives on the value of consumption in both periods. These first-order conditions along with the budget constraint (3) together determines the value of c_1, c_2 and λ.

1.3. Comparative statics. Assuming that consumption in both periods is a normal good, an increase in either y_1 or y_2 increases both c_1 and c_2.

The effects of a change in r are ambiguous. An increase in r makes consumption in period 2 relatively cheap compared to consumption in period 1. Therefore the substitution effect causes c_1 to decrease and c_2 to increase. It is traditional to decompose the income effect into two parts. First, an increase in r reduces the present value of the consumer's endowments and hence decreases his real income. Second, an increase in r, by making the consumption in period 2 cheaper increases his real income.[1] The sign of the resultant of these two effects on consumption depends on whether the consumer is a net lender in period 1 and a net borrower in period 2 or vice-versa. In case the consumer is a net lender in period 1 and a net borrower in period 2 the net income effect is positive. Assuming the consumption in both periods in a normal good, this means that the substitution effect and the income effect act in opposite directions on c_1 in this case leading to an ambiguous effect.

[1] For more about the Slutsky equation in the case of a consumer with fixed endowments of goods see section 9.1 in Varian's *Microeconomic Analysis*, 3rd ed.

2. Many periods

Assume that rather than just living for two periods the consumer lives for $T+1$ periods. Further assume that the real rate of interest takes a constant value r over the consumer's lifetime. For convenience we define $\delta = 1/(1+r)$. It is also convenient to start time from period 0 rather than period 1.

2.1. Budget constraint. Arguing as before, the consumer's budget constraint is

$$\sum_{i=0}^{T} \delta^i c_i = \sum_{i=0}^{T} \delta^i y_i \tag{6}$$

2.2. Utility function. We could proceed as before by assuming a utility function $U(c_0, \ldots, c_T)$ and deriving the first order conditions. However, because the marginal utility in each period depends on consumption in all periods it is hard to draw any sharp conclusions at this level of generality. Therefore we need to impose some restrictions on the form of the utility functions.

Suppose, for example we assume that the utility function is additively separable, i.e.

$$U(c_0, \ldots, c_T) = v_0(c_0) + v_1(c_1) + \cdots + v_T(c_T) \tag{7}$$

Then the first-order conditions take the form

$$v_i'(c_i) = \delta^i \lambda \qquad i = 0, \ldots, T \tag{8}$$

where, as before, λ is the Lagrange multiplier corresponding to the budget constraint.

Sometimes we want to restrict the consumers preferences even further, by assuming that the different v_i differ from each other by only a geometric discounting factor.

$$U(c_0, \ldots, c_T) = \sum_{i=0}^{T} \beta^i u(c_i) \tag{9}$$

where β is a constant, referred to as the subjective rate of discount, such that $0 < \beta < 1$.

In this case the first-order conditions take the particularly simple form

$$u'(c_i) = \left(\frac{\delta}{\beta}\right)^i \lambda \qquad i = 0, \ldots, T \tag{10}$$

In case $\delta = \beta$, this implies that $u'(c_i)$ is the same for all i, which, assuming that $u'(\cdot)$ is a strictly decreasing function, means that c_i is constant for all i. The present period's income does not influence the present period's consumption at all. Consumption is determined solely by lifetime resources as given by (6).

The case $\delta \neq \beta$ is also instructive. Suppose $\delta > \beta$. In this case it follows from (10) that consumption decreases over time. Formally, this is because if $\delta > \beta$ then by (10) $u'(c_i)$ increases over time, and since $u'(c)$ is a decreasing function of consumption, this implies that c decreases over time.

The economic logic behind this result is that δ is the number of units of consumption we have to give up at present in order to purchase one more unit of consumption next period, whereas β is the number of units of marginal utility we are willing to give up at present in order to have one more unit of marginal utility in the next period. Suppose we start with the same consumption c in this period and the next. If we reduce consumption in the next period by a small amount Δc then at the prevailing market prices we can increase present consumption by $\delta \Delta c$. The increase in utility from the increase in present consumption is approximately $u'(c)(\delta \Delta c)$.[2] The decrease in utility from the reduction in next period's consumption is approximately $\beta u'(c)(\Delta c)$. The net change in utility would be $(\delta - \beta)u'(c)(\Delta c)$ which is positive when $\delta > \beta$. Thus it is beneficial to increase present consumption and reduce future consumption if we are starting from a position of equality. Indeed, it will be optimal to increase consumption in the present period (say period i) and decrease consumption in the next period (period $i+1$) till the following equality between the MRS and the price ratio is satisfied,

$$\frac{u'(c_{i+1})}{u'(c_i)} = \frac{\delta}{\beta}$$

If δ/β is close to 1 then c_{i+1} is close to c_i and we can use Taylor's Theorem from calculus to the above equation to the above equation

[2]We are using Taylor's theorem: $u(c + \delta\Delta c) - u(c) \approx u'(c)(\delta\Delta c)$

to get a useful approximation.

$$\frac{u'(c_{i+1})}{u'(c_i)} = \frac{\delta}{\beta}$$
$$\frac{u'(c_{i+1}) - u'(c_i)}{u'(c_i)} = \frac{\delta}{\beta} - 1$$

Applying Taylor's Theorem

$$\frac{u''(c_i)(c_{i+1} - c_i)}{u'(c_i)} \approx \frac{\delta}{\beta} - 1$$

Defining $\Delta c = c_{i+1} - c_i$, and dropping the subscript i,

$$\left(\frac{u''(c)c}{u'(c)}\right)\left(\frac{\Delta c}{c}\right) \approx \frac{\delta}{\beta} - 1$$

The quantity $\sigma = -u'(c)/cu''(c)$ is known as the *intertemporal elasticity of substitution* and captures the sensitivity of marginal utility of changes in consumption. It is positive since marginal utility decreases with consumption.

$$\left(\frac{\Delta c}{c}\right) \approx \sigma\left(1 - \frac{\delta}{\beta}\right)$$

The formula confirms our earlier reasoning that consumption decreases over time if $\delta > \beta$. Moreover, it shows that the sensitivity of the growth of consumption on the rate of return depends on the intertemporal elasticity of substitution. This is because the intertemporal elasticity of substitution is the reciprocal of the elasticity of marginal utility with respect to the level of consumption. The more elastic is marginal utility to consumption, the smaller is the deviation in consumption from a constant path that is required the equate the ratio of marginal utilities in consecutive time periods to δ/β.

2.3. Exogenous variables. It is possible to unify (7) and (9) by writing

$$v_i(c_i) = \beta^i u(c_i, \xi_i)$$

where ξ_i is an exogenous variable such a the consumer's age or the number of members in the household. In this case the first-order conditions become

$$u'(c_i, \xi_i) = \left(\frac{\delta}{\beta}\right)^i \lambda \qquad i = 0, \ldots, T$$

Knowing how ξ affects the marginal utility would now let us make some predictions regarding the path of consumption.

2.4. Comparative statics. Assuming that consumption in every period is a normal good, an increase in y_i increases every c_i.

The effect of an increase in r, or equivalently, a decrease in δ remains ambiguous because of the same income and substitution effects as discussed earlier. But for the utility function given by (9), we can say a little more. From (10) we can see that a decrease in δ means that the *growth rate* of consumption speeds up. Remember that even in this case we do not have any information regarding the *level* of consumption in any period since the level would depend on λ which in turn depends on δ.

Exercises

1. A consumer consumes a single good in two periods—period 1 and period 2. Let her consumption in the two periods be denoted by c_1 and c_2 respectively. The consumer has an endowment of e_1 and e_2 units of consumption in the two periods respectively. The consumer has no other sources of income or wealth. Assume that the *money* price of the consumption good in the two periods is P_1 and P_2 respectively and the nominal interest rate between the two periods is i.
   (a) Write down the consumer's budget constraint.
   (b) Write down an exact (not approximate) formula for the real interest rate in this setting in terms of P_1, P_2 and i. (Think of how many additional units of consumption you can get in period 2 if you give up one unit of consumption in period 1).
   (c) Argue that the consumer's budget set depends only on the real rate of interest, i.e., combinations of changes in prices and the nominal interest rate which leave the real interest rate unchanged also leave the consumer's budget set unchanged.

(d) Assume that the consumer's utility function is given by:

$$U(c_1, c_2) = c_1 c_2$$

Calculate the amount consumed and the amount saved by the consumer in period 1 as a function of the real interest rate. Sketch rough graphs of these two functions.

2. Consider a consumer who lives from period 0 to T, has an initial wealth w and no other sources of income. Suppose that the consumer has additive separable preferences with the felicity function,

$$v_i(c_i) = \beta^i c_i^{1-\rho}, \quad 0 < \rho < 1.$$

The consumer can lend and borrow any amount she wishes at the real rate of interest r.

(a) What is the intertemporal elasticity of substitution corresponding to this consumer's felicity function?

(b) Use the first-order conditions of the conumer's utility maximization problem (not a linear approximation) to show that the optimal consumption path chosen by this consumer shows a constant rate of growth of consumption. Derive an expression for the growth rate of consumption in terms of β, r and ρ.

3. Consider a consumer who lives for two periods and must decide on how much to spend on a durable good in each of the two periods. The consumer's utility function is given by

$$U(x_1, x_2) = [x_1^{1-\rho} + x_2^{1-\rho}]/(1-\rho), \qquad \rho > 0$$

where x_1 and x_2 is the stock of durable goods held by the consumer in the two periods.

The stocks are related to the consumer's expediture c_1 and c_2 in the two periods by

$$x_1 = c_1$$
$$x_2 = \gamma x_1 + c_2, \qquad 0 < \gamma < 1$$

where $(1-\gamma)$ is the rate at which the stock of the durable depreciates.

The consumer has an initial wealth w and no other source of income. She is free to lend and borrow at the interest rate r.

Under what conditions on γ, R and ρ will the consumer not spend anything in the second period? Give an economic interpretation for your result.

4. A and B are two agents who derive satisfaction from the consumption of leisure and apples over a number of periods. For A leisure and apples are substitutes whereas for B leisure and apples are complements. Suppose an exogenous shock reduces the leisure available to both in a given period without affecting their incomes. How will the consumption of apples in that period change for each?

5. Consider a consumer who lives for two periods. The consumer has a real earning of y_1 and y_2 in the two periods respectively and must choose his level of real consumption c_1 and c_2 in the two periods.

   The consumer can lend and borrow at the real rate of interest r. However, the consumer cannot borrow more than a fraction θ ($0 < \theta < 1$) of the present value of his second period earnings. That is, if b_1 is the amount borrowed by the consumer in the first period then it must be the case that

$$b_1 \leq \frac{\theta y_2}{1+r}.$$

   There are no restrictions on the amount that the consumer can lend.

   (a) Sketch the consumer's budget set in the c_1, c_2 plane.

   (b) Suppose that the consumer's utility function is given by

$$U(c_1, c_2) = \ln c_1 + \ln c_2.$$

   Calculate the consumer's first-period consumption demand (c_1) as a function of y_1, y_2, θ and r. [Hint: Take the possibility of a corner solution into account.]

   (c) Calculate this consumer's first-period marginal propensity to consume $\partial c_1/\partial y_1$. How does this marginal propensity to consume change with changes in y_1? Explain your answer in economic terms.

6. Consider the following example of a two-period utility function with habit formation:

$$U(c_1, c_2) = \frac{1}{1-\rho}[c_1^{1-\rho} + (c_2 - \gamma c_1)^{1-\rho}], \qquad \gamma > 0, \rho > 0$$

(a) Suppose a consumer with these preferences has an initial wealth w, no other sources of income and can freely lend and borrow at the interest rate $r > 0$. What will be the consumer's optimal choice of c_1 and c_2?

(b) An economist wishes to use observed data on c_1, c_2 and r to estimate the parameter ρ. But the economist mistakenly assumes that there is no habit formation. That is, the economist mistakenly assumes that the consumer's preferences are

$$U(c_1, c_2) = \frac{1}{1-\rho}[c_1^{1-\rho} + c_2^{1-\rho}], \qquad \rho > 0.$$

Will this economist's estimate of ρ be higher or lower than the true value?

CHAPTER 4

The Envelope Theorem

1. Parametrised optimisation problems

Let's think of unconstrained problems first. Every optimisation problem has an objective function. It is the function that we are trying to maximise or minimise (henceforth maximise). Some of the variables entering the objective function are *choice variables*, variables whose values we are free to choose in order to maximise the objective function. But all the variables entering into the objective function need not be choice variables. The value of the objective function may also depend on the value of other variables which we are not free to choose. We call these the *parameters* of the optimisation problem.

EXAMPLE 4.1. Consider the short-term profit maximising problem of a firm that produces according to the production function

$$y = f(L, K) = L^{1/2} K^{1/2}$$

In the short-run the capital stock of the firm is fixed at some value $\bar{K}$ and the firm can only choose the labour input L. If the firms buys labour and capital in perfectly competitive labour market at prices w and r respectively and sells its output in a perfectly competitive market at the price p then its profits are:

$$\pi(L, \bar{K}) = py - wL - r\bar{K} = pL^{1/2}\bar{K}^{1/2} - wL - r\bar{K}$$

For the short-run profit maximising problem $\pi(L, \bar{K})$ is the objective function, with L as a choice variable and $\bar{K}$ as a parameter.[1]

[1] In fact p, r and w are also parameters in the profit function. But we shall ignore this fact for now since we will not be looking at the effects of changes in these variables.

Denoting the optimal amount of labour input by L^*, the first-order condition for profit maximisation is,

$$\begin{aligned}\frac{\partial \pi}{\partial L} &= 0\\ \frac{1}{2}pL^{*-1/2}\bar{K}^{1/2} - w &= 0\\ L^* &= \bar{K}(p/2w)^2\end{aligned}$$

You should check that $\pi(L, \bar{K})$ is a concave function of L and therefore the first-order condition is sufficient to give us a global maximum. The profit earned by the firm at the optimal point is,

$$\begin{aligned}\pi^* &= \pi(L^*, \bar{K})\\ &= p[\bar{K}^{1/2}(p/2w)]\bar{K}^{1/2} - w[\bar{K}(p/2w)^2] - r\bar{K}\\ &= \bar{K}(p^2/2w) - \bar{K}(p^2/4w) - r\bar{K}\\ &= \bar{K}(p^2/4w) - r\bar{K}\end{aligned}$$

We see that both the amount of labour input chosen by the firm and the maximum profit it earns are functions of the value of the parameter $\bar{K}$. The function mapping the parameter values to the maximum (or minimum) value of the objective function is called the *value function.* In this case, denoting the value function by $V(\cdot)$ we have,

$$V(\bar{K}) = \pi^* = \bar{K}(p^2/4w) - r\bar{K}$$

□

2. The envelope theorem

How does the optimal value change when we change the parameters? In our example since we have an explicit formula for the value function we can calculate its value directly

$$V'(\bar{K}) = (p^2/4w) - r\bar{K}$$

Even when we do not have an explicit formula for the value function, there is an interesting relationship between the partial derivatives of the objective function and the partial derivatives of the value function.

Consider the general problem of maximising the objective function

$$\phi(x_1, \ldots, x_n; c_1, \cdots, c_m)$$

where the x_i are choice variables and c_i are parameters.

The first order conditions for the problem are,

$$\frac{\partial \phi}{\partial x_i}(x_1, \ldots, x_n; c_1, \ldots, c_m) = 0 \qquad i = 1, \cdots, n \tag{11}$$

Just as in the example, the optimal values of the choice variables, denoted by x_i^*, will be functions of the parameters $c_1, \ldots, c_m$. The value function will be given by

$$V(c_1, \ldots, c_m) = \phi(x_1^*, \ldots, x_n^*; c_1, \ldots, c_m)$$

Suppose we want to calculate the partial derivative of the value function with respect to one of the parameters, say c_j. In doing so we have to take into account the fact that the optimal value of each of the choice variables would also be a function of c_i. If we assume that the mapping from the c_i to the optimal values of the choice variables is differentiable, we can use the chain rule,

$$\frac{\partial V}{\partial c_j} = \frac{\partial \phi}{\partial x_1}\frac{\partial x_1^*}{\partial c_j} + \cdots + \frac{\partial \phi}{\partial x_n}\frac{\partial x_n^*}{\partial c_j} + \frac{\partial \phi}{\partial c_j}$$

However, from (11), we know that $\partial \phi / \partial x_i$ is 0 for all i when the partial derivatives are evaluated at the optimal values. So we have,

$$\frac{\partial V}{\partial c_j} = \frac{\partial \phi}{\partial c_j} \tag{12}$$

This remarkably is the same result that we would have got if we had treated each of the x_i^* as a constant. But that would not have been justified since the choice variables do vary when parameters are varied. That is, $\partial x_i^* / \partial c_j$ is generally not zero. It is just that when we are starting from an optimal point then the marginal impact on this variation on the objective function (i.e., $\partial \phi / \partial x_i$ is zero and therefore we can ignore the changes in the choice variables.

Equation (12) is known as the "Envelope Theorem".

3. Geometric Interpretation

Figure 4.1 illustrates the envelope theorem in the case of Example 4.1. Each of the coloured curves shows the level of profit for a given level of L and for different values of K. Let's call them "profit

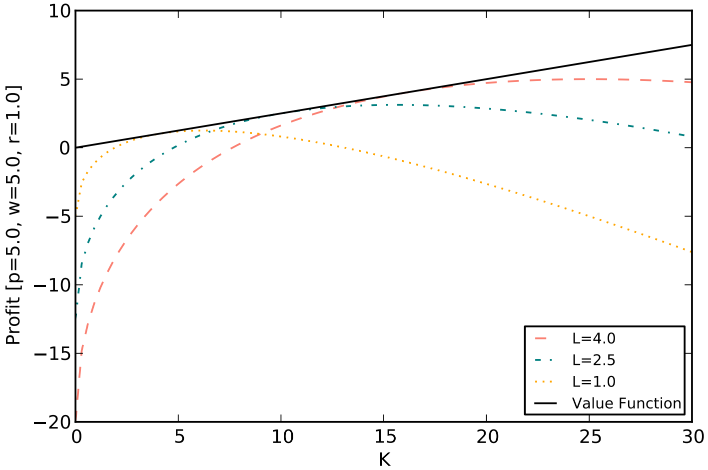


FIGURE 4.1. The Envelope Theorem

curves".[2] We have drawn only three of these curves but you should imagine there to be one curve for each possible value of L. Now, since our purpose is to maximise profit for a given value of K, we move along a vertical line for our particular value of K and choose that L whose profit curve is the highest at that value of K.

Thus, for example, at $K = 4.0$ we would choose $L = 4.0$ whereas at $K = 10.0$ we would choose $L = 2.5$.

The value of the highest profit curve for a given K gives us the highest profit we can obtain when K takes on that value. But that is precisely the definition of the value function. Therefore the graph of the value function touches the highest of the profit curves at each K. Or, in other words, the graph of the value function (the black line in the figure) must be the upper envelope of the graphs of the profit functions for given values of L.

Since the value function is the upper envelope of the profit curves, no profit curve can ever cross it. But at each value of K one of the

[2]This is not standard terminology and you must remember that these curves are not graphs of the full profit function since we are holding L constant on each of them.

profit curves, corresponding to the optimal L, touches it. The only way two graphs can touch without crossing is if they are tangent to each other. The slope of the graph of the value function is $\partial V/\partial K$ whereas the slope of the profit curves is $\partial \pi/\partial K$. Tangency of the two graphs implies that these slopes should be equal, which is precisely what our the envelope theorem in eq. (12) also says when applied to this example.

Now you know what the envelope theorem is called by that name.

4. Constrained Optimisation

So far we have discussed unconstrained problems. There is also a version of the envelope theorem for constrained optimisation problems. Suppose our problem is to maximise

$$\phi(x_1, \ldots, x_n; c_1, \ldots, c_m)$$

subject to the constraint

$$h(x_1, \ldots, x_n; c_, \ldots, c_m) = 0 \tag{13}$$

Here we have allowed both the objective function and the constraint to depend on a set of parameters.

The first-order condition for this problem is

$$\frac{\partial \phi}{\partial x_i} = \lambda \frac{\partial h}{\partial x_i} \qquad i = 1, \ldots n \tag{14}$$

where λ is a Lagrange multiplier.

As before, if the problem has a solution the optimal values of the choice variables, the x_i^*, will be functions of the parameters of the problem. Once again we look at the case where this mapping is differentiable.[3] Also as before, we can define the value function as

$$V(c_1, \ldots, c_m) = \phi(x_1^*, \ldots, x_n^*; c_1, \ldots, c_m)$$

Differentiating the value function with respect to c_j gives us,

$$\frac{\partial V}{\partial c_j} = \frac{\partial \phi}{\partial x_1}\frac{\partial x_1^*}{\partial c_j} + \cdots + \frac{\partial \phi}{\partial x_n}\frac{\partial x_n^*}{\partial c_j} + \frac{\partial \phi}{\partial c_j} \tag{15}$$

To simplify this we need to digress a bit. The optimal values of the choice variables must satisfy the constraint (13) for all values of

[3]For sufficient conditions that this be so see [**LY08**, Section 11.7].

the parameters, so we have

$$h(x_1^*, \ldots, x_n^*; c_, \ldots, c_m) = 0.$$

Differentiating this with respect to c_j we get

$$\frac{\partial h}{\partial x_1}\frac{\partial x_1^*}{\partial c_j} + \cdots + \frac{\partial h}{\partial x_n}\frac{\partial x_n^*}{\partial c_j} + \frac{\partial h}{\partial c_j} = 0$$

Substituting the first-order conditions (14) we have,

$$\frac{1}{\lambda}\frac{\partial \phi}{\partial x_1}\frac{\partial x_1^*}{\partial c_j} + \cdots + \frac{1}{\lambda}\frac{\partial \phi}{\partial x_n}\frac{\partial x_n^*}{\partial c_j} + \frac{\partial h}{\partial c_j} = 0$$

or,

$$\frac{\partial \phi}{\partial x_1}\frac{\partial x_1^*}{\partial c_j} + \cdots + \frac{\partial \phi}{\partial x_n}\frac{\partial x_n^*}{\partial c_j} = -\lambda\frac{\partial h}{\partial c_j}$$

Now this can be substituted in (15) to give us

$$\frac{\partial V}{\partial c_j} = -\lambda\frac{\partial h}{\partial c_j} + \frac{\partial \phi}{\partial c_j} \tag{16}$$

Equation (16) is the envelope theorem for the constrained case. It is similar to the unconstrained envelope theorem in that the change in the choice variables as a result of the change in the parameters drops out of the calculation. It differs in that the change in the value function as a result of a change in a parameter depends not just on the direct change in the objective function ($\partial\phi/\partial c_j$) but also the change in constraint set ($\partial h/\partial c_j$). The Lagrange multiplier λ can be interpreted as a sensitivity factor, indicating the extent to which a given change in the constraint set translates into a change in the value function.

Example 4.2. Consider the problem of maximising the utility function $U(x_1, x_2)$ subject to the budget constraint $p_1x_1 + p_2x_2 = M$. Treating x_1 and x_2 as choice variables and p_1, p_2 and M as parameters, we have the objective function

$$\phi(x_1, x_2; p_1, p_2, M) = U(x_1, x_2)$$

and the constraint function

$$h(x_1, x_2; p_1, p_2, M) = p_1x_1 + p_2x_2 - M$$

In this case the value function $V(p_1, p_2, M)$ is important enough to be given a name. It is called the *indirect utility function*.

If the value of the Lagrange multiplier at a the optimal bundle is λ, then the envelope theorem (16) tells us that,

$$\begin{aligned}\frac{\partial V}{\partial M} &= -\lambda\frac{\partial h}{\partial M} + \frac{\partial \phi}{\partial M} \\ &= -\lambda \cdot -1 + 0 \\ &= \lambda\end{aligned}$$

This gives us an economic interpretation of the Lagrange multiplier. It measures the amount by which the maximum attainable utility increases per unit increase in income. In looser phrasing, it is the "marginal utility of income".

□

CHAPTER 5

Dynamic programming

1. The setup

In a dynamic optimisation problem, our goal is to find a *path* of the choice variable which maximises the value of an objective function defined over the entire path of the choice variable. Often, there are constraints on what paths can be chosen. For example, in the consumption-saving problem we choose a path of consumption which maximises the lifetime utility function subject to a budget constraint.

The dynamic programming approach to solving dynamic optimisation problems turns this single large optimization problem into a sequence of simple optimization problems. At each point of time we try to find the best action at that particular point of time. But since this is a dynamic problem after all, this search for the best actions at a particular point of time has to be done with an eye on both the past and the future. Past events and actions[1] determine what choices can be made now. By the same token, the action that we take now will change the options available to us the in future. The value of the objective function that will be achieved will in general depend on the entire path of past, present and future actions and not just the action in any period in isolation.

In the dynamic programming framework this linkage between the past and the future is captured by the notion of the *state*. Intuitively, we can think of the state at any given point of time as a description of all the *relevant* information about the actions and events that have happened until that point. The state should contain all the information that is required from the decision-maker's history to determine

[1]We want to make a distinction between *events* which are outside of our control and *actions* which are things we choose. This distinction becomes important when we are dealing with uncertainty.

the set of available actions at future points of time and to evaluate the contribution made by future actions to the objective function.

The notion that knowing the state at a point of time is enough to know what actions are possible in the future is captured by the following definitions:

Set of states (S_t): this is the set of possible states the decision-maker can be in time t. There is one such set for each time period t. The elements of these sets, i.e. the possible states, are assumed to be vectors with real-number elements. Elements of this set are denoted by s_t.

Set of actions (A_t): this is the set of possible actions that can be taken at time t. Elements of this set are denoted by a_t. As we shall see next, all possible actions cannot be taken at all possible states.

Constraint correspondence $f_t(s_t) \subset A_t$: this tells us the subset of actions that are available in a particular state. This is not a function but a correspondence (i.e. a set-valued function) since for each element of S_t it gives us a subset and not just a single element of A_t.

Transition function $\Gamma_t(s_t, a_t) \in S_{t+1}$: This tells us our state in period $t + 1$ if we take the action a_t in state s_t in period t.

With these definitions in hand we can define the set of *feasible plans* when starting with s_t at time t, denoted by $\Phi_t(s_t)$, as the set of sequences of actions $(a_t, \ldots, a_T)$ such that

$$a_i \in f_i(s_i) \quad \text{for } i = t, \ldots, T$$

and

$$s_{i+1} = \Gamma_i(s_i, a_i) \quad \text{for } i = t, \ldots, T - 1$$

The first condition says that the action taken on each date is a feasible action given the state. The second condition says that the state at each date is derived from the state and action taken in the previous date, with the state at time t as given.

The set of feasible plans tells us about the constraints faced in our optimisation problem. What about the objective function? We assume that the objective function can be written in an additively separable form

$$U_t(a_t, s_t, \ldots, a_T, s_T) = \sum_{i=t}^{T} v_i(a_i, s_i)$$

where $v_t(a_t, s_t)$ is the *per-period payoff function* that gives the contribution of action a_t in state s_t at time t to the overall objective. Being able to write the objective function in an additively separable form is essential for us to be able to use dynamic programming.[2]

In writing the above objective function we have also assumed that there is a finite time period T at which our optimisation problem comes to an end. This assumption of what is known as a finite horizon is made just to simplify the mathematics. Dynamic programming problems with an infinite horizon are routinely used in economic modelling.

The solution to the dynamic programming problem is expressed in terms of two functions:

Policy function $g_t(s_t) \in f_t(s_t)$: The policy function tells us the best action to take in each possible state at time t among all the available actions. In general it is possible that there be two equally good actions at a particular state, in which case the policy function would have to be replaced by the policy correspondence.

Value function $V_t(s_t) \in \mathbb{R}$: The value function denotes the maximum attainable value of the objective function when starting at time t from state s_t. That is,

$$\begin{aligned} V_t(s_t) &= \max_{(a_t,\ldots,a_T)\in\Phi_t(s_t)} U_t(a_t, s_t, \ldots, a_T, s_T) \\ &= \max_{(a_t,\ldots,a_T)\in\Phi_t(s_t)} [v_t(a_t, s_t) + \cdots + v_T(a_T, s_T)] \end{aligned}$$

In applications of dynamic programming we generally want to know the optimal path starting at a specific point of time (taken to be $t = 0$ here) and from a particular state at that point of time (say $\bar{s}_0$). But we have defined the policy and value functions for all points of time and for each possible state at each of the time periods. Thus it would seem that we have multiplied our work manyfold beyond what is necessary for our original problem. But as we shall see below, being

[2]We are cheating a bit here. The assumption of additive separability can be relaxed to what is called 'recursiveness' while still allowing the use of dynamic programming.

willing to contemplate the policy and value functions for all possible time periods and states often actually simplifies the task of solving the original problem.

2. Bellman's Principle of Optimality

Suppose I am starting at some time $t < T$ from some particular state s_t and trying to find the best actions from time t to T, where 'best' means the choices of actions and consequent states which maximise

$$U_t(a_t, s_t, \ldots, a_T, s_T) = \sum_{i=t}^{T} v_i(a_i, s_i).$$

Because of the additive nature of the lifetime utility function we can rewrite the above equation as

$$U_t(a_t, s_t, \ldots, a_T, s_T) = v_t(a_t, s_t) + U_{t+1}(a_{t+1}, s_{t+1}, \ldots, a_T, s_T)$$

If we divide the plan (path of actions and corresponding states) from time t to time T into a "head" consisting of the action in period t and a "tail" consisting of actions in period $t+1$ to T then the above equation says that the lifetime utility of the plan starting at period t is the sum of the per-period payoff at time t (the value of the "head") and the lifetime utility of the remaining part of the plan from period $t+1$ onwards (the value of the "tail").

How do we find the plan which maximises U_t? Suppose we choose the action $\hat{a}_t$ in period t. This will lead us to the state $\hat{s}_{t+1} = \Gamma_t(s_t, \hat{a}_t)$ in the next period. Now we have to pick a plan from period $t+1$ onward. Now $U_t = v_t(\hat{a}_t, s_t) + U_{t+1}$ and $v_t(\hat{a}_t, s_t)$ is already fixed by our choice of action $\hat{a}_t$ in period t. Therefore in choosing our plan from period $t+1$ onward the best we can do is to pick a plan that maximises U_{t+1}. This optimal plan for the "tail" yields the value of U_{t+1} equal to $V_{t+1}(\hat{s}_{t+1})$. Thus we can evaluate each choice of action $\hat{a}_t$ in the "head" by looking at

$$\tilde{U}_t = v_t(\hat{a}_t, s_t) + V_{t+1}(\hat{s}_{t+1}), \quad \text{where } \hat{s}_{t+1} = \Gamma(s_t, \hat{a}_t)$$

We have put a tilde over U_t to remind ourselves that now we are not considering arbitrary plans starting at t but only plans where the "tail" component is optimal given the state $\hat{s}_{t+1}$ at which we find ourselves in the beginning of period $t+1$.

The optimal plan from period t involves choosing $\hat{a}_t$ which maximises the expression above. Since the value function gives the value of the objective function U_t for the optimal plan, it is therefore the case that,

$$V_t(s_t) = \max_{\hat{a}_t \in f_t(s_t)} [v_t(\hat{a}_t, s_t) + V_{t+1}(\Gamma(s_t, \hat{a}_t))], \quad \text{for } t < T \tag{17}$$

Equation (17) above which relates the value function at consecutive time periods is known as *Bellman's Equation*. The argument above, which shows that the value function must satisfy Bellman's equation is known as *Bellman's Principle of Optimality*.[3]

Intuitively Bellman's equation tells us that we can evaluate each present action by adding its contribution $v_t(\hat{a}_t, s_t)$ to the objective in the present period and the value $V_{t+1}(\hat{s}_{t+1})$ of the state in which it leaves us in the next period. Provided we know the function V_{t+1} for all possible states in the next period we can choose the best action in the current period by choosing $\hat{a}_t$ to maximise this sum. Thus we have turned the big optimisation problem of choosing an entire sequence of actions from time 0 to time T into a sequence of simple optimisation problems, one for each time period t, in each of which we choose a single action $\hat{a}_t$.

But there seems to be a chicken-and-egg problem: we cannot use Bellman's equation without knowing V_{t+1} for each t and how do we know V_{t+1} if we have not solved the optimisation problem already? Here our finite horizon assumption makes life particularly simple for us.

Since period T is the last period, our objective function in that period is

$$U_T(a_T, s_T) = v_T(a_T, s_T)$$

and the value function is simply given by

$$V_T(s_T) = \max_{a_T \in f_T(s_T)} v_T(a_T, s_T)$$

We can solve this maximisation problem and calculate V_T since $v_T(\cdot, \cdot)$ is a known function.

[3] To be complete, Bellman's principle of optimality also deals with the converse: that a function which satisfies Bellman's equation plus some other technical conditions must be the value function. This converse is not important in our current finite horizon setting.

Now consider Bellman's equation for period $T-1$:

$$V_{T-1}(s_{T-1}) = \max_{\hat{a}_{T-1} \in f_t(s_{T-1})} [v_{T-1}(\hat{a}_{T-1}, s_{T-1}) + V_T(\Gamma(s_{T-1}, \hat{a}_{T-1}))]$$

As we have calculated $V_T(\cdot)$ in the previous step, all the functions in the maximisation problem are known and we can solve the problem to calculate V_{T-1}. With this in hand we can solve Bellman's equation for period $T-2$. We keep going backward one period at a time until we have calculated the value function for all periods until period 0. At each step the value of $\hat{a}_t$, as a function of s_t, which solves the maximisation problem gives us the policy function. So by the end of our process we also have the policy function for each time period.

Now if we are given a starting state $\bar{s}_0$ in period 0 we can use the calculated policy function for period 0 to find the best action a_0 in period 0. We know from the transition function that we will end up in state $s_1 = \Gamma(\bar{s}_0, a_0)$ in the next period. The policy function for period 1 tells us the best action a_1 to take in that period. We again use the transition function to tell us the next state $s_2 = \Gamma(s_1, a_1)$. And so on until we have traced out the optimal plan to time T. Our optimisation problem is solved!

The way we have calculated the value function backwards from a known final time period is sometimes called "backward induction".

3. Example: consumption-savings with log utility

Suppose that the consumer maximises

$$\sum_{i=0}^{T} \log(c_i)$$

subject to

$$\sum_{i=0}^{T} c_i/R^i = w_0$$

Can we solve this maximisation problem using dynamic programming? The action variable in this case must be c_i since it is the variable being chosen by the decision maker. The objective function is already in an additively separable form with a per-period payoff $\log(c_i)$. But what is the state?

Since the per-period payoff depends only on the action variable c_i we do not need any notion of state to evaluate payoffs. But the choice of a consumption in each period does affect future periods through the budget. The more we consume today, the less purchasing power we have to consume tomorrow. We can capture this by rearranging the budget slightly to read,

$$\sum_{i=1}^{T} c_i/R^i = w_0 - c_0$$

Multiplying throughout by R we have

$$\sum_{i=1}^{T} c_i/R^{(i-1)} = R(w_0 - c_0)$$

Which shows us that the path of consumption from time 1 onwards follows a budget constraint of the same form as the period 0 budget constraint provided we take

$$w_1 = R(w_0 - c_0)$$

This suggests to us that we can take the wealth at the beginning of period t as our state with the transition function,

$$w_{t+1} = R(w_t - c_t), \quad \text{for } t = 0, \ldots, T$$

and the constraint function

$$c_T = w_T$$

The constraint captures the fact that there can be no outstanding debt in the last period and a consumer who has monotonic preferences would not leave any wealth unused in the last period. There is no constraint on consumption in periods other that T.[4].

You can check that we have formulated the problem right by eliminating $w_1, \ldots, w_T$ in the transition functions and constraints above to recover our original budget constraint.

Now we can use backward induction to calculate the value function and policy function for all time periods. Denoting the policy function by $g(\cdot)$ and the value function in period t by $V_t(\cdot)$, we have,

$$g_T(w) = w, \qquad V_T(w_T) = \log(g_T(w)) = \log(w) \tag{18}$$

[4]We could have imposed a non-negativity constraint but we leave it out for simplicity

Now consider period $T-1$. Bellman's principle of optimality tells us,

$$\begin{aligned} V_{T-1}(w) &= \max_c [\log(c) + V_T(R(w-c))] \\ &= \max_c [\log(c) + \log(R(w-c))] \qquad \text{[using (18)]} \end{aligned} \tag{19}$$

The first-order condition for this maximisation problem is:

$$\begin{aligned} \frac{1}{c} + \frac{-R}{R(w-c)} &= 0 \\ w - c &= c \\ c &= w/2 \end{aligned}$$

Since the objective function in (19) is concave in c (check this!), the first-order condition is sufficient and gives us our policy function:

$$g_{T-1}(w) = w/2$$

Substituting this into (19) we get the value function,

$$\begin{aligned} V_{T-1}(w) &= \log(g_{T-1}(w)) + V_T(R(w - g_{T-1}(w))) \\ &= \log(w/2) + \log(Rw/2) \\ &= \log(R) + 2\log(w/2) \end{aligned} \tag{20}$$

Now that we know V_{T-1} we could use the Bellman equation relating V_{t-2} to V_{t-1} to derive g_{T-2} and V_{t-2}. If you do this you will find,

$$g_{T-2}(w) = w/3, \qquad V_{T-2}(w) = (1+2)\log(R) + 3\log(w/3) \tag{21}$$

We could continue like this to find $V_{T-3}, \ldots, V_0$. In general this is precisely what we do. In fact, in most applications of dynamic programming it is not possible to express the value function by a formula in the state variables and the best that we can do is to use a computer to calculate the value function at a number of possible values of the state variable using Bellman's equation.

But our present problem is a particularly simple one. Looking at (20) and (21) suggests to us the guess,

$$V_{T-n}(w) = \frac{n(n+1)}{2}\log(R) + (n+1)\log\left(\frac{w}{n+1}\right) \tag{22}$$

[Remember $1 + 2 + \cdots + n = n(n+1)/2$]

How do we check that our guess is right? We will use the principle of mathematical induction. By comparing to (20) we see that (22) is correct for $n = 1$. Suppose that the equation is true for $n = k$. What then would be $V_{T-(k+1)}$? We once again write down the Bellman equation

$$\begin{aligned} V_{T-(k+1)} &= \max_c [\log(c) + V_{T-k}(R(w-c))] \\ &= \max_c \left[\log(c) + \frac{k(k+1)}{2}\log(R) + (k+1)\log\left(\frac{R(w-c)}{k+1}\right)\right] \\ &\text{[assuming (22)]} \end{aligned} \tag{23}$$

The first-order condition is:

$$\begin{aligned} \frac{1}{c} + (k+1)\left(\frac{k+1}{R(w-c)}\right)\left(\frac{-R}{k+1}\right) &= 0 \\ (k+1)\frac{1}{(w-c)} &= \frac{1}{c} \\ c &= w/(k+2) \end{aligned} \tag{24}$$

Substituting this into (23) we have

$$\begin{aligned} V_{T-(k+1)} &= \log(c) + \frac{k(k+1)}{2}\log(R) + (k+1)\log\left(\frac{R(w-c)}{k+1}\right) \\ &\text{substituting (24),} \\ &= \log\left(\frac{w}{k+2}\right) + \frac{k(k+1)}{2}\log(R) + (k+1)\log\left(\frac{Rw}{k+2}\right) \\ &= \log\left(\frac{w}{k+2}\right) + \frac{k(k+1)}{2}\log(R) + (k+1)\log(R) + (k+1)\log\left(\frac{w}{k+2}\right) \\ &= \frac{(k+2)(k+1)}{2}\log(R) + (k+2)\log\left(\frac{w}{k+2}\right) \end{aligned} \tag{25}$$

But this is the same as (22) for $n = k + 1$. We therefore conclude that if (22) is true for $n = k$ it is also true for $n = k + 1$. We have already checked that (22) is true for $n = 1$. Hence we conclude by the principle of mathematical induction that the value function for our dynamic programming problem is given by (22) for $n = 1, \ldots, T$.

Also, now that we have verified that (22) is indeed the value function of the problem, (24) gives the policy function, i.e.

$$g_{T-n}(w) = w/(n+1) \tag{26}$$

4. The Euler equation

As we discussed in the last section, for most dynamic programming problems it is not possible to compute the value and policy functions in terms of simple formulae. The best we can do is to calculate numerical values. But even if we cannot find an exact formula for the solution to our optimisation problem, it may still be possible to get some qualitative information about the problem by studying the consequences of the Bellman equation. That is the subject of this section.

Let's recall the Bellman equation,

$$V_t(w_t) = \max_{c_t}[u(c_t) + V_{t+1}(R(w_t - c_t))$$

The first order condition for this maximisation problem is:

$$u'(c_t) = RV'_{t+1}(R(w_t - c_t)) \tag{27}$$

By itself (27) does not seem very useful unless we know $V_{t+1}(\cdot)$ and can calculate its derivative. But there is a trick that we can use to eliminate this unknown derivative from (27).[5]

Let $c^*_t(w_t)$ be the optimal consumption in period t when period t wealth is w_t. From (27) we already know that,

$$u'[c^*_t(w_t)] = RV'_{t+1}[R(w_t - c^*_t(w_t))] = RV'_{t+1}(w_{t+1}) \tag{28}$$

But from the definition of the value function

$$V_t(w_t) = u[c^*_t(w_t)] + V_{t+1}[R(w_t - c^*_t(w_t))]$$

Differentiating with respect to w_t we have,

$$\begin{aligned} V'_t(w_t) &= u'[c^*_t(w)]c^{*\prime}_t(w_t) + V'_{t+1}[R(w_t - c^*_t(w_t))][R(1 - c^{*\prime}_t(w_t))] \\ &= c^{*\prime}_t(w_t)[u'(\cdot)) - RV'_{t+1}(\cdot)] + RV'_{t+1}[R(w_t - c^*_t(w_t))] \end{aligned}$$

From (28) the first terms equals 0, so,

$$V'_t(w_t) = RV'_{t+1}(w_{t+1})$$

[5]The 'trick' is a particular case of a general result known as the envelope theorem. See section M.L of Mas-Colell, Whinston and Green or some mathematical methods book for more detail.

Using (28)

$V'_t(w_t) = u'(c_t)$

The equation above was derived for arbitrary t. So it is equally good for $t+1$, i.e.

$$V'_{t+1}(w_{t+1}) = u'(c_{t+1})$$

Substituting this in (28) we have,

$$\frac{u'(c_{t+1})}{u'(c_t)} = \frac{1}{R} = \delta \tag{29}$$

This condition is known as the Euler[6] equation for our dynamic programming problem. We can alternatively derive it by starting out with an optimal consumption plan, increasing consumption in period t by a small amount Δc and reducing consumption in period $t+1$ by $R\Delta c$ so that wealth at the end of the period $t+1$ is once again the same as what it would have been under the optimal plan. The first-order change in utility from this deviation is

$$\Delta u = u'(c_t)[\Delta c] - u'(c_{t+1})[R\Delta c]$$

Now for the original plan to have been optimal Δu must be 0 since if $\Delta u > 0$ the deviation considered above increases total utility whereas if $\Delta u < 0$ then the opposite of the deviation considered above increases total utility. But $\Delta u = 0$ implies

$$u'(c_t) - Ru'(c_{t+1}) = 0$$

which is again our Euler equation (29).

The Euler equation also follows from the first-order conditions (8) of the Lagrange-multiplier approach, showing that we have come full circle.

The Euler equation tells us how consumption should grow or decline. It does not tell us the level of the consumption. But we can characterise the entire consumption path if we keep track of the path of wealth implied by the path of consumption and impose, in addition to the Euler equation, the conditions

$$w_0 = \overline{w_0}$$

[6] Pronounced "oiler". Named after a eighteenth-century mathematician who was among the earliest to study dynamic optimisation problems.

which comes to us as a given data and

$$w_T = 0$$

which comes to us from our no bequest, no terminal borrowing, monotonic utility assumptions about the terminal period.

CHAPTER 6

Probability

We assume that you are familiar with concepts such as *sample space*, *random variable*, *expectation* and *conditional expectation*. Here we only discuss some ideas important for macroeconomic applications. Our mathematical treatment is applicable to only to probability spaces with a finite number of outcomes but the ideas can be carried over to more general settings. References to more advanced treatments are provided at the end of the chapter.

1. Information structures

In macroeconomics we often have to deal simultaneously with randomness and time. For example, we might be interested in the entire time path of an agent's consumption expenditure when the expenditure in each period is random and may depend on other random factors as the consumer's income.

One way in which we could try to simultaneously model randomness and time would be to set up a different sample space for each time period. We would then define random variables corresponding to quantities measured in a particular period on the sample space for that period. However, if we do so we would not be able to model the dependence between quantities in different periods. Such dependence however is common. For example, receiving a promotion in period t may increase the likelihood of a high income in period $t + 1$ as well.

To capture the dependence between random events at different periods of time, we need a common sample space on which random variables corresponding to different time periods can be simultaneously define. We are thus led to construct a sample space such that each sample point contains enough information to tell us about the entire trajectory of any quantity that we may be interested in. While we continue to refer to sample points as "states of the world" in this context they are better seen as complete "histories of the world"—each sample point representing one possible history.

When modelling the combination of randomness and time we must also represent the fact that information is only gradually revealed to agents. Even though each sample point contains information regarding the entire trajectory of all variables, this full trajectory is not known to agents when they take decisions. At the very best, agents taking a decision at time t know only what has happened until time t. They have no way of looking into the future. In many economically interesting problems, there may be agents who do not even have full information regarding current and past events.

We need some way to keep track of what is known when. The way we do this is by grouping together outcomes (sample points) at each point of time. Two outcomes are placed in the same group if they cannot be distinguished based on information that is available at that point of time. Two outcomes which are not distinguishable at one point of time may become distinguishable at a later point of time when more information arrives. However, we assume that people never forget what they know—so if two outcomes were distinguishable at a point of time then they must remain distinguishable at future points of time.

Here's an example. Suppose we toss a coin thrice in a row. Each sample point must give us information regarding the full history of the tosses, so the sample space is,

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Before we have begun to toss the coins we don't know which of these trajectories our world will follow. So all the outcomes are indistinguishable from each other. Suppose that the first toss turns up heads. We know that the history of our world must be one of

$$\{HHH, HHT, HTH, HTT\}$$

but we cannot distinguish between different members of this set since they differ in the outcomes of the second and the third tosses which we have not observed yet. Similarly if the first toss had come up tails we would have known that our world must be one of

$$\{THH, THT, TTH, TTT\}$$

but again we could not make a finer distinction. So the set of indistinguishable groups is

$$\{\{HHH, HHT, HTH, HTT\}, \{THH, THT, TTH, TTT\}\}.$$

Now suppose the outcome of the second toss is revealed to us. Knowing this outcome allows us to distinguish between some of the outcomes we could not distinguish between before. The set of indistinguishable groups becomes,

$$\{\{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}\}.$$

With this example at hand we can now give the following formal definitions.

DEFINITION 6.1 (Partition). A *partition* $\mathcal{P}$ is a set of subsets A_i of Ω such that $A_i \cap A_j = \emptyset$ if $i \neq j$ and $\cup_i A_i = \Omega$.

In the language of events a partition is a collection of mutually exclusive and exhaustive events.

DEFINITION 6.2 (Fineness). Given two partitions $\mathcal{P}$ and $\mathcal{Q}$, the partition $\mathcal{P}$ is said to be *finer* than the partition $\mathcal{Q}$ if for every $A \in \mathcal{P}$ there is a $B \in \mathcal{Q}$ such that $A \subset B$.

Thus $\mathcal{P}$ is finer than $\mathcal{Q}$ if every event in P is a subset of some event in Q.

DEFINITION 6.3 (Information structure). An *information structure* is a sequence of partitions $\mathcal{P}_t$ such that if $t \geq s$ then $\mathcal{P}_t$ is finer than $\mathcal{P}_s$.

The interpretation is the same as that in our coin-tossing example. Each event in $\mathcal{P}_t$ consists of outcomes that cannot be distinguished based on information at time t. The requirement that partitions at later times be finer than partitions at earlier times says that information is not forgotten—two events that were distinguishable at time s and hence belonged to different events in $\mathcal{P}_s$ cannot become indistinguishable at time t.

2. Event tree

Event trees provide a graphical representation of information structures. Every event in $\mathcal{P}_t$ becomes a node in the tree. Since partitions must become finer over time, for every node A in $\mathcal{P}_{t+1}$ there must be a

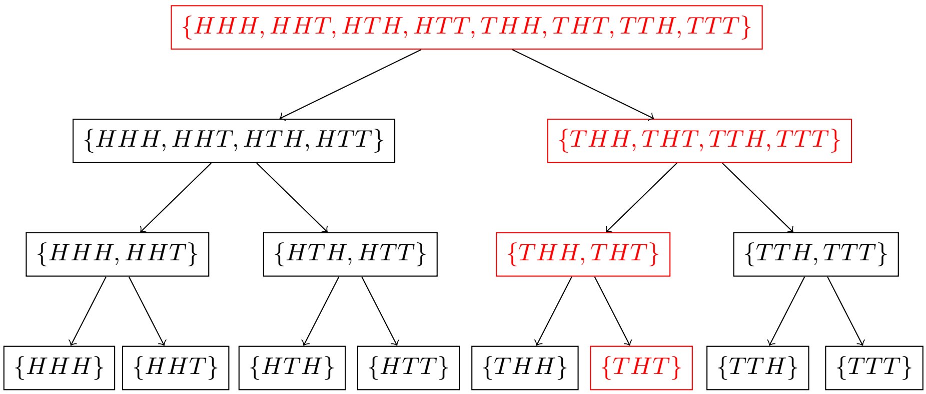


FIGURE 6.1. Event tree for the coin-tossing example. Nodes in red trace the path of the outcome THT.

node B in $\mathcal{P}_t$ such that $A \subset B$. In that case we draw an arrow from B to A. For example Figure 6.1 shows the event tree for the coin tossing example.

For each outcome in the sample space, we can think of the evolution of the system over time tracing out a path in the event tree by picking up the nodes to which the outcome belongs and following the arrows which join these nodes. Thus, for example, in Figure 6.1 the nodes in red trace out the path corresponding to the outcome THT.

3. Partitions generated by random variables

We have seen that partitions of the sample space can be used to model the state of knowledge of an agent, with all the outcomes belonging to a single event in the partition being considered to be indistinguishable.

Since the value of a random variable differs from point to point in the sample space, a random variable also conveys information about what the state of the world is. The information is not necessarily full information, since the random variable may have the same value at two different points in the sample space. In fact, for a given random variable X, if we consider all sample points at which X has the same value as indistinguishable and all points at which X has different values as distinguishable we obtain a partition of the sample space. This is called the *partition generated by* X, denoted by $\mathcal{P}(X)$.

In economic models we often want to specify that the information contained in a random variable X be no more than the information already available according to another partition $\mathcal{P}$. This is captured by the following definition.

DEFINITION 6.4 (Measurability). A random variable X is measurable with respect to a parition $\mathcal{P}$ if $\mathcal{P}$ is finer than the partition generated by X.

Note that we have only required $\mathcal{P}$ to be finer than the partition generated by X and not necessarily equal to it. This is because we want to allow for the fact that $\mathcal{P}$ contains more information than is conveyed by X so that we can distinguish between two outcomes on the basis of $\mathcal{P}$ even though we cannot distinguish them on the basis of X. All that we require is that things that we can distinguish on the basis of X be distinguishable on the basis of $\mathcal{P}$.

There are two reasons why we may want to impose the requirement of measurability on the variables of a model. First, in the case of exogenous variables we would like to impose this requirement to model the fact that the information contained in X is already captured in $\mathcal{P}$. So for example we will assume that the consumer's endowment is measurable if we want to model the fact that consumers know what their endowments are. Second, for an endogenous variable we want to model the fact that the consumer's decisions are limited by the information that they have. Thus we would require that consumption expenditure should be measurable since if a consumer cannot distinguish between two states of the world on the basis of her information, there is no way that she can choose different condumption expenditure in those two states.

A sequence of random variables X_t indexed by time is known as a *stochastic process*. A stochastic process captures the evolution of a random process over time.

We have already modelled the evolution of information over time by an information structure. The extension of the notion of measurability to this intertemporal context is given by the following.

DEFINITION 6.5 (Adapted Process). A stochastic process X_t is adapted to an information structure $\mathcal{P}_t$ if X_t is measurable with respect to P_t for each t.

Thus the knowledge of the values taken by a process adapted to $\mathcal{P}_t$ does not convey any more information at time t than what is conveyed by $\mathcal{P}_t$ itself. If we represent $\mathcal{P}_t$ by an event tree then a process X_t is adapted to $\mathcal{P}_t$ if any only if it takes the same value on all the outcomes that belong to a single node of the event tree. Because of this property of processes which are adapted to an event tree, we can think of the process as taking on a value on each of the tree nodes, rather than on each point of the sample space.

4. Conditional expectation

Suppose we do not directly observe the random variable X but have some other information represented by the partition $\mathcal{P}$. What is the best that we can say about X given this information? One answer to this is the concept of condition expectation.

Given a partition $\mathcal{P}$ and an outcome ω there is a unique event in $\mathcal{P}$ which contains ω. We denote it by $A_{\mathcal{P}}(\omega)$.

DEFINITION 6.6 (Conditional expectation). Given a random variable X and a partition $\mathcal{P}$ the conditional expectation of X with respect to $\mathcal{P}$, denoted $\mathbb{E}[X \mid \mathcal{P}]$ is a random variable defined by

$$\mathbb{E}[X \mid \mathcal{P}](\omega) = \frac{\sum_{\omega' \in A_{\mathcal{P}}(\omega)} X(\omega')\mathbb{P}(\omega')}{\sum_{\omega' \in A_{\mathcal{P}}(\omega)} \mathbb{P}(\omega')}$$

where $\mathbb{P}(\omega')$ denotes the probability of the outcome ω'.

The conditional expectation $\mathbb{E}[X \mid \mathcal{P}]$ is a random variable in its own right. Our best guess about X would depend on what information we have actually received and therefore would be different in different states of the world.

EXAMPLE 6.1. Consider a coin which is tossed twice, so that our sample space is $\Omega = \{HH, HT, TH, TT\}$. Assume that all the outcomes have an equal probability of $1/4$.

Let X be the number of heads in a toss. This is a random variable with the values $X(HH) = 2$, $X(HT) = 1$, $X(TH) = 1$ and $X(TT) = 0$.

Suppose we have observed only the first toss of the coin. Our information parition is

$$\mathcal{P} = \{\{HH, HT\}, \{TH, TT\}\}.$$

What is $\mathbb{E}[X \mid \mathcal{P}]$?

Let us begin with the outcome HH. $A_{\mathcal{P}}(HH) = \{HH, HT\}$ so we have

$$\mathbb{E}[X \mid \mathcal{P}](HH) = \frac{X(HH)\mathbb{P}(HH) + X(HT)\mathbb{P}(HT)}{\mathbb{P}(HH) + \mathbb{P}(HT)} = 1.5$$

Similarly,

$$\mathbb{E}[X \mid \mathcal{P}](HT) = \frac{X(HH)\mathbb{P}(HH) + X(HT)\mathbb{P}(HT)}{\mathbb{P}(HH) + \mathbb{P}(HT)} = 1.5$$

$$\mathbb{E}[X \mid \mathcal{P}](TH) = \frac{X(TH)\mathbb{P}(TH) + X(TT)\mathbb{P}(TT)}{\mathbb{P}(TH) + \mathbb{P}(TT)} = 0.5$$

$$\mathbb{E}[X \mid \mathcal{P}](TT) = \frac{X(TH)\mathbb{P}(TH) + X(TT)\mathbb{P}(TT)}{\mathbb{P}(TH) + \mathbb{P}(TT)} = 0.5$$

Note that $\mathbb{E}[X \mid \mathcal{P}]$ is defined for each element of the sample space, as a random variable should be. But its value is constant within each element of $\mathcal{P}$. We will see below that this is a general property of conditional expectations.

4.1. Properties. We now list some properties that will be useful for computing conditional expectations later.

4.1.1. *Measurability.*

PROPOSITION 6.1. *For any random variable X, the random variable $\mathbb{E}[X \mid \mathcal{P}]$ is measurable with respect to $\mathcal{P}$.*

PROOF. The right-hand side of the definition of conditional expectation (Definition 6.6) depends only on the set $A_{\mathcal{P}}(\omega)$. If ω and ω'' belong to the same set A in $\mathcal{P}$ then we would have $A_{\mathcal{P}}(\omega) = A_{\mathcal{P}}(\omega'')$ and hence $\mathbb{E}[X \mid \mathcal{P}](\omega) = \mathbb{E}[X \mid \mathcal{P}](\omega'')$. Since A was an arbitrary member of $\mathcal{P}$ our argument shows that $\mathbb{E}[X \mid \mathcal{P}]$ must be constant over every such set. It is therefore measurable with respect to $\mathcal{P}$. $\square$

This property is a reasonable one. If the information we have does not allow us to distinguish between the states of the world ω and ω'' then the best guess we can make about X in the two states must be the same.

In a way this is also the most important property of the conditional expectation. If our information $\mathcal{P}$ does not allow us to know the exact value of X we cannot base our decisions on that value. But we would always know the value of $\mathbb{E}[X \mid \mathcal{P}]$ and can base our decisions on it.

4.1.2. *Expectation of a measurable random variables.* If a random variable is measurable with respect to a partition then knowing where we are in the partition gives us the exact value of the random variable. No further forecasting is required. So we have,

PROPOSITION 6.2. *If the random variable X is measurable with respect to the partition $\mathcal{P}$ then*

$$\mathbb{E}[X \mid \mathcal{P}] = X.$$

The intuition behind the proposition is this: if $\mathcal{P}$ already contains the information in X then knowing the information in $\mathcal{P}$ gives us the actual value of X and no averaging is required.

Rather than proving this result we will prove the following slight generalisation,

PROPOSITION 6.3. *If the random variable X is measurable with respect to the partition $\mathcal{P}$ and Y is any random variable then*

$$\mathbb{E}[XY \mid \mathcal{P}] = X\mathbb{E}[Y \mid \mathcal{P}].$$

PROOF. From Definition 6.6 we have

$$\mathbb{E}[XY \mid \mathcal{P}](\omega) = \frac{\sum_{\omega' \in A_{\mathcal{P}}(\omega)} X(\omega')Y(\omega')\mathbb{P}(\omega')}{\sum_{\omega' \in A_{\mathcal{P}}(\omega)} \mathbb{P}(\omega')}$$

But if X is measurable with respect to $\mathcal{P}$ then X must be constant over each element of $\mathcal{P}$.

In particular it must be constant over $A_{\mathcal{P}}(\omega)$. Since $\omega \in A_{\mathcal{P}}(\omega)$, we have $X(\omega') = X(\omega)$ for all $\omega' \in A_{\mathcal{P}}(\omega)$. So we have,

$$\begin{aligned}
\mathbb{E}[XY \mid \mathcal{P}](\omega) &= \frac{\displaystyle\sum_{\omega' \in A_{\mathcal{P}}(\omega)} X(\omega)Y(\omega')\mathbb{P}(\omega')}{\displaystyle\sum_{\omega' \in A_{\mathcal{P}}(\omega)} \mathbb{P}(\omega')} \\
&= X(\omega)\left[\frac{\displaystyle\sum_{\omega' \in A_{\mathcal{P}}(\omega)} Y(\omega')\mathbb{P}(\omega')}{\displaystyle\sum_{\omega' \in A_{\mathcal{P}}(\omega)} \mathbb{P}(\omega')}\right] \\
&= X(\omega)\mathbb{E}[Y \mid \mathcal{P}]
\end{aligned}$$

□

4.1.3. *Law of iterated expectations.* We can imagine the operation of taking the conditional expectation of a random variable with respect to a partition as a ‘blurring’ operation: within each event of the partition we replace the individual values of the random variable by a common average value. The *law of iterated expectation* says that blurring on a fine grid and then further blurring the result on a coarser grid gives the same result as blurring the original variable directly on a coarse grid. More formally,

THEOREM 6.1. *Given a random variable X and two partitions $\mathcal{P}$ and $\mathcal{Q}$ where $\mathcal{Q}$ is finer than $\mathcal{P}$ it is the case that*

$$\mathbb{E}\{\, \mathbb{E}[X \mid \mathcal{Q}] \mid \mathcal{P}\,\} = \mathbb{E}[X \mid \mathcal{P}].$$

4.2. Special cases.

4.2.1. *Unconditional expectation.* Consider the trivial partition $\mathcal{T}$ which has only one element—the full sample space Ω. Then for any ω we have $A_{\mathcal{T}}(\omega) = \Omega$ and the definition of conditional expectation gives us

$$\mathbb{E}[X \mid \mathcal{T}](\omega) = \frac{\displaystyle\sum_{\omega' \in \Omega} X(\omega')\mathbb{P}(\omega')}{\displaystyle\sum_{\omega' \in \Omega} \mathbb{P}(\omega')}$$

But from the axioms of probability $\sum_{\omega' \in \Omega} \mathbb{P}(\omega') = 1$, so we have

$$\mathbb{E}[X \mid \mathcal{T}](\omega) = \sum_{\omega' \in \Omega} X(\omega')\mathbb{P}(\omega')$$

You should recognize the right-hand side above as just the unconditional expectation of X, $\mathbb{E}[X]$. Thus the unconditional expectation can be seen as a special case of the conditional expectation where the conditioning is over the trivial partition which represents complete lack of information.

4.2.2. *Expectation conditional on a random variable.* When the partition over which we are conditioning is generated by a random variable then we say and write, as a form of shorthand, that we are conditioning on the random variable. So if X and Y are two random variables then we define

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \mathcal{P}(X)].$$

Example 6.2. Consider a sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and two random varables X and Y defined on that sample space. The probabilities of the sample point and the values of the random variables at them are

ω	$\mathbb{P}(\omega)$	$X(\omega)$	$Y(\omega)$
ω_1	4/9	1	1
ω_2	2/9	1	0
ω_3	2/9	0	1
ω_4	1/9	1	0

The partition generated by the random variable Y, $\mathcal{P}(Y)$, is

$$\mathcal{P}(Y) = \{\{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}\}.$$

Then the conditional expectation $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \mathcal{P}(Y)]$ is given by

$$\begin{aligned}\mathbb{E}[X \mid Y](\omega_1) = \mathbb{E}[X \mid Y](\omega_3) &= \frac{X(\omega_1)\mathbb{P}(\omega_1) + X(\omega_3)\mathbb{P}(\omega_3)}{\mathbb{P}(\omega_1) + \mathbb{P}(\omega_3)} \\ &= \frac{1 \cdot (4/9) + 0 \cdot (2/9)}{(4/9 + (2/9)} = \frac{2}{3} \\ \mathbb{E}[X \mid Y](\omega_2) = \mathbb{E}[X \mid Y](\omega_4) &= \frac{X(\omega_2)\mathbb{P}(\omega_2) + X(\omega_4)\mathbb{P}(\omega_4)}{\mathbb{P}(\omega_1) + \mathbb{P}(\omega_3)} \\ &= \frac{1 \cdot (2/9) + 1 \cdot (1/9)}{(2/9 + (1/9)} = 1\end{aligned}$$

4.2.3. *Expectation conditional on information at time t.* When dealing with stochastic processes we often work with an information structure, that is, a sequence of partitions $\mathcal{P}_t$. In this context we sometimes write "expectation conditional on information at time t" when we condition on the partition $\mathcal{P}_t$ for a particular t. We also use the notations $\mathbb{E}_t[X]$ and $\mathbb{E}[X \mid t]$ as shorthand for $\mathbb{E}[X \mid \mathcal{P}_t]$.

5. Independence

We assume that you are familiar with the definition of the independence of a set of random variables from elementary probability theory. It is possible to extend this definition using the language of partitions we have developed above. We record only the following useful result,

PROPOSITION 6.4. *If the random variable X is independent of the random variables $Y_1, \ldots, Y_n$ then*

$$E[X \mid Y_1, \ldots, Y_n] = \mathbb{E}[X].$$

The expectation of a random variable X conditional on variables independent of it is just the unconditional expectation. This makes intuitive sense since if X is independent of the Y_i then knowing the Y_i gives us no knowledge about X.

6. Martingales

DEFINITION 6.7 (Martingale). Let X_t be a stochastic process, and $\mathcal{I}_t$ and information structure with the following properties:

(1) X_t is adapted to $\mathcal{I}_t$.
(2) $\mathbb{E}[X_{t+1} \mid I_t] = X_t$.

Then the process X_t is said to be a *martingale* with respect to the information structure $\mathcal{I}_t$.[1]

Among the defining characteristics of a martingale given above the crucial one is the last: $\mathbb{E}[X_{t+1}|\mathcal{I}_t] = X_t$. This says that conditional on the information available in period t the random variable X_t neither grows nor declines in the next period in expected value terms. In this a martingale process is like the wealth of a gambler playing a fair game. Some outcomes of the game increase the gambler's wealth, other outcomes of the game decrease the gambler's wealth, but since the game is fair the increases and decreases cancel out on an average.

We can use Proposition 6.2 and the properties of a martingale to show that

$$\mathbb{E}[X_{t+1} - X_t \mid \mathcal{I}_t] = 0$$

and therefore if Y is any variable measurable with respect to $\mathcal{I}_t$ then

$$\mathbb{E}[(X_{t+1} - X_t)Y] = 0$$

so that $X_{t+1} - X_t$ is uncorrelated with Y. (Give detailed proofs of all the claims made in this paragraph so far.) That is, the change in the value of a martingale process is uncorrelated with all the information available from the past. Speaking loosely, it is a 'surprise'. And so a martingale is a sum of surprises.

A random walk is a martingale (Prove. Be careful to specify the sequence of information sets.) This has led some economic literature to loosely use the term "random walk" when discussing martingales. This usage should be avoided since not all martingales are random walks.

The definition of a martingale tells us about the expectation of a value of the process in a period conditional on the information in the immediately preceding period. The following proposition covers the case where conditioning set and the value is separated by more than one period.

PROPOSITION 6.5. *Let X_t be a martingale with respect to the information structure $\mathcal{I}_t$. For any m and any $n > 0$,*

$$\mathbb{E}[X_{m+n} \mid \mathcal{I}_m] = X_m.$$

[1]On probability spaces which are not finite, we also need the condition $\mathbb{E}[|X_t|] < \infty$.

Proof. The proof is by mathematical induction on n.

For $n = 1$ the proof follows directly from the definition of a martingale.

Suppose the lemma is true for $n = k$. Consider the case $n = k + 1$. From the law of iterated expectations

$$\mathbb{E}[X_{m+k+1} \mid \mathcal{I}_m] = \mathbb{E}[\mathbb{E}[X_{m+k+1} \mid \mathcal{I}_{m+k}] \mid \mathcal{I}_m].$$

The martingale property, applied at time $m + k$ tells us that

$$\mathbb{E}[X_{m+k+1} \mid \mathcal{I}_{m+k}] = X_{m+k}.$$

The assumption that the lemma is true for $n = k$ gives us,

$$\mathbb{E}[X_{m+k} \mid \mathcal{I}_m] = X_m.$$

Putting everything together, we have

$$\mathbb{E}[X_{m+k+1} \mid \mathcal{I}_m] = \mathbb{E}[\mathbb{E}[X_{m+k+1} \mid \mathcal{I}_{m+k}] \mid \mathcal{I}_m] = \mathbb{E}[X_{m+k} \mid \mathcal{I}_m] = X_m$$

thus establishing the result for $n = k + 1$.

Since we have shown that the result is true for $n = 1$ and it is true for $n = k+1$ whenever it is true for $n = k$, it follows from the principle of mathematical induction that it is true for all $n > 0$. □

Exercises

1. Consider a sample space with three points. List all possible partitions of this sample space. For each pair of partitions from your list, state whether one is finer than the other.

2. Consider the sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and the two partitions

$$\mathcal{P} = \{\{\omega_1, \omega_1\}, \{\omega_3, \omega_4\}\}$$
$$\mathcal{Q} = \{\{\omega_1\}, \{\omega_2, \omega_3\}, \{\omega_4\}\}.$$

Give an example of a random variable X which is measurable with respect to $\mathcal{P}$ but not measurable with respect to $\mathcal{Q}$.

3. Let $\mathcal{P}$ and $\mathcal{Q}$ be two partitions such that $\mathcal{P}$ is finer than $\mathcal{Q}$. Argue that any random variable measurable with respect to $\mathcal{Q}$ must necessarily be measurable with respect to $\mathcal{P}$.

4. If X_t is a martingale with respect to the information structure $\mathcal{I}_t$, show giving reasons for all your steps that
$$\mathbb{E}[X_{t+1} - X_t \mid \mathcal{I}_t] = 0.$$

5. If X_t is a martingale with respect to the information structure $\mathcal{I}_t$, and Y_t is a process adapted to $\mathcal{I}_t$ show giving reasons for all your steps that
$$\mathbb{E}[(X_{t+1} - X_t)Y_t] = 0.$$
[Hint. You will need Proposition 6.3 and the law of iterated expectations.]

6. Let ϵ_t be a sequence of independent and identically distributed variables with mean 0 and variance σ^2. We define
$$Y_t = \epsilon_t + b\epsilon_{t-1}$$
for some constant b. Calculate
   (a) $\mathbb{E}[Y_t]$
   (b) $\mathrm{Var}[Y_t]$
   (c) $\mathrm{Cov}[Y_t, Y_{t-1}]$

7. Let ϵ_t be a sequence of independent and identically distributed variables with mean 0 and variance σ^2. We define
$$Y_t = \begin{cases} 0 & t = 0 \\ bY_{t-1} + \epsilon_t & t > 0 \end{cases}$$
for some constant $|b| < 1$. Calculate
   (a) $\mathbb{E}[Y_t]$
   (b) $\mathrm{Var}[Y_t]$
   (c) $\mathrm{Cov}[Y_t, Y_{t+h}]$ for arbitrary $h > 0$.
[Hint: See that $Y_1 = \epsilon_1$, $Y_2 = b\epsilon_1 + \epsilon_2$, $Y_3 = b^2\epsilon_1 + b\epsilon_2 + \epsilon_3$, etc. Use mathematical induction to show
$$Y_{t+h} = b^h Y_t + \sum_{i=1}^{h} b^{h-i}\epsilon_{t+i}.$$
Proceed from there.]

References

There are many excellent texts on elementary probability theory, for example [**CA03**] or [**Ros09**].

Our representation of information structures using partitions is satisfactory only for finite sample spaces. To deal rigorously with more general probability spaces and random variables which are not discrete we need to replace partitions by *σ-algebras* and information structures by *filtrations*. These concepts are discussed in texts on "measure-theroetic probability". Examples in order of increasing difficulty are [**JP04**], [**Wil91**] and [**Bil95**].

CHAPTER 7

Consumption: Uncertainty

1. Euler equation

In the case of uncertainty in labour incomes, but with certain interest rates, the Euler equation becomes

$$v'_t(c_t) = RE_t[v'_{t+1}(c_{t+1})] \tag{30}$$

where E_t denotes the mathematical expectation conditional on information at time t.

2. Quadratic felicity

2.1. Martingale property. Suppose the felicity (i.e. per-period utility) function is

$$v_t(c_t) = \beta^t(ac_t - 0.5c_t^2)$$

where a is some constant.

In this case (30) specialises to

$$a - c_t = R\beta(a - E_t c_{t+1})$$

If we further assume that $R = 1/\beta$ then

$$E_t c_{t+1} = c_t \tag{31}$$

that is, consumption is a martingale process.

Since c_t is part of the information set at time t, $E_t c_t = c_t$ Therefore, another way to write (31) is

$$E_t(c_{t+1} - c_t) = 0$$

which says that the change in consumption between time t and $t+1$ has no predictable direction based on information at time t.

This result is a consequence of the very special assumptions that we have made. Assuming the same felicity function for each period (apart from the discount factor β) and then assuming that the market rate of discount $(1/R)$ equals this subjective discount factor creates a situation where the consumer has no desire to have a higher consumption in any particular period of her life either to meet greater consumption needs or to take advantage of the difference between market and subjective discount rates. Unconstrained lending and borrowing mean that the consumer can actually move around her income across periods so as to achieve this perfect symmetry in her consumption in the sense of equating expected marginal utility across periods. But with quadratic felicity expected marginal utility is the same thing as expected consumption and we have our martingale result.

The long list of assumptions leading up to the martingale result means that this precise result is not very robust or realistic. Therefore rather than taking it as a property that is likely to be literally true, we should understand it as a demonstration of the tendency of the lending and borrowing behaviour of consumers to delink current consumption from current income. This tendency will be there as long as consumers have access to asset markets, though in more realistic settings it will be overlaid with factors which impart a systematic pattern to the trajectory of consumption such as a changing pattern of lifetime consumption needs or differences between the subjective and market rate of discount.

2.2. The level of consumption. The martingale property of consumption only tells us how consumption evolves from one point to the next, not the *level* of consumption. The level of consumption would depend on the consumer's resources in terms of her initial wealth and expected labour income. We now show that this is so mathematically by deriving an explicit formula for the level of consumption in the case where consumption is a martingale.

Consider a consumer who stands at period t with wealth w_t and is planning her future consumption for the periods $t, t+1, \ldots, T$. Since she cannot leave any bequests or outstanding debt in period T, it must be the case that her *realized* stream of consumption (c_t) and labour income (y_t) must satisfy,

$$\sum_{i=t}^{T} \delta^{i-t} c_i = \sum_{i=t}^{T} \delta^{i-t} y_i + w_t \tag{32}$$

Taking expectations as of time t,

$$\sum_{i=t}^{T} \delta^{i-t} E_t c_i = \sum_{i=t}^{T} \delta^{i-t} E_t y_i + w_t \tag{33}$$

From Proposition 6.5 on page 50, $E_t c_i = c_t$ for all $i > t$. And for $i = t$, $E_t c_t = c_t$ because c_t is known at time t. Hence,

$$c_t \sum_{i=t}^{T} \delta^{i-t} = \sum_{i=t}^{T} \delta^{i-t} E_t y_i + w_t \tag{34}$$

$$c_t = \frac{1}{\sum_{i=t}^{T} \delta^{i-t}} \left[\sum_{i=t}^{T} \delta^{i-t} E_t y_i + w_t \right] \tag{35}$$

Thus the level of consumption in a given period depends on the expected discounted value of the future stream of labour income over the entire remaining lifetime as well as initial wealth. This once again reiterates the idea of the permanent income hypothesis that the consumption in each period depends not just on income in that period but on the entire expected path of future income.

2.3. Increments in consumption. With an explicit formula for the level of consumption in hand, we can now try to understand the martingale result better by seeing what it is exactly that drives changes in consumption.

Rewriting (34) for period $t + 1$ we have,

$$c_{t+1} \sum_{i=t+1}^{T} \delta^{i-t-1} = \sum_{i=t+1}^{T} \delta^{i-t-1} E_{t+1} y_i + w_{t+1}$$

Substituting $w_{t+1} = (w_t + y_t - c_t)/\delta$,

$$c_{t+1} \sum_{i=t+1}^{T} \delta^{i-t-1} = \sum_{i=t+1}^{T} \delta^{i-t-1} E_{t+1} y_i + (w_t + y_t - c_t)/\delta$$

Multiplying throughout by δ,

$$c_{t+1} \sum_{i=t+1}^{T} \delta^{i-t} = \sum_{i=t+1}^{T} \delta^{i-t} E_{t+1} y_i + w_t + y_t - c_t$$

Subtracting (34) from this equation

$$(c_{t+1} - c_t) \sum_{i=t+1}^{T} \delta^{i-t} - c_t = \sum_{i=t+1}^{T} \delta^{i-t}(E_{t+1}y_i - E_t y_i) + y_t - E_t y_t - c_t$$

But $E_t y_t = y_t$ since income at time t is known at that time,

$$(c_{t+1} - c_t) \sum_{i=t+1}^{T} \delta^{i-t} = \sum_{i=t+1}^{T} \delta^{i-t}(E_{t+1}y_i - E_t y_i)$$

$$c_{t+1} - c_t = \frac{1}{\sum_{i=t+1}^{T} \delta^{i-t}} \left[\sum_{i=t+1}^{T} \delta^{i-t}(E_{t+1}y_i - E_t y_i) \right]$$

We can divide both the numerator and denominator by δ to have the factor multiplying the first term in the sums equal to one.[1] This gives us our final formula,

$$c_{t+1} - c_t = \frac{1}{\sum_{i=t+1}^{T} \delta^{i-t-1}} \left[\sum_{i=t+1}^{T} \delta^{i-t-1}(E_{t+1}y_i - E_t y_i) \right] \qquad (36)$$

What the formula above says is that changes in consumption are a result of *revisions* in expectations of future income based on the difference in information at time t and $t+1$. Therefore changes in income that were predictable at time t do not contribute to the change in consumption between time t and $t+1$. This is consistent with our assumption that consumption is a martingale but goes further by predicting the actual size of the consumption change rather than just asserting that the expected value of this change would be zero.

2.4. Specific income processes. In general the difference in expectations which occur on the right of (36) depends on all the information which becomes available to the consumer between time t and $t+1$. One special case which we now consider is when the only source of new information is the realisation of the labour income y_{t+1}. What revision this new information causes in the expectation of future labour income depends on how labour income in different periods are related.

[1] This is a purely aesthetic change and does not change any results.

As an example consider the labour income process given by the following stochastic difference equation,

$$y_{t+1} - \mu = \rho(y_t - \mu) + \epsilon_{t+1} \tag{37}$$

where ϵ_t is a white-noise process, μ and ρ are constants and $0 < \rho < 1$. This is a special case of what is known as a first-order autoregressive process (sometimes denoted as a AR(1) process). The coefficient ρ measures how persistent the deviations in y from its long-run average μ are.

Writing (37) for the period $t + 2$ we have,

$$y_{t+2} - \mu = \rho(y_{t+1} - \mu) + \epsilon_{t+2}$$

Substituting (37),

$$y_{t+2} - \mu = \rho^2(y_t - \mu) + \rho\epsilon_{t+1} + \epsilon_{t+2}$$

Carrying out successive substitutions like this, we find for any $i > t$

$$y_i - \mu = \rho^{i-t}(y_t - \mu) + \sum_{j=t+1}^{i} \rho^{i-j}\epsilon_j$$

Taking expectations conditional on the information at time t,

$$E_t(y_i - \mu) = \rho^{i-t}(y_t - \mu) + \sum_{j=t+1}^{i} \rho^{i-j}E_t\epsilon_j$$

Here we have used the fact that y_t is known at time t. We further note that since ϵ_t is IID, ϵ_j is independent of all information at time t when $j > t$ and we can replace $E_t\epsilon_j$ by $E\epsilon_j$ which is 0 by the definition of white noise. Hence we conclude,

$$E_t(y_i - \mu) = \rho^{i-t}(y_t - \mu) \qquad \text{for } i \geq t \tag{38}$$

(We have established this above for $i > t$ and it is trivially true for $i = t$.)

Using $t + 1$ in the place of t,

$$E_{t+1}(y_i - \mu) = \rho^{i-t-1}(y_{t+1} - \mu) \qquad \text{for } i \geq t + 1 \tag{39}$$

For $i \geq t+1$ both (38) and (39) hold. Subtracting the former from the latter we have,

$$E_{t+1}y_i - E_ty_i = \rho^{i-t-1}(y_{t+1} - \mu) - \rho^{i-t}(y_t - \mu)$$

Using (37),

$$
\begin{aligned}
&= \rho^{i-t-1}[\rho(y_t - \mu) + \epsilon_{t+1}] - \rho^{i-t}(y_t - \mu) \\
&= \rho^{i-t-1}\epsilon_{t+1}
\end{aligned}
$$

Substituting this in (36) we get,

$$
c_{t+1} - c_t = \left[\frac{\sum_{i=t+1}^{T} (\delta\rho)^{i-t-1}}{\sum_{i=t+1}^{T} \delta^{i-t-1}} \right] \epsilon_{t+1}
$$

So we see that for a given innovation in consumption, ϵ_{t+1}, the increment in consumption is higher the higher is the degree of persistence ρ in the income process.

In empirical applications of the model we can estimate ρ (or its analogues for more complex income processes) from data on consumers' incomes and then check if changes in consumption satisfy the forumla above. This yields a sharper test of our theory compared to just checking if consumption is a martingale.

CHAPTER 8

Neoclassical Optimal Growth Model

1. The Problem

The planner maximises

$$\sum_{t=1}^{\infty} \delta^{t-1} u(c_t)$$

where $u\colon \mathbb{R}_+ \to \mathbb{R} \cup \{-\infty\}$ is an upper semicontinuous strictly increasing function.

We are given a continuous production function f such that $f(0) \geq 0$. Define $f^t(k)$ recursively by

$$f^t(k) = \begin{cases} f(k) & t = 1 \\ f[f^{t-1}(k)] & t > 1 \end{cases}.$$

So $f^t(\cdot)$ is the t-th iterate of the function f.

A pair of sequences $(\mathbf{c}, \mathbf{k})$ is *feasible* from k is $c_t, k_t \geq 0$ and $0 \leq k_t + c_t \leq f(k_{t-1})$ for $t = 1, 2, \ldots$. The *feasible set* is $\mathbf{Y}(k_0) = \{(\mathbf{c}, \mathbf{k}) \mid (\mathbf{c}, \mathbf{k}) \text{ is feasible from } k_0\}$. The sets of feasible capital and consumption programs are $\mathbf{F}(k_0) = \{\mathbf{k} \mid (\mathbf{c}, \mathbf{k}) \in \mathbf{Y}(k_0) \text{ for some } \mathbf{c}\}$ and $\mathbf{B}(k_0) = \{\mathbf{c} \mid (\mathbf{c}, \mathbf{k}) \in \mathbf{Y}(k_0) \text{ for some } \mathbf{k}\}$

2. Existence

Lemma 8.1. *The sets $\mathbf{F}(k_0)$ and $\mathbf{B}(k_0)$ are compact for all k_0.*

Proof. Since $c_t, k_t \geq 0$ we have

$$\mathbf{Y}(k_0) \subset \prod_{t=1}^{\infty} ([0, f^t(k_0)] \times [0, f^t(k_0)]).$$

The latter set is compact by Tsychonoff's theorem. $\mathbf{Y}(k_0)$ is closed by the from the inequalities defining it and the continuity of f. Thus

$\mathbf{Y}(k_0)$ is compact. The sets $\mathbf{F}(k_0)$ and $\mathbf{B}(k_0)$ being projections of the compact set $\mathbf{Y}(k_0)$ are also compact. □

The following is the "Basic Existence Theorem" of Becker and Boyd [**BB97**, Section 4.2.3].

THEOREM 8.1. *Suppose the basic assumptions are satisfied, $u(c) \leq a + bc^\gamma/\gamma$, and $f(k) \leq \alpha + \beta k$ with $\beta \geq 1$ and $b, \alpha \geq 0$. If $\beta^\gamma \delta < 1$ and either $a = 0$ or $\delta < 1$, then an optimal path exists.*

PROOF. Let $\theta > \beta \geq 1$ with $\theta^\gamma \delta < 1$. Now $c_t \leq f^t(k_0)$ and $f^t(k_0) \leq \alpha + \alpha\theta + \cdots + \alpha\theta^{t-1} + \theta^t k_0 = \alpha(\theta^t - 1)/(\theta - 1) + \theta^t k_0$ by induction. Since $\theta > 1$, $c_t \leq [k_0 + \alpha/(\theta-1)]\theta^t$. Let $A = b[k_0 + \alpha/(\theta-1)]^\gamma/\gamma$ and let $g_t(c_t) = u(c_t) - a - A\theta^{\gamma t} \leq 0$ for feasible $\mathbf{c}$. On the feasible set, the partial sums $S_T(\mathbf{c}) = \sum_{t=1}^T \delta^{t-1} g_t(c_t)$ form a decreasing sequence of upper semicontinuous functions. Their limit, which equals their infimum, is upper semicontinuous. Now $\sum_{t=1}^\infty g_t(c_t) = U(\mathbf{c}) - a/(1-\delta) - A\theta^\gamma/(1-\theta^\gamma\delta)$. Thus $U(\mathbf{c})$ is upper semicontinuous on the compact set $\mathbf{B}(k_0)$. By the Weierstrass Theorem, an optimal path exists. □

3. Characterization

4. Dynamics

References

This chapter is based closely on [**BB97**]. They refer to the model discussed in this chapter as the "time additive separable" (TAS) model.

CHAPTER 9

Overlapping Generations

The Ramsey model is at least in two ways a well-behaved model. First, the market equilibrium in the model is always Pareto optimal. Second, the dynamics of the model is very simple. Regardless of what capital stock we start the model with, both the capital/labour ratio and consumption/labour ratio move monotonically[1] towards their steady-state values.

This simplicity of the Ramsey model is deceptive. Even within the Ramsey framework, very complex dynamics is possible if we consider multisector models rather than the single-good model that we have studied. However, equilibria are Pareto efficient in all Ramsey models, so this complex dynamics is still efficient.

This is not the case in the overlapping generations models that we consider in this chapter. These models can demonstrate both complex dynamics and Pareto suboptimal equilibria. These models are completely orthodox methodologically: all agents in these economies are optimisers and have perfect foresight and all markets are competitive and in equilibrium at all times. Therefore their behaviour is even more surprising and causes us to question our understanding of the behaviour of even idealised competitive market economies.

An overlapping generations economy is an infinite-horizon economy in which the agents can be grouped into different 'generations' such that the following properties hold.

(1) At each point of time members of different generations coexist.
(2) There are generations whose lifetime is not the same as the lifetime of the entire economy.

[1] that is, without changing their direction of movement

The above definition tries to capture the essence of the many different varieties of overlapping generations economies that have been studied in the literature. However, in this chapter we shall limit ourselves to very simple examples. We will find instances of counter-intuitive behaviour in even these simple cases.

1. Pure exchange: Incomplete Participation

Our first example is that of a pure exchange economy, i.e. an economy without production. Time is discrete and doubly infinite: there is an infinite past as well as an infinite future.

Agents have a two-period lifetime. At the beginning of each period a new generation of agents is born, which dies at the end of the next period. Following convention, we refer to the agents in the first period of their lives as 'young' and in the second period of their lives as 'old' respectively.

All agents in the same generation are identical and each generation has the same number of agents. Because of these assumptions we can replace each generation by a single representative agent.

There is a single physical good in the economy which is not storable. Each agent is endowed with some quantity of this good in each of the periods of their lives.

Even though there is only one physical commodity, agents care not only about how much of this commodity they have but also when they have it. Therefore in economic terms we need to tag the quantities of the commodities with dates on which those quantities would be available. Thus, even with a single physical commodity there is a separate economic commodity corresponding to the delivery of this physical commodity on a particular date. Since our model has infinite number of periods we therefore have an infinity of economic commodities in our model. We call these economic commodities 'dated-commodities' to indicate that along with a physical description each commodity is also marked with a delivery date.

In this section we assume that in every period markets open to allow trading in all dated-commodities. Only those agents who live in a particular period can participate in that period's market. Neither the unborn nor the dead can trade. It is because of this last assumption that the model of this section is called a model of limited participation.

To keep things simple we assume that every generation is alike in terms of having the same pattern of endowments and the same preferences over consumption streams. Note that this does not mean that the two periods of an agent's lifetime are alike. It is possible that the agents want to consume more when they are old. Or perhaps agents are richer when they are young. All we are assuming is that if one generation is richer when it is young then so are all generations and so on. This assumption of stationarity means that every period looks alike in our model. But in each period some agents are young and some are old and this is enough to generate interesting phenomena since we allow endowments and preferences to be age-dependent.

1.1. Notation. We will denote agent-specific variables by two subscripts: the first to indicate whether the variables pertains to the young (1) or old (2) and the second to indicate the date to which the variable pertains. So an agent born in period t, who lives in the periods t and $t+1$ will have the consumption stream $(c_{1,t}, c_{2,t+1})$ since the consumer will be young in period t and old in period $t+1$. On the other hand the total consumption in period t will be $c_{1,t} + c_{2,t}$ since we need to add up the consumption of the young and the old, belonging to different generations, who live in period t.

Our assumption of stationarity makes the notation simpler for endowments and utility functions. All generations have the same endowment stream (ω_1, ω_2) and utility function $U(c_{1,t}, c_{2,t+1})$.

1.2. Equilibrium. The way we have set up the model there is no opportunity to trade. Because there is only one good there are no trades within a period since there is no point buying and selling the same good. Therefore all trade must be across time, or in simpler language, the only possible trades involve borrowing and lending. Since each generation is homogeneous there cannot be any trade between members of the same generation. If one member of a generation wants to borrow then so do all members of that generation and vice versa. Finally, and most importantly, there can be no trade between generations because two agents of different generations meet only in one period. Today's old cannot trade with today's young since the former will not be around tomorrow to repay a loan they receive or collect on a loan they give. Today's young cannot trade with tomorrow's young since the latter have not been born yet.

Thus in the way we have set up the model there can be no trade and the only possible equilibrium is the autarky equilibrium in which each agent consumes her endowment.

In general equilibrium theory for economies with a finite number of goods and agents all competitive equilibrium are necessary Pareto optimal. This result is known as the "First Fundamental Theorem of Welfare Economics".[2] The following examples shows that this theorem does not necessarily hold for overlapping generations model.

EXAMPLE 9.1. Consider the economy as described above with the utility function for the initial old given by

$$u(c_{2,0}) = \ln(1 + c_{2,0})$$

and utility function for later generations given by

$$u(c_{1,t}, c_{2,t+1}) = \ln(1 + c_{1,t}) + \ln(1 + c_{2,t+1}).$$

Suppose endowments are $\omega_1 = 1, \omega_2 = 0$.

The autarkic equilibrium involves everyone consuming their endowments, i.e.

$$c_{1,t} = 1, c_{2,t} = 0$$

The utility of the initial old under this equilibrium is

$$u(c_{2,0}) = \ln(1 + 0) = 0$$

and the utility of all later generations is

$$u(c_{1,t}, c_{2,t+1}) = \ln(1 + 1) + \ln(1 + 0) = \ln 2.$$

It turns out that this allocation is not Pareto optimal. Consider the alternative allocation

$$\hat{c}_{1,t} = \hat{c}_{2,t} = 1/2$$

We first check that this allocation is feasible given the economy's resources

$$\hat{c}_{1,t} + \hat{c}_{2,t} = \omega_1 + \omega_2 = 1$$

The utility of the initial old under the hat allocation is higher

$$u(\hat{c}_{2,0}) = \ln(1 + 1/2) = \ln 3/2 > 0$$

[2] See [**MCWG95**, 16.C]. The additional technical assumption of local non-satiation of preferences is required. This is satisfied in all our examples.

The utility of all other generations under the hat allocation is also higher

$$u(\hat{c}_{1,t}, \hat{c}_{2,t+1}) = \ln(1 + 1/2) + \ln(1 + 1/2) = \ln(3/2)^2 > \ln 2.$$

Thus the hat allocation is a feasible allocation that makes all agents better off than the competitive allocation. Thus it is a Pareto improvement over the competitive allocation and its existence shows that the competitive allocation is not Pareto optimal.

□

EXAMPLE 9.2. Consider an economy with the same structure as before but in which time is doubly infinite, i.e. time neither has a beginning nor an end. Because there is no initial period there is no initial old generation to consider. All generations are alike. Each generation has the utility function

$$u(c_{1,t}, c_{2,t+1}) = \ln(1 + c_{1,t}) + \ln(1 + c_{2,t+1}).$$

and the endowment

$$\omega_1 = 0, \omega_2 = 1.$$

Once again, because any two generations meet only in a single period no trade is possible and the equilibrium allocation is the autarkic one

$$c_{1,t} = 0, c_{2,t} = 1$$

The alternative allocation

$$\hat{c}_{1,t} = 1/2, \hat{c}_{2,t} = 1/2$$

is feasible and we can check that it makes every generation strictly better off. So the autarkic equilibrium is once again not Pareto optimal.

□

The failure of the First Fundamental Theorem in these examples is since this theorem has very few explicit assumptions and a straightforward proof and therefore would be expected to hold quite generally. What is going wrong?

The infinity of time certainly has a role to play. In Example 9.1 the Pareto improvement comes about because each generation gives up half a unit of the consumption goods to its elders during its youth and in return receives half a unit of the consumption good from its juniors in its own old age. If there had been an end of time then the last generation would not have had any juniors to compensate them and would have been worse off. Then the new allocation would no longer be a Pareto improvement. In Example 9.2 the direction of transfers is the opposite and hence there would be no Pareto improvement if there had been a beginning of time.

However, the infinity of time cannot by itself explain the failure of the First Fundamental Theorem. The Ramsey model provides a counterexample of a type of economy which also has infinite time but where the First Fundamental Theorem always holds.

One property which differentiates overlapping generations models from both finite equilibrium models and the Ramsey model is the property of limited participation. Only agents who are alive on a certain date are allowed to buy and sell commodities deliverable on that date. The unborn and the dead are excluded from markets. While this is certainly reasonable from an economic point of view, can it be this exclusion which is responsible for the possibility of competitive equilibria not being Pareto optimal in this model?

2. Pure Exchange Economies: Complete Participation

To see whether it is the incompleteness of market participation which leads to non-Pareto optimal equilibria, we modify the overlapping generations by allowing all agents to trade in all dated commodities and looking for equilibrium prices. We might imagine a marketplace that stands outside of time, in which souls can trade for delivery of commodities at any point of time. We look for prices at which the demands and supplies of the souls of different generations match. We show by an example that even when completing the markets in this manner there still can be equilibria which are not Pareto optimal.

Example 9.3. We continue with the economy of Example 9.2 in which time is doubly infinite, preferences are given by

$$u(c_{1,t}, c_{2,t+1}) = \ln(1 + c_{1,t}) + \ln(1 + c_{2,t+1}).$$

and endowments are

$$\omega_1 = 0,\ \omega_2 = 1.$$

Suppose that the price of the good in period t is p_t. Then the consumer born in time t solves the problem

$$\max_{c_{1,t}, c_{2,t+1}} \ln(1 + c_{1,t}) + \ln(1 + c_{2,t+1})$$

subject to

$$p_t(\omega_1 - c_{1,t}) = p_{t+1}(c_{2,t+1} - \omega_2)$$

Assuming an interior solution the solution to this optimisation problem is[3]

$$c_{1,t} = \frac{2(p_{t+1}/p_t) - 1}{2}, \quad c_{2,t+1} = \frac{p_t}{2p_{t+1}} \tag{40}$$

Equilibrium requires

$$c_{1,t} + c_{2,t} = \omega_1 + \omega_2 = 1, \qquad \text{for all } t \tag{41}$$

There are many sequence of prices which satisfy the above equilibrium conditions. Here are two interesting ones.

First, the sequence $(\ldots, 4, 2, 1, {}^1\!/_2, {}^1\!/_4, \ldots)$. That is $p_t = 1/2^t$ for all t. With these prices we see from (40) that $c_{1,t} = 0, c_{2,t+1} = 1$ for all t which satisfies (41). This is an equilibrium in which every agent consumes their endowment—an allocation which we have shown to be Pareto inferior earlier.[4]

Second, consider the prices $(\ldots, 1, 1, 1, \ldots)$. That is $p_t = 1$ for all t. For these prices $c_{1,t} = 1/2, c_{2,t+1} = 1/2$ which again satisfies (41). This allocation is an Pareto improvement on the previous allocation and we show in the appendix to this chapter that it is in fact Pareto optimal, that is there is no possibility of a further Pareto improvement from this allocation.

□

Example 9.3 shows that Pareto suboptimal equilibria can persist in an overlapping generations economy even when we assume that all agents can participate in all markets. Thus it is not incomplete participation which is to blame for the Pareto suboptimality. Rather, it is the combination of an infinite horizon and an infinity of agents who do not live forever which is to blame for the failure of the First Welfare Theorem.

---

[3] Exercise: derive this.

[4] We derived our demand functions (40) assuming an interior solution to the consumer's maximisation problem. On the other hand, the allocation in this case lies on the boundary of the consumption set since $c_{1,t} = 0$. Why is this not a problem?

3. Money

Overlapping generations models have also been used as a tool for studying monetary economies. Modern fiat money—unlike money based on precious metals—is intrinsically worthless. It is neither useful as an object of consumption nor is it useful as a factor of production. It is demanded only in the expectation that it will be accepted in exchange for useful goods. But those who accept money must in turn expect that it will be accepted by others when they try to spend it. This dependence of the value of money on the general belief in its acceptability is captured well by overlapping generations models.

Consider an economy with singly-infinite time. The initial old have a utility function given by

$$u_0(c_{2,0})$$

and other generations have an utility function

$$u(c_{1,t}, c_{2,t+1}).$$

Endowments in the two periods are ω_1 and ω_2 respectively.

Only living people can trade. As we discussed earlier, with this assumption the only equilibrium in this model is an autarkic one. Example 9.1 shows that this autarkic equilibrium might not be Pareto optimal.

Now suppose we endow the initial old with M units of useless but durable green pieces of paper that we call 'money'. Assume that just like goods this money can also be traded by agents alive at each date. Let p_t be the price of the consumption good in terms of money at time t.

The decision problem of the initial old is simple. Since they have only one period to live and money is useless they sell the money for whatever it is worth in terms of consumption goods. So we have

$$c_{2,0} = M/p_0 \tag{42}$$

The later generations now have the choice of selling some goods in their youth and acquiring money. Since money is durable they can carry it over to their old age and spend it for consumption goods then. If we denote by M_t^d the amount of money demanded in their youth by the generation born in period t then the optimisation problem for that generation is

$$\max_{c_{1,t}, c_{2,t+1}, M_t^d} u(c_{1,t}, c_{2,t+1})$$

subject to

$$\begin{aligned} p_t(\omega_1 - c_{1,t}) &= M_t^d \\ p_{t+1}(c_{2,t+1} - \omega_2) &= M_t^d \\ M_t^d &\geq 0 \end{aligned}$$

If we define real money demand by $m_t^d = M_t^d / p_t$ and the relative increase in prices by $\pi_{t+1} = p_{t+1}/p_t$ then the above optimization problem can be rewritten as

$$\max_{c_{1,t}, c_{2,t+1}, m_t^d} u(c_{1,t}, c_{2,t+1})$$

subject to

$$\begin{aligned} \omega_1 - c_{1,t} &= m_t^d \\ c_{2,t+1} - \omega_2 &= m_t^d / \pi_{t+1} \\ m_t^d &\geq 0 \end{aligned}$$

This form of the optimisation problem makes it clear that the real money stock demanded by a young agent equals that agent's real savings and $1/\pi_{t+1}$ is the returns that the agent earns on her savings.

The solution of the optimisation problem gives us m_t^d, $c_{1,t}$ and $c_{2,t+1}$ as functions of π_{t+1}. In particular we define

$$m_t^d = L(1/\pi_{t+1})$$

Equilibrium in the money market requires

$$L(1/\pi_{t+1}) = M/p_t \tag{43}$$

Equilibrium in the goods market requires

$$c_{1,t} + c_{2,t} = \omega_1 + \omega_2 \tag{44}$$

From Walras' law we know that if one of these equilibrium conditions is satisfied then the other will be satisfied automatically.

For periods $t > 0$ we can use the consumer's budget constraint to rewrite (44) as follows

$$\begin{aligned} c_{1,t} + c_{2,t} &= \omega_1 + \omega_2 \\ c_{2,t} - \omega_2 &= \omega_1 - c_{1,t} \\ m^d_{t-1}/\pi_t &= m^d_t \\ L(1/\pi_t)\frac{1}{\pi_t} &= L(1/\pi_{t+1}) \end{aligned} \tag{45}$$

Using (43) for $t = 0$ and (45) for $t > 0$ and remembering that Walras' law allows us to use just one equilibrium condition per period we have the equilibrium conditions

$$L(1/\pi_1) = M/p_0 \tag{46}$$

$$L(1/\pi_t)\frac{1}{\pi_t} = L(1/\pi_{t+1}), \qquad t > 0 \tag{47}$$

For simplicity let us look at steady state equilibria where π_t and hence $c_{1,t}$ and $c_{2,t}$ are constants. Then the equilibrium conditions become,

$$L(1/\pi) = M/p_0 \tag{48}$$

$$L(1/\pi)(1/\pi) = L(1/\pi) \tag{49}$$

There are two solutions to these equations. The first is a value of $\pi = \pi^*$ where

$$L(\pi^*) = 0.$$

There is always such a π^* since if we take the relative price ratio between consumption in the two periods to be equal to the consumer's marginal rate of substitution at the endowment point then the consumer is happy consuming her endowment and does not want to hold any money.

If $L(\pi^*) = 0$ then (48) cannot strictly be satisfied if $M > 0$. However, loosely speaking we can say that this equation is satisfied with $p_0 = \infty$. An infinite price of good in terms of money means that it is impossible to buy even the smallest amount of goods with any amount of money however large. Money is thus valueless in this equilibrium and everyone continues to consume their endowment.[5]

[5] If you are uncomfortable with infinity as a price, you can get rid of it as follows. Rather than choosing money as the numeriare choose an abstract numeriare in terms of which μ_t is the price of money and λ_t is the price of goods in period t.

The other steady-state solution is the one with $\pi = 1$. In this case (49) is automatically satisfied and (48) gives us

$$p_0 = M/L(1)$$

If $L(1) = 0$ then this once again gives us a equilibrium with $p_0 = \infty$ where money is not valued. But if $L(1) > 1$ then we have an equilibrium in which money has a constant value in the steady state. Thus even if money is useless it is possible to have an equilibrium with valued money in an overlapping generations model.

Appendix

PROPOSITION 9.1. *In the overlapping generations model with utility function*

$$u(c_{1,t}, c_{2,t+1}) = \ln(1 + c_{1,t}) + \ln(1 + c_{2,t+1}),$$

stationary endowments $\omega_1 = 1, \omega_2 = 0$ *and doubly-infinite time, the stationary allocation* $c_{1,t} = 1/2, c_{2,t} = 1/2$ *is Pareto optimal.*

The main idea. If the allocation mentioned in the proposition is not Pareto optimal then there must be another allocation which is a Pareto improvement, i.e. which makes everyone at least as well off and at least one agent strictly better off. Let us denote this allocation by $(\hat{c}_{1,t}, \hat{c}_{2,t})$.

Since any leftover output can be given to some agent to make them better off without making anyone worse off, we assume we have already done so, so that,

$$\hat{c}_{1,t} + \hat{c}_{2,t} = \omega_1 + \omega_2 = 1 \tag{50}$$

If the allocation with the hats is different from $(1/2, 1/2)$ there must be some t for which $\hat{c}_{1,t} \neq 1/2$ or $\hat{c}_{2,t} \neq 1/2$. From (50) it then follows that for that t either $\hat{c}_{1,t} > 1/2$ or $\hat{c}_{2,t} > 1/2$. (Why?)

Assume that $\hat{c}_{2,t} > 1/2$ for some t.

(The argument remains basically unchanged when $\hat{c}_{1,t} > 1/2$ if time is doubly infinite. All that we have to do is to look at periods before t rather than at periods after t as we do below. A little more change is required if time has a beginning. Try to think through both these cases once you have studied the case where $\hat{c}_{2,t} > 1/2$.)

Arguing as above then you will find an equilibrium with $\mu_t = 0$ for all t. Our $p_t = \lambda_t/\mu_t$ and hence it turns out to be infinite in this equilibrium.

If $\hat{c}_{2,t} > 1/2$ then from (50) it must be the case that $\hat{c}_{1,t} < 1/2$. That is, the young born in period t consume less in their youth under the hat allocation than they did in the initial allocation. But if the hat allocation is to be a Pareto improvement then these period-t young have to be at least as well off under the hat allocation as under the original allocation. This can happen only if they consume more in their old age under the hat allocation than under the original allocation. So, it must be that $\hat{c}_{2,t+1} > 1/2$.

Let

$$\delta_t = 1/2 - \hat{c}_{1,t}$$
$$\phi_t = \hat{c}_{2,t+1} - 1/2$$

We are looking at the case where $\delta_t > 0$ and we have argued in the last paragraph that $\phi_t > 0$. In fact we can do better than that. Here's how. Consider the possibility that $\phi_t = \delta_t$. Then

$$\begin{aligned} u(\hat{c}_{1,t}, \hat{c}_{2,t+1}) &= u(1/2 - \delta_t, 1/2 + \phi_t) \\ &= u(1/2 - \delta_t, 1/2 + \delta_t) \\ &= \ln(1 + 1/2 - \delta_t) + \ln(1 + 1/2 + \delta_t) \\ &= \ln[(3/2 - \delta_t)(3/2 + \delta_t)] \\ &= \ln[9/4 - \delta_t^2] \\ &< \ln(9/4) = u(c_{1,t}, c_{2,t+1}) \end{aligned}$$

Thus increasing consumption in old age by the same amount as consumption is reduced in youth will make the generation-t agent strictly worse off than she was under the original $(1/2, 1/2)$ allocation. Therefore if the hat allocation is to keep the consumer at least as well off as before it must be the case that

$$\phi_t > \delta_t \tag{51}$$

But where is this extra consumption going to come from? It can only come by reducing the consumption of the young in period $t+1$. Since

$$\hat{c}_{1,t+1} + \hat{c}_{2,t+1} = 1 = 1/2 + 1/2$$

It must be the case that,

$$1/2 - \hat{c}_{1,t+1} = \hat{c}_{2,t+1} - 1/2$$

If we define $\delta_{t+1} = 1/2 - \hat{c}_{1,t+1}$ the above shows that

$$\delta_{t+1} = \phi_t > \delta_t > 0$$

Defining $\phi_{t+1} = \hat{c}_{2,t+2} - 1/2$ and applying the argument that we applied to generation t to generation $(t+1)$ we find that

$$\delta_{t+2} = \phi_{t+1} > \delta_{t+1} > 0$$

Continuing to repeat this argument we get the chain of inequalities

$$\delta_t < \delta_{t+1} < \delta_{t+2} < \cdots \tag{52}$$

The above chain of inequalities shows that δ_k increases with k. But there is an upper bound on δ_k. Since

$$\delta_k = 1/2 - \hat{c}_{1,k}$$

and $\hat{c}_{1,k} \geq 0$ it must be the case that

$$\delta_k \leq 1/2 \tag{53}$$

If we could use (52) to argue that δ_k grows without bounds then that would contradict (53) and we would have succeeded in proving that it is not possible to construct a Pareto improvement over the original allocation $(1/2, 1/2)$.

Unfortunately this is not the case. Equation (52) does not imply unbounded growth of δ_k. A sequence like

$$0.4, 0.44, 0.444, 0.4444, \ldots$$

can satisfy both (52) and (53). So in order to be successful in our proof we have to find a way to strengthen (52).

A stronger inequality. As before we define

$$\delta_k = 1/2 - \hat{c}_{1,k} > 0$$
$$\phi_k = \hat{c}_{2,k+1} - 1/2 > 0$$

Also as before feasibility requires

$$\phi_k = \delta_{k+1} \tag{54}$$

For the hat allocation to be a Pareto improvement it is necessary that

$$\begin{aligned}
u(\hat{c}_{1,k}, \hat{c}_{2,k+1}) &\geq u(c_{1,k}, c_{2,k+1}) \\
u(1/2 - \delta_k, 1/2 + \phi_k) &\geq u(1/2, 1/2) \\
\ln[(3/2 - \delta_k)(3/2 + \phi_k)] &\geq \ln(9/4) \\
\frac{9}{4} + \frac{3}{2}\phi_k - \frac{3}{2}\delta_k - \delta_k\phi_k &\geq \frac{9}{4} \\
\frac{\phi_k}{\delta_k} &\geq \frac{1}{1 - 2\delta_k/3}
\end{aligned}$$

Now we know from (52) that $\delta_t \leq \delta_k$ for $k \geq t$. Applying this to the denominator of the above expression and remembering that $\delta_t \leq 1/2$ from (53) we have

$$\frac{\phi_k}{\delta_k} \geq \frac{1}{1 - 2\delta_t/3} \geq \frac{1}{1 - 2 \cdot (1/2)/3} = 3/2$$

But $\phi_k = \delta_{k+1}$ so we have

$$\frac{\delta_{k+1}}{\delta_k} \geq 3/2 \tag{55}$$

This says that in each period δ grows at least 1.5 times. As a result it will grow unboundedly starting from any nonzero value. Formally, by starting from time t and chaining together (55) together j times we have

$$\delta_{t+j} \geq (3/2)^j \delta_t$$

Provided $\delta_t > 0$ the right-hand side becomes larger than 1/2 for large enough j and hence contradicts (53), thus proving that no Pareto improvement over our original allocation is possible.

Generalisation. The inequalities (52) and (55) were derived using a particular functional form of the utility function and a particular initial allocation. But the style of reasoning we have used above can be generalised to provide a criteria for Pareto optimality applicable to arbitrary utility functions and arbitrary allocations. See Proposition 5.6 of [**BS80**].

Bibliography

[Bar94] Robert J Barro. The aggregate-supply/aggregate-demand model. *Eastern Economic Journal*, 20(1):1–6, 1994.

[BB97] Robert A. Becker and John H. Boyd, III. *Capital Theory, Equilibrium Analysis and Recursive Utility*. Blackwell, 1997.

[Bil95] Patrick Billingsley. *Probability and Measure*. Wiley-Interscience, 3rd edition, 1995.

[BS80] Y. Balasko and K. Shell. The overlapping generations model, I: The case of pure exchange without money. *Journal of Economic Theory*, 23(3):281–306, 1980.

[CA03] K. L. Chung and Farid AitSahlia. *Elementary Probability Theory: With Stochastic Processes and an Introduction to Mathematical Finance*. Springer, 4th edition, 2003.

[Col95] David Colander. The stories we tell: A reconsideration of as/ad analysis. *The Journal of Economic Perspectives*, 9(3):169–188, 1995.

[Dav11] Paul Davidson. *Post-Keynesian Macroeconomic Theory*. Edward Elgar, 2nd edition, 2011.

[Dut02] Amitava Krishna Dutt. Aggregate demand-aggregate supply analysis: A history. *History of political economy*, 34(2):321–363, 2002.

[JP04] Jean Jacod and Philip Protter. *Probability Essentials*. Springer, 2nd edition, 2004.

[Kah13] Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2013.

[Key36] John Maynard Keynes. *The General Theory of Employment, Interest and Money*. Macmillan, 1936.

[LS12] Lars Ljungqvist and Thomas Sargent. *Recursive Macroeconomic Theory*. MIT Press, 3rd edition, 2012.

[LY08] David G. Luenberger and Yinyu Ye. *Linear and Nonlinear Programming*. Springer, 3rd edition, 2008.

[MCWG95] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

[Rom00] David Romer. Keynesian macroeconomics without the LM curve. *The Journal of Economic Perspectives*, 14(2):149–169, 2000.

[Rom11] David Romer. *Advanced Macroeconomics*. McGraw Hill, 4th edition, 2011.

[Ros09] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 10th edition, 2009.

[SLP89] Nancy Stokey, Robert E. Lucas, Jr., and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.

[Tay04] Lance Taylor. *Reconstructing Macroeconomics: Structuralist Proposals and Critiques of the Mainstream*. Harvard University Press, 2004.

[Wil91] David Williams. *Probability with Martingales*. Cambridge university press, 1991.