Notes on Dynamic Methods in Macroeconomics

Nicola Pavoni.

`This version: October 2008.`

VERY PRELIMINARY AND LARGELY INCOMPLETE.

Contents

Chapter 1

Introduction to Recursive Methods

These notes are targeted to advanced Master and Ph.D. students in economics. I believe they can also be of some use to researchers in macroeconomic theory. The material contained in these notes only assumes the reader to know basic math and static optimization, and an advanced undergraduate or basic graduate knowledge of economics.

Useful references are Stokey et al. (1991), Bellman (1957), Bertsekas (1976), and Chapters 6 and 19-20 in Mas-Colell et al. (1995), and Cugno and Montrucchio (1998).[1] For further applications see Sargent (1987), Ljungqvist and Sargent (2003), and Adda and Cooper (2003). Chapters 1 and 2 of Stokey et al. (1991) represent a nice introduction to the topic. The student might find useful reading them before the course starts.

1.1 The Optimal Growth Model

Consider the following problem

$$\begin{aligned} V^*(k_0) &= \max_{\{k_{t+1},\, i_t,\, c_t,\, n_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) \qquad (1.1) \\ & s.t. \\ k_{t+1} &= (1-\delta)k_t + i_t \\ c_t + i_t &\leq F(k_t, n_t) \\ c_t,\ k_{t+1} &\geq 0,\ n_t \in [0,1]\,;\ k_0 \text{ given.} \end{aligned}$$

where k_t is the capital stock available for period t production (i.e. accumulated past investment i_t), c_t is period t consumption, and n_t is period t labour. The utility function

[1] Cugno and Montrucchio (1998) is a very nice text in dynamic programming, written in Italian.

u is assumed to be strictly increasing, and F is assumed to have the characteristics of the *neoclassical* production function. [2] The above problem is know as the (deterministic neoclassical) optimal growth model, after the seminal papers of Frank Ramsey (1928), David Cass (1965) and Tjalling C. Koopmans (1963). The first constraint in the above problem constitutes the is *law of motion* or *transition function* for capital, which describes the evolution of the *state variable* k. The optimal value of the problem $V(k_0)$ - seen as a function of the initial state k_0 - is denominated as the *value function*.

By the strict monotonicity of u, an optimal plan satisfies the feasibility constraint $c_t + i_t \leq F(k_t, n_t)$ with equality. Moreover, since the agent does not value leisure (and labor is productive), an optimal path is such that $n_t = 1$ for all t. Using the law of motion for k_t, one can rewrite the above problem by dropping the leisure and investment variables, and obtain:

$$\begin{aligned} V^*(k_0) &= \max_{\{k_{t+1},\, c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) \\ & \quad s.t. \\ c_t + k_{t+1} &= f(k_t); \\ c_t,\; k_{t+1} &\geq 0;\; k_0 \text{ given,} \end{aligned} \tag{1.2}$$

where $f(k_t) = F(k_t, 1) + (1-\delta)k_t$ is the total amount of resources available for consumption in period t. The problem can be further simplified and written - as a function of k alone - as follows [Check it!]:

$$\begin{aligned} V^*(k_0) &= \max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(f(k_t) - k_{t+1}) \\ & \quad s.t. \\ 0 &\leq k_{t+1} \leq f(k_t);\; k_0 \text{ given.} \end{aligned}$$

This very simple framework includes already few interesting models in macroeconomics. For example, when $f(k_t) = Ak_t$ we have the following: If $A = 1$ the problem can be interpreted as the one of eating optimally a cake of initial size k_0. Since Gale (1967)

[2] F is linearly homogeneous, with $F_1 > 0,\ F_{11} < 0$, and $F_2 > 0,\ F_{22} < 0$. In addition, commonly used assumptions are complementarity $F_{12} \geq 0$ and Inada conditions: $\lim_{k \to 0} F_1 = \infty$ and $\lim_{k \to \infty} F_1 = 0$. A typical example is the Cobb-Douglas production function

$$F(k, n) = Ak^{\alpha} n^{1-\alpha} \text{ with } \alpha \in (0, 1)\,.$$

used extensively this example, our problem with $f(k_t) = k_t$ is sometimes denominated as `the Gale's Cake Eating problem`. If $A < 1$ the problem can still be interpreted as a cake eating problem, where the cake depreciates every period. However, note that when $A > 1$ the problem becomes the simplest model of endogenous growth, and it is called the `AK-model`.[3] Finally, if we set $f(k_t) = (1 + r) k_t + y$, and interpret $r \geq 0$ as the interest rate and $y \geq 0$ as labor income, the model becomes one of the simplest problems of optimal saving (borrowing is indeed not allowed since $k_{t+1} \geq 0$).

1.1.1 Sequential Formulation

We shall regard the above formulation as the *sequential form* of the problem. Notice that the objective function is an infinite summation, which must converge in order to have a well defined problem.[4] The Max operator presents similar complications.

Once the problem is well defined, one could attack it by using appropriate generalizations to the well known tools of "static" maximization (such as Lagrange and Kuhn-Tucker theorems). The fact that the choice variable belongs to an infinite dimensional space induces few technical complications, this is however a viable way.[5] If one takes the sequential approach, the infinite summation $\sum_{t=0}^{\infty} \beta^t u(f(k_t) - k_{t+1})$ is a special case of the function(al) $W(k_0, k_1, k_2, ...)$ defined on an appropriate infinite dimensional space. If the optimum is interior and W is sufficiently smooth, an optimal capital path $\left\{k_{t+1}^*\right\}_{t=0}^{\infty}$ starting from k_0 must satisfy the usual first order conditions $\frac{\partial W(k_0, k_1^*, k_2^*, \ldots)}{\partial k_t} = 0$ for $t = 1, 2, ...$. In our (time-separable) case, the first order conditions become

$$u'\left(f(k_{t-1}^*) - k_t^*\right) = \beta f'(k_t^*) u'\left(f(k_t^*) - k_{t+1}^*\right) \qquad t = 1, 2, ... \tag{1.3}$$

with k_0 given. These conditions are known as Euler equations (or Euler-Lagrange conditions). In fact, conditions (1.3) are the differentiable version of a set of necessary conditions that one obtains by following a general approach to optimization problems which is denoted as the *Euler variational approach.* Unfortunately, even when the problem is concave (i.e., when both u and f are concave) the Euler equations are not sufficient for an optimum. To detect on optimal path in our case one typically imposes the *transversality*

[3] This case can be seen as a Cobb-Douglas where the capital share α equals 1.

[4] The problem could, in fact, be tackled without imposing convergence of the series. This approach would however require the knowledge of advanced maximization concepts (such as overtaking), which are beyond the targets of these notes.

[5] An accessible text that introduces to the mathematics of optimization in infinite dimensional spaces is Luenberger (1969).

condition: $\lim_{T\to\infty}\beta^T f'(k_T^*)u'\left(f(k_T^*)-k_{T+1}^*\right)k_T^*\leq 0$ as well. This will all be explained in Chapter 3.

The economic interpretation of the Euler equation is relatively easy, and quite important. Using the definition of $f(k_t)$, and recalling that $c_t = f(k_t) - k_{t+1}$, (1.3) becomes

$$\frac{u'(c_{t-1}^*)}{\beta u'(c_t^*)} = f'(k_t^*) = \frac{\partial F(k_t^*, 1)}{\partial k_t} + (1-\delta), \tag{1.4}$$

which is a condition of dynamic efficiency. It asserts the equality between the marginal rate of substitution between $t-1$ and t consumption $\frac{u'(c_{t-1})}{\beta u'(c_t)}$, and its marginal rate of transformation $\frac{\partial F(k_t,1)}{\partial k_t}+(1-\delta)$: since consumption and investment are the same good, a marginal reduction of period $t-1$ consumption implies a one-to-one increase in i_{t-1}, which in turn increases by $\frac{\partial F(k_t,1)}{\partial k_t}+(1-\delta)$ the amount of goods available for period t consumption.

Under the assumptions mentioned in footnote 2, the economy allows for an interior steady state for capital and consumption: k^{ss}, and c^{ss}. In the steady state, the Euler equation implies $\frac{1}{\beta} = f'\left(k^{ss}\right)$, which becomes - assuming a Cobb-Douglas production function -

$$\frac{1}{\beta} = \alpha A k^{\alpha-1} + (1-\delta) \implies k^{ss} = \left(\frac{\alpha A}{\delta+\rho}\right)^{\frac{1}{1-\alpha}}$$

where $\rho = \beta^{-1} - 1$ is the agent's discount rate [What is the intuition for the above expression?].

The model allows a simple graphical analysis of the dynamics in the (k_t, c_t) space. The steady state will be described by the crossing point between two curves: the curve that describes the condition $k_{t+1} = k_t$ and that describing the condition $c_{t+1} = c_t$. From the feasibility constraint, one gets: $k_{t+1} = k_t \Leftrightarrow c_t = f(k_t) - k_t$. The concave function $c^0(k) = f(k) - k$ hence describes the curve of time constant capital. Moreover, notice that given k_t, next period capital k_{t+1} is decreasing in c_t. For consumption values above the curve c^0, capital must decrease, while when the chosen c is below this curve, one has $k_{t+1} > k_t$. If $f(0) = 0$ this curve starts at the origin of the $\mathbb{R}_+^2$ diagram and - whenever $\lim_{k\to\infty} f'(k_t) < 1$ - it intersects the zero consumption level at a finite $\bar{k} < \infty$.

The dynamics for c_t are described by condition (1.4). It is easy to see that by the strict concavity of f, and u there is only one interior point for k that generates constant consumption. This is the steady state level of capital. For capital levels above k^{ss} the value at the right hand side of the Euler equation $\alpha k^{\alpha-1} + (1-\delta)$ is lower than $\frac{1}{\beta}$, which implies a ratio $\frac{u'(c_{t-1})}{u'(c_t)}$ lower than one, hence $c_t > c_{t-1}$. The opposite happens when $k > k^{ss}$. This is hence represented by a vertical line in the (k, c) space. Of course, only one level of consumption is consistent with feasibility. But feasibility is guaranteed by the function c^0. At the crossing point of the two curves one finds the pair (k^{ss}, c^{ss}) which fully describes the steady state of the economy]

As we saw, the Euler equations alone do not guarantee optimality. One typically impose the transversality condition. It is easy to see, that any path satisfying the Euler equation and converging to the steady state satisfied the transversality condition, it is hence an optimal path [Check it].

1.1.2 Recursive Formulation

We now introduce the *Dynamic Programming* (DP) approach. To see how DP works, consider again the optimal growth model. One can show, under some quite general conditions, that the optimal growth problem can be analyzed by studying the following *functional equation*

$$V(k) = \max_{0 \leq k' \leq f(k)} u(f(k) - k') + \beta V(k'), \tag{1.5}$$

where the unknown object of the equations is the value function V. A function can be seen as in infinite dimensional vector. The number of equations in (1.5) is also infinite as we have one equation for every $k \geq 0$.

What are then the **advantages** of such a *recursive formulation* of this sort? First of all, realizing that a problem has a recursive structure quite often helps understand the nature of the problem we are facing. The recursive formulation reduces a single infinite-dimensional problem into a sequence of one-dimensional problems. This often implies several computational advantages. Moreover, the study of the above functional equation delivers much richer conclusions than the study of the solution of a specific optimal growth problem. In dynamic programming we embed the original problem within a family of similar problems, and study the general characteristics of this class of problems. Instead of deriving a specific path of optimal capital levels $\left\{k^*_{t+1}\right\}_{t=0}^{\infty}$, a solution to the recursive problem delivers a *policy function* $k' = g(k)$ which determines tomorrow's optimal capital level k' for *all possible* levels of today's capital $k \geq 0$.

Of course, solving a functional equation is typically a difficult task. However, once the function V is known, the derivation of the optimal choice k' is an easy exercise since it involves a simple 'static' maximization problem. Moreover, we will see that in order to study a given problem one does not always need to fully derive the exact functional form of V. One can then add many other advantages which go from the more applied numerical computation facilities, to the appealing philosophical idea of being able to summarize a complicate world (the infinite-dimensional problem) with a policy function, which is defined in terms of a small set of states.

In Dynamic Programming, state variables play a crucial role. But what are the state variable exactly?

> *"In some problems, the state variables and the transformations are forced upon us; in others there is a choice in these matters and the analytic solution stands or fall upon this choice; in still others, the state variables and sometimes the transformations must be artificially constructed. Experience alone, combines with often laborious trial and error, will yield suitable formulations of involved processes."* Bellman (1957), page 82.

In other terms, the detection of the state variables is sometimes very easy. In other cases it is quite complicated. Consider the following formulation of the model of optimal growth

$$
\begin{aligned}
V(k_0) &= \max_{\{i_t,\, c_t,\, n_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) \\
&\quad s.t. \\
c_t + i_t &\le F\left(\sum_{j=1}^{t} (1-\delta)^{t-j} i_{j-1} + (1-\delta)^t k_0, n_t \right) \\
c_t &\ge 0;\ \sum_{j=1}^{t} (1-\delta)^{t-j} i_{j-1} + (1-\delta)^t k_0 \ge 0;\ n_t \in [0,1]\,;\ k_0 \text{ given.}
\end{aligned}
$$

The problem is now described in terms of *controls* alone, and one can of course solve it in this form. However, "experience and trial and error" suggests that it convenient to first summarize some of the past controls $\{i_j\}_{j=0}^{t-1}$ by the state variable $k_t = \sum_{j=1}^{t}(1-\delta)^{t-j} i_{j-1} + (1-\delta)^t k_0$ and attack the problem using the (1.1)-(1.2) formulation. In this case it was also easier to detect the state as it was the predetermined variable in our problem (recall that k_0 was 'given'). Looking for predetermined variables is certainly useful, but the 'trick' does not always work.

In other cases, the purely sequential formulation can actually be very convenient. For example, one can easily check that the cake eating problem can be written in terms of controls alone as follows:

$$
\begin{aligned}
V(k_0) &= \max_{\{c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) \\
&\quad s.t. \\
\sum_{t=0}^{\infty} c_t &\le k_0;\ c_t \ge 0;\ k_0 \text{ given.}
\end{aligned}
\tag{1.6}
$$

With the problem formulated in this form the solution is quite easy to find.

Exercise 1 *Consider the above problem and assume* $u = \ln$. *Use the usual Lagrangian approach to derive the optimal consumption plan for the cake eating problem.*

However, let me be consistent with the main target of this chapter, and write the cake eating problem in recursive form. By analogy to (1.5), one should easily see that when u is logarithmic the functional equation associated to $(Cake)$ is

$$V(k) = \max_{0 \leq k' \leq k} \ln(k - k') + \beta V(k'). \tag{1.7}$$

As we said above, the unknown function V can be a quite complicate object to derive. However, sometimes one can recover a solution to the functional equation by first guessing a form for V and then verifying that the guess is right. Our guess for V in this specific example is an affine logarithmic form:

$$V(k) = A + B \ln(k).$$

Notice that the *guess and verify procedure* simplifies enormously the task. We now have to determine just two numbers: the coefficients A and B. A task which can be accomplished by a simple static maximization exercise. Given our guess for V and the form of u, the problem is strictly concave, and first order conditions are necessary and sufficient for an optimum. Moreover, given that $\lim_{k \to 0} \ln(k) = -\infty$, the solution will be and interior one. We can hence disregard the constraints and solve a free maximization problem. The first order conditions are

$$\frac{1}{k - k'} = \beta \frac{B}{k'}$$

which imply the following policy rule $k' = g(k) = \frac{\beta B}{1 + \beta B} k$. Plugging it into our initial problem $V(k) = \ln(k - g(k)) + \beta V(g(k))$, we have

$$V(k) = A + \frac{1}{1 - \beta} \ln k,$$

so $B = \frac{1}{1-\beta}$, which implies the policy $g(k) = \beta k$.[6]

Exercise 2 *Visualize the feasibility set* $0 \leq k' \leq k$ *and the optimal solution* $g(k) = \beta k$ *into a two dimensional graph, where* $k \geq 0$ *is represented on the horizontal axis. [Hint: very easy!]*

[6] Note that the actual value of the additive coefficient A is irrelevant for computing the optimal policy.

Exercise 3 *Consider the optimal growth model with* $y = f(k) = k^\alpha$, $\alpha \in (0,1)$ *(so* $\delta = 1$*) and log utility* ($u(c) = \ln c$)*. Guess the same class of functions as before:*

$$V(k) = A + B \ln k,$$

and show that for $B = \frac{\alpha}{1-\beta\alpha}$ *the functional equation has a solution with policy* $k' = g(k) = \alpha\beta k^\alpha$*. Use the policy to derive the expression for the steady state level of capital* k^{ss}*.*

Exercise 4 *Under the condition* $f(0) = 0$*, the dynamic system generated by the policy we described in the previous exercise has another non-interior steady state at zero capital and zero consumption. The non-interior steady state* $k = 0$ *is known to be a 'non-stable' stationary point. Explain the meaning of the term as formally as you can.*

Exercise 5 *Solve the Solow model with* $f(k) = k^\alpha$ *and* $\delta = 1$ *and show that given the saving rate* s*, the steady state level of capital is* $k^{ss} = s^{\frac{1}{1-\alpha}}$*. Show that the golden rule level of savings implies a policy of the form* $k' = \alpha k^\alpha$*. Compare this policy and the golden rule level of capital in the steady state, with the results you derived in the previous exercise. Discuss the possible differences between the two capital levels.*

Exercise 6 *Consider the following modification of the optimal growth model*

$$\begin{aligned} V(k_0) &= \max_{\{k_{t+1},\, c_t\}_{t=0}^{\infty}} \widetilde{u}(c_0) + \sum_{t=1}^{\infty} \beta^t u(c_{t-1}, c_t) \\ & \quad s.t. \\ c_t + k_{t+1} &= f(k_t) \\ c_t,\ k_{t+1} &\geq 0;\ k_0 \text{ given.} \end{aligned}$$

Write the problem in recursive form by detecting the states, the controls, and the transition function for the states. [Hint: you might want to assume the existence of a non negative number $x \geq 0$ *such that* $\widetilde{u}(c_0) = u(x, c_0)$ *for any* c_0*].*

1.2 Finite Horizon Problems as Guided Guesses

Quite often, an analytic form for V does not exist, or it is very hard to find (this is why numerical methods are useful). Moreover, even in the cases where a closed form does exist, it is not clear what a good guess is. Well, in these cases, the guess can often be guided by economic reasoning. For example, consider the one-period Gale's cake eating problem

$$V_1(k) := \max_{0 \le k' \le k} u\left(k - k^{'}\right). \tag{1.8}$$

The sub-index on the value function V indicates the time horizon of the problem. It says how many periods are left before the problem ends. In (1.8) there is only one period. Note importantly, that (1.8) describes a map from the initial function $V_0(k) \equiv 0$ and the new function V_1, as follows

$$V_1(k) := (TV_0)(k) = \max_{0 \le k' \le k} u\left(k - k^{'}\right) + \beta V_0(k') = \max_{0 \le k' \le k} u\left(k - k^{'}\right).$$

The operator T maps functions into functions. T is known as the *Bellman operator*. Given our particular choice for V_0, the n-th iteration of the Bellman operator delivers the function

$$V_n(k) := T^n V_0 = (TV_{n-1})(k) = \max_{0 \le k' \le k} u\left(k - k^{'}\right) + \beta V_{n-1}(k'),$$

which corresponds to the value function associated to the n-horizon problem. The solution to the functional equation (1.7) is the value function associated to the infinite horizon case, which can be see both as the limit function of the sequence $\{V_n\}$ and as a *fixed point* of the $T-$operator:

$$\lim_{n \to \infty} T^n V_0 = V_\infty = TV_\infty.$$

Below in these notes we will see that in the case of discounted sums the Bellman operator T typically describes *a contraction.* One important implication of the Contraction Mapping Theorem - which will be presented in the next chapter - is that, in the limit, we will obtain the same function V_∞, regardless our initial choice V_0.

As a second example, we can specify the cake eating problem assuming power utility: $u(c) = c^\gamma$, $\gamma \le 1$. It is easy to see that the solution to the one-period problem (1.8) gives the policy $k' = g_0(k) = 0$. If the consumer likes the cake and there is no tomorrow, she will eat it all. This solution implies a value function of the form $V_1(k) = u(k - g_0(k)) = B_1 k^\gamma = k^\gamma$. Similarly, using first order condition, one easily gets the policy $g_1(k) = \frac{k}{1+\beta^{\frac{1}{\gamma-1}}}$ as an interior solution of the following two period problem

$$\begin{aligned}
V_2(k) &= T^2 V_0 = (TV_1)(k) = \max_{0 \le k' \le k} \left(k - k^{'}\right)^\gamma + \beta V_1(k') \\
&= (k - g_1(k))^\gamma + \beta V_1(g_1(k)) \\
&= \left(1 - \frac{1}{1+\beta^{\frac{1}{\gamma-1}}}\right)^\gamma k^\gamma + \beta \left(\frac{1}{1+\beta^{\frac{1}{\gamma-1}}}\right)^\gamma k^\gamma \\
&= B_2 k^\gamma
\end{aligned}$$

and so on. It should then be easy to guess a value function of the following form for the infinite horizon problem

$$V(k) = Bk^{\gamma} + A.$$

Exercise 7 ***(Cugno and Montrucchio, 1998)*** *Verify that the guess is the right one. In particular, show that with* $B = \left(1 - \beta^{\frac{1}{1-\gamma}}\right)^{\gamma-1}$ *the Bellman equation admits a solution, with implied policy* $k' = g(k) = \beta^{\frac{1}{1-\gamma}}k$. *Derive the expression for the constant* A. *What is the solution of the above problem when* $\gamma \to 1$? *Explain.*

Bibliography

[1] Adda J., and Cooper (2003) *Dynamic Economics: Quantitative Methods and Applications,* MIT Press.

[2] Bellman, R. (1957) *Dynamic Programming.* Princeton, NJ: Princeton University Press.

[3] Bertsekas, D. P. (1976) *Dynamic Programming and Stochastic Control*, Academic Press, NY.

[4] Cass, D. (1965) “Optimum Economic Growth in an Aggregative Model of Capital Accumulation,” *Review of Economic Studies:* 233-240.

[5] Cugno, F. and L. Montrucchio (1998) *Scelte Intertemporali*, Carocci, Roma.

[6] Koopmans, T.C. (1965) *On the Concept of Optimal Growth*, Pontificia Accademia Scientiarum, Vatican City: 225-228.

[7] Mas-Colell, A., Michael D. Winston and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press.

[8] Ramsey, F. (1928) “A Mathematical Theory of Savings,” *Economic Journal*, 38: 543-549.

[9] Sargent, T. (1987) *Dynamic Macroeconomic Theory*, Harvard University Press.

[10] Sargent, T. J. and L. Ljungqvist (2000), *Recursive Macroeconomic Theory*, MIT Press.

[11] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

Chapter 2

Useful Mathematics

In order to fully enjoy the beauty of dynamic programming modeling, we must first recall some useful mathematical concepts. Classical references for the material contained in this chapter are SLP, and Rudin (1964). More advanced references are Lipschutz (1994), and Rudin (1991).

2.1 Metric and Linear Spaces

Definition 1 *A metric space* (X, d) *is a set* X, *together with a metric (or distance function)* $d : X \times X \to \mathbb{R}$, *such that for all* $x, y, z \in X$ *we have: (i)* $d(x, y) \geq 0$, *with* $d(x, y) = 0$ *if and only if* $x = y$; *(ii)* $d(x, y) = d(y, x)$; *and (iii)* $d(x, z) \leq d(x, y) + d(y, z)$.

For example, the set of real numbers $X = \mathbb{R}$ together with the absolute value $d(x, y) = |x - y|$ is a metric space. Notice indeed that (i) is trivial and also (ii) is verified. To see (iii) make a picture. Remember that the absolute value is defined as follows

$$|x - y| = \begin{cases} x - y \text{ if } x \geq y \\ y - x \text{ otherwise.} \end{cases}$$

The previous discussion can be easily generalized to any n dimensional space, with $n < \infty$. The most natural metric for these spaces is the Euclidean distance.

Exercise 8 *Show that the absolute value represents a metric on the set* $\mathbb{N}$ *of the natural numbers.*

Exercise 9 *Consider the* $\mathbb{R}^n$ *space of the vectors* $\mathbf{x} = (x_1, x_2, .., x_n)$. *Show that the Euclidean distance* $d_E(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$ *defines a metric on* $\mathbb{R}^n$.

$I\!R$ as an Ordered Field Notice we have implicitly defined a way of taking the difference between two real numbers and the fact that one real number can be greater than another one. This and other properties are common to any Ordered Field. In this section, we will also discuss what are the other properties that define $I\!R$ as an Ordered Field.

- The fact we were able to take the difference between two real numbers is a combination of two properties. First, within a **Field** we can **A**dd up two elements being sure that the resulting element still belong to the field. Moreover, the addition satisfies the following properties: (A1) $a+b=b+a$ and (A2) $a+(b+c)=(a+b)+c$. Second, within a Field we can also **M**ultiply among them two real numbers. The multiplication satisfies two properties very similar to the ones of the addition, namely: (M1) $a\cdot b=b\cdot a$ and (M2) $a\cdot(b\cdot c)=(a\cdot b)\cdot c$. Usually we do not write explicitly the multiplication, so $a\cdot b=ab$. Moreover, a Field satisfies also a **mixed** property (AM) $a(b+c)=ab+ac$. Finally, we have the zero element 0 and the one element 1 as invariants of the two mentioned operations: namely $a+0=a$ and $a\cdot 1=a$. From these invariant elements we can define other two elements which are the inverse of the operations (and have to belong to the Field). Namely, given an element a, we can define the element s_a as $a+s_a=0$ and, when $a\neq 0$ we can also define the element q_a as $a\cdot q_a=1$. They can be also denoted as $s_a=-a$ and $q_a=a^{-1}$.

Exercise 10 *Show that the set of natural numbers $I\!N=\{1,2,3,...\}$ is not a Field, according to the previous discussion.*

Linear Spaces: $I\!R$ as a Linear Space

Definition 2 *A (Real) vector space X (or* ***Linear Space****) is a set of elements on which we use the Field $I\!R$ to define two operations: addition and scalar multiplication. The important property of a Linear Space is that for any two elements $x,y\in X$ and real numbers $a,b\in I\!R$ we have that the vector $ax+by\in X$, where the vector ax is derived by the scalar multiplication between a and x, and the $+$ symbol stays for the addition law. The addition is defined as follows. For any three elements $x,y,z\in X$ and $a\in I\!R$ we have (i) $x+y=y+x$; (ii) $x+(y+z)=(x+y)+z$ and (iii) $a(x+y)=ax+ay$. So they are the usual low of algebra. The operations allowed between the scalars are the one we saw previously for a Field. Moreover, a vector space always contain the zero element θ, which is the invariant element of the sum, that is $x+\theta=x$. Finally, note that from the scalar multiplication we have that $0x=\theta$ and $1x=x$.*

Exercise 11 *(From Exercise 3.2 in SLP). Show that the following are vector spaces:*
(a) any finite-dimensional Euclidean space $I\!R^n$*;*
(b) the set of infinite sequences $(x_0,\ x_1,\ x_2,\ ...)$ *of real numbers;*
(c) the set of continuous functions on the interval $[0,1]$.
Is the set of integers $\{.....,\ -1,\ 0,\ 1,\ 2,\ ...\}$ *a vector space? Explain.*

Definition 3 *A normed linear space is a linear space* X*, together with a norm* $\|\cdot\| : X \to I\!R$*, such that for all* $x, y \in X$ *and* $a \in I\!R$ *we have: (i)* $\|x\| \geq 0$*, and* $\|x\| = 0$ *if and only if* $x = \theta$*; (ii)* $\|ax\| \leq |a| \cdot \|x\|$*; and* $\|x + y\| \leq \|x\| + \|y\|$.

Note that from a norm, we can always define a metric (or distance), as follows $d(x, y) = \|x - y\|$. It is important to notice that this is a particular case of distance function, for example such that $d(x, y) = d(x - y, \theta)$. Indeed, in Linear Spaces, the zero element is very important.

The set of real numbers can also be seen as a Normed Linear space. Notice that - since we analyze *real* vector spaces - we use IR also for the scalar multiplication. However, this last use is very much different from the first one, namely the use of IR as a particular set of vectors of a Linear Space (that is we set $X =$ IR in Definition 2). The absolute value can now be seen as a norm, and IR becomes a Normed Linear Space.

Exercise 12 *Solve Exercise 3.4 in SLP. Skip (a) and (f).*

2.2 The Infimum and the Supremum of a Set, the Completeness of IR

Once we have an ordered set we can ask: what is the largest element? Does it exist? Take for example the open interval $(0, 1)$ of real numbers. Common sense tells us that 0 should be the smallest element of the set and 1 should be the largest one. Moreover, we know that none of these two elements belongs to the set $(0, 1)$.

Definition 4 *Given a set* $S \subseteq I\!R$ *we say that* $\mu = \sup\ S$ *if (i)* $\mu \geq x$ *for each* $x \in S$ *and (ii) for every* $\varepsilon > 0$ *there exists an element* $y \in S$ *such that* $y > \mu - \varepsilon$*. The infimum is defined symmetrically as follows:* $\lambda = \inf S$ *if (i)* $\lambda \leq x$ *for each* $x \in S$*; and (ii) for every* $\varepsilon > 0$ *there exists an element* $y \in S$ *such that* $y < \lambda + \varepsilon$.

According to the above definition, in order to derive the supremum of a set we first should consider all its upper bounds

Definition 5 *A real number M is an upper bound for the set $S \subseteq \mathbb{R}$ if $x \leq M$ for all $x \in S$.*

We then choose the smallest of such upper bounds. Similarly, the infimum of a set is its largest lower bound. In the example above we decided to take the number 1 (and not 1.2 for example) as representative of the largest element of the set, perhaps precisely because it has such a desirable property.

Notice that we introduced the concepts of sup and inf for sets of real numbers. One reason for this is that we will consider these concepts only for sets of real numbers. Another important reason is that the set of real numbers satisfies a very important property that qualifies it as *Complete,* which basically guarantees that both sup and inf are well defined concepts in $\mathbb{R}$. Before defining this property, and to understand its importance, consider first the set of Rational Numbers $\mathbf{Q}$:

$$\mathbf{Q} = \left\{ q : \exists\, n, m \in \mathbf{Z} \text{ such that } q = \frac{n}{m} = n \cdot m^{-1} = n : m \right\}.$$

Where $\mathbf{Z}$ is the set of integers, that is the natural numbers $\mathbb{N}$ with sign ($+$ or $-$) and the zero. With the usual $\geq$ operator, we can see $\mathbf{Q}$ as an ordered field.[1] Now consider the following subset $B = \{b \in \mathbf{Q} : b \cdot b \leq 2\}$. It is easy to see that the supremum (sup) of this set is the square root of 2. Moreover, we all know that the square root of 2 does not belong to $\mathbf{Q}$. This is not a nice property. In fact, this problem induced mathematicians to look for a new set of numbers. The set of real numbers does not have this problem.

> **Property C**. *The Set of Real Numbers* $\mathbb{R}$ *has the Completeness property, that is, each set of real numbers which is bounded above has the least upper bound (*$\sup$*), and each set of real numbers which is bounded below has the greatest lower bound (*$\inf$*). Where, we say that a set S is bounded above when it has an upper bound; that is, there exists a $U \in \mathbb{R}$ such that $x \leq U$ for all $x \in S$. Bounded below, if $\exists\, L \in \mathbb{R} : x \geq L \;\forall\, x \in S$.*

We will take the previous statement as an axiom. In fact, the set of real numbers can be defined as an ordered field satisfying **Property C**. So, if we consider sets of real numbers, we can always be sure that each bounded set has inf and sup. Sometimes, when the set $S \subseteq \mathbb{R}$ is not bounded, we will use the conventions $\sup S = \infty$ or/and $\inf S = -\infty$.

[1] It is an useful exercise to check that it satisfies all the properties we introduced above for an Ordered Field.

2.3 Sequences: Convergence, liminf, limsup and the Cauchy Criterion

Definition 6 *A sequence from a set X is a mapping $f : I\!N \rightarrow X$, from the natural numbers to X. We will define each element $f(n) = x_n$, in turn, the whole mapping will be summarized as $\{x_n\}_{n \in I\!N}$.*

It is usual to denote it by $\{x_n\}_{n=0}^{\infty}$,[2] or, if does not create confusion, simply by $\{x_n\}$. Another important concept which is of common usage, and it is good to define formally is the one of subsequence.

Definition 7 *A sequence $\{y_k\}_{k \in I\!N}$ is a subsequence of the sequence $\{x_n\}_{n \in I\!N}$ if there exists a function $g : I\!N \rightarrow I\!N$ such that: (i) for every k we have $y_k = x_{g(k)}$, and (ii) for every n (index in the main sequence) there exists a N (index of the subsequence) such that $k \geq N$ implies $g(k) \geq n$.*

A typical notation for subsequences is $\{x_{n_k}\}$, where $n_k = g(k)$. Note that to define a subsequence we also need some sort of monotonicity property that has to be satisfied by the "renumbering" function g.

Definition 8 *A sequence of **real numbers** is said to be bounded if there exists a number M such that $|x_n| < M$ for all n.*

Definition 9 *Consider a metric space (X, d). We say that the sequence $\{x_n\}_{n=0}^{\infty}$ is convergent to $y \in X$ if for each real number $\varepsilon > 0$ there exists a natural number N such that for all $n \geq N$ we have $d(x_n, y) < \varepsilon$. And we write $x_n \rightarrow y$, or*

$$\lim_{n \rightarrow \infty} x_n = y.$$

Notice that - since the distance function maps into $I\!R_+$ - we can equivalently say that the sequence $\{x_n\}_{n=0}^{\infty}$ in the generic metric space (X, d) converges to y, if (and only if) the sequence $\{d(x_n, y)\}_{n=0}^{\infty}$ of real numbers, converges to 0 (thus, in the familiar one-dimensional space $I\!R$, with the usual absolute value as distance function). So, this concept of convergence is the most usual one, and it easy to check .... Yes, it is easy to check, once we have the candidate y. But suppose we do not know y (and actually we do not care

[2] In fact, it is probably more common to write it as $\{x_n\}_{n=1}^{\infty}$, that is starting the sequence with x_1 not with x_0. We will see below the reason for our notational choice.

about y), but we still would like to know whether a given sequence is converging.[3] Then it can be very difficult to make the guess. Long time ago, the mathematician Cauchy defined an important concept for sequences, which is somehow close to convergence.

Definition 10 *We say that a sequence $\{x_n\}_{n=0}^{\infty}$ of elements from a metric space (X, d) is Cauchy if for each $\varepsilon > 0$ there exists a natural number N such that for all $n, m \geq N$ we have $d(x_n, x_m) < \varepsilon$.*

Below we will study the relationship between this concept and convergence. Intuitively, the difference between a Cauchy and converging sequence is that in the former, as both n and m increase the values x_n and x_m get closer and closer to each other. In a converging sequence, both x_n and x_m get closer and closer to the converging point y. Notice that if both x_n and x_m get close to y they must get close to each other as well. So convergence is a stronger concept than Cauchy. You will be asked to formally show this statement in Exercise 18 below.

We can also formally define divergence as follows.

Definition 11 *We say that a sequence of* ***real numbers*** *diverges, or converges to $+\infty$, if for each real number M there exists a natural number N such that for all $n \geq N$ we have $x_n > M$. The definition of divergence to $-\infty$ is trivially symmetric to this one.*

Notice that, of course, it is **not** true that every sequence of real numbers either converges or diverges. Indeed, the points of convergence are very special ones. In particular, they are accumulation points.

Definition 12 *Given a metric space (X, d) and a set $S \subset X$ an element x is an accumulation point for the set S if for each $\varepsilon > 0$ there exists an element $y \in S$, $y \neq x$ such that $d(x, y) < \varepsilon$.*

Notice that an accumulation point does not necessary belong to the considered set. However, we have the following result.[4]

[3] For example, we will see that in infinite horizon models this is the main problem. We just want to know whether the objective function (an infinite sum) converges somewhere or not.

[4] Here is a useful result about accumulation points, which is not crucial for the following analysis.

Remark 1 *Each bounded set $E \subset \mathbb{R}$ which contains infinitely many real numbers has at least one accumulation point.*

Proposition 1 *Consider a set S in (X, d). If a point y is an accumulation point of S then there exists a sequence of $\{y_n\}$, with $y_n \in S$ for every n, that converges to y.*

Proof From Definition 12 we can always choose a sequence of $\varepsilon_n = \frac{1}{n}$ and we can be sure that there exists a $y_n \in S$ such that $d(y_n, y) < \varepsilon_n = \frac{1}{n}$. But since $\lim_{n\to\infty} \varepsilon_n = 0$, $\lim_{n\to\infty} y_n = y$ and we are done. **Q.E.D.**

Now a useful concept that may be completely new to someone.

Definition 13 *Given a bounded sequence $\{x_n\}$ of **real numbers** we say that $-\infty < s \equiv \limsup_{n\to\infty} x_n < +\infty$ if (i) For every $\varepsilon > 0$ there is a N such that for all $n \geq N$ we have $x_n \leq s + \varepsilon$; and (ii) For every $\varepsilon > 0$ and N there is an $n > N$, such that $x_n \geq s - \varepsilon$. The concept of* lim inf *is symmetric. We say $-\infty < l \equiv \liminf_{n\to\infty} x_n < +\infty$ if (i) For each $\varepsilon > 0$ there exists a N such that for every $n \geq N$ we have $x_n \geq l - \varepsilon$; and (ii) For each $\varepsilon > 0$ and N there exists an $n > N$, such that $x_n \leq l + \varepsilon$. We can write also*

$$
\begin{aligned}
s &= \inf_n \sup_{k\geq n} x_k \\
l &= \sup_n \inf_{k\geq n} x_k .
\end{aligned}
$$

And note that $s \geq l$ and that $\liminf x_n = -\limsup(-x)$.

Note that from the above definition both lim inf and lim sup must exist. For example, define $s = \inf_n y_n$ with $y_n = \sup_{k\geq n} x_k$. Since x_k are real numbers and the sequence is bounded, than by **Property C** each $y_n \in \mathbb{R}$ is well defined, an so is $s \in \mathbb{R}$. One can

---

Proof The proof is by construction. If $E \subset \mathbb{R}$ is bounded, then there exists an interval $[a, b]$ with $a, b \in \mathbb{R}$ such that $E \subset [a, b]$. First assume w.l.o.g. that $b > a$. We are going to split the interval in two and we will chose the one of the two semi-intervals that contain infinitely many points. Since the set contains infinitely many points this can always be done. So we have $\left[a, \frac{b-a}{2}\right]$ and $\left[\frac{b-a}{2}, b\right]$. We call the one we chose $[a_1, b_1]$. Notice that we can do this division infinitely many times. In this way, we generate at the same time two sequences and two sets of real numbers $A = \{a_n\}$ and $B = \{b_n\}$ such that $\sup A = \inf B = \xi$, note indeed that the way we constructed the sequence of real numbers,

$$0 \leq \inf B - \sup A \leq \frac{b-a}{2^n} \text{ for each } n.$$

Moreover it is easy to see that the real number ξ exists and has to belong to each of the intervals $[a_n, b_n]$, indeed if this were not true than $\sup A$ could not coincide with $\inf B$. Thus we have that for each $\varepsilon > 0$ we can construct an interval $I_\varepsilon(\xi)$ centered in ξ and such that $|x - \xi| < \varepsilon$ for all $x \in I_\varepsilon(\xi)$ and such that it contains infinitely many points different from ξ and belonging to E (so, we can always find at least one) and this proves that ξ is an accumulation point of E. **Q.E.D.**

actually show that $\liminf x_n$ is the smallest accumulation point of $\{x_n\}$ and $\limsup x_n$ is the greatest.

Consider now the lim sup concept in Definition 13. As a first approximation we can think it as the sup of the tail of the sequence. This is, of course not a well defined concept, since we do not know exactly what "the tail of the sequence" is yet. However, note that the number $s = \limsup$ has all the characteristics of a sup . Indeed, from (i) it is not possible to have values above it in the tail, and from (ii) for every $\varepsilon > 0$ we can find one element in the tail such that it is greater than $s - \varepsilon$. That is, we can show that the first definition implies the second. More formally, we will show that if s satisfies (i) and (ii), then $s = \inf_n \sup_{k \geq n} x_k$. First note that (i) implies $\forall \varepsilon > 0 \; \exists N : y_N = \sup_{k \geq N} \; \{x_k\} < s + \varepsilon$, so $\inf_n \; y_n \leq y_N \leq s + \varepsilon$, but since ε is arbitrary, we can say that $\inf_n \; y_n \leq s$. Now from (ii) we have that $\forall \varepsilon > 0$ and $\forall N$ we have $y_N = \sup_{k \geq N} \; \{x_k\} > s - \varepsilon$, which further implies (since it is for any N) that actually $\inf_n \; y_n \geq s - \varepsilon$ (to see it suppose that it is not true and you will have a contradiction). Again, since the choice of ε was arbitrary, we have that $\inf_n \; y_n \geq s$ which, together with the previous result, gives $\inf_n \sup_{k \geq n} x_k = \inf_n \; y_n = s \equiv \limsup_{n \to \infty} x_k$.

Since the two concept coincide one should be able to show that also converse is true.

Exercise 13 *Consider a bounded sequence $\{x_n\}$ of real numbers. Show that if a real number is such that $s = \inf_n \sup_{k \geq n} x_k$ then it has both the properties (i) and (ii) stated for s at the beginning of Definition 13.*

Definition 14 *A point y is an accumulation point (or a limit point, or a cluster point) of a sequence $\{x_n\}$ from the metric space (X, d), if for every real number $\varepsilon > 0$ and natural number N there exists some $k > N$ such that $d(x_k, y) < \varepsilon$.*

Exercise 14 *State formally the relationship between the concept of an accumulation point in Definition 14 for sequences of real numbers and that in Definition 12 for sets.*

Exercise 15 *Show that a sequence $\{x_n\}$ of real numbers has limit L if and only if $\liminf x_n = \limsup x_n = L$.*

Thus, when a sequence converges to a point then the limit is the **unique** accumulation point for the sequence.

Theorem 1 *If a sequence $\{x_n\}$ has limit L, then every possible subsequence $\{x_{n_k}\}$ of $\{x_n\}$ has limit L.*

Proof Straightforward from the definition of limit and subsequence. Indeed (sub)sequences can converge only to some accumulation point of the original sequence. **Q.E.D.**

Proposition 2 *Given a bounded sequence $\{x_n\}$ of* ***real numbers,*** *it is always possible to take a converging subsequence $\{x_{n_k}\}$ such that*

$$\lim_{n_k \to \infty} x_{n_k} = \limsup_{n \to \infty} x_n$$

Proof This is a particular case of accumulation point introduced in Definition 14. From Definition 13 we are sure that for every $\varepsilon > 0$ and $N > 0$ we can take an x_{n_k} such that $L - \varepsilon \leq x_{n_k} \leq L + \varepsilon$ (part (i) of Definition 13 is actually a stronger statement). Moreover, one can easily check that the sequence can be chosen to satisfy the monotone property for sequences requested in Definition 7. **Q.E.D.**

Exercise 16 *Following the lines of the previous proof, show Proposition 2 for the* $\liminf$.

Corollary 1 *From a bounded sequence $\{x_n\}$ of* ***real numbers,*** *it is always possible to extract a converging sequence.*

Proof This result was originally due to Weierstrass. From Proposition 2, we can always take a converging subsequence which converges to $\limsup$. **Q.E.D.**

Theorem 2 *Given a sequence $\{x_n\}$ of* ***real numbers*** *and $y \in I\!R$. If from any subsequence of $\{x_n\}$ it is possible to extract a converging sub-subsequence which converges to the same y then $\lim_{n\to\infty} x_n = y$.*

Proof From Proposition 2 (and Exercise 16) we know that among all the possible subsequences from $\{x_n\}$ there will be in particular a subsequence that converges to $\limsup \{x_n\}$ and another converging to $\liminf \{x_n\}$, but then $\liminf \{x_n\} = \limsup \{x_n\}$ and we are done by Exercise 15. **Q.E.D.**

Quick Review about Series

Definition 15 *A series is a couple of sequences $\{x_n, s_n\}_{n=0}^{\infty}$ where s_n are called the partial sums and are such that*

$$s_n = \sum_{t=0}^{n} x_t.$$

Sometimes, abusing in notation, a series is written $\sum_{t=0}^{\infty} x_t$ or even simply $\sum_t x_t$.

Definition 16 *A series is said to be convergent if exists finite the limit of the sequence of partial sums*

$$s = \lim_{n \to \infty} s_n < \infty,$$

and the number s is said to be the ***sum*** *of the series and we can write $s = \sum_{t=0}^{\infty} x_t$.*

Note that, in principle the above definition allows us to write the infinite sum only if the series converges. Actually a series can be of three *characters.* (i) Convergent, (ii) Divergent or (iii) Indeterminate. A series is *divergent* when the sequence of partial sums goes either to minus or to plus infinity.

2.4 Closed and Compact Sets: Maxima and Minima

We are now ready to define closed and compact sets. There are many concepts of closedness and compactness. Here we will use sequences to define them. So, to be precise we will speak about sequential closedness and sequential compactness, but for the spaces we will be interested in, these concepts will all coincide.

Definition 17 *A set S is closed if for each convergent sequence of elements in S, the point of convergence is in S, i.e.*

$$x_n \to y \text{ and } x_n \in S \; \forall n, \text{ implies } y \in S.$$

A set is open if its complement is closed. Where, in $\mathbb{R}$, the complement of S is defined $S^c \equiv \{x \in \mathbb{R} : x \notin S\}$.

Definition 18 *A set S is (sequentially) compact if for each sequence of elements in S we can take a subsequence which converges in S.*

Exercise 17 *Show that a set S is closed if and only if it contains all its accumulation points.*

Theorem 3 *A set $F \subset \mathbb{R}$ is (sequentially) compact if and only if it is closed and bounded.*

Proof ($\Rightarrow$) Assume F is compact. Then first it has to be limited otherwise it is possible to show that there must exist a subsequence going to either $+\infty$ or $-\infty$ (example). In this last case, by the definition of sequence converging to (say) $+\infty$ it will be impossible to find a sub-subsequence converging to any real number, since from Theorem 1 all of them will converge to $+\infty$, and this would contradict the assumption of compactness. Now we show that F has to be closed by showing that F contains all its accumulation points. So consider an accumulation point y. From Proposition 1 we know that there must be a sequence $\{y_n\}$ converging to y. Moreover, this sequence is such that $y_n \in F$ for every n. Now we know that by the definition of compactness it is possible to extract from $\{y_n\}$ a subsequence converging to a point $y' \in F$, but from Theorem 1 we know that each sub-sequence must converge to y, so $y' = y \in F$ and we are done.

($\Leftarrow$) Now assume F is closed and bounded, and consider a generic sequence $\{x_n\}$ with $x_n \in F$ for all n. Since F is bounded, this sequence must be bounded. Hence from Proposition 2 it is possible to extract a converging subsequence $\{x_{n_k}\}$. Moreover, since F is closed and $x_{n_k} \in F$ $\forall k$, the converging point must belong to F, and we are done. **Q.E.D.**

Definition 19 *An element is a maximum for S if $x = \sup S$ and $x \in S$. Similarly, y is the minimum of S if $y = \inf S$ and $y \in S$.*

Theorem 4 *Each closed and bounded set $F \subset \mathbb{R}$ has a Max and a Min.*

Proof Since it is bounded, it has $\sup F < \infty$. By the definition of $\sup F$, it is an accumulation point. Thus, by Proposition 1 we can always construct a sequence with elements in F that converges to $\sup F$. By the closedness of F the limit of this sequence must belong to F. **Q.E.D.**

2.5 Complete Metric Spaces and The Contraction Mapping Theorem

Complete Metric Spaces: $\mathbb{R}$ as a Complete Metric Space Recall that when we introduced the concept of a Cauchy sequence and we related it to the concept of a convergent sequence we also said that the latter is stronger then former. In the exercise below you are asked to show it formally.

Exercise 18 *Consider a generic metric space (X, d). Show that each convergent sequence is Cauchy.*

In fact, convergence is a strictly stronger concept only in some metric spaces. Metric spaces where the two concepts coincide are said to be *Complete*.

Definition 20 *A metric space* (X, d) *is said to be Complete if any Cauchy Sequence is convergent in* X*.*

The concept of completeness here seems very different from the one we saw in Section 2.2. However, one can show that there is a very close relationship between the two concepts of completeness.

Exercise 19 *Show that in the metric space* $(\mathbb{R}, |\cdot|)$ *every Cauchy sequence is bounded.*

Theorem 5 $(\mathbb{R}, |\cdot|)$ *is a Complete metric space.*

ProofConsider a generic Cauchy sequence $\{x_n\}$ in $\mathbb{R}$. From Exercise 19, we know it is bounded. Then from Corollary 1 of Proposition 2 there must be a subsequence converging to $y = \limsup x_n$. Using the triangular inequality, it is not difficult to see that if a Cauchy sequence has a converging subsequence, it is convergent. Hence we are done. **Q.E.D.**

To have an example of a non complete metric space consider again the set of real numbers, with the following metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ \max\left\{\frac{1}{1+|x|}, \frac{1}{1+|y|}\right\} & \text{otherwise.} \end{cases}$$

[Check that d is actually a metric!]. Now consider the sequence $\{x_n\}_{n=0}^{\infty}$ of integers $x_0 = 0,\ x_1 = 1,\ x_2 = 2,\ x_3 = 3,\ ...\ x_n = n,\ ...$. It is easy to see that as m and n increase the distance $d(x_n, x_m) = d(n, m)$ goes to zero.[5] Hence the sequence is Cauchy. However, it is easy to see that the sequence $\{x_n\}_{n=0}^{\infty} = \{n\}_{n=0}^{\infty}$ does not converge to any real number x, since for any fixed $x < \infty$ we have $d(x, n) \geq \frac{1}{1+|x|} > 0$.

Another possibility is to keep the metric of the absolute value and change X. Assume for example $X = \mathbf{Q}$, the set of rational numbers. Consider now a sequence of rational numbers that converges to $\sqrt{2}$. It is clear that this sequence would satisfy the Cauchy criterion, but would not converge in $\mathbf{Q}$ by construction. Each time we guess a $q \in \mathbf{Q}$ at which the sequence converges, since $q \neq \sqrt{2}$, we must have $\left|q - \sqrt{2}\right| = \varepsilon > 0$ for some $\varepsilon > 0$, hence we can find a contradiction each time the elements of the sequence are sufficiently close to $\sqrt{2}$.

Exercise 20 *Show that if* (X, d) *is a complete metric space and* S *is a closed subset of* X*, then* (S, d) *is still a complete metric space.*

[5] If $n, m \geq N$, then $d(n, m) \leq \frac{1}{1+N}$.

Banach Space (Complete Normed Linear Space): $\mathbb{R}$ and C^1 as Banach Spaces

Definition 21 *A Complete Normed Linear Space is called Banach Space.*

Directly, from the definition is clear that $\mathbb{R}$ is a Banach space.

Theorem 6 *Let $X \subset \mathbb{R}$ and $\mathcal{C}(X)$ the set of bounded and continuous*[6] *real valued functions $f : X \to \mathbb{R}$, together with the "sup" norm $\|\cdot\|_\infty : \mathcal{C}(X) \to \mathbb{R}$ such that $\|f\|_\infty \equiv \sup_t |f(t)|$ is a Banach Space.*

Proof From Exercise 12 we know that this space is a normed linear space. We want to show that it is complete. Here is a sketch of the proof based on Theorem 3.1 in SLP (Page.47-49). One has to show that for any Cauchy sequence $\{f_n\}$ of functions there exists a limit function f^* such that $f_n \to f^*$, and $f^* \in \mathcal{C}(X)$. Our candidate f^* is defined as follows: $f^*(x) = \lim_n f_n(x)$ for any given x. Notice that f^* is well defined since for any given x, the sequence of real numbers $y_n = f_n(x)$ is Cauchy and, from the completeness of $\mathbb{R}$ it must converge. This type of convergence is called pointwise.

We have to first show that $\{f_n\}$ converge to our candidate in sup norm, or uniformly. Second, that f^* is bounded and continuous (we just show that f^* is real valued). The first part requires that for any $\varepsilon > 0$ exists a N_ε such that for all $n \geq N_\varepsilon$ we have

$$\sup_x |f_n(x) - f^*(x)| \leq \varepsilon.$$

Notice that if N_ε is such that for all $n, m \geq N_\varepsilon$ we have $\|f_n - f_m\| \leq \frac{\varepsilon}{2}$, then for any given x it must be that for $n \geq N_\varepsilon$

$$\begin{aligned} |f_n(x) - f^*(x)| &\leq |f_n(x) - f_{m_x}(x)| + |f_{m_x}(x) - f^*(x)| \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where $m_x \geq N_\varepsilon$ is possibly a different number for each x, but m_x must exist since we saw that for all x $f_m(x) \to f^*(x)$. For each x, there is a function $f_{m_x}(\cdot)$ that can be used as pivotal to show that f_n and f^* are close to each other at the point x. This function never appears in the left hand side however. Hence we are done. To show that f^* is bounded and continuous, let x be given and consider an arbitrary y in an proper neighborhood of

[6] Continuity is intended with respect to the topology induced by the Euclidean norm in X.

x. If n is such that $\|f^* - f_n\| < \frac{\varepsilon}{3}$, we have

$$\begin{aligned} |f^*(x) - f^*(y)| &\leq |f^*(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f^*(y) - f_n(y)| \\ &\leq 2\|f^* - f_n\| + |f_n(x) - f_n(y)| \\ &< 2\frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

This is true for all y belonging to the appropriate neighborhood of x. The second term in the second line can then be chosen small enough since f_n is continuous. **Q.E.D.**

The sup norm defines in the obvious way the sup distance function, which will be sometimes denoted by d_∞.

The Contraction Mapping Theorem

Definition 22 *Let (X, d) a metric space and $T : X \to X$ a function mapping X into itself. T is a Contraction (with modulus β) if for some $\beta < 1$ we have*

$$d(Tx, Ty) \leq \beta d(x, y), \text{ for all } x, y \in X.$$

To understand the idea, make a picture of a contraction from $\mathbb{R}_+$ to $\mathbb{R}_+$. For example draw a line with slope $\beta \in [0, 1)$ and nonnegative intercept.

Theorem 7 *If (X, d) is a complete metric space and $T : X \to X$ is a contraction with modulus β, then (i) T has exactly one fixed point x^* in X, i.e. $x^* = Tx^*$, and (ii) for all $x_0 \in X$ we have $d(T^n x_0, T^n x^*) \leq \beta^n d(x_0, x^*)$, $n = 0, 1, 2, ...$*

Proof The proof goes as follows. Start with a generic $x_0 \in X$ and construct a sequence as follows: $x_n = T^n x_0$. Since X is complete, to show existence in (i), it suffice to note that the so generated sequence is Cauchy. Let's show it.

Notice first that by repeatedly applying our map T one gets, $d(x_{n+1}, x_n) \leq \beta^n d(x_1, x_0)$. Consider now, w.l.o.g., $m = n + p + 1$; we have

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + ... + d(x_{n+1}, x_n) \\ &\leq \beta^p d(x_{n+1}, x_n) + \beta^{p-1} d(x_{n+1}, x_n) + ... + d(x_{n+1}, x_n) \\ &\leq \frac{1}{1-\beta} d(x_{n+1}, x_n) \leq \frac{\beta^n}{1-\beta} d(x_1, x_0) \end{aligned}$$

where, for the first inequality we used property (iii) in Definition 1, for the second and fourth we used the property of our sequence. The third inequality is trivial, since $d(x_{n+1}, x_n) \geq 0$ and $\beta < 1$.

As a consequence, for each $\varepsilon > 0$, we can choose an index N large enough and have $\frac{\beta^N}{1-\beta} d(x_1, x_0) < \varepsilon$ for all $n, m \geq N$ as it is required by the definition of Cauchy sequence. Since (X, d) is complete this sequence must converge, that is, there must exist a x^* such that

$$\lim_{n\to\infty} x_n = x^*.$$

By the continuity of T (in fact T is uniformly continuous) we have that the limit point of the sequence is our fixe fixed point of T[7]

$$Tx^* = T \lim_{n\to\infty} x_n = \lim_{n\to\infty} Tx_n = \lim_{n\to\infty} x_{n+1} = x^*.$$

It remains to show that the limit of the sequence $x^* = Tx^*$ is unique. Suppose the contrary, and call $x^{**} = Tx^{**}$ the second fixed point. Note that

$$d\left(x^{**}, x^*\right) = d\left(Tx^{**}, Tx^*\right) \leq \beta d\left(x^{**}, x^*\right),$$

which is a contradiction as long as $d\left(x^{**}, x^*\right) > 0$, hence we must have $d\left(x^{**}, x^*\right) = 0$, that is $x^*$and x^{**} must in fact be the same point. The (ii) part of the proof is simple and left as an exercise, see also Theorem 3.2 in SLP (page 50-52). **Q.E.D.**

Exercise 21 *Notice that we allow for $\beta = 0$. Draw a mapping $T : \mathbb{R}_+ \to \mathbb{R}_+$ which is a contraction of modulus $\beta = 0$ and show graphically that T must have a unique fixed point. Now formally show the statement, that is, show that if T is a contraction of modulus $\beta = 0$ in a complete metric space then T admits a unique fixed point. Is the completeness assumption on (X, d) crucial in this case?*

The following result is due to Blackwell and provides a couple of easy-to-verify conditions for a contraction.

Theorem 8 *Let $X \subset \mathbb{R}^l$, and let $\mathbf{B}(X)$ the space of bounded functions $f : X \to \mathbb{R}$, with the sup norm. Let $T : \mathbf{B}(X) \to \mathbf{B}(X)$ be an operator satisfying: (i) $f, g \in \mathbf{B}(X)$ and $f(x) \leq g(x)\ \forall x \in X$, implies $(Tf)(x) \leq (Tg)(x)\ \forall x \in X$, and (ii) there exists some $0 \leq \beta < 1$ such that*

$$[T(f + a)](x) \leq (Tf)(x) + \beta a,\ \forall f \in \mathbf{B}(X),\ a \geq 0,\ x \in X.$$

Then T is a contraction with modulus β.

[7] See also Exercise 25.

Proof Notice that for any two functions f and g in $\mathbf{B}(X)$ the sup norm $d_\infty(f,g) \equiv \sup|f(x) - g(x)|$ implies that $g(x) \geq f(x) - d_\infty(f,g)$ for all x; or $g + d_\infty(f,g) \geq f$. From the monotonicity of the operator T (property (i)) we have $T[g + d_\infty(f,g)](x) \geq Tf(x)$ for all x. Now using property (ii) (discounting) with $a \equiv d_\infty(f,g) \geq 0$, we have $Tg + \beta d_\infty(f,g) \geq Tf$ or $\beta d_\infty(f,g) \geq (Tf)(x) - (Tg)(x)$ for all x. But this implies $d_\infty(Tf,Tg) \leq \beta d_\infty(f,g)$, hence T is a contraction with modulus β. **Q.E.D.**

Assumptions (i) and (ii) in the above theorem are usually called *Blackwell sufficient conditions* for a contraction.

Exercise 22 *(a) Show that the Bellman operator of the Optimal Growth Model satisfies Blackwell's sufficient conditions if u is a bounded function. (b) One of the most commonly used utility functions in growth theory is the CRRA utility $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$; $\sigma \geq 0$. We know that in this case u is not a bounded function. Could you suggest a way of showing that Bellman operator of the optimal growth model with no technological improvement is a contraction when $\sigma < 1$? [Hint: Use the Inada conditions.]*

2.6 Continuity and the Maximum Theorem

Continuity and Uniform Continuity: The Weierstrass's Theorem We start by reviewing one of the most basic topological concepts: Continuity.

Definition 23 *A real valued function $f : X \to I\!R$ is continuous at $x \in X$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$d(x,y) < \delta \text{ implies } f(x) - \varepsilon < f(y) < f(x) + \varepsilon.$$

A function is continuous in a set $S \subset X$ if it is continuous at every point $x \in S$.

Definition 24 *A real value function is said to be **uniformly** continuous in $S \subset X$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for any two points $x, y \in S$ such that $d(x,y) < \delta$ we have $f(x) - \varepsilon < f(y) < f(x) + \varepsilon$.*

Notice that uniform continuity is stronger than continuity. Indeed, for any $\varepsilon > 0$ uniform continuity requires to find a δ, which is the same for each point $x \in X$. While the usual concept of continuity allows you to choose a different δ for each x.

Exercise 23 *Show that if f is uniformly continuous in E, then f is also continuous in E.*

Exercise 24 *Show that a continuous function on a compact set is uniformly continuous. [A bit difficult].*

Exercise 25 *Show that a real valued function* $f : I\!R \to I\!R$*, is continuous in* x *if and only if for every sequence* $\{x_n\}$ *converging to* x *we have that the implied sequence* $\{y_n = f(x_n)\}$ *converges to* $y = f(x)$.

We are now ready to show one of the most well known theorems in Real analysis.

Theorem 9 *(Weierstrass) A continuous real valued function* f *defined on a compact set* S *has maximum and minimum.*

Proof The simplest way to show this result is using sequences. We already saw that the supremum of a set of real numbers is an accumulation point. As a consequence, Proposition 1 guarantees that there exists a sequence y_n that converges to $y^* = \sup_x f(x)$. Since f is continuous, Exercise 25 implies that the induced sequence x_n such that $y_n = f(x_n)$ must also converge (say to the point x^*), hence $y^* = f(x^*)$. By the compactness of S, x^* must belong to it, so x^* is the maximum. **Q.E.D.**

Exercise 26 *Show that a continuous function on a compact set is bounded, i.e.* $\sup_x f(x) < \infty$.

Exercise 27 *(**Brower Fixed Point**) Show that a continuous function defined on the compact set* $[0, 1]$ *and mapping values into* $[0, 1]$ *has a fixed point, that is, a point* x^* *such that* $f(x^*) = x^*$.

Correspondences: Some Basic Concepts Correspondences are more complicated concepts than functions but the idea is similar.

Definition 25 *A map* $\Gamma : X \to Y$ *is said a correspondence if for any* $x \in X$ *assigns a set* $\Gamma(x) \subset Y$.

Sometimes many concepts are easy to understand if we consider the case where it fails to be satisfied.

Definition 26 *A non empty correspondence* Γ *is said to be **not** lower hemi-continuous (**not l.h.c.**) at* x *if for at least one converging sequence* $x_n \to x$*, it is* not *possible to reach a point* $y \in \Gamma(x)$ *with a converging sequence of points such that* $y_n \in \Gamma(x_n)$.

Definition 27 *A non empty and compact-valued correspondence* Γ *is said to be* ***not*** *upper hemi-continuous (**not u.h.c.**) at* x *if for at least one converging sequence* $x_n \to x$, *it is possible to find a converging (sub)sequence* $y_k \in \Gamma(x_{n_k})$, *(with* x_{n_k} *a subsequence of* x_n*), whose limit point* y*is such that* $y \notin \Gamma(x)$.

We can have an idea of these concepts by drawing graphs. If we try to visualize when one of the two concepts fail we will immediately understand that l.h.c. does not allows for "discontinuities" that appear as "explosions" in the set of points, whereas the u.h.c. does not allows for "discontinuities" that appear as "implosions".[8]

Definition 28 *A Correspondence* Γ *which is both u.h.c. and l.h.c. at any* x *is said to be continuous.*

Note that a single valued correspondence is actually a function.

Exercise 28 *Show that a single valued correspondence is l.h.c. if and only if it is u.h.c.*

Exercise 29 *Is a continuous correspondence also a continuous function ? Is a continuous function also a continuous correspondence ? Try to justify formally your answer.*

Exercise 30 *Show the following useful result. Let* f_i, g_i $i = 1, 2, ...N$ *be continuous real valued functions such that* $f_i \geq g_i$*for all* i. *Define* $\Gamma(x) = \{y \in I\!R^N : \; g_i(x) \leq y_i \leq f_i(x),$ $i = 1, 2, ...N\}$. *Then* Γ*is a continuous correspondence.*

From Exercise 25 we saw that an elementary way of defining continuity of f at a point is to guarantee that if $x_n \to x$ then $f(x_n) \to f(x)$. Here below we provide some generalizations based on this definition:

Definition 29 *A function* $f : X \to I\!R$ *is upper (lower) semi-continuous at* x *if for all converging sequences* $x_n \to x$, $\limsup_{n\to\infty} f(x_n) \leq f(x)$ $(\liminf f(x_n) \geq f(x))$.

There is a way of seeing continuity of a correspondence very similar to that for functions, as specified in the above definition. We just need to define the appropriate extension of Definition 13 for sets.

[8] In general, since the correspondence is (more or less implicitly) assumed to map into a compact Hausdorff space, u.h.c. is equivalent to the *graph* of Γ is closed, where the graph of a correspondence is defined as $Gr(\Gamma) = \{(x, y) \in X \times Y : \; y \in \Gamma(x)\}$ (see Theorem 14.12 in Aliprantis et al., 1994). While l.h.c. is *implied* by the fact that $Gr(\Gamma)$ is an open set (but the converse is not necessarily true: for example, a continuous correspondence is obviously l.h.c. but its graph is closed).

Definition 30 *Let* $\{A_n\}_{n=1}^{\infty}$ *a sequence of sets in* $I\!R^n$. *We say that* $x \in \liminf A_n$ *if every neighborhood* I_x *of* x *intersects all* A_n *for a* n *sufficiently large, i.e. for each* I_x *there is a* N *such that for all* $n \geq N$ $I_x \cap A_n \neq \varnothing$. *We say* $x \in \limsup A_n$ *if every neighborhood of* x *intersects infinitely many* A_n. *Clearly* $\liminf A_n \subset \limsup A_n$. *We say that* $A_n \rightarrow A$ *or* $\lim A_n = A$ *if limsup and liminf are the same set.*

In words, the set $\liminf \Gamma(x_n)$ is the set of all possible limit points of sequences $\{y_n\}_n$ such that $y_n \in \Gamma(x_n)$ for all n, while $\limsup \Gamma(x_n)$ is the set of all cluster points of such sequences.

Definition 31 *Let* $\Gamma(\cdot)$ *be a correspondence which maps points* X *into subsets of* $I\!R^n$, *and let* $\{x_n\}_n$ *be a sequence converging to* x. *We say that* Γ *is* ***u.h.c.*** *(resp.* ***l.h.c.*** *) [resp. continuous] if* $\Gamma(x) \supset \limsup \Gamma(x_n)$ *(resp.* $\Gamma(x) \subset \liminf \Gamma(x_n)$ *), [resp.* $\Gamma(x) = \lim \Gamma(x_n)$ *].*

Exercise 31 *Show that when we consider compact valued correspondences the last definitions and those in Definitions 26 and 27 are equivalent.*

The definition of u.h.c. in Definition 27 is hence more restrictive than that in Definition 2.6 in that the later does not require to work on compact spaces. In contrast, the correspondence between the two concepts of l.h.c. is perfect. Note however, that it can be the case that a correspondence fails to be u.h.c. at a point x according to Defintion 27 while $\Gamma(x) = \lim \Gamma(x_n)$. That is, strictly speaking, we should have used a qualification such as *'weak'* for u.h.c. and continuity properties in Definition . Since in the theorems below we will assume compact valued correspondences this distinction will not matter.

Exercise 32 *Let* $f : X \times Z$, *and define the graph of* f *given* z *as*

$$Grf(z) = \{(x, y) \in X \times Y : -B \leq y \leq f(x, z)\},$$

where $B < \infty$ *guarantees to have a compact valued correspondence for each given* z. *Show that if* $Grf(z)$ *is a continuous correspondence then* f *is jointly continuous in* (x, z). *What are the properties of* f *if* $Grf(z)$ *is a upper (lower) hemi-continuous correspondence?*

The Maximum Theorem This is not a simple result, but it is one of the most important ones.

Theorem 10 *Let $X \subset \mathbb{R}^l$ and $Y \subset \mathbb{R}^m$, let $f : X \times Y \to \mathbb{R}$ be a continuous function, and let $\Gamma : X \twoheadrightarrow Y$ be a compact valued and continuous correspondence. Then (i) the function $h : X \to \mathbb{R}$ defined as*

$$h(x) = \max_{y \in \Gamma(x)} f(x, y)$$

is continuous, and the policy correspondence $G : X \to Y$ defined as

$$G(x) = \{y \in \Gamma(x) : f(x, y) = h(x)\}$$

is (ii) nonempty, (iii) compact valued and (iv) upper hemi-continuous.

Proof (sketch): (i) Recall the 'implosions and explosions interpretation' for correspondences. It f is continuous it cannot have jumps. So the maximum value h can have jumps (i.e., it can be discontinuous) only if there are some implosions or explosions in the feasible set. In particular, with implosions we can have a sharp reduction of the sup, with explosions h can jump upward. The continuity of Γ guarantees that there are no such implosions and explosions in the feasible set, hence h must vary continuously with x.

Now, from the Weierstrass Theorem a maximum exists for any x, so (ii) is immediate. (iii) To show that G is compact valued, note first that $G(x) \subset \Gamma(x)$ hence G is bounded. To see that G maps closed sets for all x, note that for any convergent sequence y_n with $y_n \in G(x)$ it must be that $f(x, y_n) = h(x)$ so from the continuity of f and Exercise 25, we have that

$$\lim_{n\to\infty} f(x, y_n) = f(x, \lim_{n\to\infty} y_n) = f(x, y) = h(x),$$

where the last equality comes for the fact that $f(x, y_n) = h(x)$ for all n, which implies $\lim_{n\to\infty} f(x, y_n) = h(x)$. Hence $y \in G(x)$ and, according to Definition 17, we have shown that $G(x)$ is closed for all x.

(iv) We now show that the policy correspondence cannot fail to be u.h.c.. Recall Definition 27. In order for a correspondence not to be u.h.c. at a point x one must be able to find two converging sequences $x_n \to x$ and $y_n \in G(x_n) \to y$ such that $y \notin G(x)$. However, if we recall the definition of the policy $G(x) = \{y \in \Gamma(x) : f(x, y) = h(x)\}$, it is easy too see that as long as Γ is continuous this failure surely cannot happen because of $y \notin \Gamma(x)$. Moreover, because of continuity of both f and h for any couple of such converging sequences we must have that

$$f(x, y) = f(\lim x_n, \lim y_n) = \lim_n f(x_n, y_n) = \lim_n h(x_n) = h\left(\lim_n x_n\right) = h(x)$$

where in the first equality we used the definition of x an y as limit points, in the second we used the joint continuity of f; in the third we used the fact that $y_n \in G(x_n)$ for all n, and in the penultimate equality we used again the continuity of h. Hence, we have just seen that this cannot happen at all.[9] **Q.E.D.**

Notice that the policy correspondence is "only" upper hemi-continuous. Hence we can have explosions in the set that describes the optimal points, even tough the feasibility set cannot.

Exercise 33 *How can we have explosions in the policy correspondence if both f and Γ are continuous ?*

Exercise 34 *Solve exercise 3.16 of SLP.*

Theorem 10 is due to C. Berge (1959) and can be somehow extended as follows:

Theorem 11 *Let $\Gamma : X \twoheadrightarrow Y$ a u.h.c. correspondence with non-empty and compact values. And let $f : X \to \mathbb{R}$ be upper semi-continuous. Then the "value function" $h : X \to \mathbb{R}$ defined by*

$$h(x) = \max_{y \in \Gamma(x)} f(x, y)$$

is upper semi-continuous.

Proof: See Lemma 14.29, page 773 in Aliprantis et al. (1994). **Q.E.D.**

When f is u.s.c. and X is compact, a direct extension of the Weierstrass theorem guarantees the existence of a maximal point of f in X. As expected, if we allow both the feasibility set and the functions to jump upward, the value function may jump upward as well. Notice that downward jumps are not always preserved by the max operator. These are preserved by the min. Indeed, the same theorem guarantees that whenever $-f$ is upper semi-continuous (i.e., f is lower semi-continuous) then h is lower semi-continuous since $-h = -\min f$ must be upper semi-continuous. Do not get confused, Γ is always required to be u.h.c. [Can you explain intuitively why?].

[9] Of course, the tricky part is to show that $G(x)$ is u.h.c. and that h is continuous simultaneously. In the proof of Theorem 3.6, SLP show that G is u.h.c. only using the continuity of f.

Bibliography

[1] Aliprantis, C. D. and K.C. Border (1994) *Infinite Dimensional Analysis*, Springer-Verlag, Berlin.

[2] Berge, C. (1959) *Espaces Topologiques, Fonctions Multivoques*, Dunop, Paris. English translation by E. M. Patterson (1963), *Topological Spaces,* Macmillan, New York.

[3] Lipschutz, S. (1994) *General Topology,* McGraw Hill Inc.

[4] Rudin, W. (1964) *Principles of Mathematical Analysis*, second edition, McGraw-Hill, New York.

[5] Rudin, W. (1991) Functional *Analysis*, McGraw Hill, New York.

[6] Sargent, T. (1987) *Dynamic Macroeconomic Theory*, Harward University Press.

[7] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

Chapter 3

Deterministic Dynamic Programming

3.1 The Bellman Principle of Optimality

Richard Bellman (1957) states his Principle of Optimality in full generality as follows:

> *"An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."* Bellman (1957), page 83.

An *optimal policy* is a rule making decisions which yield an allowable sequence of decisions and maximize a preassigned function of the final state variables.

The Bellman Principle of Optimality (BPO) is essentially based on the following property of the real valued functions.

Lemma 1 *Let f be a real valued function defined on $X \times Y$, then*

$$V^* = \sup_{(x,y) \in X \times Y} f(x,y) = \sup_{x \in X} \left\{ \sup_{y \in Y} f(x,y) \right\},$$

that is, if for any $x \in X$ we define the value function

$$W(x) = \sup_{y \in Y} \; f(x,y)$$

then

$$V^* = \sup_{x \in X} \; W(x).$$

Proof. First notice that if we fix x, we have $V^* \geq f(y, x)$ for all y, hence $V^* \geq W(x)$ for all x. As a consequence

$$V^* \geq \sup_{x \in X} W(x).$$

Now we need to show the inverse inequality. We can have two cases. *(a)* $V^* < \infty$. In this case, by the definition of sup in V^* we have that for every $\varepsilon > 0$ there exists a couple $(x', y') \in X \times Y$ such that $f(x', y') + \varepsilon > V^*$. In addition, we know that $W(x') = \sup_y f(x', y) \geq f(x', y')$, hence

$$\sup_x W(x) + \varepsilon \geq W(x') + \varepsilon > V^*.$$

Since the inequality $\sup_x W(x) + \varepsilon > V^*$ must be true for all $\varepsilon > 0$, it must be that $\sup_x W(x) \geq V^*$ (otherwise it is easy to see that one obtains a contradiction). *(b)* If $V^* = \infty$ we have to show that for all real numbers $M < \infty$ there is a $x \in X$ such that $W(x) > M$. Assume it is not the case and let $\bar{M} < \infty$ such that $W(x) \leq \bar{M}$, for all x. Since for any x we have $W(x) \geq f(x, y)$ for all y, it must be that $\infty > \bar{M} \geq f(x, y)$ for all x, y, but this implies that $\bar{M}$ is an upper bound for f. Since V^* is the least upper bound we have a contradiction. **Q.E.D.**

Using the *infinite penalization approach*[1] the same result can be stated for the case where the choice (x, y) is restricted to a set $D \subset X \times Y$. In this case, one must be able to decompose the feasibility set in an appropriate way. In these notes, we will always

[1] The infinite penalization approach basically reduces a constrained maximization problem into a free maximization one. For example, consider the constrained problem

$$\sup_{(x,y) \in D \subset X \times Y} f(x, y).$$

If we define a new function f^* as follows

$$f^*(x, y) = \begin{cases} f(x, y) \text{ if } (x, y) \in D \\ -\infty \quad \text{otherwise} \end{cases}$$

then it should be easy to see that

$$\sup_{(x,y) \in D} f(x, y) = \sup_{(x,y) \in X \times Y} f^*(x, y).$$

analyze environments where this decomposition can be done.[2,3] Moreover, we will typically consider environments where the objective function f is a sum of terms. In this case, the following property of the supremum becomes quite useful.

Lemma 2 *Let a, and b two real numbers, if $b > 0$, then*

$$\sup_{x \in X} \; a + bf(x) = a + b \sup_{x \in X} \; f(x),$$

if $b < 0$, then

$$\sup_{x \in X} \; a + bf(x) = a + b \inf_{x \in X} \; f(x).$$

Proof. We will show the result when $b > 0$, assuming that the sup takes a finite value. Let $f^* = \sup_{x \in X} \; f(x)$, and $V^* = \sup_{x \in X} \; a + bf(x)$. First, we show $V^* \leq a + bf^*$. Note that for all $x \in X$ we have $a + bf^* \geq a + bf(x)$, that is, $a + bf^*$ is an upper bound for the set

$$\{y : y = a + bf(x) \text{ for some } x \in X\}.$$

As a consequence, its least upper bound V^* must be such that $a + bf^* \geq V^* = \sup_{x \in X} a + bf(x)$. To show the converse, note that from the definition of f^* as a supremum, we have that for any $\varepsilon > 0$ there must exist a $\bar{x}_\varepsilon \in X$ such that $f(\bar{x}_\varepsilon) > f^* - \varepsilon$. Hence $a + bf(\bar{x}_\varepsilon) > a + bf^* - b\varepsilon$. Since $\bar{x}_\varepsilon \in X$, it is obvious that $V^* \geq a + bf(\bar{x}_\varepsilon)$. Hence $V^* \geq a + bf^* - b\varepsilon$. Since ε was arbitrarily chosen, we have our result: $V^* \geq a + bf^*$. **Q.E.D.**

Notice that in economics the use of the BPO is quite common. Consider for example the typical profit maximization problem

$$\begin{aligned} \pi^*(p, w) &= \max_{z,y} \; py - wz \\ s.t. \; y &\leq f(z), \end{aligned}$$

[2] Hence, $(x, y) \in D$ will always be equivalent to $x \in dom\Gamma$ and $y \in \Gamma(x)$ for some correspondence Γ, where the domain of a correspondence is the set of values for which it is non empty, i.e.

$$dom\Gamma = \{x : \Gamma(x) \neq \emptyset\}.$$

[3] The constrained maximization version of Lemma 1 is

$$\sup_{(x,y) \in D} \; f(x, y) = \sup_{x \in dom\Gamma} \left\{ \sup_{y \in \Gamma(x)} \; f(x, y) \right\}.$$

where y is output and z is the vector of inputs; and p and w are prices. Using Lemma 1 and 2, the problem can be decomposed as follows:

$$\pi^*(p,w) = \max_y \; py - C(y;w),$$

where C is the cost function, and for any given y is defined

$$\begin{aligned} C(y;w) &= \inf_z \; wz \\ s.t.\; y &\leq f(z). \end{aligned}$$

Let me now introduce some notation. To make easier the study of the notes I follow closely SLP, Chapter 4. Consider again the optimal growth problem. In the introductory section we defined the problem as follows

$$\begin{aligned} &\sup_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u\left(f(k_t) - k_{t+1}\right) \\ s.t.\; k_0 &\geq 0 \\ 0 &\leq k_{t+1} \leq f(k_t) \quad \text{for all } t. \end{aligned}$$

In general, the class of dynamic problems we are going to consider is represented by

$$\begin{aligned} V^*(x_0) &= \sup_{\{x_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t F\left(x_t, x_{t+1}\right) \\ s.t.\; x_0 &\in X \\ x_{t+1} &\in \Gamma(x_t) \quad \text{for all } t. \end{aligned} \tag{3.1}$$

One key ingredient is the time invariant function F of the present and the future states, whose discounted sum describes the objective function of the problem. The time invariant correspondence Γ describing feasibility, and $\beta \in (0,1)$ is the discount factor. Finally, notice that we denoted the *true value function* V^* with an asterisk. This is done in order to make the distinction between the true value function V^* and a specific solution V to the Bellman functional equation implied by (3.1). We will indeed see that the two concepts are closely related but quite different.

Exercise 35 *Show that the general formulation in (3.1) can be specified to describe the optimal growth problem defined above.[Hint: very easy!]*

From these primitives, the problem can be rewritten in a more compact way. For any sequence $\mathbf{x} = \{x_{t+1}\}_{t=0}^{\infty}$ with initial value x_0 define the set

$$\Pi(x_0) = \left\{ \{x_{t+1}\}_{t=0}^{\infty} \text{ such that } x_{t+1} \in \Gamma(x_t) \quad \text{for all } t \right\}.$$

If $\mathbf{x} \in \Pi(x_0)$ we say $\mathbf{x}$ is a *feasible plan* (with respect to x_0).

We now make the following assumption on our primitive Γ.

Assumption 4.1 $\Gamma(x)$ is non-empty for any $x \in X$.

The relationship between the two definitions of feasibility is clarified by the following exercise.

Exercise 36 *Show that if Assumption 4.1 is true then $\Pi(x_0)$ is non-empty for each $x_0 \in X$.*

For any sequence $\mathbf{x}$ we define the intertemporal payoff function as follows

$$\mathbf{U}(\mathbf{x}) = \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1})$$

so we can equivalently write the problem (3.1) in a compact way

$$V^*(x_0) = \sup_{\mathbf{x} \in \Pi(x_0)} \mathbf{U}(\mathbf{x}). \tag{3.2}$$

We allow the problem to have an unbounded value, so we write the infinite sum even when the series is divergent. What we will never do is to consider infinite sums when the series have *indeterminate* character.

Assumption 4.2 For all $x_0 \in X$ and $\mathbf{x} \in \Pi(x_0)$, $\lim_{n\to\infty} \mathbf{U}_n(\mathbf{x}) = \lim_{n\to\infty} \sum_{t=0}^{n} \beta^t F(x_t, x_{t+1})$ exists although it might be plus or minus infinity.

Exercise 37 *Show that if Assumption 4.2 is satisfied, we can write $\mathbf{U}(\mathbf{x}) = \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1})$ as follows*

$$\mathbf{U}(\mathbf{x}) = F(x_0, x_1) + \beta \mathbf{U}(\mathbf{x}')$$

for each feasible path $\mathbf{x} \in \Pi(x_0)$ with $x_0 \in X$.

We are now ready to state the BPO for our class of problems.

(Infinite) Bellman Principle If $\mathbf{x}$ is optimal then $\mathbf{x}'$ is optimal, where $\mathbf{x} = (x_0, x_1, ...)$ and $\mathbf{x}' = (x_1, x_2, ...)$ is the "one-step ahead" sequence.

The BPO principle is equivalent to the possibility of writing the value function V^* for our infinite dimensional problem in the form of a functional equation. Which is called the *Bellman Equation*[4]

$$V(x_0) = \sup_{x_1 \in \Gamma(x_0)} F(x_0, x_1) + \beta V(x_1), \tag{3.3}$$

and this for any t. More formally, we have

[4] The corresponding finite horizon BPO is as follows

Theorem 12 *Let Assumptions **4.1** and **4.2** be satisfied by our problem. Then the function* $V^*(x_0) = \sup_{\mathbf{x}\in\Pi(x_0)} \mathbf{U}(\mathbf{x})$ *satisfies the functional equation (3.3) for all* $x_0 \in X$. *Moreover, if a feasible plan* $\mathbf{x}^* \in \Pi(x_0)$ *attains the supremum in (3.2) then the maximal plan* $\mathbf{x}^*$ *must satisfy (3.3) with* $V = V^*$, *i.e.*

$$V^*(x_t^*) = F(x_t^*, x_{t+1}^*) + \beta V^*(x_{t+1}^*),\ t = 0, 1, 2, ... \tag{3.4}$$

Proof What we have to show for the first part of the theorem is the following:

$$V^*(x_0) \equiv \sup_{\mathbf{x}\in\Pi(x_0)} \mathbf{U}(\mathbf{x}) = \sup_{x_1\in\Gamma(x_0)} F(x_0, x_1) + \beta V^*(x_1).$$

As a preliminary step, use Assumption **4.2** (and Exercise 37) to rewrite

$$\sup_{\mathbf{x}\in\Pi(x_0)} \mathbf{U}(\mathbf{x}) = \sup_{\mathbf{x}\in\Pi(x_0)} F(x_0, x_1) + \beta\mathbf{U}(\mathbf{x}').$$

By Lemma 1 (in its constrained version), we can decompose the sup operator as follows

$$\sup_{\mathbf{x}\in\Pi(x_0)} F(x_0, x_1) + \beta\mathbf{U}(\mathbf{x}') = \sup_{x_1\in\Gamma(x_0)} \sup_{\mathbf{x}'\in\Pi(x_1)} F(x_0, x_1) + \beta\mathbf{U}(\mathbf{x}').$$

The relationship between the correspondence Γ and the feasible set $\Pi(x_0)$ guarantees (by definition) that Assumption **4.1** (together with Exercise 36) suffices to allow us to do the decomposition.

The final step is to use Lemma 2 to pass through the second sup operator. That is, applying Lemma 2 to the second sup operator with $F(x_0, x_1) = a$ and $\beta = b > 0$ we have

$$\sup_{x_1\in\Gamma(x_0)} \sup_{\mathbf{x}'\in\Pi(x_1)} F(x_0, x_1) + \beta\mathbf{U}(\mathbf{x}') = \sup_{x_1\in\Gamma(x_0)} F(x_0, x_1) + \beta \sup_{\mathbf{x}'\in\Pi(x_1)} \mathbf{U}(\mathbf{x}').$$

One must keep in mind that this last step can only be done because of the specific characteristics of the sum. First of all, the discounted summation satisfies an obvious monotonicity assumption since $\beta > 0$. Second, it also satisfies an important property

Finite Bellman Principle If a path $(x_0, x_1, ..., x_T)$ is optimal for the T horizon problem. Then the path $(x_1, x_2, ..., x_T)$ is optimal for the $T-1$ horizon problem.

That is, we can write the value function $V_T^*(x_0)$ for the T-horizon problem in terms of the $T-1$ horizon value function V_{T-1}^* as follows

$$V_T^*(x_0) = \sup_{x_1\in\Gamma(x_0)} F(x_0, x_1) + \beta V_{T-1}^*(x_1).$$

of continuity (as we saw a sum it is actually a linear mapping).[5] The last part of the proposition is easily derived by the fact that $\mathbf{x}^*$ reaches the supremum, i.e.

$$\mathbf{U}(\mathbf{x}^*) = V^*(x_0) = \max_{\mathbf{x} \in \Pi(x_0)} \mathbf{U}(\mathbf{x}),$$

and is left as an exercise. For an alternative proof see SLP Theorems 4.2 and 4.4. Finally, note that we used $\beta > 0$, however the case $\beta = 0$ is really easy, and again left as an exercise. **Q.E.D.**

Let's have another look at conditions (3.4). The key idea of the Bellman principle is that we can simply check for "one stage deviations". An optimal plan $\mathbf{x}^*$ has the property that for any t - once the past is given by x_t^*, and the effect of your choice x_{t+1} on the future returns is summarized by $V^*(x_{t+1})$ - there is no incentive to choose a different x_{t+1} from

[5] Monotonicity alone is not enough, as the following counterexample from Bertsekas et al. (1978) shows. In this example there are no states, and the problem is in two-periods. The agent must choose the control $c \in (-1, 0]$ and the continuation value is $J_0 = 0$. All the complications are due to the properties of the aggregator $H : U \times \mathcal{J} \to \mathcal{J}$, $\mathcal{J} \subset \mathbb{R}$. It is such that

$$H(u, J) = \begin{cases} u & \text{if } J > -1 \\ u + J & \text{if } J \leq -1. \end{cases}$$

One should easily see that the true optimal values are

$$J_1^* = \inf_u H(u, J_0) = -1$$

and

$$\begin{aligned} J_2^* &= \inf_{u_1, u_2} H(u_1, H(u_2, J_0)) = \\ &= \inf_{u_1, u_2} H(u_1, u_2) = -1. \end{aligned}$$

However, if one uses the recursive formulation and the Bellman operator gets something different. In particular

$$J_1 = T^1(J_0) = \inf_{u_2} H(u_2, J_0) = -1$$

and

$$\begin{aligned} T^2(J_0) &= T^1(J_1) = \inf_{u_1} H(u_1, J_1) \\ &= \inf_{u_1} H(u_1, -1) = -2. \end{aligned}$$

To have a sufficient condition we might require either continuity, or $\exists \alpha \geq 0 : \forall\, r \geq 0$, and J

$$H(u, J) \leq H(u, J + r) \leq H(u, J) + \alpha r.$$

that implied by the optimal plan x_{t+1}^*. That is, there is no incentive to revise the plan. That is why economists often call the paths generated by the BPO as *time consistent*.

The above theorem states a set *necessary* conditions of an optimal path in terms of the functional equation. It is important to notice that regarding these properties the distinction between finite horizon and infinite horizon is merely technical. For example, for the finite periods version of the above theorem we do not need Assumption **4.2** since finite sums will always be well defined. In the infinite horizon, such assumptions are only needed to have a well defined continuation problem in each period.

If one looks for *sufficient* conditions the distinction between finite and infinite horizon becomes much more important. In particular, the finite version of Theorem 12 is also a sufficient condition for an optimum. That is, a path that satisfies the BPO is an optimal plan. The idea is simple. The BPO states that an optimal path is such that the agent does not have incentives to deviate for one period from his maximizing behavior and then reverting to an optimal behavior (summarized by $V^*(x)$). By induction on this principle, one can show that the agent has never incentive to deviate for finitely many periods either. In contrast, the BPO cannot say anything about infinite deviations. That is behaviors that never revert to the optimizing behavior any more. As a consequence, in order to use the BPO to detect optimal plans one must induce some additional structure on the problem so that the agent cannot gain from infinite deviations either. In term of the objective function $\mathbf{U}(\mathbf{x})$ one typically requires the so called *continuity at infinity* (see for example Fudenberg and Tirole, 1991, Chapter 5). Here below we follow SLP and state the additional condition in terms of the (true) value function V^*.

Theorem 13 *Assume **4.1** and **4.2**. Let a feasible path* $\mathbf{x}^* \in \Pi(x_0)$ *from* x_0 *satisfying (3.4), and such that*

$$\limsup_{t\to\infty} \beta^t V^*(x_t^*) \leq 0.$$

Then $\mathbf{x}^*$ *attains the supremum in (3.2) when the initial state is* x_0.

Proof. First notice that since $\mathbf{x}^*$ is feasible its value cannot be greater than the supremum, that is $V^*(x_0) \geq \mathbf{U}(\mathbf{x}^*)$. We have to show the inverse inequality. If we apply repeatedly (3.4) we get

$$V^*(x_0) = \mathbf{U}_n(\mathbf{x}^*) + \beta^{n+1} V^*(x_{n+1}^*) \text{ for } n = 1, 2, ...$$

now using $\limsup_{n\to\infty} \beta^{n+1} V^*(x_{n+1}^*) \leq 0$ we have

$$V^*(x_0) = \mathbf{U}_n(\mathbf{x}^*) + \beta^{n+1} V^*(x_{n+1}^*) \leq \lim_{n\to\infty} \mathbf{U}_n(\mathbf{x}^*) \equiv \mathbf{U}(\mathbf{x}^*).$$

Q.E.D.

The limit value condition imposed above can be interpreted as a transversality condition for the BPO. The same idea can be applied to the value function as in the first part of Theorem 12. In this case, we obtain a sufficient condition for a given function to be the true value function.

Theorem 14 *Assume 4.1 and 4.2. If* V *is a solution to the functional equation (3.3) and satisfies*

$$\lim_{t\to\infty} \beta^t V(x_t) = 0 \text{ for all } \mathbf{x} \in \Pi(x_0) \text{ and all } x_0 \in X,$$

Then V *is the true value function, i.e.* $V = V^*$.

Proof. The proof here is basically on the lines of the previous theorem. The only additional complication is that we are dealing with the sup instead of the Max. In particular, we are not sure of the existence of an optimal plan. However, notice that one key aspect of V^* to be verified for V is that

$$V^*(x_0) \geq U(\mathbf{x}) \text{ for all } \mathbf{x} \in \Pi(x_0) \text{ and all } x_0 \in X.$$

Now since V solves (3.3) for all t we have that

$$\begin{aligned} V(x_0) &\geq F(x_0, x_1) + \beta V(x_1) \geq F(x_0, x_1) + \beta F(x_1, x_2) + \beta^2 V(x_2) \geq ... \\ &\geq \sum_{t=0}^{T-1} \beta^t F(x_t, x_{t+1}) + \beta^T V(x_T) \text{ for all } \mathbf{x} \in \Pi(x_0). \end{aligned}$$

Hence, as long as $\beta^T V(x_T) \to 0$ we have the desired property for V. See also Theorem 4.3 in SLP. **Q.E.D.**

The above theorem also suggests that the "guess and verify" procedure we discussed in the introductory section simply provides *one* solution to the functional equation (3.3). However (3.3) might have multiple solutions, and we are obviously looking for the right value function V^*. Theorem 14 guarantees that a bounded solution V to (3.3) is actually the "right" value function.

3.2 The BPO under Bounded Returns: Continuity, Concavity and Differentiability of the Value Function

In this section we will specify the problem used to study the BPO by imposing additional restrictions on the primitives F and Γ so that to be able to show some properties for the value function V^*. Following SLP, we will heavily use the contraction mapping theorem (Theorem 7). To show that the Bellman operator is a contraction we will use the Blackwell sufficient conditions (Theorem 8), so we will work in the space of bounded functions with the sup norm.

Continuity To show continuity we will use the theorem of the maximum (Theorem 10). Here are the necessary assumptions to use it.

Assumption 4.3 $\Gamma(x)$ is a non-empty, compact valued and continuous correspondence, and $X \subseteq \mathbb{R}^l$.

Assumption 4.4 F is bounded and continuous and $\beta \in [0, 1)$.

Theorem 15 *Assume 4.3 and 4.4 and consider the metric space $(\mathcal{C}(X), d_\infty)$ of bounded and continuous functions with the sup norm. Then the Bellman operator T defined by*

$$(TV)(x) = \max_{x' \in \Gamma(x)} F(x, x') + \beta V(x') \tag{3.5}$$

(i) maps $\mathcal{C}(X)$ into itself; (ii) has a unique fixed point $V \in \mathcal{C}(X)$; for all $V_0 \in \mathcal{C}(X)$ we have

$$d_\infty(T^n V_0, V) \leq \beta^n d_\infty(V_0, V), \text{ for any } n = 0, 1, ...$$

(iii) and the policy correspondence

$$G(x) = \{x' \in \Gamma(x) : V(x) = F(x, x') + \beta V(x')\}$$

is non empty, compact valued, and upper semi-continuous.

Proof. (i) If f is continuous and F is continuous the objective function of the maximization problem (3.5) is continuous. This, together with the properties of the correspondence Γ imply that we can directly apply Theorem 10 to show that the value function of this problem Tf is also continuous. The fact that F is bounded also implies that if f is bounded then Tf will be bounded too. So *(i)* is shown. To show *(ii)* and *(iii)* we use Theorem 7. We need to show that T describes a contraction and that the metric space $(\mathcal{C}(X), d_\infty)$ is complete. The completeness of $(\mathcal{C}(X), d_\infty)$ have been presented in Theorem 6. To show that T is a contraction we can use Theorem 8. In Exercise 22 we have precisely shown monotonicity and discounting for the optimal growth model, however, one can closely follow the same line of proof and prove the same statement in this more general case. So we have *(ii)*. Finally, since we have shown that the fixed point V is continuous, we can apply the maximum theorem again and show *(iii)*. **Q.E.D.**

Notice first, that when F is a bounded function also V^* must be bounded, and from Theorem 14 the unique bounded function V satisfying the Bellman equation must be the true value function, i.e. $V = V^*$. Moreover, from *(iii)* above we are guaranteed we can construct an optimal plan by taking a selection from the policy correspondence G as follows: start form x_0 an then for any $t \geq 1$ set $x_t \in G(x_{t-1})$. But then we have also shown existence without using possibly complicated extensions of the Weierstrass theorems in infinite dimensional spaces.

Concavity and Differentiability

Assumption 4.7 Γ has a convex graph, i.e. for each two $x_1, x_2 \in X$ and corresponding feasible $x_1' \in \Gamma(x_1)$, $x_2' \in \Gamma(x_2)$ we have

$$[\theta x_1' + (1-\theta)x_2'] \in \Gamma\left(\theta x_1 + (1-\theta)x_2\right) \text{ for any } \theta \in [0,1]\,.$$

Assumption 4.8 F is concave and if $\theta \in (0,1)$ and $x_1 \neq x_2$ we have

$$F\left(\theta x_1 + (1-\theta)x_2,\ \theta x_1' + (1-\theta)x_2'\right) > \theta F\left(x_1, x_1'\right) + (1-\theta)F\left(x_2, x_2'\right).$$

Now we are ready to show our result.

Theorem 16 *Assume 4.3, 4.4, 4.7 and 4.8. (i) Then the fixed point V is strictly concave and the policy G is a continuous function g. (ii) Moreover, if F is differentiable then V is continuously differentiable and*

$$V'(x) = \frac{\partial F(x, g(x))}{\partial x} = F_1(x, g(x))$$

for all $x \in intX$ such that the policy is interior, i.e. $g(x) \in int\Gamma(x)$.

Proof. The proof of (i) uses the fact that under **4.7** and **4.8** the operator T maps continuous concave function into concave functions, and the space of continuous concave functions is a closed subset of the metric space $(\mathcal{C}(X), d_\infty)$. As a consequence we can apply Exercise 20 to be sure that the space of continuous and bounded functions in the sup norm is a complete metric space and apply the contraction mapping theorem. Under 4.8 it is easy to show that V is actually *strictly* concave. Since a strictly concave problem has a unique maximum, the policy correspondence must be single valued, hence g is a continuous function (see Exercises 28 and 12.3). Part *(ii)* will be shown graphically in class. The interested reader can see the proof in SLP, Theorem 4.11, page 85. **Q.E.D.**

The key element of the proof of part *(ii)* above is the following Lemma of Benveniste and Sheinkman (1979):

Lemma 3 *Let $x_0 \in intX$ and let D a neighborhood of x. If there exists a concave and differentiable function $W : D \rightarrow \Re$ such that for $x \in D$ $W(x) \leq V(x)$ and $V(x_0) = W(x_0)$ then V is differentiable at x_0 and $V'(x_0) = W'(x_0)$.*

Proof. If $p \in \partial V(x_0)$ then $p \in \partial W(x_0)$ since the subgradient inequality carries over. But W is differentiable, hence p is the unique subgradient of W, which implies that also V has only one subgradient at x_0. V is concave, and since any concave function with only one subgradient is differentiable V is differentiable. This last statement is not easy to show, see Rockafellar (1970). **Q.E.D.**

Monotonicity When $F(x, x')$ is monotone in x and the feasibility set $\Gamma(x)$ widens with x, it is easy to show that $V(x)$ is increasing.

Differentiability using Boldrin and Montrucchio (1998) Boldrin and Montrucchio (1998) use the properties of the contraction, the uniform convergence of the policy of a class of finite period problem to the policy of the infinite horizon problem, and a well known approximation theorem (Dieudonné, *Foundations of Mathematics* No. 8.6.3) to show differentiability.[6] Their result does not use concavity, hence their method can be used to study parameter sensitivity as well.

Differentiability of the policy under C^2 Differentiability of the policy is strictly linked to the second order differentiability of the value function. Montrucchio (1997) shows that under some conditions the Bellman operator is a contraction also in the C^2 topology. Santos (1991) shown the same result using a different methodology. Recently, Santos (2003) shows how this result can be profitably used to compute bounds on the approximation errors in the numerical discretization procedures.

[6] The key Lemma they use guarantees that if $V_T'(k) \rightarrow \phi(k)$ uniformly as $T \rightarrow \infty$, and $V_T \rightarrow V$, with V_T differentiable for all T then $V'(k) = \phi(k)$.

Unbounded returns In most cases the complication of unbounded F is solved on an ad hoc basis. Alvarez and Stokey (1998) analyze the case where both F and Γ are homogeneous functions. Montrucchio and Privilegi (1998) and, more recently, Rincón-Zapatero and Rodriguez-Palmero (2003) generalize Boyd (1990) and use the contraction mapping theorem to show existence and uniqueness of V in a large class of Bellman equation problems with unbounded returns. Streufert (1990, 1998) analyze a large class of capital accumulation problems and uses the monotonicity property of the Bellman operator to show that V is upper hemi-continuous. He defines the notion of admissibility and uses biconvergence to show uniqueness and continuity.

3.3 The Euler's Variational Approach and the Bellman Principle of Optimality

To get the key idea of the Euler's Variational approach we should recall the problem we analyzed in Lemma 1, at the beginning of this section

$$\max_{(x,y)\in X\times Y} f(x,y).$$

To keep the analysis as simple as possible, assume the existence of a maximum couple (x^*, y^*), and define

$$V^* = f(x^*, y^*) \geq f(x,y) \text{ for all } (x,y) \in X \times Y. \tag{3.6}$$

According to the *Bellman's principle (*Lemma 1) the problem can be solved in two steps:

$$V^* = \max_{x\in X} W(x),$$

with

$$W(x) = \max_{y} f(x,y);$$

and vice versa: *any pair* that solves such a two steps problem is an optimal one. The *Euler's variational approach* starts by the observation that the optimal pair (x^*, y^*) satisfies (among other things)

$$\begin{aligned} f(x^*, y^*) &\geq f(x, y^*) \text{ for all } x \in X, \text{ and} &(3.7)\\ f(x^*, y^*) &\geq f(x^*, y) \text{ for all } y \in Y. &(3.8) \end{aligned}$$

Notice the key difference from the Bellman principle: an optimal pair (x^*, y^*) has to be such that *along the optimal path* there is no incentive to deviate from it "unilaterally," that is only in one direction.[7] Notice that we can equivalently write the BPO two step procedure as

$$f(x^*, y^*) \geq f(x, y^*(x)) \text{ for all } x \in X$$

where (assuming there is only one optimum)

$$y^*(x) = \arg\max_{y \in Y} f(x, y).$$

That is, crucially $y^*(x)$ is not a fixed number, it is a function of x.

In our notation, the Euler variational approach translates in the observation that an optimal plan $\mathbf{x}^*$ for any t must satisfy

$$\begin{aligned} F(x_{t-1}^*, x_t^*) + \beta F(x_t^*, x_{t+1}^*) &\geq F(x_{t-1}^*, x_t) + \beta F(x_t, x_{t+1}^*) \\ \text{for all } x_t &\in \Gamma(x_{t-1}^*) \text{ such that } x_{t+1}^* \in \Gamma(x_t). \end{aligned} \tag{3.9}$$

That is, one property of the optimal plan is that the agent cannot gain by deviating (in a feasible fashion) from the optimal path in any period, taking the optimal path as given. This is again a one stage deviation principle. However, the key distinction with the BPO is that the deviation considered here does not take into account the future effects of such a deviation, but *takes as given both the past and the future.* Recall that the equivalent to condition (3.9) for the BPO is

$$\begin{aligned} F(x_{t-1}^*, x_t^*) + \beta V(x_t^*) &\geq F(x_{t-1}^*, x_t) + \beta V(x_t) \\ \text{for all } x_t &\in \Gamma(x_{t-1}^*). \end{aligned}$$

In other terms, the interpretation of the Euler condition under differentiability is that one-period *reversed arbitrage,* an arbitrage that immediately returns to the original path, is not profitable on an optimal path. This means that the cost calculated at $t = 0$ from acquiring an extra unit of capital at time t, $\beta^t u'(c_t^*) = \beta^t F_2(x_t^*, x_{t+1}^*)$, is at least as great as the benefit realized at time $t+1$ discounted back to period $t = 0$, from selling that additional unit of capital at $t+1$ for consumption. The extra unit of capital yields in utility terms $\beta^{t+1} F_1(x_{t+1}^*, x_{t+2}^*) = f'(k_{t+1})\beta^{t+1} u(c_{t+1}^*) : f'(k_{t+1})$ units of consumption good at time $t+1$, and each unit of that good is worth $\beta^{t+1} u(c_{t+1}^*)$ utils in period 0. Hence we have

$$u'(c_t^*) = \beta f'(k_{t+1}) u'(c_{t+1}^*).$$

The Variational approach uses one somehow weak property of optimality. As a consequence, in general the Euler's conditions are far from being sufficient for determining an

[7]Unilateral deviations of this sort are both in the concepts of saddle points and Nash equilibria.

optimal point. So the question is, suppose we have as a candidate the couple (x^{**}, y^{**}) such that satisfies (3.7) and (3.8). What are the additional conditions we need in order to guarantee that (x^{**}, y^{**}) is an optimal point? It turns out that a sufficient condition for an interior optimum is that at (x^{**}, y^{**}) f is subdifferentiable.[8] The idea is simple: if (x^{**}, y^{**}) satisfies (3.7) and (3.8) and at (x^{**}, y^{**}) f is subdifferentiable, then the vector $(0, 0)$ must belong to the subdifferential of f at (x^{**}, y^{**}). This property is at the core of the sufficiency of the first order conditions when f is concave.[9]

Example 1 *Let* $f(x, y) = (x - 1)^2 (y - 1)^2$ *with* $x, y \in [0, 3]$. *Obviously,*

$$\max_{(x,y)\in[0,3]\times[0,3]} f(x, y) = 4,$$

with solution $x^* = y^* = 3$, *however it is easy to see that the pair* $x^{**} = y^{**} = 1$ *satisfies (3.7) and (3.8) as well [check it!].*

Exercise 38 *Do the same exercise with* $f(x, y) = \sqrt{x}\sqrt{y}$ *and* $x, y \in [0, 3]$. *Notice that now we can use the first order conditions to find the 'unilateral' maximum for* x *given* $y = \bar{y}$, *since the problem is concave in* x *alone given* $\bar{y}$. *The same is true for* y *given* $x = \bar{x}$. *Is this form of concavity enough? Explain.*

Transversality So, Euler equations are necessary conditions for an optimum. We also said above that exactly following the same logic used in static maximization, when the problem is concave an interior optimal can be detected by the Euler's equations. However, this principle works only for finite horizon problems. When the time horizon is infinite Euler equations are not enough, and we need an additional restriction to detect optimal programs.

The reason is the same as the one for the BPO. The Euler conditions control for only one-stage deviations, which can be extended by induction to any finite stages deviation. But they cannot tell us anything about infinite period deviations.

---

[8] A real valued function $f : X \times Y \to \mathbb{R}$ is subdifferentiable (in the concave sense) at (x_0, y_0) if there exists a vector $p = (p_x, p_y)$ such that

$$f(x, y) \leq f(x_0, y_0) + p_x.(x - x_0) + p_y.(y - y_0) \text{ for any } x \in X, y \in Y,$$

and p is the subdifferential of f at (x_0, y_0).

[9] A function f defined in a convex X is concave if for any two $x, y \in X$ and any $\lambda \in [0, 1]$ we have

$$f(\lambda x + (1 - \lambda) y) \geq \lambda f(x) + (1 - \lambda) f(y).$$

By the separation theorem, any concave function is subdifferentiable.

Proposition 3 *Assume F is bounded, continuous, concave, and differentiable. Moreover assume Γ has a compact and convex graph. (i) If the (interior) sequence $\{x_t^*\}_{t=1}^{\infty}$ with $x_{t+1}^* \in int\Gamma(x_t^*)$ for any $t = 0, 1, 2, ...$ satisfies*

$$F_2(x_t^*, x_{t+1}^*) + \beta F_1(x_{t+1}^*, x_{t+2}^*) = 0 \text{ for } t = 0, 1, ... \tag{3.10}$$

and for any other feasible sequence $\{x_t\}_{t=0}^{\infty} \in \Pi(x_0)$ we have

$$\lim_{T\to\infty} \beta^T F_1(x_T^*, x_{T+1}^*)(x_T - x_T^*) \geq 0, \tag{3.11}$$

then $\{x_t^*\}_{t=1}^{\infty}$ is an optimal sequence. (ii) If in addition $F_1(x, x') > 0$ for all $x, x' \in intX$ and $X \subseteq I\!R_+^l$, the condition (3.11) can be substituted by

$$\lim_{T\to\infty} \beta^T F_1(x_T^*, x_{T+1}^*) x_T^* \leq 0.$$

Proof. *(i)* We are done if we can show that for any feasible $\{x_t\}_{t=1}^{\infty} \in \Pi(x_0)$ we have

$$\lim_{T\to\infty} \sum_{t=0}^{T} \beta^t F(x_t^*, x_{t+1}^*) \geq \lim_{T\to\infty} \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}),$$

where both limit exist and the inequality has a meaning since F is bounded. Now, notice that from the concavity and differentiability of F we have that

$$F(x_t, x_{t+1}) \leq F(x_t^*, x_{t+1}^*) + F_1(x_t^*, x_{t+1}^*)(x_t - x_t^*) + F_2(x_t^*, x_{t+1}^*)(x_{t+1} - x_{t+1}^*)$$

multiplying by β^t and summing up the first T terms one gets

$$\sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}) \leq \sum_{t=0}^{T} \beta^t F(x_t^*, x_{t+1}^*) + D_T, \tag{3.12}$$

where

$$D_T = \sum_{t=0}^{T} \beta^t \left[F_1(x_t^*, x_{t+1}^*)(x_t - x_t^*) + F_2(x_t^*, x_{t+1}^*)(x_{t+1} - x_{t+1}^*) \right].$$

Since $\sum_{t=0}^{T} \beta^t F(x_t, x_{t+1})$ converges for any sequence $\{x_t\}$, one can show that (3.12) implies that D_T must converge as well.[10] It then suffices to show that

$$\lim_{T\to\infty} D_T \leq 0$$

[10] See Mitchell (1990), page 715.

Notice that we can rearrange the terms in D_T and obtain[11]

$$D_T = \sum_{t=0}^{T-1} \beta^t \left[F_2(x_t^*, x_{t+1}^*) + \beta F_1(x_{t+1}^*, x_{t+2}^*) \right] (x_{t+1} - x_{t+1}^*) - \beta^T F_1(x_T^*, x_{T+1}^*)(x_T - x_T^*).$$

Euler conditions (3.10) guarantee that the fist $T-1$ terms go to zero, hence

$$\lim_{T\to\infty} D_T = - \lim_{T\to\infty} \beta^T F_1(x_T^*, x_{T+1}^*)(x_T - x_T^*) \leq 0$$

where the last inequality is implied by the transversality condition (3.11). In order to show *(ii)* notice that if $F_1 > 0$ and $x_T \geq 0$,

$$\lim_{T\to\infty} \beta^T F_1(x_T^*, x_{T+1}^*)(x_T - x_T^*) \geq - \lim_{T\to\infty} \beta^T F_1(x_T^*, x_{T+1}^*) x_T^* \geq 0$$

and we are done. **Q.E.D.**

The convexity assumption on Γ is required only in order to define concavity of F. In fact, the theorem only uses the subdifferentiability property of F, so it will remain true as long as F is subdifferentiable along the optimal trajectory. Notice moreover that feasibility is used very little (only at $t = 0$). This is so since we assumed interiority, hence the sequence $\{x_{t+1}^*\}$ is in fact unconstrained optimal. That is, the optimal sequence dominates *all* other sequences. The sequence of subdifferentials are also called *supporting prices*. The reason will become clear in the next paragraph.

The Euler equations (3.10) are the usual first order conditions for x_{t+1}^* to be an interior maximum *given* x_t^* and x_{t+t}^* of the problem described in (3.9). The transversality condition (3.11) has the following interpretation. Notice first that F_1 is the marginal return of the state, for example, in the optimal growth model $F_1(x_t^*, x_{t+1}^*) = u'(c_t^*) f'(k_t^*)$ is the price of capital.[12] Since in the optimal growth model we have both $x_t \geq 0$ and $F_1 > 0$, the

[11] In more detail, consider the first few terms

$$\begin{aligned} & F_1(x_0^*, x_1^*)(x_0 - x_0^*) + F_2(x_0^*, x_1^*)(x_1 - x_1^*) \\ & + \beta \left[F_1(x_1^*, x_2^*)(x_1 - x_1^*) + F_2(x_1^*, x_2^*)(x_2 - x_2^*) \right] \\ = & \left[F_2(x_0^*, x_1^*) + \beta F_1(x_1^*, x_2^*) \right] (x_1 - x_1^*) \\ & + \beta \left[F_2(x_1^*, x_2^*) + \beta F_1(x_2^*, x_3^*) \right] (x_2 - x_2^*) \\ & - \beta^2 F_1(x_2^*, x_3^*)(x_2 - x_2^*) \end{aligned}$$

where the term $F_1(x_0^*, x_1^*)(x_0 - x_0^*)$ disappears since feasibility implies $x_0 = x_0^*$; and we have added and subtracted the term $\beta^2 F_1(x_2^*, x_3^*)(x_2 - x_2^*)$.

[12] The value of "one" unit of k_t^* is the price of the consumption goods $u'(c_t^*)$ times the amount of consumption goods that can be produced by one unit of capital.

transversality condition requires $\lim_{T\to\infty}\ \beta^T u'(c_T^*) f'(k_T^*) k_T^* \leq 0$. It is clear that for a finite horizon problems if $\beta^T u'(c_T^*) f'(k_T^*) > 0$ the agent will not maximize lifetime utility by ending the last period with a positive amount of capital k_T^*. The transversality condition states this intuitive argument in the limit. If $\lim_{T\to\infty}\ \beta^T u'(c_T^*) f'(k_T^*) k_T^* > 0$ the agent is holding valuable capital, and perhaps he can increase the present value of its utility by reducing it.

More in general, the transversality condition (3.11) requires any alternative trajectory $\{x_t\}$ satisfying

$$\lim_{t\to\infty} \beta^t F_1(x_t^*, x_{t+1}^*)(x_t - x_t^*) < 0$$

to be infeasible. That is, the transversality condition means that if given $\{x_t^*\}$ it is impossible to reduce the limit value of the optimal stock (considered in discounted terms) by choosing $x_t \neq x_t^*$ (except perhaps for incurring in an infinite loss because $\{x\}$ is not feasible) then the value of the capital has been exhausted along the trajectory, and $\{x_t^*\}$ must be optimal as long there are no finite period gains (the Euler condition).

Exercise 39 *Reproduce the proof of the sufficiency of the Euler plus transversality conditions for the optimal growth model. That is, show the following statement. Assume that a consumption path* $\mathbf{c}^*$ *solves* $u'(c_t^*) = \beta f'\left(k_{t+1}^*\right) u'\left(c_{t+1}^*\right)$ *for all* t*, that*

$$\lim_{T\to\infty} \beta^T u'(c_T^*)\, k_{T+1}^* = 0,$$

and that both u *and* f *are concave functions (with the usual interpretations). Then* $\mathbf{c}^*$ *is an optimal path for the optimal growth problem.*

Necessity of the Transversality Condition. A typical situation where the transversality is a necessary condition is when the capital stocks are bounded in the optimal growth model. One can of course derive a general proof of it. We will just provide the intuition behind it in a special case. Recall that in the optimal growth model the transversality condition is

$$\lim_{T\to\infty} \beta^T u'(c_T^*) k_{T+1}^* = 0.$$

In 'finance' terms, the Euler conditions state the unprofitability of reversed arbitrages; while the transversality condition defines a no-arbitrage condition for *unreversed arbitrages:* arbitrages which never return to the original path (Gray and Salant, 1983). Suppose (c, k) is an optimal path and suppose the agent decides to increase consumption in period 0, this is possible if he/she foregoes one unit of capital to be used in next period production. The marginal gain in period zero is $u'(c_0^*)$. Now let T any natural number and define a *T-period reversed arbitrage* the case where the planner reacquire the unit of capital foregone in period 0 only at time $T+1$. After period $T+1$ we are back to the original path.

This deviation from the optimal path $\{c_t^*, k_{t+1}^*\}_{t=0}^{\infty}$ generates two costs. First, the direct cost of reacquiring the capital in period $T+1$, which in period zero utility terms is $\beta^{T+1}u'(c_{T+1}^*)$. The indirect cost arises since the net marginal product of capital that unit of capital is lost every period between $t=1$ and $t=T+1$; this is a foregone shadow interest loss. The indirect cost at time t in period zero utils is

$$\beta^t u'(c_t^*)\left(f'(k_t^*)-1\right).$$

Adding to those losses the direct cost and equating to the marginal benefit yields the following zero marginal profit condition:

$$u'(c_0^*) = \sum_{t=1}^{T+1} \beta^t u'(c_t^*)\left(f'(k_t^*)-1\right) + \beta^{T+1}u'(c_{T+1}^*). \tag{3.13}$$

Notice that for $T=0$ the expression reduces to the Euler equation. It is also clear from this condition, that the unprofitability of one-period reversed arbitrage expressed via the Euler equations implies the unprofitability of any $T-$period reversed arbitrage [just rearrange terms to obtain a sum of Euler equations].

However, this is not the end of the story. The infinite horizon implies that the agent should contemplate also the possibility of an unreversed arbitrage, in which a the unit of capital is permanently sacrificed at $t=0$. Of course, there are not repurchase costs associated with this deviation. Hence the zero marginal profit condition for the unreversed arbitrage is

$$u'(c_0^*) = \sum_{t=1}^{\infty} \beta^t u'(c_t^*)\left(f'(k_t^*)-1\right)$$

but this equation is compatible with (3.13) as $T\to\infty$ only if

$$\lim_{T\to\infty} \beta^T u'(c_T^*) = 0,$$

which, in the bounded k_{T+1}^* case, implies the transversality condition. Thus, the transversality condition expresses the zero marginal profit condition for the open-ended arbitrages which are only admissible in the infinite horizon context. Hence both Euler equations and transversality are necessary for optimality. We of course know, that when the problem is concave, the Euler equation together with the transversality are sufficient conditions for optimality.

3.4 Optimal Control and the Maximization Principle of Pontryagin

Under the name "optimal control" one typically refer to an approach strictly linked with the Euler's approach and developed by Pontryagin et al. (1962).

In fact, one can study an optimal control problem with both the recursive and the variational techniques. The basic discrete-time optimal control problem consists in maximizing and an objective function of the form

$$\mathbf{U}_T = \phi(x_{T+1}) + \sum_{t=0}^{T} \beta^t u\left(x_t, c_t\right).$$

subject to a dynamic system

$$x_{t+1} = f(x_t, c_t), \tag{3.14}$$

where f describes how today's *control* $c_t \in C$ affects future state x_{t+1}, given today's state x_t, and $\phi(x_{T+1})$ summarizes the final effect of the state. The set C is any direct restriction on controls other than the law of motion. As usual, we have the initial condition on state $x_0 = \overline{x}$. One can use the natural extension of the Bellman approach to study such a problem. For example, the Bellman equation for this problem is

$$V_{t+1}(x) = \sup_{c \in C} \; u(x, c) + \beta V_t(f(x, c))$$

with $V_0(x_{T+1}) = \phi(x_{T+1})$.

One can see the problem in the joint space of control and state plans and apply a generalized version (for infinite dimensional spaces) of the Lagrange multiplier theorem to it. Another approach is, however, to note that (3.14) uniquely determines the path of states $\mathbf{x} = \{x_t\}$ once the path of controls $\mathbf{c} = \{c_t\}$ is specified and hence we really have to select $\mathbf{c}$ with objective

$$\mathbf{U}_T = \mathbf{U}_T(\mathbf{c}) = \phi(x_{T+1}(\mathbf{c})) + \sum_{t=0}^{T} \beta^t u\left(x_t(\mathbf{c}), c_t\right).$$

Still another approach is to view the problem in the space of states by considering the implicitly defined set of all trajectories that can be obtained by application of admissible controls. Each of these approaches has theoretical advantages for the purpose of deriving necessary conditions and practical advantages for the purpose of developing computational procedures for obtaining solutions.

The Maximum Principle of Pontryagin The Maximum Pontryagin's principle is typically associated to optimal control problems since it emphasizes the role of controls alone. It gives a set of *necessary* conditions for optimality. It is assumed that f is such that given x_0, a given path $\mathbf{x}$ is uniquely determined by $\mathbf{c}$, hence the objective functional can be considered to be dependent only on $\mathbf{c}$. It is typically assumed that both u and f have partial derivatives with respect to x which are jointly continuous in (c, x). We will further assume that the optimal trajectory is interior with respect to C.

Theorem 17 *Let $\{x_{t+1}, c_t\}_{t=0}^{T}$ the optimal control and state trajectory for the above optimal control problem given x_0. Then under some regularity conditions there is an adjoint trajectory $\{\lambda_{t+1}\}_{t=0}^{T}$, such that, given x_0, $\{x_{t+1}, c_t, \lambda_{t+1}\}_{t=0}^{T}$ satisfy:*

$$\begin{aligned} x_{t+1} &= f(x_t, c_t) && \textit{system equation} && (3.15)\\ \lambda_t &= H_x(\lambda_{t+1}, x_t, c_t) && \textit{adjoint equation} && (3.16)\\ \lambda_{T+1} &= \phi_x(x_{T+1}) && \textit{adjoint final equation} && (3.17)\\ 0 &= H_c(\lambda_{t+1}, x_t, c_t) && \textit{variational condition} && (3.18) \end{aligned}$$

where H is the Hamiltonian function

$$H(\lambda_{t+1}, x_t, c_t) = u\left(x_t, c_t\right) + \beta\lambda_{t+1} f(x_t, c_t).$$

Proof (Sketch). We do not investigate what are the regularity conditions that guarantee the existence of the multipliers $(\lambda_{t+1})_{t=0}^{T}$. Under weaker conditions one can show that a similar result can be shown by using an extended Hamiltonian of the form $\hat{H} = p_t u\left(x_t, c_t\right) + \beta\lambda_{t+1} f(x_t, c_t)$ for some sequence $\{p_t\}$. The core part of the proof has two main ingredients. The adjoint equations and the adjoint final condition on one hand, together with the fact that we can approximate arbitrarily well the value of a deviation by a first order Taylor's expansion when the deviation is *small*. Notice indeed that the proof considers only deviations around the optimal path, and that the statement has a local nature. Specifically, let us write the Lagrangian associated with the above problem

$$L(\lambda, \mathbf{x}, \mathbf{c}) = \sum_{t=0}^{T} \beta^t u\left(x_t, c_t\right) + \phi(x_{T+1}) + \sum_{t=0}^{T} \beta^{t+1} \lambda_{t+1} \left[f(x_t, c_t) - x_{t+1}\right].$$

The key point of the approach is that once we take into account the adjoint equation, and if we consider local variations, we can check only deviations of the controls. More precisely, from the regularity properties of the problem one can show that along a plan that satisfies the adjoint equation the total value of the program can be approximated in terms of first order by a variation of the Lagrangian in only the controls, keeping states as constant. That is, suppose $(\mathbf{x}^*, \mathbf{c}^*, \lambda^*) = (x_{t+1}, c_t, \lambda_{t+1})_{t=0}^T$ is a triplet with $\mathbf{c} \in C$, and satisfying the adjoint and the final adjoint equations, given x_0, we have:

$$\mathbf{U}_T(\mathbf{c}^*) - \mathbf{U}_T(\mathbf{c}) = L(\lambda^*, \mathbf{x}^*, \mathbf{c}^*) - L(\lambda^*, \mathbf{x}^*, \mathbf{c}) + o\left(\|\mathbf{c}^* - \mathbf{c}\|\right),$$

with $\lim_{\mathbf{c} \to \mathbf{c}^*} \frac{o(\|\mathbf{c}^* - \mathbf{c}\|)}{\|\mathbf{c}^* - \mathbf{c}\|} = 0$. The last step is to realize that if states have second order effects (along the plan satisfying the adjoint equation) deriving deviations of the Lagrangian we can ignore any component that does not depend explicitly from the controls $\mathbf{c}$. This implies that it is enough to consider deviations in the Hamiltonian alone. Indeed we can rewrite the Lagrangian emphasizing the Hamiltonian as follows

$$\begin{aligned} L(\lambda, \mathbf{x}, \mathbf{c}) &= \phi(x_{T+1}) + \sum_{t=0}^{T} \beta^t \left[u\left(x_t, c_t\right) + \beta \lambda_{t+1} f(x_t, c_t) - \beta \lambda_{t+1} x_{t+1}\right] \\ &= \phi(x_{T+1}) + \sum_{t=0}^{T} \beta^t \left[H(\lambda_{t+1}, x_t, c_t) - \beta \lambda_{t+1} x_{t+1}\right], \end{aligned}$$

and notice that the remaining addends $\beta \lambda_{t+1} x_{t+1}$ can be ignored if $\mathbf{x}$ and λ satisfy the adjoint equations. See also Luenberger (1969), page 262. **Q.E.D.**

Notice that the variational condition in the above theorem is expressed in terms of stationary point for H. In continuous time this condition requires c^* to maximize the Hamiltonian each period. In fact, this distinction represents an important difference between the continuous time and the discrete time versions of the Maximum Principle. The idea behind this fact is simple. In continuous time one can construct "small" deviation in controls by varying a lot the path, but for a very short period of time: $\int_0^T |c(t) - u(t)|\, dt < \varepsilon$. This is not possible in discrete time, where to have a "small" deviation one must remain close to the optimal path for any t. As a consequence, the Pontryagin's Maximum Principle in much more powerful in continuous time than when the time is discrete.

In continuous time, Mangasarian (1966) and Arrow and Kurz (1970) derived sufficient conditions for optimality. Mangasarian showed that if H is concave in (x, c) (and C convex), the necessary conditions of the Pontryagin's maximum theorem become also sufficient for an optimum. The discrete time version of the sufficiency theorem would be as follows[13]

Proposition 4 *Let* $\left\{x_{t+1}^*, c_t^*, \lambda_{t+1}^*\right\}_{t=0}^T$ *a sequence satisfying all the conditions of Theorem 17 above. Moreover assume that* $\lambda_{t+1}^* \geq 0$ *for all* t*, and that both* u*,* f *and* ϕ *are concave in* (x_t, c_t)*, then the sequence* $\left(x_{t+1}^*, c_t^*\right)$ *is a global optimum for the problem.*

[13] See Takayama (1985), especially pages 660-666, for a didactical review in continuous time.

Proof. The proof uses the fact that a concave function is subdifferentiable to show a sequence of key inequalities using a similar derivation to that in Proposition 3. The condition $\lambda_{t+1}^* \geq 0$ only guarantees that when both u and f are concave then H is concave in (x_t, c_t). When f is linear it can be dispensed. **Q.E.D.**

Static Maximization and Pontryagin We have already mentioned that the maximum principle is basically an extension of the Lagrangian theorem. It improves at least in two direction. First, the theorem is particularly suitable for infinite dimensional spaces. The infinite dimensional version of the Lagrange theorem uses the same line of proof of the usual Lagrange theorem in finite dimensional spaces. However, the generalized Inverse Function Theorem of Liusternik is by no mean a simple result.[14] In addition, in continuous time, the theorem is not stated in terms of derivatives with respect to c, hence it allow for non differentiable cases. For example, the method allows for both corner and bang-bang solutions.

Consider the following exercise.

Exercise 40 *(i) Write the neoclassical growth model in terms of an optimal control problem. That is, distinguish states x from controls c, and specify f, u and C for this problem. [Hint: you might want to write the feasible set for controls C as a function of the state]. (ii) Next, derive the Euler equations from the Pontryagin maximum principle, and interpret economically the adjoint variables λ_t.*

Using the Pontryagin maximum principle one can deal perhaps with a larger class of problems than the one covered by the Euler's variational approach. In both cases however, when the horizon is infinite, one need to derive appropriate transversality condition.

[14] See Chapter 9 of Luenberger (1969).

Bibliography

[1] Aliprantis, C. D. and K.C. Border (1994) *Infinite Dimensional Analysis*, Springer-Verlag, Berlin.

[2] Arrow, K.J., and M. Kurz (1970) *Public Investment, the Rate of Return, and Optimal Fiscal Policy*, John Hopkins Press, Baltimore.

[3] Bellman, R. (1957) *Dynamic Programming.* Princeton, NJ: Princeton University Press.

[4] Bertsekas, D. P. (1976) *Dynamic Programming and Stochastic Control*, Academic Press, NY.

[5] Bertsekas, D. P., and S. E. Shreve (1978) *Stochastic Control: The Discrete Time Case.* New York: Academic Press.

[6] Cugno, F. and L. Montrucchio (1998) *Scelte Intertemporali*, Carocci, Roma.

[7] Luenberger, D. G. (1969) *Optimization by Vector Space Methods*, John Wiley, USA.

[8] Mangasarian, O. L. (1966) "Sufficient Conditions for the Optimal Control of Non-linear Systems," Journal of SIAM Control, 4, February.

[9] Mas-Colell, A., Michael D. Winston and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press.

[mopri] Montrucchio, L., and F. Privileggi (1998), Fast Growing Stochastic Dynamic Models, *22nd A.M.A.S.E.S. Conference, Atti.*

[10] Rockafellar, R. T. (1970) *Convex Analysis,* Princeton, N.J.: Princeton University Press.

[11] Rudin, W. (1991) *Functional Analysis*, McGraw Hill Inc.,USA.

[12] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

Chapter 4

Stochastic Dynamic Programming

The aim of this chapter is to extend the framework we introduced in Chapter 3 to include uncertainty. To evaluate decisions, we use the well known expected utility theory.[1] With uncertainty we will face Bellman equations of the following form

$$V(x,z) = \sup_{x' \in \Gamma(x,z)} F(x,x',z) + \beta \mathbf{E}\left[V(x',z') \mid z\right], \tag{4.1}$$

where z is a stochastic component, assumed to follow a (stationary) first order *Markov Process.* A first order Markov process is a sequence of random variables $\{z_t\}_{t=0}^{\infty}$ with the property that the conditional expectations depend only on the last realization of the process, that is if C is a set of possible values for z, then

$$\Pr\left\{z_{t+1} \in C \mid z_t, z_{t-1}, ..., z_0\right\} = \Pr\left\{z_{t+1} \in C \mid z_t\right\}.$$

To make the above statements formally meaningful we need to review some concepts of Probability Theory.

4.1 The Axiomatic Approach to Probability: Basic Concepts of Measure Theory

I am sure you are all familiar with the expression $\Pr\left\{z_{t+1} \in C \mid z_t\right\}$ for conditional probabilities, and with the conditional expectation operator $\mathbf{E}\left[\cdot \mid z\right]$ in (4.1). Probability theory is a special case of the more general and very powerful *Measure Theory,* first formulated in 1901 by Henri Léon Lebesgue.[2]

[1] For a review on the theories of decisions under uncertainty see Machina (1987).

[2] This outstanding piece of work appears in Lebesgue's dissertation, *Intégrale, Longueur, Aire*, presented to the University of Nancy in 1902.

We first introduce a set Z which will be our *sample space.* Any subset E of Z, will be denoted as an *event.* In this way, all results of set theory - unions, intersections, complements, ... - can be directly applied to events as subsets of Z. To each event we also assign a "measure" $\mu(E) = \Pr\{E\}$ called *probability of the event.* These values are assigned according to the function μ which has by assumption the following properties (or axioms):

1. $0 \leq \mu(E) \leq 1$;
2. $\mu(Z) = 1$;
3. For any finite or infinite sequence of disjoint sets (or mutually exclusive events) $E_1, E_2, ...$; such that $E_i \cap E_j = \emptyset$ for any i, j, we have

$$\mu\left(\cup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} \mu\left(E_i\right) \text{ where } N \text{ possibly equals } \infty.$$

All properties 1-3 are very intuitive for probabilities. Moreover, we would intuitively like to consider E as any subset of Z. Well, if Z is a finite or countable set then E can literally be *any* subset of Z. Unfortunately, when Z is a uncountably infinite set - such as the interval $[0, 1]$ for example - it might be impossible to find a function μ defined on all possible subsets of Z and at the same time satisfying all the three axioms we presented above. Typically, what fails is the last axiom of additivity when $N = \infty$. Lebesgue managed to keep property 3 above by defining the measure function μ only on the so-called *measurable sets* (or events). This is not an important limitation, as virtually all events of any practical interest tuned out to be measurable. Actually, in applications one typically considers only some class of possible events. A subset of the class of all measurable sets.

The reference class of sets $\mathcal{Z}$ represents the *set of possible events,* and will constitute a σ-algebra.[3] Notice that $\mathcal{Z}$ is a set of sets, hence an event E is an element of $\mathcal{Z}$, i.e. in contrast to $E \subset Z$ we will write $E \in \mathcal{Z}$. The pair $(Z, \mathcal{Z})$ constitutes a measurable space while the turple $(\mu, Z, \mathcal{Z})$ is denotes as a measured (or probability) space.

[3] A family $\mathcal{Z}$ of subsets of Z is called a σ algebra if: *(i)* both the empty set $\emptyset$ and Z belong to $\mathcal{Z}$; *(ii)* If $E \in \mathcal{Z}$ then also its complement (with respect to Z) $E^c = Z \backslash E \in \mathcal{Z}$; and *(iii)* for any sequence of sets such that $E_n \in \mathcal{Z}$ for all $n = 1, 2, ...$ we have that the set $(\cup_{n=1}^{\infty} E_n) \in \mathcal{Z}$. It is easy to show that whenever $\mathcal{Z}$ is a σ-algebra then $(\cap_{n=1}^{\infty} E_n) \in \mathcal{Z}$ as well. When Z is a set of real numbers, we can consider our set of possible events as the *Borel* σ-algebra. Which is σ-algebra 'generated' by the set of all open sets.

I am sure it is well known to you that the expectation operator $\mathbf{E}[\cdot]$ in (4.1) is nothing more than an integral, or a summation when z takes finitely or countably many values. For example, assume p_i is the probability that $z = z_i$. The expectation of the function f can be computed as follows

$$\mathbf{E}[f(z)] = \sum_{i=1}^{N} p_i f(z_i).$$

One of the advantages of the Lebesgue theory of integration is that, for example, it includes both summations and the usual concept of (Riemann) integration in an unified framework. We will be able to compute expectations[4]

$$\mathbf{E}[f(z)] = \int_Z f(z) d\mu(z)$$

no matter how Z is and no matter what is the distribution μ of the events. For example, we can deal with situations where Z is the interval $[0, 1]$ and the event $z = 0$ has a positive probability $\mu(0) = p_0$. Since the set of all measurable events $\mathcal{Z}$ does not include all possible subsets of Z, we must restrict the set of functions f for which we can take expectations (integrals) as well.

Definition 32 *A real valued function f is measurable with respect to $\mathcal{Z}$ if for every real number x the set*

$$E_f^x = \{z \in Z : f(z) \geq x\}$$

belongs to the set of events $\mathcal{Z}$.

Sometimes we do a sort of inverse operation. We have in mind a class of real valued functions $\mathcal{F}$, each one defined over the set of events Z. We define a σ-algebra $\mathcal{Z}_{\mathcal{F}}$ so that to have every $f \in \mathcal{F}$ measurable, and any such function $f \in \mathcal{F}$ is called as *random variable*.

Definition 33 *The Lebesgue integral of a measurable positive function $f \geq 0$ is defined as follows*

$$\int_Z f(z) d\mu(z) = \sup_{0 \leq \phi \leq f} \int_Z \phi(z) d\mu(z) = \inf_{\phi \geq f \geq 0} \int_Z \phi(z) d\mu(z).$$

[4] When at z the measure μ has a density, the notation $d\mu(z)$ corresponds to the more familiar $f_\mu(z)\, dz$. When μ does not admits density, $d\mu(z)$ it is just the notation we use for its analogous concept.

In the definition, ϕ is any *simple (positive) function* (in its standard representation), that is, ϕ is a finite weighted sum of indicator functions[5]

$$\begin{aligned}\phi(z) &= \sum_{i=1}^{n} a_i I_{E_i}(z);\ a_i \geq 0;\ \text{and its integral is} \\ \int_Z \phi(z) d\mu(z) &= \sum_{i=1}^{n} a_i \mu(E_i),\end{aligned}$$

where for each i, j $E_i \cap E_j = \emptyset$; and $\cup_{i=1}^{n} E_i = Z$.

The Lebesgue integral of f is hence (uniquely) defined as the supremum of integrals of nonnegative dominated simple functions ϕ : such that for all z, $0 \leq \phi(z) \leq f(z)$; which in turn coincides with the infimum over all the dominating simple functions: $\phi \geq f$. We do not have space here to discuss the implications of this definition[6] however, one should recall from basic analysis that the Riemann integral, that is the "usual" integral we saw in our undergraduate studies, can be defined in a similar way; where instead of simple functions one uses step functions. One can show that each function f which is Riemann integrable it is also Lebesgue integrable, and that there are simple examples where the converse is false.[7]

4.2 Markov Chains and Markov Processes

Markov Chains We now analyze in some detail conditional expectations for the simple case where Z is finite. So, assume that the stochastic component z can take finitely many values, that is $z \in Z \equiv \{z_1,\ z_2,\ ...,\ z_N\}$, with corresponding *conditional* probabilities

$$\pi_{ij} = \Pr\{z' = z_j \mid z = z_i\},\ i, j = 1, 2, ..., N.$$

[5]The indicator function of a set E is defined as

$$I_E(z) = \begin{cases} 1 \text{ if } z \in E \\ 0 \text{ otherwise.} \end{cases}$$

[6]See for example SLP, Ch. 7.

[7]One typical counter-example is the function $f : [0, 1] \to [0, 1]$ defined as follows

$$f(z) = \begin{cases} 1 \text{ if } z \text{ is rational} \\ 0 \text{ otherwise.} \end{cases}$$

This function is Lebesgue integrable with $\int f(x)dx = 0$, but it is not Riemann integrable.

Since π_{ij} describes the probability of the system to move to state z_j if the previous state was z_i, they are also called *transition probabilities* and the stochastic process form a *Markov chain.* To be probabilities, the π_{ij} must satisfy

$$\pi_{ij} \geq 0, \text{ and } \sum_{j=1}^{N} \pi_{ij} = 1 \text{ for } i = 1, 2, ..., N,$$

that is, they must belong to a $(N-1)-$dimensional simplex Δ^N. It is typically convenient to arrange the transition probabilities in a square array as follows

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \cdots & \cdots \\ \cdots & \cdots & \pi_{ij} & \cdots \\ \pi_{N1} & \cdots & \cdots & \pi_{NN} \end{bmatrix}$$

Such an array is called *transition matrix* or *Markov matrix*, or *stochastic matrix.* If the probability distribution over the state in period t is $p^t = (p_1^t, p_2^t, ...p_N^t)$, the distribution over the state in period $t+1$ is $p^t\Pi = \left(p_1^{t+1}, p_2^{t+1}, ...p_N^{t+1}\right)$, where

$$p_j^{t+1} = \sum_{i=1}^{N} p_i^t \pi_{ij}, \ j = 1, 2, ..., N.$$

For example, suppose we want to know what is the distribution of the next period states if in the current period the is z_i. Well, this means that the initial distribution is a degenerate one, namely $p^t = e_i = (0, \ ..., \ 1, \ ..., \ 0)$. As a consequence, the probability distribution over the next period state is the $i-$th row of $\Pi : e_i\Pi = (\pi_{i1}, \ \pi_{i2}, ...\pi_{iN})$. Similarly, if p^t is the period t distribution, then by the properties of the matrix multiplication, $p^t\Pi^n = p(\Pi \cdot \Pi \cdot ...\Pi)$ is the $t+n$ period distribution p^{t+n} over the states. It is easy to see that if Π is a Markov matrix then so is Π^n. A set of natural question then arises. Is there a stationary distribution, that is a probability distribution p^* with the property $p^* = p^*\Pi$? Under what conditions can we be sure that if we start from any initial distribution p^0, the system converges to a unique limiting probability $p^* = \lim_{n\rightarrow\infty} \{p^0\Pi^n\}$?

The answer to the first question turns out to always be affirmative for Markov chains.

Theorem 18 *Given a stochastic matrix* Π, *there always exists at least one stationary distribution* p^* *such that* $p^* = p^*\Pi$, *with* $p_i^* \geq 0$ *and* $\sum_{i=1}^N p_i^* = 1$.

Proof. Notice that a solution to the system of equations $p^* = p^*\Pi$ corresponds to solving $p^*(I - \Pi) = 0$, where I is the N dimensional identity matrix. Transposing both sides of the above equation gives

$$(I - \Pi')p^{*\prime} = 0.$$

So p^* is a nonnegative eigenvector associated with a unit eigenvalue of Π', normalized to satisfy $\sum_i p_i^* = 1$. So we can use linear algebra to show this result. Thanks to the Leontief's Input-Output analysis, during the 50s and 60s economics (re)discovered many important theorems about matrices with nonnegative elements. Any matrix with nonnegative elements has a Frobenius root $\lambda \geq 0$ with associated a nonnegative eigenvector. This existence result is the most difficult part of the proof and is due to Frobenius (1912) (See also Takayama, 1996, Theorem 4.B.2, pp. 375). Fisher (1965) and Takayama (1960) showed that when the elements of each column of the a matrix with nonnegative elements sum to one then its Frobenius root equals one, i.e. $\lambda = 1$ (Takayama, 1996, Theorem 4.C.11, pp. 388). The proof of this last statement is simple: let $p^* \geq 0$ the eigenvector associated with λ. By definition $\lambda p^* = \Pi' p^*$, that is, $\lambda p_i^* = \sum_j \pi'_{ij} p_j^*$, $i = 1, 2, ..., N$. Summing up over i, we obtain

$$\lambda \sum_{i=1}^{N} p_i^* = \sum_{i=1}^{N} \sum_{j=1}^{N} \pi'_{ij} p_j^* = \sum_{j=1}^{N} p_j^* \left(\sum_{i=1}^{N} \pi'_{ij} \right)$$

since Π' is the transpose of Π, $\sum_{i=1}^{N} \pi'_{ij} = \sum_{j=1}^{N} \pi_{ij} = 1$. Hence $\lambda = \frac{\sum_{j=1}^{N} p_j^*}{\sum_{i=1}^{N} p_i^*} = 1$. **Q.E.D.**

Consider now the second question. Can we say that p^* is unique? Unfortunately, in order to guarantee that the sequence of matrices converges to a unique matrix P^* with identical rows p^*, (so that for any p we have $pP^* = p^*$), we need some further assumptions, as the next exercises shows.

Exercise 41 *Assume that a Markov chain (with $Z = \{z_1, z_2\}$) is summarized by the following transition matrix*

$$\Pi = \begin{bmatrix} \pi_{11} & 1 - \pi_{11} \\ 1 - \pi_{22} & \pi_{22} \end{bmatrix}.$$

A stationary distribution is hence a vector $(q^*, 1 - q^*)$ with $1 \geq q^* \geq 0$ such that $(q^*, 1 - q^*) \cdot \Pi = (q^*, 1 - q^*)$. We know from above that at least one such q^* must exists.
(a) *Using simple algebra show that q^* solves $(2 - \pi_{22} - \pi_{11}) q^* = (1 - \pi_{22})$, and discuss conditions where q^* might take multiple values.*
(b) *Now set $\pi_{11} = \pi_{22} = \pi$ and state conditions for q^* to be unique.*

Here is a set of sufficient conditions for uniqueness.

Theorem 19 *Assume that* $\pi_{ij} > 0$ *for all* $i, j = 1, 2, ...N$. *There exists a limiting distribution* p^* *such that*

$$p_j^* = \lim_{n\to\infty} \pi_{ij}^{(n)},$$

where $\pi_{ij}^{(n)}$ *is the* (i, j) *element of the matrix* Π^n. *And* p_j^* *are the unique nonnegative solutions of the following system of equations*

$$\begin{aligned} p_j^* &= \sum_{k=1}^{N} p_k^* \pi_{kj}; \text{ or } p^* = p^*\Pi; \text{ and} \\ \sum_{j=1}^{N} p_j^* &= 1. \end{aligned}$$

Proof. See below. **Q.E.D.**

The application of the transition matrix on a probability distribution p can be seen as a mapping of the $(N-1)-$dimensional simplex into itself. In fact, under some conditions, the operator

$$\begin{aligned} T_\Pi &: \Delta^N \to \Delta^N \\ T_\Pi p &= p\Pi \end{aligned} \tag{4.2}$$

defines a contraction on the metric space $\left(\Delta^N, |\cdot|_N\right)$ where

$$|x|_N \equiv \sum_{i=1}^{N} |x_i| .$$

Exercise 42 *(i) Show that* $\left(\Delta^N, |\cdot|_N\right)$ *is a complete metric space. (ii) Moreover, show that if* $\pi_{ij} > 0$, $i, j = 1, 2, ...$; *the mapping* T *in (4.2) is a contraction of modulus* $\beta = 1-\varepsilon$, *where* $\varepsilon = \sum_{j=1}^{N} \varepsilon_j$ *and* $\varepsilon_j = \min_i \pi_{ij} > 0$.

When some $\pi_{ij}^{(n)} = 0$, we might loose uniqueness. However, following the same line of proof one can show that the stationary distribution is unique as long as $\varepsilon = \sum_{j=1}^{N} \varepsilon_j > 0$. Could you explain intuitively why this is the case?

Moreover, from the contraction mapping theorem, it is easy to see that the above proposition remains valid if the assumption $\pi_{ij} > 0$ is replaced with: there exists a $n \geq 1$ such that $\pi_{ij}^{(n)} > 0$ for all i, j. (see Corollary 2 of the contraction mapping Theorem (Th. 3.2) in SLP).

Notice that the sequence $\{\Pi^n\}_{n=0}^{\infty}$ might not always converge. For example, consider $\Pi = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. It is easy to verify that the sequence jumps from $\Pi^{2n} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and

$\Pi^{2n+1} = \Pi$. However, the fact that in a Markov chain the state space is finite implies that the long-run averages

$$\left\{ \frac{1}{T} \sum_{t=0}^{T-1} \Pi^t \right\}_{T=1}^{\infty}$$

do always converge to a stochastic matrix P^*, and the sequence $p^t = p^0 \Pi^t$ converges to

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} p^t = p^0 P^*.$$

In the example we saw above one can easily verify that $\frac{1}{T}\sum_{t=0}^{T-1} \Pi^t \to P^* = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$, and the unique stationary distribution is $p^* = (1/2, 1/2)$.

In other cases, the rows of the limit matrix P^* are not necessarily always identical to each other. For example, consider now the transition matrix $\Pi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. It is obvious that in this case $P^* = \Pi$, which has two different rows. It is also clear that both rows constitute a stationary distribution. This is true in general: any row of the limit matrix P^* is an invariant distribution for the transition matrix Π.

What is perhaps less obvious is that any convex combination of the rows of P^* constitute a stationary distribution, and that all invariant distributions for Π can be derived by making convex combinations of the rows of P^*.

Exercise 43 *(i) Consider first the above example with* $P^* = \Pi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. *Show that any vector* $p^*_\lambda = (\lambda, 1-\lambda)$ *obtained as a convex combination of the rows of* P^* *constitutes a stationary distribution for* Π. *Provide an intuition for the result.* (ii) *Now consider the general case, and let* p^* *and* p^{**} *two stationary distributions for a Markov chains defined by a generic stochastic matrix* Π. *Show that any convex combination* p_λ *of* p^* *and* p^{**} *constitute a stationary distribution for* Π.

Markov Processes The more general concept corresponding to a Markov chain, where Z can take countably or uncountably many values, is denoted as a Markov Process. Similarly to the case where Z is finite, a Markov process is defined by a *transition function* (or kernel) $Q : Z \times \mathcal{Z} \to [0,1]$ such that: *(i)* for each $z \in Z$ $Q(z, \cdot)$ is a probability measure; and *(ii)* for each $C \in \mathcal{Z}$ $Q(\cdot, C)$ is a measurable function.

Given Q, one can compute conditional probabilities

$$\Pr\{z_{t+1} \in C \mid z_t = c\} = Q(c, C)$$

and conditional expectations in the usual way

$$\mathbf{E}[f \mid z] = \int_Z f(z')dQ(z,z').$$

Notice that Q can be used to map probability measure into probability measures since for any μ on $(Z, \mathcal{Z})$ we get a new $\mu\prime$ by assigning to each $C \in \mathcal{Z}$ the measure

$$(T_Q\mu)(C) = \mu'(C) = \int_Z Q(z,C)d\mu(z),$$

and T is denoted as *Markov operator.*

We now define a very useful property for Q.

Definition 34 *Q has the **Feller property** if for any bounded and continuous function f the function*

$$g(z) = (\mathbf{P}_Q f)(z) = \mathbf{E}[f \mid z] = \int f(z')dQ(z,z') \text{ for any } z$$

is still bounded and continuous.

The above definition first of all shown another view of Q. It also defines an operator (sometimes called transition operator) that in general maps bounded and measurable functions into bounded measurable functions. When Q has the feller property the operator $\mathbf{P}_Q$ preserves continuity.

Technical Digression (optional). It turns out that the Feller property characterizes continuous Markov transitions. The rigorous idea is simple. Let $\mathbf{M}$ be the set of all probability measures on Borel sets $\mathcal{Z}$ over a metrizable space Z, and for each z, let $Q(z,\cdot)$ a member of $\mathbf{M}$. The usual topology defined in the space of Borel measures is the topology of convergence in distribution (or weak topology).[8] It is now useful to make pointwise considerations. For each z the probability measure $Q(z,\cdot)$ can be seen as a linear mapping from the set of bounded and measurable functions into the real numbers according to $x = \langle f, Q(z,\cdot)\rangle = \int f(z')dQ(z,z')$.

It turns out that a transition function $Q : Z \to \mathbf{M}$ is continuous if and only if it has the Feller property. The fact that a continuous Q has the Feller property is immediate: By definition of the topology defined on $\mathbf{M}$ (weak topology), via the map $F_f(\mu) = \langle f, \mu\rangle$ each

[8] In this topology, a sequence $\{\mu_n\}$ in M converges to μ if and only if $\int f d\mu_n \to \int f d\mu$ for all continuous and bounded functions f.

continuous and bounded function $f : Z \to \mathbb{R}$ defines a continuous real valued function $F_f : \mathbf{M} \to \mathbb{R}$.[9] Now note that when μ is $Q(z, \cdot)$ we have $F_f(Q(z, \cdot)) = (\mathbf{P}_Q f)(z)$. Now, continuity of Q means that as $z_n \to z$ we have $Q(z_n, \cdot) \to Q(z, \cdot)$ in $\mathbf{M}$. Equivalently, if we let $\mu_n(\cdot) = Q(z_n, \cdot)$ and use the usual topology on $\mathbf{M}$, continuity of Q means that $F_f(Q(z_n, \cdot)) \to F_f(Q(z, \cdot))$ (interpreted now as sequence of real numbers). We have hence established that $(\mathbf{P}_Q f) = g$ is a continuous function in Z, i.e. that Q has the Feller property. In order to show rigorously that the Feller property implies continuity - although it is intuitive - one needs some more work.[10]

We can now study the issue of existence and uniqueness of a stationary distribution. A stationary distribution for Q is a measure μ^* on $(Z, \mathcal{Z})$ such that for any $C \in \mathcal{Z}$

$$\mu^*(C) = \int_Z Q(z, C) d\mu^*(z),$$

that is μ^*, is a fixed point of the Markov operator T_Q. There are many results establishing existence and uniqueness of a stationary distribution. Here is a result which is among the easiest to understand, and that uses the Feller property of Q.

Theorem 20 *If Z is a compact set and Q has the Feller property then there exists a stationary distribution $\mu^* : \mu^* = T_Q \mu^*$, where $\mu = \lambda$ if and only if $\int f d\mu = \int f d\lambda$ for each continuous and bounded function f.*

Proof. See SLP, Theorem 12.10, page 376-77. The basic idea of the proof can also be get as an application of one of the infinite dimensional extensions of the Brower fixed point theorem (usually called Brower-Shauder-Tyconoff fixed point). We saw above that whenever Q has the Feller property, the associated Markov operator T_Q is a continuous map from the compact convex (locally convex Hausdorff) space of distributions Λ into itself. [See Aliprantis and Border (1994), Corollary 14.51, page 485] **Q.E.D.**

Similarly to the finite state case, this invariant measure can be obtained by looking at the sequence $\left\{ \frac{1}{T} \sum_{t=1}^{T-1} T_Q^t \lambda_0 \right\}_{T=1}^{\infty}$ of T-period averages.

When the state space is not finite, we may define several different concepts of converge for distributions. The most known ones are weak convergence (commonly denoted convergence in distribution) and strong convergence (or convergence in total variation norm, also denoted as setwise convergence). We are not dealing with these issues in these class notes. The concept of weak convergence is in most cases all that we care about in the

[9] Let $x_{\mu_n} = F_f(\mu_n)$. By definition of weak topology, if $\mu_n \to \mu$ then $F_f(\mu_n) \to F_f(\mu)$.

[10] The interest reader can have a look at Aliprantis and Border (1994), Theorem 15.14, page 531-2.

context of describing the dynamics of an economic system. Theorem 20 deals with weak convergence. The most known results of uniqueness use some monotonicity conditions on the Markov operator, together with some mixing conditions. For a quite general treatment of monotonic Markov operators, with direct applications to economics and dynamic programming, see Hopenhayn and Prescott (1992).

If we require strong convergence, one can guarantee uniqueness under conditions similar to those of Theorem 19, using the contraction mapping theorem. See Chapter 11 in SLP, especially Theorem 11.12.

4.3 Bellman Principle in the Stochastic Framework

The Finite Z case. When the shocks belong to a finite set all the results we saw for the deterministic case are true for the stochastic environment as well. The Bellman Principle of optimality remains true since both Lemma 1 and 2 remain true. Expectation are simply a weighted sums of the continuation values. In this case Theorem 12 remains true under the same conditions as in the deterministic case. From the proof of Theorem 13 and 14 it is easy to see that also the verification and sufficiency theorems can easily be extended to the stochastic case with finite shocks. We just need to require boundedness to be true for all z. Even the Theorems 15 and 16 are easily extended to the stochastic case following the same lines of proof we proposed in Chapter 3.1. In order to show you that there is practically no difference between the deterministic and the stochastic case when Z is finite, let me be a bit boring and consider for example the stochastic extension of Theorem 15. Assume w.l.o.g. that z may take N values, i.e. $Z = (z_1,\ z_2,\ ...,\ z_N)$. We can always consider our fixed point

$$V(x, z_i) = \sup_{x' \in \Gamma(x, z_i)} F(x, x', z_i) + \beta \sum_{j=1}^{N} \pi_{ij} V(x', z_j),\ \forall i$$

in the space $\mathcal{C}_N(X)$ of vectors of real valued functions:

$$\mathbf{V}(x) = (V(x, z_1), ..., V(x, z_N)) = (V_1(x), ..., V_N(x))$$

which are continuous and bounded in X with the metric d_∞^N, where[11]

$$d_\infty^N(\mathbf{V},\mathbf{W}) = \sum_{i=1}^N d_\infty(V_i, W_i) = \sum_{i=1}^N \sup_x |V(x,z_i) - W(x,z_i)|.$$

One can easily show that such metric space of functions is complete, and that the same conditions for a contraction in the deterministic case can be used here to show that the operator

$$\begin{aligned} T &: \mathcal{C}_N(X) \to \mathcal{C}_N(X) \\ T\mathbf{V}(x) &= \begin{cases} \sup_{x'\in\Gamma(x,z_1)} F(x,x',z_1) + \beta \sum_{j=1}^N \pi_{1j} V(x',z_j) \\ \sup_{x'\in\Gamma(x,z_2)} F(x,x',z_2) + \beta \sum_{j=1}^N \pi_{2j} V(x',z_j) \\ \dots \\ \sup_{x'\in\Gamma(x,z_N)} F(x,x',z_N) + \beta \sum_{j=1}^N \pi_{Nj} V(x',z_j) \end{cases} \end{aligned}$$

is a contraction with modulus β. It is easy to see that both boundedness and - by the Theorem of the Maximum - continuity is preserved under T. Similarly, given that (conditional) expectations are nothing more than convex combinations, concavity is preserved under T, and the same conditions used for the deterministic case can be assumed here to guarantee the stochastic analogous to Theorem 16.

The General case When Z is continuous, we need to use measure theory. We need to assume some additional technical restrictions to guarantee that the integrals involved in the expectations and the limits inside those integrals are well defined.

Unfortunately, these technical complications prevent the possibility of having a result on the lines of Theorem 12. The reason is that we one cannot be sure that the true value function is measurable. As a consequence, the typical result in this case are in form of the verification or sufficiency theorems. Before stating formally the result we need to introduce some notation.

Definition 35 *A plan π is an initial value $\pi_0 \in X$ and a sequence of (h^t-measurable) functions*[12]

$$\pi_t : H^t \to X$$

[11] Another possibility is to use $d_\infty^{\max}$

$$d_\infty^{\max}(\mathbf{V},\mathbf{W}) = \max_i \{d_\infty(V_i,W_i)\} = \max_i \left\{ \sup_x |V(x,z_i) - W(x,z_i)| \right\}.$$

[12] A function is said to be h^t-measurable when it is measurable with respect to the $\sigma-$algebra generated by the set of all possible h^t histories H^t.

for all $t \geq 1$, *where* H^t *is the set of all length-t histories of shocks:* $h^t = (z_0, z_1, ...z_t)$, $z_t \in Z$.

That is, $\pi_t(h^t)$ is the value of the endogenous state x_{t+1} that is chosen in period t, when the (partial) history up to this moment is h^t. So, in a stochastic framework agents are taking contingent plans. They are deciding what to do for any possible history, even though some of these histories are never going to happen. Moreover, for any partial history $h^t \in H^t$ one can define a probability measure $\mu^t : \ \mu^t(C) = \Pr\{h^t \in C \subseteq H^t\}$. In this environment, feasibility is defined similarly to the deterministic case. We say that the plan π is feasible, and write $\pi \in \Pi(x_0, z_0)$ if $\pi_0 \in \Gamma(x_0, z_0)$ and for each $t \geq 1$ and h^t we have $\pi_t(h^t) \in \Gamma(\pi_{t-1}(h^{t-1}), z_t)$. We will always assume that F, Γ, β and μ are such that $\Pi(x_0, z_0)$ is nonempty for any $(x_0, z_0) \in X \times Z$, and that the objective function

$$\begin{aligned} \mathbf{U}(\pi) &= \lim_{T\to\infty} F(x_0, \pi_0, z_0) + \sum_{t=1}^{T} \beta^t \int_{H^t} F\left(\pi_{t-1}(h^{t-1}), \pi_t(h^t), z_t\right) d\mu^t(h^t) \\ &= \lim_{T\to\infty} F(x_0, \pi_0, z_0) + \sum_{t=1}^{T} \beta^t \mathbf{E}_0 \left[F\left(\pi_{t-1}(h^{t-1}), \pi_t(h^t), z_t\right)\right] \end{aligned}$$

is well defined for any $\pi \in \Pi(x_0, z_0)$ and (x_0, z_0) . Similarly to the compact notation for the deterministic case, the true value function V^* is defined as follows

$$V^*(x_0, z_0) = \sup_{\pi \in \Pi(x_0, z_0)} \mathbf{U}(\pi). \tag{4.3}$$

Let me first state a verification theorem for the stochastic case.

Theorem 21 *Assume that* $V(x, z)$ *is a measurable function which satisfies the Bellman equation (4.1). Moreover, assume that*

$$\lim_{t\to\infty} \beta^{t+1} \mathbf{E}_0 \left[V(\pi_t(h^t), z_{t+1})\right] = 0$$

for every possible contingent plan $\pi \in \Pi(x_0, z_0)$ *for all* $(x_0, z_0) \in X \times Z$; *and that the policy correspondence*

$$G(x, z) = \left\{x' \in \Gamma(x, z) : \ V(x, z) = F(x, x', z) + \beta \int_Z V(x', z') dQ(z, z')\right\} \tag{4.4}$$

is non empty and permits a measurable selection. Then $V = V^*$ *and all plans generated by* G *are optimal.*

Proof. The idea of the proof follows very closely the lines of Theorems 13 and 14. A plan that solves the Bellman equation and that does not have any left-over value at infinity, is optimal. Of course, we must impose few additional technical conditions imposed by measure theory.[13] For details the reader can see Chapter 9 of SLP. **Q.E.D.**

In order to be able to recover Theorem 12 we need to make an assumption on the *endogenous* V^* :

Theorem 22 *Let F be bounded and measurable. If the value function $V^*(x_0, z_0)$ defined in (4.3) is measurable and assume that the correspondence analogous to (4.4) admits a measurable selection. Then $V^*(x_0, z_0)$ satisfies the functional equation (4.1) for all (x_0, z_0), and any optimal plan π^* (which solves (4.3)) also solves*

$$V^*(\pi^*_{t-1}(h^{t-1}), z_t) = F(\pi^*_{t-1}(h^{t-1}), \pi^*_t(h^t), z_t) + \beta \int V^*(\pi^*_t(h^t), z_{t+1}) dQ\left(z_t, z_{t+1}\right),$$

$\mu^t(\cdot)$ almost surely for all t and h^t emanating from z_0.

Proof. The idea of the proof is similar to that of Theorem 12. For the several details however, the reader is demanded to Theorem 9.4 in SLP. **Q.E.D.**

Let finally state the corresponding of Theorems 15 and 16 for the stochastic environment allowing for continuous shocks.

Theorem 23 *Assume F is continuous and bounded; Γ compact valued and continuous; Q possesses the Feller property, $\beta \in [0, 1)$ and X is a closed and convex subset of $I\!R^l$. Then the Bellman operator T*

$$(TW)(x, z) = \max_{x' \in \Gamma(x,z)} F(x, x', z) + \beta \int_Z W(x', z') dQ(z, z')$$

has a unique fixed point V in the space of continuous and bounded functions.

Proof. Once we have noted that the Feller property of Q guarantees that if W is bounded and continuous function then $\int_Z W(x', z') dQ(z, z')$ is also bounded and continuous for all (x', z), we can apply basically line by line the proof of Theorem 15. **Q.E.D.**

Theorem 24 *Assume F is concave continuous and bounded; Γ is continuous and with convex graph; Q possesses the Feller property, $\beta \in [0, 1)$ and X is a closed and convex subset of $I\!R^l$. Then the Bellman operator has a unique fixed point V in the space of concave, continuous and bounded functions.*

[13] For example, the policy correspondence G permits a measurable selection if there exists a function $h : X \times Z \to X$, such that $h(x, z) \in G(x, z)$ for all $(x, z) \in X \times Z$.

Proof. Again the proof is similar to the deterministic case. Once we have noted that the linearity of the integral preserves concavity (since $\int_Z dQ(z,z') = 1$) we can basically apply line by line the proof of Theorem 16. **Q.E.D.**

It is important to notice that whenever the conditions of Theorem 23 are met, the boundedness of V and an application of the Maximum Theorem imply the conditions of Theorem 21 are also satisfied, hence $V = V^*$ which is a continuous function (hence measurable). In this case the Bellman equation fully characterizes the optimization problem also with uncountably many possible levels of the shock.

4.4 The Stochastic Model of Optimal Growth

Consider the stochastic version of the optimal growth model

$$\begin{aligned} V(k_0, z_0) &= \sup_{\{k_{t+1}\}_{t=0}^{\infty}} \mathbf{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(f(z_t, k_t) - k_{t+1}\right) \right] \\ s.t.\ 0 &\leq k_{t+1} \leq f(z_t, k_t) \quad \text{for all } t \\ k_0 &\in X,\ z_0 \in Z, \end{aligned}$$

where the expectation is over the sequence of shocks $\{z_t\}_{t=0}^{\infty}$. Assume that $\{z_t\}_{t=0}^{\infty}$ is an i.i.d. sequence of shocks, each drawn according to the probability measure μ on $(Z, \mathcal{Z})$.

Exercise 44 *Let $u(c) = \ln c$ and $f(z,k) = zk^{\alpha}$, $0 < \alpha < 1$ (so $\delta = 1$). I tell you that the optimal policy function takes the form $k_{t+1} = \alpha\beta z_t k_t^{\alpha}$ for any t and z_t. (i) Use this fact to calculate an expression for the optimal policy $\pi_t^*(h^t)$ [recall that $h^t = (z_0, ...z_t)$] and the value function $V^*(k_0, z_0)$ for any initial values (k_0, z_0), and verify that V^* solves the following Bellman equation*

$$V(k,z) = \max_{0 \leq k' \leq zk^{\alpha}} \ln(zk^{\alpha} - k') + \beta \mathbf{E}\left[V(k', z')\right].$$

(ii) Now show that a solution to the above functional equation is

$$V(k,z) = A(z) + \frac{\alpha}{1 - \beta\alpha} \ln k,$$

and discuss the relationship between V^* and V.

This model can be extended in many directions. This model with persistent shocks and non inelastic labor supply has been used in the Real Business Cycles literature to study the effects of technological shocks on aggregate variables like consumption and employment. This line of research started in the 80s, and for many macroeconomists is still the building block for any study about the aggregate real economy. RBC will be the next topic of these notes. Moreover, since most interesting economic problem do not have closed forms, you must first learn how to use numerical methods to approximate V and perform simulations.

Bibliography

[1] Bertsekas, D. P. (1976) *Dynamic Programming and Stochastic Control*, Academic Press, NY.

[2] Bertsekas, D. P., and S. E. Shreve (1978) *Stochastic Control: The Discrete Time Case.* New York: Academic Press.

[3] Rudin, W. *Principles of Mathematical Analysis*, McGraw Hill.

[4] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

Chapter 5

A Quick Introduction to Numerical Methods

One of the main advantages of the recursive approach is that we can use the computer to solve numerically interesting models. There is a wide variety of approaches. Each method uses a different characteristic of optimality involved in the dynamic programming method. In this chapter, we only consider - briefly - a couple of approaches which both use very heavily the Bellman equation (BE) we studied in the previous chapters. For a more exhaustive analysis we suggest Marimon and Scott (1999), Judd (1998), and Santos (1999).

5.1 Value Function Iteration

Let X the state space. Recall that the BE in the deterministic case takes the following form:

$$V(x) = \max_{x' \in \Gamma(x)} F(x, x') + \beta V(x').$$

The theoretical idea behind the value function iteration approach is to use the contraction mapping generated by the Bellman operator $\mathbf{T}$ associated to the dynamic programming problem to derive the value function $V = \mathbf{T}V$.

The algorithm recalls the 'guided guesses' approach we saw in Chapter 1, and it works as follows:

1. Pick an initial guess V_0 for the value function: $V_0 : X \to \mathbb{R}$.

2. Apply the $\mathbf{T}$ operator to it and get a new function $V_1 = \mathbf{T}V_0$

3. Then apply again $\mathbf{T}$ to V_1 and get $V_2 = \mathbf{T}V_1$ and so on until the fixed point of the operator: $V^* = \mathbf{T}V^*$ is reached.

Note few things:

First, the maximization required by $\mathbf{T}$ should be done by using *numerical algorithms.* There are very many numerical algorithms available, each with its own advantages and disadvantages. We will not analyze this specific aspect of the problem here. We demand the interested reader to Judd (1998) and the literature cited there.

Second, the operator $\mathbf{T}$ works in the space of functions. Each time the state space is not a finite set the space of functions constitute an infinite dimensional space. Clearly, a computer cannot deal with infinite dimensional spaces. We are hence forced to reduce the mapping $\mathbf{T}$ to map finite dimensional spaces into itself. This is done by *approximating the (value) function.* The two most commonly used methods of dealing with such infinities are the discretization and the smooth approximation methods. We will briefly analyze both of them in turn. We will then analyze another method of approximating the value function which uses directly the functional equation: the collocation method.

Third, the computer cannot perform an infinite amount of iterations in finite time. The exact fixed point condition $V^* = \mathbf{T}V^*$ is however very difficult to obtain in finitely many iterations. We are hence forced to allow for a degree of *tolerance.*

5.1.0.1 Discretization

The idea of this method is to substitute the value function with a discrete version of it by discretizing the state space. In this way, we solve two problems at once. First, the (approximated) Bellman operator $\hat{\mathbf{T}}$ maps finite dimensional vectors into finite dimensional vectors. Second, the maximization procedure at each step is particularly simple. The algorithm works as follows:

1. Take the state space X, and discretize it, say $\hat{X} = \left[x^1,\ x^2,\ x^3, ..., x^N\right]$.
2. Pick an initial guess V_0 by associating one value to each state level, say $\hat{V}_0 = \left[v_0^1,\ v_0^2, ..., v_0^N\right]$.
3. For each $x^i \in \hat{X}$, look for the $x^j (= x') \in \hat{X}$ which solves

$$\max_{x^j \in \hat{X},\ x^j \in \hat{\Gamma}(x^i)} F\left(x^i, x^j\right) + \beta v_0^j$$

   where $\hat{\Gamma}$ is possibly an approximated version of the feasibility due to discretization. Sometimes we do not need to relax Γ as for example we can simply substitute it into

F and solve for an unconstrained problem. Notice that this a very simple procedure as it involves the choice of the maximal element among finitely many.

4. Denote v_1^i the value associated to $F\left(x^i, x^{j^*}\right)+\beta v_0^{j^*}$ where j^* is the index that delivers the maximal value at the previous stage.

5. If we do it for all x^i we get a new N dimensional vector of values which constitutes our new (discretized) function $\hat{V}_1 = \left[v_1^1,\ v_1^2, ..., v_1^N\right]$.

6. And so on until the vectors $\hat{V}_n$ and $\hat{V}_{n+1}$ are 'close' enough to each other. Where 'close' is defined according to some metric. For example:

$$d\left(\hat{V}_n, \hat{V}_{n+1}\right) = \sum_i \varpi^i \left|v_n^i - v_{n+1}^i\right| < \varepsilon,$$

where ϖ^i are pre-specified weights and ε is the tolerance level.

Example: The Deterministic Optimal Growth Model In order to see more precisely how this method works, we will consider its application to the optimal growth model.

First of all, we need to parametrize the problem. We use the Cobb-Douglas production function and a CRRA utility:

$$f(k) = k^\alpha + (1-\delta)\,k \ \text{ and } u\,(c) = \frac{c^{1-\sigma}}{1-\sigma},$$

where δ is the depreciation rate, α is the capital share, and $\frac{1}{\sigma}$ is the intertemporal elasticity of substitution.

We will consider the special case where $\sigma = 1$. In this case, $u\,(c) = \ln c$. In order to compare our approximated solution to the exact one, we start by assuming a 100% depreciation rate, i.e. $\delta = 1$.

The problem to be solved is hence

$$V\,(k) = \max_{0\leq k'\leq k^\alpha} \ln\left(k^\alpha - k'\right) + \beta V\left(k'\right),$$

We know from Chapter 1 that in this case $g(k) = \alpha\beta k^\alpha$. We now compute it by using the discretization method. We shall open the `Matlab` program and write our computation code:[1]

[1] I thank Liam Graham for lending me his simple Matlab code.

Initialize the problem

```
clear all;
close all;
```

Define parameters

`beta=.9; %` $\beta = .9$

`alpha=.35; %` $\alpha = .35$

```
NumPoints =100;
```

Discretize the state space around the steady state capital stock

`k_bar = (alpha*beta)^(1/(1-alpha)); % Recall that` $k^* = (\alpha\beta)^{\frac{1}{1-\alpha}}$

```
k_lo = k_bar*0.5;
k_hi = k_bar*2;
step = (k_hi-k_lo)/NumPoints;
K = k_lo:step:k_hi;
n=length(K);
```

Since loops in `Matlab` are very slow, while matrix manipulations are very fast, we first build an $n \times n$matrix whose columns are output at each value of k

```
Y= K.^alpha;
YY = ones(n,1)*Y;
```

Then another $n \times n$matrix whose columns are capital

```
KK = ones(n,1)*K;
```

Consumption at each level of $k'$is then given by

```
C=YY-KK';
```

Calculate the utility arising from each level of consumption

```
U=log(C);
```

Take an initial guess at the value function

```
V = zeros(n,1);
```

Apply the operator:

$$W = U + \beta \begin{bmatrix} 0 \\ 0 \\ .. \\ 0 \end{bmatrix} [1\ 1\ .....1]$$

```
VV=V*ones(1,n);
W=U+beta*VV;
```

Given a k, we want to find the k' that solves

$$\begin{bmatrix} TV(k^1) \\ TV(k^2) \\ .. \\ TV(k^N) \end{bmatrix} = \max\ W$$

```
V=max(W)';
```

Main iteration loop for the value function.

```
flag=1;
while (flag > 10^ (-2))
    VV=V*ones(1,n);
    W=U+beta*VV;
    V1=max(W)';
    flag = max(abs(V1-V));
    V=V1;
end
```

When the value function has converged, find the policy function i.e. the k' that gives the maximum value of the operator for each k. In order to accomplish this, we first find the vector of indices where W takes its maximum value:

```
[val,ind]=max(W);
```

Then we use the indices to pick out the corresponding values of k.

```
k_star = K(ind);
```

Finally, let's keep track of the analytical optimal policy for comparison purposes:

```
k_true = K.*alpha*beta;
```

We can then plot the two policies in the same figure.[2]

Exercise 45 *Open the file* `discrete.m` *in* `Matlab`*. Compute and plot the policy for increasing levels of the discount factor, say for β between .9 and .99. Comment on the different computation times needed for convergence. Now set back $\beta = .9$ and increase the range of values of k between $0.4k^*$ and $5k^*$ and perform the same computations. Comment your results. [Hint: Be careful that a too wide grid might create negative consumption, and obviously you do not want that!] Now modify the code so that to allow for a depreciation rate δ below 100%. Produce figures for different values of α and δ : Say, α between .2 and .5 and δ between .05 and .3. Comment your results from an economic point of view.[Warning: Recall that consumption must be nonnegative.]*

---

[2] Creating plots with `Matlab` is quite easy. Have a look at the file: `discrete.m`

5.1.0.2 Smooth Approximation

This method reduces substitutes the value function V with a parametrized one V_θ where $\theta \in \Theta$ and Θ is a subset of an $k-$dimensional space.

In order to have a theoretically justified procedure, we require that V_θ is 'potentially' able to approximate V very well. Formally, we require:

$$\lim_{k\to\infty} \inf_{\theta\in\Theta\subset\Re^k} \|V_\theta - V\|_\infty = 0$$

where $\|\cdot\|_\infty$ is the sup norm, that is, $\|V\|_\infty = \sup_{x\in X} |V(x)|$. This kind of approximation are said *dense.*

One of the most commonly used dense sets is the set of polynomials, which is dense in the space of continuous functions by the Weierstrass theorem. In this case, V is approximated (or interpolated) by

$$V_\theta(x) = \sum_{i=1}^{k} \theta^i p^i(x), \text{ for all } x \in X,$$

where $p^i(x)$ is the $i-th$ order polynomial,[3] or the Chebyshev polynomial $p^i(x) = \cos(i\ arc\cos(x))$, Legendre, Hermite polynomials, Splines, etc... . The number of (independent) polynomials is called the *degree of approximation.*

We will see, that also in this case we use a discretized version $\hat{X}$ of the state space. In this case, for a different propose. The value function will indeed be defined on all the original state space X.

The numerical algorithm works as above:

1. Discretize the state space to $\hat{X}$, fix the size k and the type p^i of polynomials you want to consider (say the set of the first 20 Chebyshev polynomials). Where $\sharp\hat{X} > k+1$.

2. Start with a vector of weights $\theta_0 = \left[\theta_0^1,\ \theta_0^2,\ \theta_0^3,\ ....,\ \theta_0^N\right]$. This gives you the initial guess: $\hat{V}_0(x) = \sum_{i=1}^k \theta_0^i p^i(x)$ for all $x \in X$.

3. For each $x^i \in \hat{X}$, define

$$\widehat{\mathbf{T}}\left(V_{\theta_0}(x^i)\right) = \max_{x'\in\Gamma(x^i)} F\left(x^i, x'\right) + \beta\hat{V}_0\left(x'\right)$$

   where Γ is the original correspondence since x' is now allowed to vary in the whole X. In this case, the maximization stage requires the use of a numerical algorithm

[3] A i-th order polynomial takes the form $p^i(x) = \sum_{s=1}^i a_s x^s$, where the superscript s indicates the power s of x.

as x varies over a continuum. This is obviously more demanding. The advantage is that, all values of x are somehow evaluated when obtaining the value function.

4. Compute the new vector of weights θ_1 by minimizing some error function, for example a weighted least square criterion as follows:

$$E_N(\theta_1;\theta_0) \equiv \sqrt{\sum_{i=1}^{N} \varpi^i \left| V_{\theta_1}(x^i) - \widehat{\mathbf{T}}\left(V_{\theta_0}(x^i)\right)\right|^2} \tag{5.1}$$

where $x^i \in \hat{X}$ belong to the pre-specified grid of points, and ϖ^i are appropriate weights. Some points in the grid might indeed be more important the others (for example, consider the points close to the steady state for optimal growth models).

5. Do so until a vector of weights θ_n is reached such that for example $\sum_{i=1}^{k}\left|\theta_n^i - \theta_{n-1}^i\right| < \varepsilon$, where ε is the tolerance level. We then can evaluate the approximation error via $E_N(\theta_n;\theta_n)$.

5.2 Solving Directly the Functional Equation: Projection Methods

We saw that the BE also constitutes a functional equation of the form: $\mathbf{T}V - V = 0$. Another possibility is hence to directly look for a solution of the functional equation.

Again, the Bellman functional equation typically imposes an infinite number (in fact possibly a continuum) of conditions, namely:

$$(\mathbf{T}V)(x) - V(x) = 0 \text{ for all } x \in X.$$

And a computer cannot deal with such huge number of equations. One must therefore settle for an approximate solution that satisfies the functional equation closely. Projection methods approximate the function V with V_θ and then look for the vector of parameters θ^* which minimizes the distance between $V_\theta(x) - \widehat{\mathbf{T}}\left(V_\theta(x)\right)$, such us the error $E_N(\theta;\theta)$ for example. Note however that there is another complication involved with the above equation. The function $\widehat{\mathbf{T}}\left(V_\theta(x)\right)$ is not easy to compute. In particular, there is no analytical way of getting it. This is the reason why Projection methods are combined with policy function iteration.

The Euler equation though is also a characteristic of the optimal program. In particular, if $g(x)$ is the value function of our problem, then in the optimal growth model, with

$u = \ln$ we have

$$\frac{1}{k^\alpha - g(k)} = \beta\alpha\left(g\left(k\right)\right)^{\alpha-1}\frac{1}{\left(g\left(k\right)\right)^\alpha - g\left(g\left(k\right)\right)}.$$

In general we have a functional equation of the form

$$H\left(x, g\left(x\right)\right) = 0$$

where the unknown g is our target, while $H : \mathbb{R}^2 \to \mathbb{R}$ is a known function.

Projection methods then approximate the function g with $\hat{g}\left(x;\theta\right)$ and look for the vector of weights θ^* that minimizes a given error function based on H and on a nonnegative function ϕ:

$$\theta^* \in \arg\min_\theta \int \phi\left(x\right)\left|H\left(x, \hat{g}\left(x;\theta\right)\right)\right|,$$

with $\phi\left(x\right) \geq 0$ for all x. When ϕ takes positive values only at finitely many points in X, we obviously only evaluate the function at few points in a given grid $\hat{X}$ ad we get an error function such as that in (5.1). In this case, the method is called *collocation method.*[4] In practice, these techniques constitute other methods to approximate functions. The collocation method for example, is a generalization of the so called *interpolation methods.*

Example The collocation method applied to the optimal growth model looks for parameters $\theta = (\theta^1, \theta^2, ...\theta^q)$ that minimize

$$\sqrt{\sum_{i=1}^N \varpi^i \left|\frac{1}{k_i^\alpha - \hat{g}(k_i;\theta)} - \beta\alpha\left(\hat{g}\left(k_i;\theta\right)\right)^{\alpha-1}\frac{1}{\left(\hat{g}\left(k_i;\theta\right)\right)^\alpha - \hat{g}\left(\hat{g}\left(k_i;\theta\right);\theta\right)}\right|^2},$$

where k_i, ϖ^i $i = 1, 2, ..N$, $N > q+1$ are the points in a pre-specified grid of capital levels and given weights respectively, and $\hat{g}$ is for exmple a polynomial of the form

$$\hat{g}(k;\theta) = \theta^0 + \theta^1 k + \theta^2 k^2 + ... + \theta^q k^q.$$

[4] See McGrattan's chapter 6 in Marimon and Scott's book for several other methods.

Bibliography

[1] [AC] Adda J., and Cooper (2003) *Dynamic Economics: Quantitative Methods and Applications,* MIT Press.

[2] Judd, K.L. (1998), *Numerical Methods in Economics*, Cambridge, MA, MIT Press.

[3] Marimon, R., and A. Scott (1999) *Computational Methods for the Study of Dynamic Economies*, Oxford University Press.

[4] Santos, M. S. (1999) "Numerical Solutions of Dynamic Economic Models," in J. B. Taylor, and M. Woodford eds., *Handbook of Macroeconomics*, Vol. 1A, Elsevier: 311-386.

[5] Sargent, T. (1987) *Dynamic Macroeconomic Theory*, Harward University Press.

[6] [LS] Ljungqvist, L., and T. J. Sargent (2004), *Recursive Macroeconomic Theory*, Second Edition, MIT Press.

[7] [SLP] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

Chapter 6

Risk Aversion, Insurance and Asset Pricing in Macroeconomics

In this chapter we would like to first review the concept risk and insurance, and then study general equilibrium under uncertainty.

6.1 Risk and Risk Aversion

What is risk and risk aversion?

Consider a decision maker possessing a utility function u defined over the interval C of real numbers: $u : C \to \mathbb{R}$. We say that the decision maker is *risk averse* if for every probability distribution on C such that expectation are finite we have

$$\mathbf{E}\left[u(c)\right] \leq u\left(\mathbf{E}\left[c\right]\right).$$

We will see below that risk aversion, is also equivalently defined as u to be concave. When u is liner the agent is said to be *risk neutral*. The utility function u defined on the consumption outcomes C is sometimes called 'Bernoulli' utility. We say that agent i is *more risk averse* than agent k if the Bernoulli u_i is an increasing and concave transformation of u_k. That is, if it exists an increasing and concave function $\psi : \mathbb{R} \to \mathbb{R}$ such that $u_i\left(c\right) \equiv \psi\left(u_k\left(c\right)\right)$.

We would now like to have a measure of risk aversion. To this extent consider a gamble $\mathbf{c}$ over the set of real numbers, i.e. a random variable, with given probability distribution and expected value $\bar{c} = \mathbf{E}c$. Let y the amount of money the agent is willing to pay in order to avoid the gamble and get the expected value for sure, that is[1]

$$u(\bar{c} - y) = \mathbf{E}\left[u(c)\right].$$

Now take the Taylor series expansion around $\bar{c}$ and get

$$u(\bar{c} - y) = u(\bar{c}) - yu'(\bar{c}) + o(y). \tag{6.1}$$

Also we have

$$\begin{aligned}\mathbf{E}\left[u(c)\right] &= \mathbf{E}\left[u(\bar{c}) + (c-\bar{c})\,u'(\bar{c}) + \frac{1}{2}\left(c-\bar{c}\right)^2 u''(\bar{c}) + o\left((c-\bar{c})^2\right)\right] \\ &= u(\bar{c}) + \frac{1}{2}\sigma^2 u''(\bar{c}) + \mathbf{E}\left[o\left((c-\bar{c})^2\right)\right],\end{aligned}$$

where σ^2 is the variance of the gamble. Finally, relating the above expression with (6.1), we have

$$yu'(\bar{c}) = -\frac{1}{2}\sigma^2 u''(\bar{c}) + o(y) + \mathbf{E}\left[o\left((c-\bar{c})^2\right)\right]$$

or

$$y = \frac{\sigma^2}{2}\left(-\frac{u''(\bar{c})}{u'(\bar{c})}\right) + \frac{o(y)}{u'(\bar{c})} + \frac{\mathbf{E}\left[o\left((c-\bar{c})^2\right)\right]}{u'(\bar{c})}.$$

This equation says that the amount insurance or risk premium y that the agent is willing to pay is proportional (to first order) to the index of absolute risk aversion

$$R^A(\bar{c}) - \frac{u''(\bar{c})}{u'(\bar{c})},$$

thus justifying the use of R^A as a measure of local risk aversion.

Definition 36 *Assume the Bernoulli utility u is twice continuously differentiable. Then*

$$R^A(x) = -\frac{u''(x)}{u'(x)}$$

is called the deFinetti (1952), Pratt (1964) and Arrow (1970) absolute risk aversion measure.

Kenneth Arrow also showed that if an utility function presents decreasing absolute risk aversion over the whole domain of R^A, then the risky asset is a normal good. That is, the (dollar) demand for the risky asset increases as the agent's wealth increases.

[1] The value $\bar{c} - y$ is also denoted by *certainty equivalent,* see Definition 37 below.

The coefficient of *relative* risk aversion R^R measures the same premium for multiplicative shocks, i.e. shocks that induce *proportional* increments or decrements of wealth. If x is the initial wealth of the agent, such shocks can be represented by a random variable t, where $u(c) = u(tx)$, and $t = 1$ corresponds to the initial position. If, given x, we make the change in variable $\hat{u}(t) \equiv u(tx)$, then multiplicative shocks for agent u are captured by additive shocks for agent $\hat{u}$. We can hence recall the above derivation which says that for small risks around $t = 1$ the degree of risk aversion is well captured by the ratio $\hat{u}''(1)/\hat{u}'(1)$. Noting that $\hat{u}''(1)/\hat{u}'(1) = \frac{u''(x)x}{u'(x)}$, we define the coefficient of relative risk aversion as

$$R^R(x) = -\frac{u''(x)\,x}{u'(x)}.$$

Definition 37 *Given a Bernoulli utility function u we define the following concepts. (i) The Certainty Equivalent, denoted $ce(F,u)$, is the amount of money for which the individual is indifferent between the gamble F and the certainty amount $ce(F,u)$; that is*

$$u(ce(F,u)) = \int u(x)dF(x).$$

(ii) For any fixed amount of money x and a positive number ε, the Probability Premium denoted by $pp(x,\varepsilon,u)$, is the excess in winning probability over fair odds that makes the individual indifferent between the certain outcome x and a gamble between the two outcomes $x+\varepsilon$ and $x-\varepsilon$. That is

$$u(x) = \left(\frac{1}{2} + pp(x,\varepsilon,u)\right)u(x+\varepsilon) + \left(\frac{1}{2} - pp(x,\varepsilon,u)\right)u(x-\varepsilon).$$

It can be shown that a risk averse agent can be equivalently defined as (i) an agent with u concave, (ii) $ce(F,u) \leq \int xdF(x) = E_F(x)$ for any F, or (iii) $pp(x,\varepsilon,u) \geq 0$ for all x, ε.

Now we are ready to state the following result.

Proposition 5 *Assume we have two agents with utilities u_i and u_k. The the following statements are equivalent. (i) $R_i^A(x) \geq R_k^A(x)$ for every x. (ii) there exists an increasing concave function $\psi(.)$ such that $u_i(x) = \psi(u_k(x))$ for all x. (iii) $ce(F,u_i) \leq ce(F,u_k)$ for any F. (iv) $pp(x,\varepsilon,u_i) \geq pp(x,\varepsilon,u_k)$ for any x and ε and (v) whenever u_i finds a lottery F at least as good as a riskless outcome x^*, then u_k also finds F at least as good as x^*. That is*

$$\int u_i(x)dF(x) \geq u_i(x^*) \text{ implies } \int u_k(x)dF(x) \geq u_k(x^*)$$

in other words, any risk that u_i would accept starting from a position of certainty would also be accepted by u_k.

The equivalence between *(i)* and *(ii)* clarifies to what extent we can rank risk by using the coefficient of absolute risk aversion. **To see the idea of the proof,** assume both u_k, u_i, and ψ are twice continuously differentiable. From $u_i(x) = \psi(u_k(x))$, differentiating once and twice gives

$$u_i'(x) = \psi'(u_k(x))\, u_k'(x) \quad \text{and} \quad u_i''(x) = \psi'(u_k(x))\, u_k''(x) + \psi''(u_k(x))\,(u_k'(x))^2 .$$

Dividing both sides of the second expression by $u_k'(x) > 0$ and using the first expression, we obtain

$$R_i^A(x) = R_k^A(x) - \frac{\psi''(u_k(x))}{\psi'(u_k(x))} u_k'(x),$$

which implies that $R_i^A(x) \geq R_k^A(x)$ for all x if an only if $\psi''(u_k(x)) \leq 0$ for all $u_k(x)$. **Q.E.D.**

From *(v)* in the above Proposition, we can see that the absolute risk aversion parameter allows us to produce clear (even if not simple) comparative static results in comparing a lottery with a risk-less lottery. However, R^A is a too weak concept when we want to derive comparative static results between different lotteries (or risky assets). Thus sometimes we define a stronger concept.

Definition 38 *(Ross 1981) We will say that agent i is strongly more risk averse than individual k if*

$$\inf_x \frac{u_i''(x)}{u_k''(x)} \geq \sup_x \frac{u_i'(x)}{u_k'(x)}.$$

Note that this strong requirement implies that for any x we have $\frac{u_i''(x)}{u_k''(x)} \geq \frac{u_i'(x)}{u_k'(x)}$. Rearranging this implies

$$R_i^A(x) = -\frac{u_i''(x)}{u_i'(x)} \geq -\frac{u_k''(x)}{u_k'(x)} = R_k^A(x).$$

It turns out to be a stronger concept. Indeed, we can construct a counterexample with two CRRA utility functions as follows. Assume $u_i(x) = -e^{-ax}$ and $u_k(x) = -e^{-bx}$. Of course, if $a > b$ then agent i is more risk averse then k in the sense of Arrow-Pratt. However, we can have two points x_1 and $x_2 > x_1$ such that

$$\left(\frac{a}{b}\right)^2 e^{-(a-b)x_2} = \frac{u_i''(x_2)}{u_k''(x_2)} < \frac{u_i'(x_1)}{u_k'(x_1)} = \frac{a}{b} e^{-(a-b)x_1}.$$

moreover, in general we cannot define such a detailed description of preferences. So two important concepts have to be considered. The first concept is called First Order stochastic dominance, and defines preferences over lotteries simply assuming non-satiation. The second concept is the Second Order stochastic dominance and defines preferences simply assuming risk aversion.

The second concept is very useful when we want to compare the riskiness of any two lotteries. However, it is important to notice that S.O.s.d. does not define a complete ordering over the different lotteries (or among different risky assets).

Definition 39 *Consider two probability distributions η and λ. We say that η first order stochastically dominate λ and we write $\eta \geq \lambda$ if for any nondecreasing and continuous function $f : I\!R \to I\!R$ we have*

$$\int f(x)d\eta \geq \int f(x)d\lambda.$$

A first implication is that if $\eta \geq \lambda$ then $\mathbf{E}_\eta(x) = \int x d\eta \geq \int x d\lambda = \mathbf{E}_\lambda(x)$, this is obvious since the linear function $f(x) = x$ is in particular increasing and continuous. Thus a any (risk neutral) non-satiable agent with continuous utility will surely prefer lottery η to λ. Of course, the reverse is not necessary true. What is true is the following graphically useful result.

Proposition 6 *If λ and η are defined over the interval $I = [\underline{x}, \overline{x}]$, then the following statement are equivalent. (i) $\eta \geq \lambda$, (ii) The implied distribution functions are such that $F_\lambda(x) \geq F_\eta(x)$ for any x in I.*

Where, for example, $F_\lambda(x) = \int_{\underline{x}}^{x} d\lambda$. thus if we draw the distributions F_λ and F_η then the F_λ graph is uniformly above the other.

Definition 40 *(Mas-Colell et al.) Consider two probability measures μ and λ. We say that μ second order stochastically dominate λ and we write $\mu \geq^2 \lambda$ if (i) $\mathbf{E}_\mu(x) = \mathbf{E}_\lambda(x)$ and (ii) for any **non-decreasing concave** function $f : I\!R_+ \to I\!R$ we have*

$$\int f(x)d\mu \geq \int f(x)d\lambda.$$

It is not pleasant however to be able to only compare lotteries delivering the same mean. Moreover, in many macroeconomic applications, lotteries are defined over real numbers. It can hence be of some use to know that the above definition is equivalent to the following.

Proposition 7 *(Definition in Huang et al.) Consider two probability measures* μ *and* λ *defined over an interval* I*. Then we can say that* μ *second order stochastically dominate* λ *if for any* ***concave and differentiable*** *function* $f : \mathbb{R}_+ \to \mathbb{R}$ *with first derivative which is continuous except in a countable set of point in* I *we have*

$$\int f(x)d\mu \geq \int f(x)d\lambda.$$

Proof In order to see that the two definitions are equivalent first note that, since monotonicity is not required we can take the two (weakly) concave functions $g(x) = x$ and $f(x) = -x$ to show that $\mathbf{E}_\mu(x) = \mathbf{E}_\lambda(x)$ as required at *(i)* in Definition 40. It is now clear that statement *(ii)* in Definition 40 must be satisfied since the set of functions considered in the latter is smaller due to monotonicity. So we have proved that the above definition implies $\mu \geq^2 \lambda$ in Definition 40. We now want to show the opposite. Assume Definition 40 is true for μ and λ. The key to notice is that in the proposition above we do not require monotonicity but we require differentiability and the fact that the probability distributions are defined on an interval.

The fact that the two definitions are equivalent is shown by the next proposition, whose proof is left as an exercise.

Proposition 8 *If* λ *and* μ *are defined over the interval* $I = [\underline{x}, \overline{x}]$*, (and they have the same mean), then the following statement are equivalent. (i)* $\mu \geq^2 \lambda$*, (ii) The implied distribution functions are such that* $\int_{\underline{x}}^{y} F_\lambda(x)dx \geq \int_{\underline{x}}^{y} F_\mu(x)dx$ *for any* y *in* I*. (iii)* F_λ *is a mean-preserving spread of* F_μ*.*

Where we say that G is a mean-preserving spread of F if we can derive it with a compound lottery such that for each possible outcome x we define the final payoff as $x + z$ where z has distribution $H_x(z)$ with zero mean (i.e. $\int z dH_x(z) = 0$).

Exercise 46 *Complete the proof of Proposition 7. You are allowed to use some of the results in Proposition 8. Now show at least one of the implications in Proposition 8.*

To conclude this digression on ordering of probability distributions we introduce the following concept, which is stronger than SOsd.

Definition 41 *Consider two probability measures* ϕ *and* λ*. We say that* ϕ *second order* ***monotonic*** *stochastically dominate* λ *and we write* $\phi \geq^{M2} \lambda$ *if for any non-decreasing and concave function* $f : \mathbb{R}_+ \to \mathbb{R}$ *we have*

$$\int f(x)d\phi \geq \int f(x)d\lambda.$$

The idea here is that lotteries are ranked so that every risk averse and non-satiable individual prefers lottery ϕ to lottery λ.

Proposition 9 *If λ and ϕ are defined over the interval $I = [\underline{x}, \overline{x}]$ then the following statement are equivalent. (i) $\mu \geq^{M2} \lambda$, (ii) $\mathbf{E}_\phi(x) \geq \mathbf{E}_\lambda(x)$ and the implied distribution functions are such that $\int_{\underline{x}}^{y} F_\lambda(x)dx \geq \int_{\underline{x}}^{y} F_\phi(x)dx$ for any y in I.*

From this statement we can see that the second order monotonic stochastic dominance is weaker than the FOsd. Graphically the difference is such that First order stochastic dominance requires that $F_\lambda(x) \geq F_\phi(x)$ for any x (which obviously implies $\mathbf{E}_\phi(x) \geq \mathbf{E}_\lambda(x)$); while SOMsd allows for points where $F_\lambda(x) < F_\phi(x)$, as long as the area below the two curves satisfies the relation in (ii).

6.2 Arrow-Debreu Equilibrium

We now study the concept of competitive equilibrium in a risky environment. The American economist Kenneth J. Arrow and the French economist and mathematician Gérard Debreu had a pretty simple and powerful idea: Just consider goods consumed or produced at different contingencies as completely different goods (just like apples and bananas are different goods) and then apply the standard concept of competitive equilibrium.

It is now intuitive why the definition of an Arrow-Debreu Equilibrium requires the existence of a full set of markets at the beginning of the world (before uncertainty is revealed). One market for each possible state $s \in S$. The paradigm here is that at date $t = 0$ each agent chooses how much to buy, sell and consume of every good. At date $t = 1$, after the realization of uncertainty, each consumer just 'mechanically' perform the stated trades. There are no decisions at this stage, agents act as 'automata'. Clearly, this simple concept extends to multiperiod setting exactly in the same manner, where each state s corresponds to a full history of shocks: at date $t = 0$ each agent makes his purchases and sells and then at each node the period zero agreements are mechanically implemented with no further decisions to be taken.

Consider a pure exchange economy (with no production) with only one good in each state and I consumers.

Definition 42 *An AD equilibrium is an allocation $c = \left(c^1(s), c^2(s), ..., c^I(s)\right)_{s\in S}$ and a system of prices $p = (p(s))_{s\in S}$ such that for each consumer $i = 1, 2, ...I$; $c^i = (c^i(s))_{s\in S}$*

solves the following utility maximization problem

$$\max_{c^i \in C^i} U_i(c^i)$$

s.t.

$$\sum_{s\in S} p(s)c^i(s) \leq \sum_{s\in S} p(s)\omega^i(s),$$

where $\omega^i(s)$ is endowment of agent i in state s. And markets clear: for all $s \in S$ we have

$$\sum_{i\in I} c^i(s) = \sum_{i\in I} \omega^i(s).$$

Notice that this definition has nothing special about uncertainty. The idea of AD is to treat the same physical good in a different state as a different good, and then define the equilibrium concept in this extended goods space exactly in the same way as that under the standard situation.

It is however common usage to assume that the utilities $U_i(c^i)$ have a special form, based on some standard axioms about the choice under uncertainty (e.g. see Mas-Colell et al., Ch. 6). Those axioms imply that U_i takes the form of *expected utility:*

$$U_i(c^i) = \sum_{s\in S} \pi^i(s)u_i\left(c^i(s)\right).$$

In the most standard equilibrium problems agents are assumed to have common probability distributions on the states, that is, the only heterogeneity is on the Bernoulli utilities u_i and on endowments:

$$U_i(c^i) = \sum_{s\in S} \pi(s)u_i\left(c^i(s)\right).$$

Finally, in several macroeconomic applications the focus is on the effect of having heterogeneity only on the endowments $\omega^i(s)$, i.e. $u_i = u_j = u$ for all i, j.

It is well known that in such an economy, each equilibrium is Pareto efficient (First Welfare Theorem). Moreover, under the assumptions of convexity of preferences and feasible set - a property satisfied in our one-good for state economy with expected utility with C^i substituted by $c^i \geq 0$ - also the Second Welfare Theorem guarantees that each Pareto efficient (PE) allocation can be obtained as an Equilibrium with transfers (see Mas-Colell et al., 1995, pages 551-557 for details).

More formally, a PE allocation is a feasible allocation c such that there is no other feasible allocation $\hat{c}$, for which there is at least one agent i such that $U(\hat{c}^i) > U(c^i)$ and all the remaining agents are at least indifferent, that is $U(\hat{c}^j) \geq U(c^j)$ for any $j \in S\backslash i$. In our context, the Second Welfare Theorem states that any such allocation can be obtained by designing a transfer scheme $\tau = (\tau^i)_{i \in I}$ such that

$$\sum_{i \in I} \tau^i = 0,$$

and allowing the agents to freely trade in the market. This will lead to a price system p and an allocation c such that, each agent solves

$$\max_{c^i \in C^i} u_i(c^i)$$

s.t.

$$\sum_{s \in S} p(s) c^i(s) \leq \sum_{s \in S} p(s) \omega^i(s) + \tau^i.$$

The technical condition that $\omega^i(s) + \tau^i > 0$ for all i and all s guarantees that the equilibrium will have a price system not identically to zero ($p \neq 0$).

6.2.1 Properties of the efficient allocation: Insurance

It is well known that in an interior Pareto optimal allocation **c** consumers' indifference curves through **c** must be tangent. That is, the marginal rate of substitution (i.e. the slope of the line tangent to the indifference curve) must be equated.

Proposition 10 *An (internal) Pareto optimal allocation c is always such that the consumers' indifference curves through c must be tangent. That is, the marginal rate of substitution have to be equated:*

$$-\frac{\frac{\partial U_i}{\partial c^i(s)}}{\frac{\partial U_i}{\partial c^i(s')}} = -\frac{\frac{\partial U_j}{\partial c^j(s)}}{\frac{\partial U_j}{\partial c^j(s')}} \tag{6.2}$$

Notice that this result does not depend on the absence of aggregate shock nor on other homogeneities in preferences or probabilities. This condition must always be true as long we can freely re-allocate goods across agents to improve welfare.

Exercise 47 *Carefully explain why an efficient allocation must satisfy condition (6.2).*

Heuristically, one can see why this must be the case by computing the first order conditions for each agent in equilibrium. Let's obtain these conditions for the case where agents' preferences are of the expected utility form. If we assume U_i takes the expected utility form, we have

$$\pi^i(s)u_i'\left(c^i(s)\right) = \lambda^i p(s), \tag{6.3}$$

where λ^i is the multiplier associated to agent i budget constraint. We hence obtain the above result by considering any two agents i, j and, for each agent, taking the ratio of conditions (6.3) for states s and :

$$\frac{\pi^i(s)u_i'(c^i(s))}{\pi^j(s)u_j'(c^j(s))} = \frac{\lambda^i}{\lambda^j} = \frac{\pi^i(s')u_i'(c^i(s'))}{\pi^j(s')u_j'(c^j(s'))},$$

constant for all s. In the case where the agent have common probabilities ($\pi^i(s) = \pi^j(s)$ for all s) we get:

Proposition 11 *If agents have common probability, then in a Pareto optimal allocation we have*

$$\frac{u_i'(c^i(s))}{u_j'(c^j(s))} = \frac{\lambda^i}{\lambda^j},$$

which is a constant number for all s.

Assume now for simplicity that we have only two agents (with common probabilities), using the market clearing condition we get

$$\frac{u_i'(c^i(s))}{u_j'(\omega(s) - c^i(s))} = \frac{\lambda^i}{\lambda^j} = \frac{u_i'(c^i(s'))}{u_j'(\omega(s') - c^i(s'))}. \tag{6.4}$$

It is hence easy to see that when $\sum_i \omega^i(s) = \sum_i \omega^i(s')$, agents have common probabilities, and u_i are concave we must have $c^i(s') = c^i(s)$ for any i and s, s'. This is so since concavity implies that $u_i'(c^i(s'))$ in (6.4) are decreasing function. As a consequence, whenever $\omega(s') = \omega(s)$ any $c^i(s) \neq c^i(s')$ contradicts (6.4).

This can easily shown formally in the general case.

Proposition 12 *When there are no aggregate shocks (i.e. $\sum_i \omega^i(s) = \omega$ for all s) and common probability then in a Pareto optimal allocation the agents full insure themselves, i.e. for all i we have $c^i(s) = c^i(s')$ for all s, s'.*

Notice that agents get fully-insured even when they have different utilities. What matter is the common probability assumption, and of course the fact that there is not aggregate uncertainty.

Exercise 48 *Show Proposition 12 allowing for $u_i \neq u_j$. Explain intuitively why this is the case as long as the assessment probabilities $\pi(s)$ are homogeneous across consumers.*

6.3 Asset Pricing Under Complete Markets

6.3.1 Radner Sequential Trade Equilibrium

The Radner sequential equilibrium is the equilibrium concept used in most financial applications and in dynamic macroeconomics. This is the formal framework that will allow us to talk properly about assets. In the Radner sequential trade framework at the beginning of the world (period 0), each agent can only trade claims $z^i(s)$ over 'wealth' in state s.[2] The market at period zero is called *forward market.* The today's price of a claim over one unit of wealth in period $t = 1$ will be denoted by $q(s)$. When s is realized at $t = 1$, agents are allowed to trade goods in a *spot market*, and are *committed* to pay their claims $z(s)$. Each agent hence faces the following optimization problem

$$\max_{c^i \in C^i, z^i \in \Re^S} U_i(c^i) \tag{6.5}$$

s.t.

$$\sum_{s \in S} q(s) z^i(s) \leq 0 \tag{6.6}$$

$$p(s) c^i(s) \leq p(s)\omega^i(s) + z^i(s) \text{ for each } s. \tag{6.7}$$

Notice that $z^i \in \Re^S$, hence we allow for $z^i(s) < -p(s)\,\omega^i(s)$. When this happen, we say the agent is selling short. In period 1, the agent will have to actually buy the good at the prevailing spot price in order to fulfill the commitment. Of course, the agent cannot take the appropriate decisions in period 0, regarding the z^i trades, without having a pretty good idea regarding the *spot* price $p(s)$ that will prevail at period 1 under each contingency s. Agents will form *expectations* regarding the prices $p(s)$, $s \in S$. The most common assumption is that of *rational expectations:*

> "consumers' expectations are *self-fulfilled* or *rational;* that is, we require that consumers' expectations of the prices that will clear the spot markets for the different states s do actually clear them once date $t = 1$ has arrived and a state s is revealed." Mas-Colell et al., 1995, page 696.

We are now ready to specify the concept Radner equilibrium.

[2] Alternatively, we could have considered assets paying in real terms and choose one good as 'numeraire' and set z^i denominated in such good. Obviously, this distinction is however relevant only in environments where for each state s one have more than one consumption good.

Definition 43 *A Radner equilibrium is an allocation* $c = \left(c^1(s), c^2(s), ..., c^I(s)\right)_{s\in S}$ *and* $z = \left(z^1(s), z^2(s), ..., z^I(s)\right)_{s\in S}$, *and a system of prices* $p = (p(s))_{s\in S}$ *and* $q = (q(s))_{s\in S}$, *such that for each consumer* $i = 1, 2, ...I$; (c^i, z^i) *solve the utility maximization problem in (6.5)-(6.7) and such that markets clear, that is,*

$$\sum_i z^i(s) \leq 0, \ \ and \ \sum_i c^i(s) \leq \sum_i \omega^i(s) \quad for\ each\ s.$$

Notice that the two market clearing conditions, can be rewritten as follows

$$\sum_i z^i(s) \leq 0 \leq \sum_i \omega^i(s) - \sum_i c^i(s),$$

It is not very difficult to show that the AD and Radner equilibrium allocations coincide. The technical condition for the equivalence is that $q, p > 0$. Below we will state the precise equivalence result (see also Mas-Colell et al., 1995, page 697).

6.3.2 Introducing Proper Assets

We now introduce a more general sequential equilibrium environment, with proper assets, that allows for incomplete markets. *An asset (or security)* z is a title to receive an amount $r(s)$ of wealth at date $t = 1$ if state s occurs. An asset is therefor characterized by a *return vector* $r = (r(s))_{s\in S}$. Assuming that there are K assets in the economy, a compact way of stating the returns of the whole set of available assets is the *return matrix*

$$R = \begin{bmatrix} r^1(1) & r^2(1) & ... & ... & r^K(1) \\ ... & ... & ... & ... & ... \\ ... & ... & r^k(s) & ... & ... \\ ... & ... & ... & ... & ... \\ r^1(S) & r^2(S) & ... & ... & r^K(S) \end{bmatrix}$$

which displays the return vector associated to each asset in each row. The definition of Radner equilibrium when we consider such assets instead of generic claims is as follows.

Definition 44 *A Radner equilibrium is an allocation* $c = \left(c^1(s), c^2(s), ..., c^I(s)\right)_{s\in S}$ *and* $z = \left(z_k^1, z_k^2, ..., z_k^I\right)_{k\in K}$, *and a system of prices* $p = (p(s))_{s\in S}$, *and* $q = (q_k)_{k\in K}$, *such that, for each consumer* $i = 1, 2, ...I$; c^i, z^i *solve the following utility maximization problem*

$$\max_{c^i\in C^i, z^i\in\Re^K} U_i(c^i)$$

s.t.

$$\sum_{k\in K} q_k z_k^i \leq 0$$
$$p(s)c^i(s) \leq p(s)\omega^i(s) + \sum_k z_k^i r^k(s) \text{ for each } s.$$

and such that market clears, that is for any k, and s

$$\sum_i z_k^i \leq 0, \text{ and } \sum_i c^i(s) \leq \sum_i \omega^i(s).$$

The key novelty in this definition consists is the return vector, which implicitly restricts the consumers' choices, by requiring a proportionally in each state: Each agent can choose how much to buy for each asset, however, once an amount z_k of this asset has been decided in period zero, the return in $t = 1$ in each state is restricted to be $z_k r^k(s)$, where $r^k(s)$ is obviously exogenously to the agent. In contrast, in the previous definition of the sequential equilibrium the agent is allowed to buy *independently* the quantities $z(s)$ for all s. He hence can transfer wealth across states by choosing an appropriate portfolio of $z(s)'s$.

The wealth allocation across states can be seen as a point in $\mathbb{R}^S$. By appropriately choosing the $z(s)'s$ the agent is hence able to generate any point $\mathbb{R}^S$ (when appropriately scaled down so that to satisfy the budget constraint). Another way of seeing the securities $z(s)$ is as S orthonormal vectors. It is well known, from linear algebra, that a space of dimension S can be completely mapped by a linear combination of $K \geq S$ vectors, where at least S of them are linearly independent. It is also well known that if either $K < S$ or there are less then S linearly independent vectors then we cannot map all $\mathbb{R}^S$ by linear combining them. A compact way of static this condition is to require that $rank(R) = S$. When this condition is met, we say *the asset structure is complete,* and we have the following fundamental result.

Theorem 25 *Suppose that the asset structure is complete. Then (i) if the consumption plan c and the price vector $p > 0$ constitutes an AD equilibrium, then there are asset prices $q > 0$ and portfolio plans z such that the allocation c, z, the asset prices q and the spot prices q constitute a Radner Equilibrium. (ii) Conversely, if the allocation c, z and prices $p, q > 0$ constitute a Radner equilibrium, then there are multipliers $(\mu(s))_{s\in S} > 0$ such that the consumption plans c and the contingent commodity prices $(\widetilde{p}(s) = \mu(s)p(s))_{s\in S}$ constitute an AD equilibrium.*

Since the agent's preferences over allocations are obviously the same, the proof consists in showing that the budget sets of the agents are the same in the two specifications of the environment (AD with only period zero trades vs Radner sequential equilibrium with forward and spot markets). The reader is demanded to the proof of Proposition 19.E.2 in Mas-Colell et al. (1995).

From the optimality condition with respect to z_k^i in the Radner equilibria, we get[3]

$$\delta^i q_k = \sum_s \phi^i(s) r^k(s) = \sum_s \pi(s) u_i'\left(c^i(s)\right) \frac{r^k(s)}{p(s)}, \tag{6.8}$$

where δ^i and $\phi^i(s)$ are the multipliers associated to agent i's forward and state-s spot budget constraint respectively; and the last equality is obtained from the optimality conditions with respect to $c^i(s)$, which says: $\pi(s) u_i'(c^i(s)) = \phi^i(s) p(s)$. Condition (6.8) says that the vector of marginal utilities of the K assets must be proportional of the vector of asset prices. If we set

$$\mu^i(s) = \frac{\pi(s) u_i'(c^i(s))}{\delta^i p(s)}$$

we can compute the price of each asset k with return $r^k = \left(r^k(s)\right)_{s\in S}$, $r^k(s) \geq 0$ and $r^k \neq 0$, as follows

$$q_k = \sum_{s\in S} \mu^i(s) r^k(s) \tag{6.9}$$

It is easy to see that since q_k and $r^k(s)$ do not depend on i, and the above condition holds for all k, the theorem implies that $\mu^i(s) = \mu(s)$ for all i [Check it]. This property of the asset prices in a framework with complete markets is widely used in Finance literature and the mapping $\mu : S \to \mathbb{R}_+$is denoted as the (unique) *pricing rule*. The idea is that in a complete market framework, because of efficiency, the value of the wealth at each state is equated *ex-ante* across agents, hence it is identical for each of them and equal to $\mu(s)$. But then $\mu(s)$ becomes a public $t = 0$ value or a "market" value of state contingent wealth: the multipliers $\mu(s)$ represents the value, at $t = 0$, of 1 dollars at $t = 1$ and state s.[4] As a consequence, the price of one asset can be computed using a pure arbitrage argument, by

---

[3]Notice, that since z_k^i is allowed to take negative values, the solution is always interior.

[4]From the envelope theorem, the quantity $\frac{\pi(s)u_i'(c^i(s))}{p(s)} = \phi^i(s)$ represents the marginal value of wealth in state s for consumer i.

Since the allocation in the Radner and AD equilibria is the same, from the FOC in AD and the relationship between spot prices in the two definitions, we have

$$\pi(s) u_i'(c(s)) = \lambda^i \widetilde{p}(s) = \lambda^i \mu(s) p(s)$$

hence $\mu(s) = \frac{\phi^i(s)}{\lambda^i}$.

equating the monetary cost of this asset q_k with its value in dollar terms $\sum_{s\in S}\mu(s)r^k(s)$. In matrix notation we have $q = \mu R$ and μ is unique since by the full rank condition, we can invert $R$and write it as the unique solution to the system of equations $\mu = R^{-1}q$.

When the asset structure is not complete we say we are in an *incomplete markets* framework. In this situation, the solution to the above equation $\mu = R^{-1}q$ is not unique, hence the pricing operator μ is indeterminate. Moreover, both welfare theorems fail in general. In our specific case however, where at period $t = 1$ we have only one consumption good, any Radner equilibrium is constrained Pareto optimal. That is, a Planner who is as constrained as the agents in transferring wealth across states cannot do any better than the allocation induced by the Radner Equilibrium.

To see it, note that given R and the endowments w, if the assets z were expressed in real terms, the consumption $c^i(s)$ in state s is unambiguously give by the portfolio $z^i = (z_k)_{k\in K}$. Indeed $c^i(s) = \omega^i(s) + \sum_k z_k r^k(s)$, hence we rewrite the agent's problem as follows:

$$\max_{z^i} U_i\left(\omega^i(1) + \sum_k z_k r^k(1),\ \dots, \omega^i(S) + \sum_k z_k r^k(S)\right) = \widetilde{U}_i(z_1, z_2,\ \dots, z_K)$$

s.t.

$$\sum_{k\in K} q_k z_k^i \leq 0$$

but then if $\sum_i z_k^i \leq 0$ this agent's maximizing problem defines a Radner Equilibrium, since we can always normalize the prices $p(s) = 1$ in each state. However, notice that this problem is also a well defined problem with K commodities, hence we can apply the First Welfare Theorem. **Q.E.D.**

Finally, note that in a complete market framework, the price formula (6.9) for assets implies that the price of a risk free bond, i.e. and with $r^b(s) = 1\ \forall s$, is

$$q^b = \sum_{s=1}^{S} \mu(s).$$

6.3.2.1 Equilibrium price of time

Let us now twist the model a bit by considering the possibility of consuming in the first period. Moreover, assume the following time separable utility for all agents

$$U_i(c) = u_i\left(c^i(0)\right) + \beta\sum_{s=1}^{S}\pi(s)u_i\left(c^i(s)\right).$$

The price formula must hold again. But what is the price of an Arrow security or that of a bond in this case?

After normalizing $p(0) = 1$, and expressing arrow securities $z(s)$ in real terms, the agent's problem is:

$$\max_{c,z} \; u_i\left(c^i(0)\right) + \beta \sum_s \pi(s) u_i\left(c^i(s)\right)$$

$$\begin{aligned} \text{s.t.} &\quad : \\ c^i(0) + \sum_s q(s) z^i(s) &\leq 0 \qquad \phi^i(0) \\ p(s)\left(\omega^i(s) + z^i(s)\right) &\geq p(s)\, c^i(s), \qquad \phi^i(s) \end{aligned}$$

which gives the usual condition:

$$q(s) = \frac{p(s)\,\phi^i(s)}{\phi^i(0)} = \frac{\beta\pi(s) u_i'\left(c^i(s)\right)}{u_i'\left(c^i(0)\right)}. \tag{6.10}$$

The above condition implies that the ratio $\frac{u_i'\left(c^i(s)\right)}{u_i'\left(c^i(0)\right)}$ must be constant across i. Moreover, whenever the aggregate endowment $\omega(s) = \sum_i \omega^i(s) = \omega$ is constant across states, we saw above that $c^i(s) = c^i(s')$ for all s, s'. If we denote the common (across agents) and state-independent ratio of marginal utilities by $\frac{u'(c_1)}{u'(c_0)}$, the price of a discount bond becomes

$$q^b = \sum_{s=1}^{S} \beta\pi(s) \frac{u'(c_1)}{u'(c_0)} = \beta \frac{u'(c_1)}{u'(c_0)}.$$

The price of a bond is increasing on the agent's patience and decreasing in the growth rate of consumption (note that since u_i is concave $c_1 > c_0$ implies $\frac{u'(c_1)}{u'(c_0)} < 1$). The latter effect is due to the desire for *intertemporal consumption smoothing*. The price of the Arrow security can hence be computed as follows:

$$q(s) = \beta\pi(s) \frac{u'(c_1)}{u'(c_0)} = \pi(s) q^b,$$

where the 'intertemporal preferences' of the economy are summarizes by q^b.

When consumption is stochastic (for example because there is aggregate uncertainty) from condition (6.10) we have that

$$q^b = \sum_s q(s) = \sum_s \frac{\beta\pi(s) u_i'\left(c^i(s)\right)}{u_i'\left(c^i(0)\right)}.$$

Now notice that as long as the marginal utility u_i' is a convex function of $c^i(s)$, from Jensen's inequality we have that

$$\sum_s \frac{\pi(s)u_i'(c^i(s))}{u_i'(c^i(0))} > \frac{u_i'\left(\sum_s \pi(s)\, c^i(s)\right)}{u_i'(c^i(0))},$$

that is, $q^b > \beta$ even though $\sum_s \pi(s)\, c^i(s) = c^i(0)$, that is, even though consumption does not increase on average. This is a dual way of seeing the *precautionary motive for savings:* the bond is quite expensive since 'certeris paribus', the agent likes wealth for a precautionary motive, in order to get partially insured against consumption variations.

In this case, the price of an Arrow security

$$q(s) = \frac{\beta\pi(s)u_i'(c^i(s))}{u_i'(c^i(0))}$$

will crucially depend on $u_i'(c^i(s))$, in particular, it will tend to be higher than $\pi(s)q^b$ for states where $u_i'(c^i(s))$ is high, and vice versa for states associated to low marginal utilities. This is very intuitive.

Exercise 49 *Explain intuitively the previous result, that is, explain why the price of an Arrow security tends to be higher than $\pi(s)q^b$ when the security pays in states associated to a high marginal utility of consumption, and vice versa for securities paying in states associated to low marginal utilities.*

Finally, consider the case where there is no uncertainty but states are represented by different dates alone. We get the standard infinite time horizon model under certainty

$$\max \sum_{t=0}^{\infty} \beta^t u(c_t^i) \quad \text{s.t.}$$

$$\sum_{t=0}^{\infty} p_t c_t^i \leq \sum_{t=0}^{\infty} p_t \omega_t^i. \qquad \lambda^i$$

The budget constraint written is this (AD) form is called consolidate or intertemporal budget constraint. It should be easy to see that if we use the utility $\hat{u} = \frac{u}{1-\beta}$ instead of u the problem becomes identical to that under uncertainty where the probability of state t is $\frac{\beta^t}{1-\beta}$. We can hence apply all above derived results for the two period stochastic environment to this case, and interpret them within our purely dynamic framework with no risk.

Let's keep the time dimension interpretation and consider the agent's fist order conditions

$$\beta^t u'(c_t^i) = \lambda^i p_t,$$

where λ^i is the multiplier associated to agent i budget constraint. For agents i and j we must hence have

$$\frac{u'(c_t^i)}{u'(c_t^j)} = \frac{\lambda^i}{\lambda^j}.$$

And we saw above that whenever $\sum \omega_t^i = \omega$ we must get $c_t^i = \bar{c}^i$. As a consequence, if we normalize $p_0 = 1$, we get $\lambda^i = u'(\bar{c}^i)$ and $p_t = \beta^t$. From the agent's budget constraint we obtain

$$\bar{c}^i = (1-\beta) \sum_{t=0}^{\infty} \beta^t \omega_t^i,$$

that is the agent consumes the *annuity* of the present value of his total income (his *human capital wealth*).

The assets used in the most straightforward definition of the Radner equilibrium are bonds with maturity t, whose price will be denoted as q_t^0.[5] In a more common definition of the sequential equilibrium, the agent will trade each period one-period bonds, and face a sequence of budget constraint of the form

$$c_t + q_{t+1}^t b_{t+1} = b_t + \omega_t,$$

where q_{t+1}^t represents the period t price of a (one-period) bond with maturity $t+1$. When the agent is allowed to trade intertemporally at each date, one typically imposes the so called **no-Ponzi-game (NPG) condition**. In this case it will be

$$\lim_{T\to\infty} \beta^T b_{T+1} \geq 0 \text{ for all otherwise feasible sequences } \{b_{t+1}\}_{t=0}^{\infty}.$$

We now briefly discuss such condition.

Transversality has nothing to do with the No Ponzi Game (NPG) Condition. In Chapter 3 we learnt that we must be careful in defining transversality conditions when the state can take negative values. An example of such as situation is when the state defines the level of assets, and the agent is allowed to borrow. In this same example, it is important to notice that the transversality condition has nothing to do with the so called NPG condition. We saw that the former complements the Euler's equation to describe

[5] Obviously, in this framework the AD prices are $p_t = q_t^0$ for all t.

sufficient conditions for optimality. The NPG condition is typically used in situations where the agent can borrow and save and simply guarantees that the agent's intertemporal budget constraint is satisfied, when a sequence of per period budget constraints is used instead of only one intertemporal budget constraint in period zero.

A typical example is the following. Consider the problem of a consumer who is facing a deterministic path of income and aims at smoothing consumption across time, by taking an appropriate sequence of borrowing and lending decisions. When both the interest rate r_t and income y_t are time constant, with $\beta(1+r) = 1$, that is, the period zero price of consumption in period t is $p_t = \beta^t$ [Check this!], her problem becomes

$$\begin{aligned} & \max_{\{c_t\}_{t=0}^{\infty}} \sum_t \beta^t u(c_t) \\ \text{s.t.} \quad : & \quad c_t \geq 0, \text{ for all } t, \text{ and} \\ \sum_t p_t c_t & \leq \sum_t p_t y, \end{aligned}$$

which, as long as u is strictly concave delivers the plan $c_t^* = y$ for all t as the only optimal solution.

Alternatively, one can denote by b_t the level of debt at period t, and solve the sequential equilibrium problem

$$\begin{aligned} & \max_{\{c_t, b_{t+1}\}_{t=0}^{\infty}} \sum_t \beta^t u(c_t) \\ \text{s.t.} \quad : & \quad c_t \geq 0, \\ c_t + q_{t+1}^t b_{t+1} & = b_t + y, \\ \text{for all } t, \text{ with } b_0 & = 0. \end{aligned}$$

Where - since markets are complete - the price of one period bond is $q_{t+1}^t = \beta$.[Explain why this is the case!]

If we rearrange the sequence of per period budget constraints one gets that for all T

$$\sum_{t=0}^{T} p_t c_t \leq \sum_{t=0}^{T} p_t y - p_T b_{T+1}.$$

It is hence clear that a necessary condition for having the same solution as above is that $\lim_{T\to\infty} p_T b_{T+1} = 0$ for any sequence. Since the agent would never over save, it would be enough to require

$$\lim_{T\to\infty} p_T b_{T+1} = \lim_{T\to\infty} \beta^T b_{T+1} \geq 0, \tag{6.11}$$

which is the usual form of the NPG. And *it must be imposed as additional condition on the last problem.* It is called NPG condition since if it is not imposed on the sequence of $\{b_t\}_{t=0}^{\infty}$ the agent is tempted to consume much more than y by rolling over the debt. For example, consider the possibility of consuming $c_0 = y + \varepsilon$ and $c_t = y$ for all $t \geq 1$. Recall that $p_t = \beta^t$. This would generate an initial debt of $b_1 = \varepsilon$ which will growth precisely at a rate β^{-1} generating $\lim_{T\to\infty} \beta^T b_{T+1} = \varepsilon$. Clearly the agent will use these strategies as long as they are allowed, and perhaps drive $c_0 \to \infty$.

Now let's compare it with the **transversality condition**. Which is an *optimality condition, not a constraint.* Since in our simple problem, the optimal path implies $b_t^* = 0$ for all t, the transversality condition takes the form

$$\lim_{T\to\infty} \beta^T u'\left(c_T^*\right)\left(b_{T+1} - b_{T+1}^*\right) = \lim_{T\to\infty} \beta^T u'\left(c_T^*\right) b_{T+1} \geq 0, \tag{6.12}$$

for all feasible sequences $\{b_t\}_{t=0}^{\infty}$.

Exercise 50 *Show that our solution $c_t^* = y$ and $b_t^* = 0$ for all t solves the transversality condition (6.12).[Hint: Mind the definition of feasibility]*

One important difference between the NPG condition and the transversality condition is that the former does not contemplate the marginal utility. However, sometimes you might get confused either because when u' is finite (at least in the limit) one can disregard it, and the transversality may take the form of the last expression in (6.11).

Another difference between the two is that the NPG conditions refers to the budget constraint hence to the market price p_T. However again, under complete markets and homogeneous agents, the equilibrium price of period t consumption goods takes the value $p_t = \beta^t \frac{u'(c_t^*)}{u'(c_0)}\mu$, for some constant $\mu > 0$. It is however common practice - and we followed it in our example as well - to normalize p_0 to one, so that $\mu = u'(c_0)$. In this case again, condition (6.12) and (6.11) are indistinguishable, although they are objects with a completely different economic nature.

Exercise 51 *An economy consists of two infinitely lived consumers named $i = 1, 2$. There is one non storable consumption good. Consumer i consumes c_t^i at time t. Consumer i ranks consumption streams by*

$$\sum_{t=0}^{\infty} \beta^t u(c_t^i),$$

where $\beta \in (0, 1)$ and u is increasing, strictly concave, and twice continuously differentiable. Consumer 1 is endowed with a stream of consumption good $\omega_t^1 = 1, 0, 1, 0, 1, ...$

Consumer 2 is endowed with a stream of the consumption good $\omega_t^2 = 0, 1, 0, 1, 0, ...$ *Notice that* $\omega_t^1 + \omega_t^2 = 1$ *for any* t. *Assume that there are complete markets with time-0 trading.*
(a) *Define a AD competitive equilibrium, being careful to name all the objects of which it consists.*
(b) *Compute the competitive equilibrium allocation and prices.*
(c) *Recall that a one-period risk free bond is an asset that costs* q_t *in period* t *and pays one unit of consumption goods in the next period. What would the price of that asset be? Argue that in our environment* q *is time constant.*
(d) *Choose any of the two agents, and denote by* b_t *her/his stock of bonds in period* t. *State her/his problem in a (Radner) sequential equilibrium framework. Do not forget to state the No-Ponzi-Game condition.*
(e) *Show that the No-Ponzi-Game condition is equivalent to imposing a couple of constraints on the consumer asset holdings, requiring that they never fall below the fixed amounts* B^0 *and* B^1, *where* B^0, B^1 *are allowed to be negative. In other words, show that there is a couple of borrowing limits* (B^0, B^1) *such that the set of feasible consumption plans defined by the sequential budget constraint and the No-Ponzi-Game restriction is identical to the set of feasible consumption plans by the sequential budget constraints and the borrowing constraints: for all* t, $b_{t+1} \geq B^0$ *(if* $\omega_t = 1$*) or* $b_{t+1} \geq B^1$ *(if* $\omega_t = 0$*). Express* B^0 *and* B^1 *in terms of* q. *[Hint: Imagine a consumer who has zero consumption and whose asset holdings do not change over time]*
(f) *State the agent's problem in recursive form (i.e., write down the Bellman equation for this problem), stating carefully what are the states variables and the controls, and specify the law of motion and the initial conditions for the states.*

Exercise 52 *An economy consists of two infinitely lived consumers named* $i = 1, 2$. *There is one non storable consumption good. Consumer* i *consumes* c_t^i *at time* t. *Consumer* i *endowment in period* t *can take two values:* $\omega_t^i \in \{\underline{\omega}, 1 - \underline{\omega}\}$. *Whenever period* t *consumer* 1 *endowment is* $\omega_t^1 = \underline{\omega}$ *consumer* 2 *period* t *endowment is* $\omega_t^2 = 1 - \underline{\omega}$ *and vice versa. The uncertainty in this economy is described by a shock* $z_t \in \{1, 2\}$ *with probability* π *and* $1 - \pi$ *respectively, whose realization indicates which consumer have the endowment* $\underline{\omega}$ *(e.g. if* $z_t = 2$ *then* $\omega_t^1 = 1 - \underline{\omega}$ *and* $\omega_t^2 = \underline{\omega}$*). Since* $\omega_t^1 + \omega_t^2 = 1$ *for any* t, *consumers face idiosyncratic risk but there is no aggregate uncertainty in the economy. Define the state of the economy in period* t *as* $z^t = (z_1, z_2, ..., ..., z_t)$, *the sequence of past and present* z'*s. If we denote by* $\mu_t(z^t)$ *the probability associated to state* z^t, *agent* i *preferences over* *consumption plans can be specified as follows*

$$\mathbf{E}_0\left[\sum_{t=1}^{\infty}\beta^{t-1}u(c_t^i)\right]=\sum_{t=1}^{\infty}\beta^{t-1}\sum_{z^t}\mu_t\left(z^t\right)u\left(c_t^i\left(z^t\right)\right),$$

where $\beta\in(0,1)$ *and* u *is increasing, strictly concave, and twice continuously differentiable. Assume that there are complete markets with time-0 trading.*
(a) *Define a competitive equilibrium, being careful to name all the objects of which it consists. [Hint. You may want to use the notation* $\omega_t^i\left(z^t\right)\in\{\underline{\omega},1-\underline{\omega}\}$ *for agent* i *endowment in state* z^t*.]*
(b) *Compute a competitive equilibrium and argue that the equilibrium allocation is Pareto efficient.*
(c) *Price a derivative asset which pays .5 units of consumption with certainty in each period. What is the price of a discount bond with maturity* T*?*

6.4 The Lucas Tree Model

In the previous section, we used the properties of efficiency to derive implications for the allocation of consumption. Then we were able to price each redundant security. This procedure can be done only when we are able to compute the equilibrium allocation and markets are complete (or when we know the allocation is constrained efficient). In more complicated situations with aggregate uncertainty or in situations with incomplete markets, this procedure might not be viable. Here below we borrow the strong symmetry across consumers of the Lucas' tree paper, to study the representative agent, and show how several interesting characteristics the price of an assets can be derived from first order conditions, even without fully solving for the equilibrium allocation. We will see that the model can still drive testable restrictions on prices without the need of computing the actual prices (e.g., by only considering arbitrage free conditions such as in the consumption asset pricing model CAPM). We will see that such minimal prediction are actually not always supported by the data (e.g., the so called equity premium puzzle).

Consider a large number of identical consumers, with v-M-N utility. Each agent has exactly the same preferences and owns shares of $k\geq 1$ productive assets in fixed supply (the trees). All trees are identical and produce random quantities $\{d_t\}$ of a single perishable consumption in all time periods (the dividends). There is hence a common (aggregate) shock driving dividends which is uninsurable. The agent can trade shares of the trees (the risky asset), and one bond.

The problem faced by the representative agent is

$$\max_{\{c_t s_t, b_t\}} \quad \mathbf{E}_0\left[\sum_{t=0}^{\infty}\beta^t u(c_t)\right] \quad \text{s.t.}$$
$$c_t + q_t b_t + p_t s_t \leq b_{t-1} + (p_t + d_t)\, s_{t-1},$$
$$s_t \geq 0, \text{with} s_{-1} \text{given, and } b_{-1} = 0;$$
$$\text{and the NPG condition: } \lim_{T\to\infty}(\Pi_{s=0}^{T} q_s) b_T \geq 0 \text{ with probability one.}$$

In the above formulation, the value $r_{t+1} := q_t^{-1} - 1$ corresponds to the risk free interest rate, which is assumed to be known with certainty at period t. For all variable, the time t subscript also indicates that the variable is chosen/known in period t, it is h^t-measurable.[6] The first order conditions yield

$$q_t u_t'(c_t^*) = \mathbf{E}_t\left[\beta u'(c_{t+1}^*)\right] \tag{6.13}$$
$$p_t u'(c_t^*) = \mathbf{E}_t\left[\beta\left(p_{t+1} + d_{t+1}\right) u'(c_{t+1}^*)\right] \tag{6.14}$$

The first condition is the standard Euler equation with respect the risk free asset, and it is well known since Hall (1978). It roughly states that intertemporal efficiency requires marginal utility to follow a martingale. Rearranging (6.14) yields

$$p_t = \mathbf{E}_t\left[\left(p_{t+1} + d_{t+1}\right)\frac{\beta u'(c_{t+1}^*)}{u'(c_t^*)}\right] = \mathbf{E}_t\left[\left(p_{t+1} + d_{t+1}\right) SDF_{t+1}\right].$$

The term $\frac{\beta u'(c_{t+1}^*)}{u'(c_t^*)}$ is often denoted as the stochastic discount factor (SDF), and from (6.13) we obviously have

$$q_t = \mathbf{E}_t\left[SDF_{t+1}\right].$$

[6] A more extensive (and more rigorous) version of the problem is as follows. Define $h^t := (d_0, d_1, ..., d_t)$ the past history of dividends, let $\pi\left(h^t\right)$ the probability that history h^t occurs. We restrict attention to situations where both interest rates and the prices for shares p_t are function of h^t only. Then let

$$(\mathbf{c}, \mathbf{s}, \mathbf{b}) : = \left\{\mathbf{c}_t\left(h^t\right), \mathbf{s}_t\left(h^t\right), \mathbf{b}_t\left(h^t\right)\right\}_{t=0}^{\infty}$$

be the agent's contingent plan. The agent solves the following problem

$$\max_{\mathbf{c},\mathbf{s},\mathbf{b}} \quad \sum_{t=0}^{\infty}\beta^t\left[\sum_{h^t\in H^t}\pi\left(h^t\right) u(\mathbf{c}_t\left(h^t\right))\right]$$
$$\text{s.t.} \quad s_{-1} \text{ given, and } b_{-1} = 0; \text{ and for all histories } h^t$$
$$\mathbf{c}_t\left(h^t\right) + q_t(h^t)\mathbf{b}_t\left(h^t\right) + p_t\left(h^t\right)\mathbf{s}_t\left(h^t\right) \leq b_{t-1}\left(h^{t-1}\right) + \left(p_t\left(h^t\right) + d_t\right) s_{t-1}\left(h^t\right),$$
$$\text{and } \lim_{T\to\infty}(\Pi_{s=0}^{T} q_t(h^t))\mathbf{b}_t\left(h^t\right) \geq 0, \text{with probability one.}$$

Using the properties of expectation operator we can express the price of the tree as

$$p_t = \mathbf{E}_t\left(p_{t+1}+d_{t+1}\right)\mathbf{E}_t\frac{\beta u'(c^*_{t+1})}{u'(c^*_t)} + Cov_t\left(\left(p_{t+1}+d_{t+1}\right),\ \frac{\beta u'(c^*_{t+1})}{u'(c^*_t)}\right).$$

The covariance component of the price suggests that the price of one asset is large when is positively correlated with the marginal rate of substitution, that is when c^*_{t+1} is low with respect to c^*_t. The reason being that in this case the asset is very useful for insurance, as it provides the agent with some income precisely when consumption tend to be low.

6.4.1 The Consumption CAPM and the Security Market Line

Recall again

$$p_t = \mathbf{E}_t\left(p_{t+1}+d_{t+1}\right)\mathbf{E}_t\frac{\beta u'(c^*_{t+1})}{u'(c^*_t)} + Cov_t\left(p_{t+1}+d_{t+1},\ \frac{\beta u'(c^*_{t+1})}{u'(c^*_t)}\right).$$

If we define the return of the risky asset: $\frac{p_{t+1}+d_{t+1}}{p_t} = 1+\rho_{t+1}$, and using the definition of the interest rate r_{t+1} we obtain

$$\begin{aligned}\mathbf{E}_t\left[(1+\rho_{t+1})u'(c^*_{t+1})\right] &= \mathbf{E}_t\left[(1+r_{t+1})u'(c^*_{t+1})\right] = \frac{u'(c^*_t)}{\beta}\\ &\Rightarrow \mathbf{E}_t\left[(\rho_{t+1}-r_{t+1})u'(c^*_{t+1})\right] = 0\end{aligned}$$

Since r_{t+1} is known in period t, this condition can equivalently be written as:

$$\mathbf{E}_t\rho_{t+1} = r_{t+1} - \frac{Cov_t\left(\rho_{t+1},\ u'(c^*_{t+1})\right)}{\mathbf{E}_t u'(c^*_{t+1})}. \tag{6.15}$$

Now suppose there is an asset (m) denote as “the market” which is perfectly negatively correlated with $u'(c^*_{t+1})$, i.e. $u'(c^*_{t+1}) = -\nu\rho^m_{t+1}$. Intuitively, this asset is driven by the aggregate shock, which hence drives aggregate consumption. This way we can talk about exogenous and non-observable factors that move the returns (such as the aggregate shock) by using endogenous variable such as those created by by forming the portfolio we denotes as market.

Then, we obviously have

$$Cov_t\left(\rho_{t+1},\ u'(c^*_{t+1})\right) = -\nu Cov_t(\rho_{t+1},\rho^m_{t+1}),$$

hence, from the first order conditions we have

$$\mathbf{E}_t\rho^m_{t+1} = r_{t+1} + \nu\frac{Var_t\left(\rho^m_{t+1}\right)}{\mathbf{E}_t u'(c^*_{t+1})}.$$

Rearranging terms and using (6.15) we get

$$\mathbf{E}_t\rho_{t+1} - r_{t+1} = \frac{Cov_t(\rho_{t+1}, \rho_{t+1}^m)}{Var_t\left(\rho_{t+1}^m\right)}\left(\mathbf{E}_t\rho_{t+1}^m - r_{t+1}\right).$$

This is the condition that has often been used in empirical finance literature. By regressing $\mathbf{E}_t\rho_{t+1} - r_{t+1}$, the expected return of a risky asset in excess of the save rate, on $\left(\mathbf{E}_t\rho_{t+1}^m - r_{t+1}\right)$: the expected return on the market asset in excess of the safe rate, one obtains the beta coefficient $\beta^\rho = \frac{Cov_t(\rho_{t+1},\rho_{t+1}^m)}{Var_t\left(\rho_{t+1}^m\right)}$ which has the usual interpretation of the regression coefficient 'beta'. Notice that we do not require the specification of agent's preferences and risk aversion. However, his theoretical foundation is based on the so called *two-fund separation theorem*, which states condition in which all agents hold a portfolio of only a risk free asset and a share of all risky assets in the economy. And the theorem does require assumptions on agents' preferences about risk.[7] The SML verifies a powerful concept. Individuals who hold a portfolio will value assets for the marginal contribution they will make to the portfolio. If an asset has a positive beta with respect to that portfolio, then adding it to the portfolio will increase the volatility of the portfolio's returns. To compensate a risk averse investor for this increase in volatility, in equilibrium such asset must have a positive excess expected return, and the SML verifies that this excess return will, in fact, be proportional to the beta. Note however that we have just shown that this intuition is true in a multiperiod setting where agents may adjust their portfolios. This is because we assumed that agent's preferences are additive separable and that the agent only cares about consumption. The mode we presented is also denotes as the Consumption Beta Model.

6.4.2 Pricing assets from fundamentals

Finally, we use our conditions to price assets in terms of the fundamentals. Here - in following Lucas (1978), we use heavily the symmetry across consumers. In order to rule out bubbles, we focus on equilibria such that some strong transversalities are satisfied,

[7] In particular, it must be that the absolute risk tolerance is linear with common slope. That is, if x is the wealth level of agent i, u_i are such that the ratios

$$-\frac{u_i'(x)}{u_i''(x)} = a_i + bx$$

with b the same across i (Gollier, 2001, page 314).

and obtain

$$p_t = \mathbf{E}_t \left[\sum_{n=1}^{\infty} \frac{\beta^n u'(c^*_{t+n})}{u'(c^*_t)} d_{t+n} \right]$$

Now use the law of iterate expectations in (6.13) and get the period $t+1+n$ return for the bond

$$\frac{1}{1+r_{t+1+n}} = \mathbf{E}_t \frac{\beta u'(c^*_{t+n+1})}{u'(c^*_{t+n})} = \mathbf{E}_t \left[SDF_{t+1+n} \right].$$

We can now use such condition to obtain the price reflecting the *fundamental* value of the tree:

$$\begin{aligned} p_t &= \mathbf{E}_t \left[\sum_{n=1}^{\infty} \frac{d_{t+n}}{\prod_{k=1}^{n} (1+r_{t+k})} \right] \\ &= \mathbf{E}_t \left[\sum_{n=1}^{\infty} \frac{d_{t+n}}{(1+r)^n} \right] \text{ when } r_t = r. \end{aligned}$$

This is the well known result that the price of an assets is represented by the expected discounted value of its stream of future dividends.

In order to find the final expression for prices one need to compute the interest rate. This can be done by using the market clearing condition and the extreme symmetry across consumers (that actually generate a no-trade equilibria) $c^*_t = d_t$.

Exercise 53 *Use the first order conditions of the agent to describe the stochastic process of the interest rate when d_t is i.i.d.*

6.5 The Equity Premium Puzzle

Assume CRRA/CES utility[8]

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

that the interest rate r is time constant, and write the discount factor as $\beta = \frac{1}{1+\theta}$. Notice that $\frac{1}{\gamma}$ represents the intertemporal elasticity of substitution of consumption.

Let's investigate first what intertemporal substitution implies, that is, the *time dimension* of the problem. The Euler equation for the risk free asset imposes an upper bound for γ since

$$1 = \beta(1+r) \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} = \frac{1+r}{1+\theta} \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}$$

[8] The known evidence suggests agents have decreasing absolute risk aversion (DARA), but not necessarily for CRRA.

which can be written as

$$r \simeq \theta + \gamma g \tag{6.16}$$

where g is the growth rate of consumption: $\frac{c_{t+1}-c_t}{c_t}$. The intuition for such condition is simple. The remuneration for waiting (r) must compensate impatience (θ) and the desire for intertemporal consumption smoothing (γg). With an interest rate not very far from $r = 5\%$, a growth rate $g = 2\%$, and a positive $\theta \geq 0$ (impatience); we have that in order to satisfy the Euler equation γ should be less than 2.5. This is sometimes denoted as the *risk-free puzzle,* in the sense that if we accept the existing estimates for γ - which are all below 2 - one must have a $\theta < 0$. If we want to remain agnostic about the unobservable parameters of the representative consumer, we should first investigate the cross-sectional dimension of the problem, driven by risk aversion.

However, this seems to contrasts with the *risk dimension.* Recall again our optimality conditions:

$$\mathbf{E}_t \rho_t = r_t + \gamma Cov_t\left(\rho_t, g\right).$$

Mehra and Prescott (1985) found that in the data the premium between stock and bond was 6% (Short-Term Debt $\mathbf{E}r_t = 1\%$ and Stock Exchange Index $\mathbf{E}\rho_t = 7\%$ period: 1889-1978). With Covariance between g and ρ of a bit more than 0.002, γ should be at least 25. The *equity premium puzzle* refers to the contrast between the relationship that links the real risk free return on capital r, risk aversion and growth rate (6.16), and the fact that in order to explain the difference between risky and risk free returns with the standard asset pricing model you need a much larger risk aversion parameter γ (sometimes between 50 and 100).

Kreps-Porteus (1978) preferences allow for the difference between intertemporal substitutability and risk aversion. See Weil (1990) for a survey of the implications of Kreps-Porteus preferences and non-separabilities for macroeconomics.

In the past, several authors tried to solve the Equity premium puzzle in many different ways. One of the first attempts was to introduce incomplete markets and investigate the implications for the representative agent. It seems that *as long as idiosyncratic shocks are orthogonal to aggregate shocks*, incomplete insurance markets can only relax the prediction on the interest rate, i.e., they might only help addressing the risk free puzzle. The reader is demanded to a new interesting working paper by Kruger and Lustig (2006).

Bibliography

[1] Cooley, T. F., and E. C. Prescott (1995) "Economic Growth and Business Cycles," in T. F. Cooley ed., *Frontiers of Business Cycle Research*, Princeton University Press: 1-38.

[2] King, R. G., and S. T. Rebelo (1999) "Resuscitating Real Business Cycles," in J. B. Taylor, and M. Woodford eds., *Handbook of Macroeconomics*, Vol. 1B, Elsevier: 927-1007.

[3] Mas-Colell, A., Michael D. Winston and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press.

[4] Sargent, T. (1987) *Dynamic Macroeconomic Theory*, Harward University Press.

[5] Sargent, T. J. and L. Ljungqvist (2000), *Recursive Macroeconomic Theory*, MIT Press.

[6] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

Chapter 7

Introduction to Consumption Theory

- Only one risk free asset b but uncertain income y
- Intertemporal preferences with forward looking individuals

$$\max \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \quad \text{s.t.} \quad c_t + q_t b_t \leq b_{t-1} + y_t$$

  with $q_t = \frac{1}{1+r_t}$. (Partial equilibrium approach).

- Key optimality condition: the usual Euler equation

$$q_t u'(c_t) = \beta \mathbf{E}_t [u'(c_{t+1})].$$

- It is an intertemporal optimality condition.
- Because of incomplete markets the agent cannot equate marginal utility across states. Marginal utility is hence a martingale.

7.1 Quadratic Utility

- When $u(c_t) = Bc_t - \frac{1}{2}c_t^2$ is quadratic the Euler equation becomes

$$q_t c_t = (q_t - \beta)B + \beta \mathbf{E}_t [c_{t+1}]$$

  or - if we denote by η_{t+1} the innovation in consumption and $q_t = \beta$

$$c_{t+1} = c_t + \eta_{t+1},$$

  that is, consumption itself is a martingale (possibly with deterministic drift if $q_t \neq \beta$).

- This prediction of the intertemporal maximization problem with rational expectations has been widely tested in the literature. (Hall, Ec.metrica 1978; Flavin, JPE 1981; ....)

- Earlier literature finds that aggregate consumption is excessively sensitive: c_{t+1} reacts to date t or earlier variables other than c_t.

7.1.1 Liquidity Constraints: Excess Sensitivity

- Several authors interpreted the excess sensitivity of consumption as evidence of the presence of liquidity constraints (e.g., Zeldes, JPE 1989; Deaton, Ec.metrica 1991).

- Liquidity constraints can be empirically detected as the violation of the Euler equation
$$q_t u'(c_t) > \beta \mathbf{E}_t [u'(c_{t+1})].$$

- Since the agent cannot borrow more than a certain amount $b_t \geq -B$, when the liquidity constraint is binding he/she is forced to consume a lower c_t than that desired.

- Other authors argue that once u, demographic variables, and labour supply have been properly modeled there is no evidence of a failure of the Euler equation at individual level (e.g. Attanasio, Handbook 2000; or Attanasio and Weber, REStud 1993).

- Consumption in period t may still react to predictable changes in permanent income just because when the information has been revealed the consumer was unable to borrow against an income level below the permanent income. Consumption should not over react to bad news as they require to save rather than borrow (Shea, AER 1995). He does not find evidence of Liquidity constraints.

- Liquidity constraints can rigorously be micro-founded with to commitment problems: Default models.

- There is now a new steam of literature that aim at testing the predictions of the default model.

- MRS as a new state variable (Kocherlakota, RES 1996; and Thomas and Worral, RES 2002), liquidity constraints should be less stringent as the variance of temporary income increases (Attanasio and Albarran, mimeo 2003a,b; also Krueger and Perri, JEEA 2004), ...

7.1.2 Quadratic Utility: Certainty Equivalence

- The intertemporal maximization problem with quadratic utility has a much sharper prediction for consumption.

- To see it notice that since the agent can only use risk free bonds, the budget constraint (together with the NPG condition) implies (assuming $q_t = q = \beta$)

$$\sum_{s=0}^{\infty} q^s c_{t+s} = b_t + \sum_{s=0}^{\infty} q^s y_{t+s},$$

for all future income histories $\{y_{t+s}\}_{s=0}^{\infty}$. Which implies

$$\mathbf{E}_t \sum_{s=0}^{\infty} q^s c_{t+s} = b_t + \mathbf{E}_t \sum_{s=0}^{\infty} q^s y_{t+s} = b_t + H_t$$

where H_t is denominated human capital wealth.

- From the Euler equation and the law of iterated expectations we can simplify the left hand side of the budget constraint as well:

$$\mathbf{E}_t \sum_{s=0}^{\infty} q^s c_{t+s} = \sum_{s=0}^{\infty} q^s c_t = \frac{c_t}{1-q}.$$

Hence we have

$$c_t = (1-q)\left[b_t + H_t\right].$$

If we denote by $A_t = b_t + H_t$ the total wealth of the agent we get the permanent income hypothesis equation

$$c_t = (1-q)\, A_t = y_t^p,$$

that is, the agent consumes the annuity value of its total wealth, his/her permanent income.

7.2 Permanent Income Hypothesis (PIH)

- The difference version of this hypothesis is

$$c_{t+1} - c_t = \Delta c_t = \Delta y_t^p = (1-q)\left(\mathbf{E}_{t+1} - \mathbf{E}_t\right) \sum_{s=1}^{\infty} q^s y_{t+s}$$

  that is, consumption growth should react one-to-one with innovations in the permanent income.

- More than today's income, what matters for consumption variation is the informational content of today's income shock.

- Consumption does not change a lot with today's income if there is no persistence.

- Campbell and Deaton, RES 1989; Gali', AER 1991 and others find that consumption is excessively smooth, i.e., consumption growth is smoother than permanent income.

- Is it because agents have superior information? Saving reveals agent's (superior) information: saving for the 'rainy days'.

7.2.1 Excess Smoothness versus Excess Sensitivity

- Excess sensitivity and excess smoothness are not tightly linked. Excess sensitivity refers to how consumption reacts to past (predictable) income shocks whereas excess smoothness refers to how consumption reacts to present (unpredictable) income shocks.

- Micro data display no-excess sensitivity but some excess smoothness. Recall that excess smoothness result regards both the Euler equation and the budget constraint.

- Full-insurance (complete-markets) models can do the job, but seem to be inconsistent with the data (e.g., Attanasio and Davis, 1995).

- In Attanasio and Pavoni (2005) we propose a model with informational asymmetries (endogenous incomplete markets) where the Euler equation is satisfied but consumption displays excess smoothness because of the failure of the intertemporal budget constraint with a single asset.

- Campbell and Deaton, RES 1989; Gali', AER 1991 and others find that consumption is excessively smooth, i.e., consumption growth is smoother than permanent income.

- The usual confusion arises here. Obviously consumption can display at the same time excess sensitivity and excess smoothness. Excess sensitivity refers to how consumption reacts to past (predictable) income shocks whereas excess smoothness refers to how consumption reacts to present (unpredictable) income shocks.

- However notice that the excess smoothness result regards both the euler equation and the budget constraint.

- In Attanasio and Pavoni (2005) we propose a model with informational asymmetries (endogenous incomplete markets) where the euler equation is satisfied but consumption displays excess smoothness because of the failure of the intertemporal budget constraint with a single asset.

Chapter 8

Theory of Fixed Investment and Employment Dynamics

Investment is expenditures by firms on equipment and structures. Business (fixed) investment is commonly held to be an important determinant of an economy's long-run growth. While the significance of short-term changes in business investment is less widely recognized, the importance of such changes for the business cycle has been known to economists since the beginning of the last century. For example, many believe that the US record expansion in the 90s had been driven, at least in part, by strong investment in computers and related equipment.

For individual plants, investment is simply the expenditure required to adjust its stock of capital. Capital includes all equipment and structures the plant uses. The plant combines capital with other inputs, such as labor and energy, to produce goods or services. When an extraction company acquires diesel engines, it is investing in equipment. When an automobile manufacturer builds a new warehouse, it is investing in structures.

Since investment spending raises future capital and thus the quantity of goods and services that may be produced in the future, plants will tend to adjust their investment levels in response to forecasted changes in the market's demand for their own output. Changes in productivity will also tend to increase investment. For example, if the efficiency with which inputs may be combined to produce output increases, the firm may be able to sell more of its product, since it can offer it at a more attractive price. The firm may then expand and more workers may be hired. These workers will need equipment, and, as a result, investment will rise.

8.1 The Value of the Firm

We denote by V_t^* the value of the firm. If he stock market works efficiently, V_t^* should correspond to the expected discounted present value (DPV) of all future profits π_{t+j}^* , $j = 1, ..., \infty$ from period t onward. But at what rate should firm discount cash flows? Recall the Lucas's tree model, where the consumer trades a risk free bond and and a risky asset (the trees). If we interpret the risky asset as shares of our firm, the first order conditions of the consumer are

$$\begin{aligned} u'(c_t^*) &= \mathbf{E}_t \left[\beta(1+r_t) u'\left(c_{t+1}^*\right) \right] \\ V_t^* u'(c_t^*) &= \mathbf{E}_t \left[\beta \left(V_{t+1}^* + \pi_{t+1}^* \right) u'\left(c_{t+1}^*\right) \right], \end{aligned}$$

where $SDF_{t+1} = \frac{\beta u'\left(c_{t+1}^*\right)}{u'(c_t^*)}$ is the stochastic discount factor, and $\frac{V_{t+1}^* + \pi_{t+1}^*}{V_t^*}$ represents the rate of return from holding the firm. When $r_t = r$ for all t, the law of iterated expectations implies $\frac{1}{1+r} = \mathbf{E}_t \left[SDF_{t+j+1} \right]$ for all $j \geq 0$.

If we focus on equilibria with no-bubbles, unraveling the second conditions and using the law of iterated expectations, yields

$$V_t^* = \mathbf{E}_t \left[\sum_{j=1}^{\infty} \frac{\beta^j u'(c_{t+j}^*)}{u'(c_t^*)} \pi_{t+j}^* \right].$$

Rearranged, this condition implies:

$$V_t^* = \mathbf{E}_t \left[\sum_{j=1}^{\infty} \left(\frac{1}{1+r} \right)^j \pi_{t+j}^* \right].$$

Cash flow should hence be discounted by at the real risk free rate.[1]

The above evaluation of the firm also applies to the case where agents were heterogeneous, facing idiosyncratic shocks, as long as the asset market is *complete,* that is, the existing securities would span all states of nature. In a world with incomplete asset markets things can get very complicated. The bottom line is that prices are not longer uniquely determinated. There are however cases where we can still use the above definition to evaluate the value of a firm or of a project. This is the situation where the profits of the firms are zero in all states of nature not spanned by the securities. In this case, the existing securities will still span all states of nature which are relevant for the firm.

[1]It is easy to see that when r_t changes with time, we have $\frac{1}{1+r_{t+n}} = \mathbf{E}_t \left[SDF_{t+n+1} \right]$ and

$$V_t^* = \mathbf{E}_t \left[\sum_{j=1}^{\infty} \frac{\pi_{t+j}^*}{\Pi_n^j \left(1 + r_{t+n}\right)} \right].$$

8.2 Problem of the Representative Firm

Consider the problem of an infinitely lived firm that in every period chooses how much to invest, i.e. how much to add to its stock of productive capital.

Since we aim at studying the behavior of *aggregate* investment, we assume that the firm owns capital. We could have assumed that the firm hires capital from consumers or from a firm who produces it. This requires a second agent and the distinction between internal and external adjustment costs. At the aggregate level, internal and external adjustment costs have equivalent implications (see Sala-i-Martin, 2005).[2]

This firm has hence a dynamic choice. Because it takes time to manufacture, deliver, and install new capital goods, investment expenditures today do not immediately raise the level of a plant's capital. So investment involves a dynamic trade-off: by investing today, the firm foregoes current profits to spend resources in order to increase its stock of future capital and raise future production and future profits.

Clearly, every period the firm will also choose labor input n_t, but we abstract from this static choice, and assume the firm has already maximized with respect to n_t when is called to make the optimal investment decision.[3]

The law of capital is the usual one

$$k_{t+1} = (1 - \delta) k_t + i_t.$$

In each period, the firm produces with the stock of capital k_t, which hence partially depreciates (δ), it then makes the (gross) investment decision i_t that will determine the capital stock in place for next period's production k_{t+1}.

[2] We will assumed that the interest rate and prices are exogenously given to the firm. This seems a reasonable assumption if we want to think about the behavior of individual firms. *Aggregate* investment however, both depends upon and affects the interest rate of the economy. That is, in the aggregate the interest rate is endogenous. One can endogenize the real interest rate by embodying the individual neoclassical firms we will describe in a general equilibrium model where there are also consumers and the interest rate is determined by the equalization between the desired investment by firms and desired savings by households. This will give raise to the Neoclassical model of economic growth we saw in Chapter 1.

[3] Recalling the analysis of Chapter 1, an alternative possibility would be to assume that n_t is supplied inelastically by individuals. Hence if we normalize the aggregate labor supply to one, market clearing always requires $n_t = 1$ for all levels of capital. In other terms, for each k the market wage fully adjust - since there is a vertical labor supply - to

$$w(k) = F_n(k, 1).$$

.

For notational reasons, in what follow we abuse a bit in notation and denote by V_t^* the values of the firm *including* period t dividends π_t^*. Given k_0, and a sequence of profit functions $\{\Pi_t\}_{t=0}^{\infty}$ the sequential problem the firm is facing in period $t = 0$ can be formulated as follows

$$\begin{aligned} V_0^* &= \max_{\{i_t, k_{t+1}\}} \mathbf{E}_0 \left[\sum_{t=0}^{\infty} \left(\frac{1}{1+r} \right)^t \Pi_t \left(k_t, k_{t+1}, i_t \right) \right] \\ & s.t. \\ k_{t+1} &= (1-\delta) k_t + i_t \\ k_{t+1} &\geq 0, \ , \text{ for all } t; \ k_0 \text{ given.} \end{aligned}$$

8.3 The Neoclassical Theory of Investment

We did not specify the cash flow (or profit) functions Π_t yet. The traditional neoclassical theory of investment uses a very simple formulation of the problem: Denote the production function of the firm by $f(k_t)$, the level of technology of the firm at time t by z_t, and the price of a unit of investment good or the unit price of capital goods as p_t;[4] we have

$$\Pi_t \left(k_t, k_{t+1}, i_t \right) = \Pi \left(k_t, k_{t+1}, i_t; p_t, z_t \right) = z_t f \left(k_t \right) - p_t i_t,$$

hence the optimal profit in each period is $\pi_t^* = z_t f \left(k_t^* \right) - p_t i_t^*$.

Consider now the deterministic version of the model. Given k_0 and the sequential of prices and shocks $\{p_t, z_t\}_{t=0}^{\infty}$ problem specializes to

$$\begin{aligned} V_0^* &= \max_{\{i_t, k_{t+1}\}} \sum_{t=0}^{\infty} \left(\frac{1}{1+r} \right)^t \left[z_t f \left(k_t \right) - p_t i_t \right] \\ & s.t. \\ k_{t+1} &= (1-\delta) k_t + i_t; \\ k_{t+1} &\geq 0, \text{ for all } t; \ k_0 \text{ given.} \end{aligned} \tag{8.1}$$

We now derive the Euler equation for the problem. We are hence looking for a feasible deviation from the optimal *interior* program $\{i_t^*, k_{t+1}^*\}_{t=0}^{\infty}$, where interiority simple requires $k_{t+1}^* > 0$ for all t. In the spirit of the Euler variational approach, the perturbation is aimed at changing k_{t+1}^* (and i_t^*, i_{t+1}^*), while keeping unchanged all k_s^* for $s \neq t+1$, in particular both k_t^* and k_{t+2}^*.

[4] Notice that this is a price relative to the price of the final good, which is normalized to one as usual.

Let ε any real number (positive and negative) in an open neighborhood O of zero. Such neighborhood is obviously constructed to maintain feasibility. For each ε, the perturbed plan $\{\hat{\imath}_t^\varepsilon, \hat{k}_{t+1}^\varepsilon\}_{t=0}^\infty$ is constructed from $\{i_t^*, k_{t+1}^*\}_{t=0}^\infty$ as follows: $\hat{k}_{t+1}^\varepsilon = k_{t+1}^* + \varepsilon$, and $\hat{k}_s^\varepsilon = k_s^*$ for $s \neq t+1$. It is easy to check from the law of motions that such perturbation to the optimal plan is achieved by modifying the investment plan as follows: $\hat{\imath}_t^\varepsilon = i_t^* + \varepsilon$ and $\hat{\imath}_{t+1}^\varepsilon = i_{t+1}^* - (1-\delta)\,\varepsilon$ and $\hat{\imath}_s^\varepsilon = i_s^*$ for $s \neq t, t+1$. If we denote by $\hat{V}_0(\varepsilon)$ the value associated to the perturbed plan for each $\varepsilon \in O$, the optimality of the original plan implies $\hat{V}_0(\varepsilon) \leq V_0^*$ for all $\varepsilon \in O$, and $\hat{V}_0(0) = V_0^*$. Stated in other terms, $\varepsilon = 0$ is the optimal solution to

$$\max_{\varepsilon \in O} \hat{V}_0(\varepsilon).$$

The necessary first order condition of optimality is hence $\hat{V}_0'(0) = 0$. Since k_s^* are untouched, both for $s \leq t$ and $s \geq t+2$ the derivative with respect to ε of all terms are zero but period t and $t+1$ returns. We hence have:[5]

$$\begin{aligned} &(1+r)^t\,\hat{V}_0'(\varepsilon) = \\ = \;& \frac{d}{d\varepsilon}\left[z_t f\left(k_t^*\right) - p_t\left(i_t^* + \varepsilon\right) + \left(\frac{1}{1+r}\right)\left(z_{t+1} f\left(k_{t+1}^* + \varepsilon\right) - p_{t+1}\left(i_{t+1}^* - (1-\delta)\,\varepsilon\right)\right)\right]. \end{aligned}$$

The FOC condition $\hat{V}_0'(0) = 0$ hence delivers the following Euler equation:

$$p_t = \frac{1}{1+r}\left[z_{t+1} f'\left(k_{t+1}^*\right) + p_{t+1}(1-\delta)\right]. \tag{8.2}$$

This condition determines the optimal level of next period capital (hence the optimal investment decision, given k_t). It states that the marginal cost of a unit of investment equals the marginal benefit. The marginal cost is the price of capital p_t. The marginal benefit accrues next period, so it's discounted by $(1+r)$. The next period marginal benefit is composed by two terms: *(i)* the increase in production associated to the higher stock of capital $z_{t+1}f'(k_{t+1})$ and *(ii)* the market value of one unit of capital after production. Equation (8.2) is sometimes called the Jorgenson's optimal investment condition, from the name of the Harvard's economist who advanced this theory.

If we assume that $p_t = 1$, i.e. that the price of capital is constant at one, the Euler equation (8.2) becomes

$$r + \delta = z_{t+1} f'\left(k_{t+1}^*\right),$$

[5] In fact, it is as if we solved the local problem:

$$\max_{\varepsilon}\; z_t f\left(k_t^*\right) - p_t\left(i_t^* + \varepsilon\right) + \left(\frac{1}{1+r}\right)\left(z_{t+1} f\left(k_{t+1}^* + \varepsilon\right) - p_{t+1}\left(i_{t+1}^* - (1-\delta)\,\varepsilon\right)\right)$$

which is the usual condition one gets in the static model of the firm. When the firm rents capital (instead of purchasing it), $r+\delta$ represents the *user cost of capital.* In words, the previous equality says that firms invest (purchase capital) up to the point where marginal product of capital (net of depreciation δ) equals the return on alternative assets, the real interest rate.[6]

Exercise 54 *Set $p_t = p$ and $z_t = z$ for all t, and state the problem in recursive form, carefully specifying what are the controls and the states of the dynamic problem, and what are the law of motion for the states. Now compute the FOC with respect to k_{t+1}, and the envelope condition. By rearranging terms, you should get the Euler equation for this problem. Perform the same exercise assuming that productivity $z \in \{z_h, z_l\}$ follows a Markov chain with transition matrix $\begin{bmatrix} \pi & 1-\pi \\ 1-\pi & \pi \end{bmatrix}$, while the price of capital is fixed at p.*

When f is concave $f'(k_{t+1})$ is a decreasing function, hence invertible. If we denote by h the inverse function of $f'(k_{t+1})$, condition (8.2) can be written as

$$k^*_{t+1} = h\left(\frac{p_t\,(1+r) - p_{t+1}\,(1-\delta)}{z_{t+1}}\right), \tag{8.3}$$

with $h(\cdot)$ a decreasing function. We hence have that the optimal level of next period capital (hence investment as k_t is given) is increasing in z_{t+1} and p_{t+1}, while it decreases in p_t, r, and δ.

Exercise 55 *Explain intuitively, in economic terms, why according to (8.3) investment is increasing in z_{t+1} and p_{t+1}, and decreasing in p_t, r, and δ.*

Exercise 56 *(i) Assume $p_t = \bar{p}$ and $z_t = \bar{z}$ for all t and derive the steady state level of capital and investment. (ii) Now state the transversality condition problem (8.1) and verify that the optimal path converging to the steady state satisfies the transversality condition.*

8.4 Convex Adjustment Costs: The q-Theory of Investment

The neoclassical model has a couple of drawbacks. First, consider the case where firms are heterogeneous, say they have different marginal product of capital. As long as all

[6] See also Abel and Blanchard (1983).

firms face the same interest rate and prices for investment goods, all investment in the economy will take place in the firm with the highest marginal product of capital. This is clearly a counterfactual implication of the model.[7]

Another potential source of unrealistic behavior is that current investment is independent of future marginal products of capital. Recall that the equalization of marginal product to interest rate yields the desired level of capital and that investment is then equal to the difference between the existing and the desired capital stocks. Hence, investment is a function of both the existing capital stock and the real interest rate, but is independent of *future* marginal products of capital. If firms know that the marginal product will increase at some point T in the future, their best strategy is not to do anything until that moment arrives at which point they will discretely increase the amount of capital to the new desired level. In other words, because firms can discretely get the desired capital level at every moment in time, it does not pay them to plan for the future since future changes in business conditions will be absorbed by future discrete changes in capital stocks. Economists tend to think that future changes in business conditions have effects on today investment decisions. To get rid of this result we need a theory that makes firms willing to smooth investment over time. One way of introducing such a willingness to smooth investment is to make it costly to invest or disinvest large amounts of capital at once. This is the idea behind the concept of *adjustment costs.*

We will now imagine that firms behave exactly as just described, except that there are some installation or adjustment costs. By that we mean that, like in the neoclassical model, p units output can be transformed into one unit of capital. This capital (which we will call "uninstalled capital") is not useful until it is installed. Unlike the neoclassical model, firms have to pay some installation or adjustment costs in order to install or uninstall capital. These adjustment costs are foregone resources within the firm: for example computers can be purchased at price p but they cannot be used until they have been properly installed. The installation process requires that some of the workers stop working in the production line for some of the time. Hence, by installing the new computer the firm foregoes some resources, which we call *internal adjustment costs.*

The (cash-flow or profit) function Π_t will be modified as follows:

---

[7] A similar type of situation arises when we consider the world economy where all countries face the same "world real interest rate" but different countries have different levels of capital (so the poorest country has the highest marginal product of capital). If capital is free to move across borders, the neoclassical model of investment predicts that *all* the investment in the world will take place in the poorest country.

$$\Pi\left(k_t, k_{t+1}, i_t; p_t, z_t\right) = z_t f\left(k_t\right) - p_t\left(i_t + \phi\left(i_t, k_t\right)\right).$$

The only difference with respect to the Neoclassical model is hence the introduction of adjustment costs via the function $\phi\left(i_t, k_t\right)$. Since ϕ is multiplied by p_t, it is defined in physical units, just like its arguments i and k. We will assume that for all k, $\phi_1'\left(\cdot, k\right), \phi_{11}''\left(\cdot, k\right) > 0$, with $\phi\left(0, k\right) = \phi_1\left(0, k\right) = 0$. Intuitively, ϕ should decrease with k as congestion costs tend to be more proportional to the ratio i/k rather than the absolute value of i.

The problem of the firm hence specializes to

$$\begin{aligned} V_0^* &= \max_{\{i_t, k_{t+1}\}} \sum_{t=0}^{\infty}\left(\frac{1}{1+r}\right)^t\left[z_t f\left(k_t\right) - p_t\left(i_t + \phi\left(i_t, k_t\right)\right)\right] \\ & \quad s.t. \\ k_{t+1} &= \left(1-\delta\right) k_t + i_t; \qquad \left(\frac{\lambda_t}{\left(1+r\right)^t}\right) \\ k_{t+1} &\geq 0, \text{ for all } t;\ k_0 \text{ given.} \end{aligned}$$

We now compute the optimal program using (somehow heuristically) the standard Kuhn-Tucker theory. The first order conditions are

$$i_t : \quad p_t\left(1 + \phi_1'\left(i_t^*, k_t^*\right)\right) = \lambda_t^* \tag{8.4}$$

$$k_{t+1} : \quad \frac{1}{1+r}\left[z_{t+1} f'\left(k_{t+1}^*\right) - p_{t+1}\phi_2'\left(i_{t+1}^*, k_{t+1}^*\right) + \left(1-\delta\right)\lambda_{t+1}^*\right] = \lambda_t^*. \tag{8.5}$$

The (costate) variable λ_t^* represents the *present (i.e., at period* t*) value* of the marginal contribution of capital to profits (the period t shadow price). Condition (8.4) hence just equates costs (to the left hand side) to returns (to the left hand side) of a marginal unit of investment. Now define $q_t = \frac{\lambda_t^*}{p_t}$ the same marginal value normalized by the market price of capital. From (8.4) we obtain

$$1 + \phi_1'\left(i_t^*, k_t^*\right) \equiv g\left(i_t^*, k_t^*\right) = q_t.$$

Since $\phi_{11}'' > 0$, given k_t^* both ϕ_1' and g are increasing functions in of i_t^*. Denoting by h the inverse function of g conditional on k, we obtain

$$i_t^* = h\left(q_t, k_t^*\right),$$

with $h\left(1, k\right) = 0$ (since $\phi_1\left(0, k\right) = 0$). This is a very important relationship. First, since k_t^* is given, it means that the *only* thing that firms need to observe in order to make investment decisions in period t is q_t, the shadow price of investment. In other words, q_t is a "sufficient statistic" for fixed investment.[8] Second, the firm will make positive investment if and only if $q_t > 1$. The intuition is simple: When $q > 1$ (hence $\lambda^* > p$) capital is worth more inside the firm than in the economy at large; it is hence a good idea to increase the capital stock installed in the firm. Symmetrically, when $q < 1$ it is a good idea to reduce capital. Third, how much investment changes with q depends on the slope of h, hence on the slope of g. Since $g' = \phi''$ such slope is determined by the *convexity of the adjustment cost function* ϕ.

We now analyze the analogy of (8.5) to the Jorgenson's optimal investment condition (8.2). If we assume again $p_t = 1$, we obtain

$$\frac{1}{1+r}\left[z_{t+1}f'\left(k_{t+1}^*\right) - \phi_2'\left(i_{t+1}^*, k_{t+1}^*\right) + (1-\delta)\, q_{t+1}\right] = q_t. \tag{8.6}$$

The analogy with the standard condition is quite transparent. The introduction of adjustment costs into the neoclassical model creates a discrepancy between the market cost of investment p and the internal value λ of *installed* capital. The *shadow* (as opposed to market) marginal cost of capital is hence q_t (as opposed to one). The (discounted and deflated by $p_{t+1} = 1$) marginal benefit is again composed by two terms, where the first term now includes the additional component $-\phi_2'\left(i_{t+1}^*, k_{t+1}^*\right)$ (≥ 0) since capital also reduces adjustment costs. The internal value of one unit of capital in the next period, after production, is $(1-\delta)\, q_{t+1}$.[9]

Let's fix again p_t at one and z_t at a constant level z. The steady state level of investment will obviously be $i^{ss} = \delta k^{ss} > 0$, which implies that $q^{ss} = \lambda^{ss} > 1$. Since $\lambda^{ss} = q^{ss}$ is uniquely defined by (8.4), according to

$$\lambda^{ss} = g\left(\delta k^{ss}, k^{ss}\right),$$

[8] If h is linearly homogeneous in k, we have

$$\frac{i}{k} = \hat{h}(q) = h(q, 1),$$

that is, q is a sufficient statistic for the investment rate.

[9] When we allow time variations in prices, this condition becomes:

$$\frac{p_{t+1}}{p_t}\frac{1}{1+r}\left[\frac{z_{t+1}f'\left(k_{t+1}^*\right)}{p_{t+1}} - \phi_2'\left(i_{t+1}^*, k_{t+1}^*\right) + (1-\delta)\, q_{t+1}\right] = q_t,$$

which has the same interpretation as the above condition, with the additional deflationary term $\frac{p_{t+1}}{p_t}$ that keeps the benchmark value for q at one.

we only need to compute the steady state level of capital. From (8.5) (or (8.6)), we get

$$\frac{1}{1+r}\left[zf'\left(k^{ss}\right)-\phi_2'\left(\delta k^{ss},k^{ss}\right)\right]=\frac{r+\delta}{1+r}g\left(\delta k^{ss},k^{ss}\right).$$

Whenever the left hand side decreases with k^{ss} while the right hand side increases with k^{ss} (with at least one of the two conditions holds strictly) and some limiting conditions for $k=0$ and $k\to\infty$ are satisfied, there exists *one and only one* solution to this equation.

8.4.1 Marginal versus Average Tobin's q

Hayashi (1982) showed that under four key conditions the shadow price q_t (the *marginal* q) corresponds to the ratio between the value of the firm V_t^* divided by the replacement cost of capital $p_t k_t$. The latter ratio is often called Tobin's *average* q. Such conditions are: *(i)* the production function and the adjustment cost function are homogeneous of degree one, i.e. they display constant returns to scale; *(ii)* the capital goods are all homogeneous and identical; and *(iii)* the stock market is efficient, i.e. the stock market price of the firm equals the discounted present value of all future dividends; *(iv)* and the firms operates in a competitive environment, i.e. it takes as given prices and wages;

The intuition for such conditions is as follows. The first condition is a necessary condition as otherwise we obviously have a discrepancy between the returns of capita ad different firm's dimensions. The homogeneity of capital goods is also required since the marginal q refers to the last, newly installed (or about to be installed), capital, while the average also considers the value of all previously installed capital. If there is a large discrepancy between the two, because of the price of old equipment (say computers for example) decrease sharply, the average q tends to be well below the marginal q. Finally, the inefficiency of the stock market is clearly important. Recall that the marginal q considers the marginal value of future profits. The average q does not considers the average value of profits directly, it computes the ratio $\frac{V_t^*}{p_t k_t}$. Consider now phenomena that bring the value of the firm away from the fundamentals (such as some type of bubbles for example), then the two values (average and marginal q) can match only by chance. Finally, the competitive assumption is obviously important to maintain the linearities induces by the homogeneity of degree one. If a larger firm could grasp more profits by a stronger market power, this should be included while computing the marginal return to new installed capital.

To see it more formally, assume that $zf(k)=zk$ and that $\phi_2'(i,k)=0$ then from (8.6) we obtain:

$$\frac{1}{1+r}\left[z_{t+1}+(1-\delta)\,q_{t+1}\right]=q_t, \tag{8.7}$$

which implies

$$q_t = \sum_{s=0}^{\infty} \left(\frac{1-\delta}{1+r}\right)^s \frac{z_{t+s}}{1+r}. \tag{8.8}$$

That is, as long as productivity and adjustment costs do not depend on k, the shadow value of the the marginal unit of installed capital does not depend on the size of the firm k, i.e., the size of the firm is irrelevant at the margin.

Hayashi shows that the same idea holds true more generally, whenever the size of the firm is irrelevant at the margin.

Exercise 57 *Let $f(k) = F(k,1)$, assume inelastic labor supply normalized to one. Assume that both $F(k,n)$ and $\phi(i,k)$ are linearly homogeneous in their arguments, and amend the profit function to $\Pi = zf(k_t) - w_t - p_t(i_t + \phi(i_t,k_t))$. Show that under the stated assumptions the average and marginal q are equivalent.[Hint: Notice that $w_t = zF_n(k_t,1)$. Moreover, since F is homogeneous of degree one, we have $zf'(k)k + w = zF_k(k,1) + zF_n(k_t,1) = zf(k)$.]*

8.5 Linear Adjustment Costs and Employment Dynamics under Uncertainty

Following Bagliano and Bertola (2004), we now specify our model to address the issue of employment dynamics. The state variable will now be the stock of workers in a firm n_t, an we will completely abstract from capital. The evolution of the employment in a firm can be stated as follows

$$n_{t+1} = (1-\delta)n_t + h_t,$$

where δ indicates an exogenous separation rate, say due to worker quitting the firm for better jobs. The variable h_t indicates the *gross* employment variation in period t.

The cash flow function Π_t will be

$$\Pi(n_t, n_{t+1}, h_t; w_t, z_t) = z_t f(n_t) - w_t n_t - \phi(h_t),$$

where

$$\phi(h) = \begin{cases} hH & \text{if } h > 0 \\ 0 & \text{if } h = 0 \\ -hF & \text{if } h < 0. \end{cases}$$

The function $\phi(\cdot)$ represents the cost of hiring and firing, or *turnover,* which depends on gross employment variation in period t, but not on voluntary quits. In a stochastic environment the problem of the firm hence becomes:

$$\begin{aligned} W_0^* &= \max_{\{n_{t+1},h_t\}} \mathbf{E}_0 \left[\sum_{t=0}^{\infty} \left(\frac{1}{1+r} \right)^t \left[z_t f(n_t) - w_t n_t - \phi(h_t) \right] \right] \\ & \quad s.t. \\ n_{t+1} &= (1-\delta) n_t + h_t; \\ n_{t+1} &\geq 0, \text{ for all } t; \; n_0 \text{ given.} \end{aligned}$$

The analogy with the theory of investment is transparent. Notice however that w_t is not the 'price' of labor, it is a flow payment, rather than a stock payment such as p_t in the previous model. In fact, since it multiplies the stock of labor n_t, the wage is analogous to the *user or rental cost* of capital $(r+\delta)$ in the previous model.

If we denote by λ_t^* the shadow value of labor, defined as the marginal increase in discounted cash flow of the firm if it hires an additional unit of labor. When a firm increases the employment level by hiring an infinitesimal amount of labor while keeping the hiring and firing decisions unchanged, from the envelope conditions we have (have a look at how we derived (8.7) from (8.8)):

$$\lambda_t^* = \mathbf{E}_t \sum_{s=0}^{\infty} \left(\frac{1-\delta}{1+r} \right)^s \frac{\left[z_{t+s} f'(n_{t+s}) - w_{t+s} \right]}{1+r},$$

which can be written similar to our Euler equation as

$$\lambda_t^* = \frac{z_t f'(n_t) - w_t}{1+r} + \frac{1-\delta}{1+r} \mathbf{E}_t \left[\lambda_{t+1}^* \right]. \tag{8.9}$$

Give the structure of turnover costs the optimality condition for h gives $\lambda_t^* \in \partial\phi(h_t^*)$,[10] which implies

$$-F \leq \lambda_t^* \leq H$$

with $\lambda_t^* = H$ if $h_t > 0$ and $\lambda_t^* = -F$ if $h_t < 0$.[11] The idea is simple: the firms is actively changing the employment stock (on top of the exogenous separation rate δ) only when the marginal return compensates the cost, and when it his doing it, h will change so that to exactly equate turnover costs to returns.

[10]Recall from that the symbol '$\partial\phi(h)$' represents the subgradient of the function ϕ at point h.

[11]Note that $h = 0$ is the only point of non-differentiability, and $\partial\phi(0) = [-F, H]$.

Now we assume that $w_t = \bar{w}$ and that z_t follows a two states Markov chain with transition matrix Π. Denote by z_h, and z_l respectively the state with high and low productivity respectively. We are looking for a steady state distribution such that $\lambda_h^* = H$ while $\lambda_l^* = -F$. If for $i \in \{h, l\}$ we denote by $\mathbf{E}[\lambda'; i] = \pi_{ih}\lambda_h^* + \pi_{il}\lambda_l^*$ From (8.9) we get the stationary levels of n from

$$\lambda_h^* = H = \frac{z_h f'(n_h) - \bar{w}}{1+r} + \frac{1-\delta}{1+r}\mathbf{E}[\lambda'; h] \tag{8.10}$$

$$\lambda_l^* = -F = \frac{z_l f'(n_l) - \bar{w}}{1+r} + \frac{1-\delta}{1+r}\mathbf{E}[\lambda'; l]. \tag{8.11}$$

and the hiring/firing decision h can take four values, which solve

$$n_j = (1-\delta) n_i + h_{ij} \qquad \text{for } i, j \in \{h, l\}.$$

Exercise 58 *Consider the case with $\delta > 0$. What is the value for h_{ij} when $i = j$? Find conditions on the transition matrix Π so that a two-states steady state distribution exists for the model we just presented.*

Exercise 59 *Assume that the parameters of the model are such that a stationary steady state exists, set $\delta = 0$, and derive the values for $f'(n_h)$ and $f'(n_l)$ as functions of the parameters of the model: $H, F, \bar{w}, z, r$ and the entries of the matrix Π, which is assumed to take the form: $\begin{bmatrix} \pi & 1-\pi \\ 1-\pi & \pi \end{bmatrix}$. Now perform the same calculations assuming that $F = H = 0$. Comment in economic terms your results in the two cases.*

8.5.1 The Option Value Effect

Bibliography

[1] Abel, A. B., and O. J. Blanchard (1983), "An Intertemporal Model of Saving and Investment," *Econometrica*, 51(3): 675-692.

[2] Abel, A. and J. Eberly, "A Unified Model of Investment Under Uncertainty,“ American Economic Review, 94 (1994), 1369-84.

[3] Bagliano, F.C., and G. Bertola (2004) *Models for Dynamic Macroeconomics,* Oxford University Press.

[4] Bertola, G. and R. Caballero, "Kinked Adjustment Costs and Aggregate Dynamics,“ NBER Macroeconomics Annual, (1990), 237-87.

[5] Caballero, R. and E. Engel, "Microeconomic Adjustment Hazards and Aggregate Dynamics,“ *Quarterly Journal of Economics,* 108 (1993), 359-84.

[6] Gilchrist, S. and C. Himmelberg, "Evidence on the role of Cash Flow for Investment," *Journal of Monetary Economics,* 36 (1995), 541-72.

[7] Hayashi, F. (1982), "Tobin's Marginal Q and Average Q: A Neoclassical Interpretation," *Econometrica*, 50: 215-224.

[8] Sala-i-Martin, X. (2005), "Internal and External Adjustment Costs in the Theory of Fixed Investment," *mimeo*, UPF and Columbia University.

Chapter 9

Introduction to Search Theory

9.1 Search and the labor market

- A good macroeconomic model of the labor market should explain why in the data we have people declaring that they are actively searching for jobs but still unemployed.
- The Competitive Paradigm with flexible prices is unable to deliver such prediction: if productivity is low and firms are reluctant to hire, the wage should simply decrease so that to restore full employment
- Is the model we saw last week able to generate unemployment?
- Well, that model regards the demand-side of the labor market
- This week we will analyze a simple supply-side explanation for the presence of unemployment

What is Search Theory about?

- The idea behind the theory of search is simple: workers are looking for jobs but they cannot observe all available offers at once (there are search frictions).
- At a given period they might have some job offer but they might reject it
- The reason why they might reject such offer is because they are planning to search further and believe to be able to find a better job
- In this way the market generates frictional unemployment

- This model also provides a useful tool to analyze the effect of unemployment insurance benefits on unemployment rate
- This is a perfect complement to the previous model of labor demand where we were able to introduce employment protection

9.2 The Representative Worker

- Consider an unemployed worker who samples wage offers on the following terms:
- Each period, with probability $p \in (0,1)$, she receives no offer;
- With probability $1-p$ she receives an offer to work for the wage w forever, where w is drawn from a cumulative distribution $F(w)$ with compact support $[0,\bar{w}]$.
- The worker chooses a strategy to maximize

$$\mathbf{E}\sum_{t=0}^{\infty}\beta^t u(c_t), \text{ where } \beta \in (0,1),$$

$c_t = w$ if the worker is employed, and $c_t = b$ if the worker is unemployed. u is increasing and concave, with $u(0) = 0$, and b is the unemployment compensation $(0 \leq b < \bar{w}$).

The Recursive Formulation

- For simplicity we assume that, having accepted a job offer at wage w, the worker stays in the job forever.
- Moreover, we assume that successive drawings across periods are independently and identically distributed.
- It is possible, but awkward, to set up the problem in sequence form $\Rightarrow$ Exercise.
- The recursive formulation for this problem is quite simple:
- Let $V(w)$ be the expected value of an unemployed worker who has offer w in hand and who behaves optimally.

$$\begin{aligned} V(w) &= \max\left\{\frac{u(w)}{1-\beta}, U\right\} \text{ with} \\ U &= u(b) + \beta\left[pU + (1-p)\int_0^{\bar{w}} V(w')\, f(w')dw',\right] (= V(0)) \end{aligned}$$

where $f > 0$ is the density of F.

Exercise 60 *Show that the Bellman operator associated to the search problem is a contraction in the space of bounded and continuopus functions with the sup norm.*

Reservation Wage Strategy

- It is now easy to guess that the optimal policy should be defined by a the wage level w^*, such that the worker accepts all wage offers $w \geq w^*$, and she reject all offers below w^*

- Given U, by definition, the expression for reservation wage is

$$\frac{u(w^*)}{1-\beta} = U \Rightarrow w^* = u^{-1}\left((1-\beta)\, U\right).$$

- And V has the following form

$$V(w) = \left\{ \begin{array}{cc} U & \text{if } w < w^* \\ \frac{u(w)}{1-\beta} & \text{if } w \geq w^* \end{array} \right\}$$

- This is the solution V to the Bellman Equation in terms of U and w^*. The value of U has in turn be given in terms of V above, while the value of w^* is from the reservation wage equation

- Using the Bellman Equation and the form of the value function V, we obtain

$$U = \frac{u(b) + \beta\,(1-p) \int_{w^*}^{\bar{w}} \frac{u(w)}{1-\beta} f(w)dw}{1 - \beta p - \beta(1-p)F(w^*)}.$$

- The above equation together with the reservation wage allows to eliminate U and obtain

$$(1 - \beta p - \beta(1-p)F(w^*))\, u(w^*) = (1-\beta)\, u(b) + \beta\,(1-p) \int_{w^*}^{\bar{w}} u(w) f(w)dw.$$

- One can show that there is a unique value for w^*.

Exercise 61 *Show that w^* exists and it is unique.[Hint: Recall that F is continuous as it admits a density f, that $f > 0$; and that u is a continuous and increasing function]*

- For future use, note that

$$F(w^*)u(w^*) = u(w^*) - \int_{w^*}^{\bar{w}} u(w^*) f(w)dw$$

9.3 Comparative Statics

Finally, let's investigate the effect of some key parameters. Manipulating the above expression, one obtains

$$\left(\frac{1}{\beta}-1\right)(u(w^*)-u(b))=(1-p)\int_{w^*}^{\overline{w}}\left[u(w)-u(w^*)\right]f(w)dw.$$

Now is easy to see that:

1. if b increases w^* must increase (the usual economic interpretation of the reservation wage and UI benefits);
2. if p increases w^* must decrease since the agent is more likely to not receive offers;
3. if β increases w^* increase since it is more patient (recall the intertemporal trade-off in the search model: worker wait for the good offers).
4. **Q:** What is the effect of a (linear) wage tax: $w \Rightarrow w(1-\tau)$ for all w?

A: Typically $\uparrow w^*$ (it reduces the value of employment for all w)

The Option Value Effect

- What is the effect of a Mean preserving spread in the wage distribution?

$$\left(\frac{1}{\beta}-1\right)(u(w^*)-u(b))=(1-p)\int_{w^*}^{\overline{w}}\left[u(w)-u(w^*)\right]f(w)dw.$$

- If the variance in F increases, given w^*, $\int_{w^*}^{\overline{w}}\left[u(w)-u(w^*)\right]f(w)dw$ tends to increases since mean preserving spreads put more weight on the tail of the wage distribution and $u(w)-u(w^*)\geq 0$ and increases with $w\geq w^*$.
- This means that the reservation wage tends to increase.
- Note that average unemployment duration is the inverse of the exit rate from unemployment: $d=\frac{1}{(1-p)(1-F(w^*))}$ hence an increase in the reservation wage increase duration and unemployment rate

Welfare Effects of an increase in wage dispersion

- Notice moreover that when w^* increases, the ex-post value increases:

$$V(w)=\max\left\{\frac{u(w)}{1-\beta},\ \frac{u(w^*)}{1-\beta}\right\}.$$

- However, since u is concave it is not clear what is the effect of wage dispersion on the ex-ante welfare of the worker:
$$W = p\left[u(b) + \beta W\right] + (1-p)\int \left[\max\left\{\frac{u(w)}{1-\beta},\ u(b) + \beta W\right\}\right] f(w)dw.$$
- Homework: when u is linear we have an unambiguous effect, both in the reservation wage and the welfare of the worker.

9.4 Search and Asset Accumulation

- Consider now exactly the same search problem as above, with the same characteristics on wage offers and wage distributions, and constant unemployment compensation.
- Assume now that the agent can save but cannot borrow.
- An unemployed worker maximizes
$$\mathbf{E}\sum_{t=0}^{\infty}\beta^t u\left(c_t\right)$$
subject to the budget constraint
$$c_t + k_{t+1} \leq Rk_t + b,\ k_t \geq 0.$$
- If the worker has a (permanent) job that pays w, her budget constraint is
$$c_t + k_{t+1} \leq Rk_t + w,\ k_t \geq 0.$$
- The Bellman equation for this problem is
$$\begin{aligned} V_e(w,k) &= \max_{k'}\ u\left(Rk + w - k'\right) + \beta V_e(w,k') \\ V_u(w,k) &= \max\left\{V_e(w,k),\ U(k)\right\} \\ U(k) &= \max_{k'}\ u\left(Rk + b - k'\right) + \beta\left[pU(k') + (1-p)\int V_u(w',k')f(w')dw'\right]. \end{aligned}$$
- Clearly both $U(k)$ and $V_e(w,k)$ are increasing in k. And, in general, they do not increase at the same speed. Hence the reservation wage function defined by $w^*(k)$
$$V_e(w^*(k),k) = U(k)$$
depends on k. The problem is not simple since not even V_e is easy to derive. $\Rightarrow$ Exercise.

- Danforth (1979) showed that with permanent jobs $w^*(k)$ can be characterized by the risk aversion parameter of the worker.

Chapter 10

General Equilibrium, Growth, and Real Business Cycles

THIS CHAPTER IS AT A VERY PRELIMINARY STAGE

10.1 Business Cycles

In their seminal treatise on "Measuring Business Cycles," Burns and Mitchell (1946) define business cycle as follows:

> *A cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own (Burns and Mitchell, 1946, p. 3).*

Lucas provided a much more concise definition:

> *Business cycles are recurrent fluctuations of output and employment about a trend (Lucas 1977).*

In few words, the real business cycle (RBC) project aim at using the neoclassical growth framework to study business cycle fluctuations. Technological shocks or Government spending shocks are both *real*, hence the name RBC. For example, New Keynesian theories have non-Walrasian features and monetary shocks. We also have real non-Walrasian theories. Technological improvement is widely accepted as a key force generating long run growth. By identifying technological shocks as one main cause for business cycle fluctuations the RBC project tries hard to deliver a unified theory able to explain both the short and the long run (this idea refers back to Knut Wicksell, 1907).

The RBC methodology adopts a *deductive or quantitative theoretic inference*, where the model is not the product but the tool or the measurement instrument to deduce implications of the theory (Prescott, 1998). This contrasts with *inductive* or empirical inference, where the model or the low is the product of the research.[1]

[1]**A quick refresh on methodology**. Friedman is an instrumentalist and thinks that the usefulness of a theory is to make predictions. If one has a phenomenon to be explained and a model, he is allowed to make assumptions somehow arbitrarily, to make the model simple and get predictions (implications) which are accurate enough to be 'useful.' Hypothesis should never be tested. One should test only the predictions. Examples: 1)- Leaves are distributed on a tree so that to maximize sunlight exposure. The importance is not whether the leaves solve a maximization problem. It is the prediction that matters. 2)- Expert billiard players shoot as if they calculate precisely the ball's trajectory. We do not want to describe whether they actually do such calculations, we are interested in the predictions of such a theory. This approach has been widely criticized by several methodologists mainly on the following grounds. First, there is no clear distinction between assumptions and predictions. You can always see an assumption as a prediction based on other assumptions. The second problem with this methodology is that you might believe that a true theory actually exists. That is you are a realist (as for example Popper is). Or at least you are a conventionalist and you at least want to speak as if a true theory did exist. The third problem is that sometimes the phenomena to be explained are too narrow so that the theory is not very useful to explain other phenomena. Or, simply, the model is not very useful to explain the same phenomena in different contests or situations. The leading example is the Lucas critique and the policy implications that a theory might want to have. We might want to understand the mechanism behind the phenomenon. For this reason you might want to use data unrelated to the phenomena you ought to explain in order to make a stronger test of the theory. These data can be used to test the assumptions. The model than has the potentials to explain phenomena previously unobserved.

If anything, Prescott is most closely to an instrumentalist. He does not want to see whether the dynamic competitive market model is true. Probably he does no even believe in the possibility of constructing a true model. He just want to use this model to make quantitative assessments. Friedman would do the same. In this sense,Prescott only suggested a disciplined way of making the specific assumptions and the calibration. On the other hand, his extreme confidence that the split of the time series in trend plus cycle is the correct one and in the neoclassical growth model might seem to bring him more toward a conventionalist or even realist attitude. His attitude toward models and statistical tests refers back to Kuhn (1962). In Kydland and Prescott (1996, page 73) they say that: 'Using probabilistic verification theories that ask us to compare a given scientific theory with all others that might fit the same data is a futile effort.' In addition they are not interested in the mechanism behind one effect. For example, they

More in detail, the methodology proposed by the RBC project works as follows: The researcher should first pose a quantitative question (such as for example: measure the welfare consequences of changes in the tax system, or quantify the magnitude and nature of business cycle fluctuations induced by different types of shocks). Then she should use a well tested theory (such as for example the neoclassical growth theory) to construct the model economy. 'Calibrate' the model economy and, finally, use the calibrated model economy to answer quantitatively to the question.

Finn Kydland and Edward Prescott (two very recent Nobel laureates) referred to the whole procedure as '*the computational experiment.*' They regarded it as something different (or broader) then the 'usual' econometric exercise. In the sense that deriving quantitative implications of an economic theory sometimes differs from measuring economic parameters. On the other hand, econometricians criticized the computational experiment methodology as Bayesian or likelihood based inference where the loss function is not well specified and standard errors are not displayed (see Hansen and Heckman, 1996; and Sims, 1996). Below we will discuss a bit more in detail this debate. Let us start with a brief historical overview.

10.1.1 A bit of History

The statistical definition of business cycles of Burns and Mitchell (1946) did not prove useful. In contrast, Kaldor '50 found some regularities in long-term economic data (similar regularities were also found by Mitchell and Kuznets), which are quite useful. First, each country tend to experience a roughly constant growth rate of GDP (Y). Second, employment (N) does not change much in the long run. Third, the capital/GDP ratio (K/Y) is also roughly constant. However, (fourth) the growth rate of GDP per capita (Y/N) varies quite a lot across countries. And, (fifth) across countries, high capital share implies high investment. Finally, (sixth) labour and investment shares appear to fluctuate around a constant mean. These facts guided researchers in the development of Growth Theory (Solow '56 and Cass-Koopmans '65)

Hodrick and Prescott (1980) find some empirical regularities consistent with growth theory (hence Kaldor's facts) and Lucas definition of business cycles. In particular they confirm Lucas idea that "business cycles are all alike," suggesting that one can hope to construct a unified theory of the business cycle. Their technique is based on the

do not care about 'whether and why' an increase in income taxes leads to lower hours worked, they only care about 'by how much they change,' they aim at appropriately quantifying the effect, not at explaining the mechanism behind it.

assumption that time series can be described by a trend plus a fluctuation around the trend (they say that this is an 'operational definition'). Their definition of trend is flexible. The now notorious Hodrick-Prescott (HP) filter is nothing more than a way of identifying low frequencies movements. A linear trend is obviously a specific case of a low frequency movement. In essence, this method involves defining the cyclical component y_t^c of output as the difference between current output y_t and a measure of trend output y_t^g. With y_t^g being a weighted average of past, current and future observations

$$y_t^c = y_t - y_t^g = y_t - \sum_{n=-N}^{N} a_n y_{t-n}. \tag{10.1}$$

To be more specific, the HP filter is derived by solving the following minimization problem

$$\min_{\{y_t^g\}_{t=0}^{\infty}} \sum_{t=1}^{\infty} \left\{ (y_t^c)^2 + \lambda \left[\left(y_{t+1}^g - y_t^g \right) - \left(y_t^g - y_{t-1}^g \right) \right]^2 \right\}.$$

For quarterly data, the standard value chosen for the smoothing parameter λ is 1600. Notice that when $\lambda = \infty$ the solution of the problem is a linear trend, while with $\lambda = 0$ the trend coincides with the original series. In a finite sample context, the weights a_n in Equation (10.1) depend on the length of the sample, so that the text expression is a simplification.

Of curse, there are other filtration techniques. The Band-pass technique for example leads to roughly the same cyclical component as the HP filter technique. The HP technique is designed to eliminate all stochastic components with periodicities greater than 32 quarters (notice that 8 years is the longest reference cycle that Burns and Mitchell uncovered using very different methods). The main critique to this approach is that it seems that output time series display slow-moving *stochastic* components, which are omitted by this definition of business cycle.

Here is a summary of their early findings. First, C (non-durable consumption) is strongly procyclical and fluctuates less than Y, about one third (however, durables are more volatile than output). Second, Investment (I) is strongly procyclical and fluctuates more than Y, about three times as much. Third, 2/3 of $var(Y)$ is accounted by $var(N)$ and 1/3 by $var(TFP)$. While capital input (K) is unimportant in describing output fluctuations, that is (fourth) K is acyclical. Fifth, they found a moderately high degree of persistence in Y fluctuations, that is, in the component of Y which are deviations from the trend. Sixth, labour productivity (Y/N) is moderately procyclical but orthogonal to N. Seventh, G is basically acyclical and less volatile than output.

Brock and Mirman (1972) provide the stochastic growth model. In 1982 Kydland and Prescott first performed the computation experiment based on a calibrated neoclassical stochastic growth model. In their seminal paper, Kydland and Prescott ask themselves the following quantitative question: "Can specific parametric descriptions of technology and preferences be found such that the moment induced in output, consumption, employment, and other series in such a model with only technological shocks resemble to the time series behavior of the observed counterpart to these series in the postwar, US economy?"

They calibrate the 'deep' parameters to "match" either cross sectional or lung-run moments, and find that the basic neoclassical model essentially generates too flat employment, (together with too volatile consumption and too smooth investment).

They hence extend labour preferences allowing people to value leisure more if they have worked in the past. This feature increased the inter-temporal elasticity of substitution of leisure and generated employment fluctuations similar to the one of the aggregate data. They also clarify that this feature should not be part of the measurement instrument (or model) when the question is: how volatile would have been output if TFP shocks were the only shocks. Unless there is (micro) evidence that preferences are indeed of this sort. They now believe that such evidence never materialized. And that Hansen (1985) Indivisible Labour model is most promising.

Hansen (1985) provides evidence that 2/3 of N fluctuations are not in hours but in the number of people that work in a given week. He hence postulates a complete markets model where agent can only work a given amount of hours. Agents trade lotteries over employment and the utility function of the representative consumer becomes linear in N. This obviously increases the reaction of N to productivity shocks since inter-temporal elasticity of substitution of leisure of the representative consumer is much larger than the one of the individuals.

10.2 The Neoclassical Growth Model Again

RBC theory assumes that economic outcomes arise as the equilibrium outcomes of fully rational agents. The growth model we saw in previous chapters identifies Pareto efficient allocations. However, recall that the 1st Welfare Theorem asserts that with complete markets, a competitive equilibrium allocation is Pareto optimal. The 2nd Welfare Theorem asserts that in a convex economy any Pareto optimal allocation can be achieved as a competitive equilibrium allocation (with initial transfers). So as long as the economy is convex and markets are complete we can use the stochastic growth model to describe a competitive equilibrium allocation. Let's start with the optimal (for simplicity deterministic) growth model

$$V(k_0) = \max_{\{k_{t+1},\, c_t,\, n_t\}} \sum_{t=0}^{\infty} \beta^t u(c_t).$$

Recall that at the optimum we have

$$c_t + k_{t+1} = F(k_t, n_t) + (1-\delta)k_t$$

and $n_t = 1$; and the Euler equation

$$u'(c_t) = \beta\left(F_1(k_t, 1) + (1-\delta)\right)u'(c_{t+1}).$$

10.3 Competitive Equilibrium and the Welfare Theorems

Which is the economy underlying the optimal growth model? One possibility is the following. The economy is formed by a representative firm which maximize profits choosing factor inputs

$$\pi = \max_{\{k_t^d, n_t^d\}} \sum_t p_t\left[F(k_t^d, n_t^d) - r_t k_t^d - w_t n_t^d\right]$$

From the first order conditions we have factor demand schedules:

$$\begin{aligned} w_t &= F_2(k_t^d, n_t^d) \text{ (Real wage)} \\ r_t &= F_1(k_t^d, n_t^d) \text{ (Rental price of Capital)} \end{aligned}$$

Notice that since F is CRS $\Rightarrow \pi = 0$ and that the representative firm can be seen as the aggregation of many identical firms with convex productions functions.

The economy is populated by households, who own capital and labour and maximize lifetime utility supplying labour and capital subject to the usual budget constraint:

$$\max_{\{k_{t+1}^s,\, c_t,\, n_t^s\}} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

s.t.

$$\sum_t p_t\left[c_t + k_{t+1}^s\right] \le \sum_t p_t\left[w_t + \left(r_t + (1-\delta)\right)k_t^s\right].$$

Optimality implies

$$\begin{aligned} \frac{p_t}{p_{t+1}} &= \frac{u_1(c_t)}{\beta u_1(c_{t+1})} = (1-\delta) + r_t \\ &= (1 + \text{Interest Rate}) \\ \text{and } (n_t^s &= 1). \end{aligned} \tag{10.2}$$

Competitive Equilibrium (C.E.)

Definition 45 *A C.E. is a set of prices* $\mathbf{q} = \{p_t, w_t, r_t\}_{t=0}^{\infty}$ *and an allocation* $\mathbf{f} = \{k_t^d, n_t^d\}$ *for the typical firm and* $\mathbf{h} = \{k_{t+1}^s,\ c_t,\ n_t^s\}$ *for the households , such that given* $\mathbf{q}$*: (i)* $\mathbf{f}$ *solves the firm profit maximization problem; (ii)* $\mathbf{h}$ *solves the household maximization problem; and (iii) all markets clear, i.e.* $k_t^d = k_t^s$; $n_t^d = n_t^s$; *and* $c_t + k_{t+1}^s - (1-\delta)k_t^s = F(k_t^d, n_t^d) \equiv y_t$.

Definition 46 *A feasible allocation* $\mathbf{x} = \{k_{t+1},\ c_t,\ n_t\}_{t=0}^{\infty}$ *is Pareto optimal if there is no other feasible allocation that yields higher utility to the agent.*

Theorem 26 *[**1st Welfare**] A C.E. Allocation is Pareto optimal.*

Proof: Assume $\mathbf{x}$ is a CE (with prices $\mathbf{q}$) and assume there exists a feasible $\mathbf{x}'$ such that gives higher utility to the agent. Since, given $\mathbf{q}$, $\mathbf{x}$ was maximizing household's utility, it must be that $\mathbf{x}'$ violates household's budget constraint, i.e.

$$\sum_t p_t \left[c_t' + k_{t+1}'\right] > \sum_t p_t \left[w_t n_t' + (r_t + (1-\delta))\, k_t'\right].$$

Moreover, since $\mathbf{x}'$ is feasible it must satisfy $c_t' + k_{t+1}' - (1-\delta)k_t' = F(k_t', n_t')$. But then, rearranging terms we have that

$$\pi' = \sum_t p_t \left[F(k_t', n_t') - w_t n_t' - r_t k_t'\right] > 0.$$

This contradicts that $\mathbf{x}$ is a C.E. since $\mathbf{x}$ does not maximizes profits given $\mathbf{q}$ (recall that $\pi = 0$). **Q.E.D.**

Theorem 27 *[**2nd Welfare**] A Pareto optimal allocation* $\mathbf{x}$ *can always be obtained as a C.E. for some prices.*

Proof: Since the planner problem is concave it has a unique solution and necessary and sufficient conditions for an efficient allocation are, feasibility and FOC of the planning problem. We need to show that $\mathbf{x}$ can be seen as a CE allocation, that is that there are prices $\mathbf{q}$ such that given $\mathbf{q}$, $\mathbf{x}$ maximizes firm and household objectives and markets clear. To this end we can use planner's first order conditions to define prices as follows:

$$r_t = F_1(k_t, 1);$$

$$\begin{aligned} w_t &= F(k_t, 1) - k_t F_1(k_t, 1); \\ \frac{p_{t-1}}{p_t} &= F_1(k_t, 1) + (1 - \delta) \text{ with } p_0 > 0 \text{ arbitrary.} \end{aligned}$$

It is easy to see that this these prices together with $\mathbf{x}$ define a CE. **Q.E.D.**

Exercise 62 *Use Definition 45 and show that* $\mathbf{q}$ *derived in the proof above together with* $\mathbf{x}$ *define a CE.*

Notice that we implicitly assumed that capital is owned by households. However, this is non influent, and it is good exercise to show that the allocation does not change.

Exercise 63 *Formulate the simple neoclassical general equilibrium model assuming that firms own capital instead, and households only own shares of the representative firm. Derive the competitive equilibrium allocation and prices, and argue that property of capital does not affect the CE allocation.*

10.4 Calibration and Estimation: The RBC Methodology

Kydland and Prescott (1996,1998) argue that the 'computational exercise' is in some broad (and old) sense estimation: determination of the approximated quantity of something. *Calibration is graduation* of measuring instruments (thermometers). Since models are measuring instruments, they must be calibrated. The calibration is done according to a model in which we have confidence. The parameters of the model are varied so that some known answers happen to be true. However, since the task is not an attempt to assessing the size of a parameter, it is not estimation in the narrow sense. They suggest that calibration should be done in order to reproduce the key long-term or growth relations among model aggregates and panel averages.

Principle 1: Any modification of the standard model of growth theory to address a business cycle question, should continue to display the growth facts.

Principle 2: The model economy being used to measure something should not have a feature which is not supported by other evidence even if its introduction results in the economy model better mimicking the reality. *Example:* Hansen (1985) introduced the indivisible labour feature by displaying empirical evidence that justifies of incorporating this feature.

Principle 3: A model that better fits the data may be a worse measurement instrument. Indeed a model matching the data on certain dimensions can be the basis for rejecting the model economy as being a useful instrument for estimating the quantity of interest. *Example:* We will see that one key dimension where the RBC models fail dramatically to replicate real world data is the correlation between hours worked and average labour productivity: this measure is close to one in the standard RBC model and approximately zero in U.S. postwar data. Kydland and Prescott (1991) modify Hansen's model by introducing variable workweek length and moving costs. They calibrate moving costs so that employment and workweek length match observations. Then they use the model and estimate that the US postwar economy would have been 70% as volatile if TFP shocks were the only disturbance. The model also predicts that if one generates the further 30% of variability by introducing shocks which are orthogonal to TFP then N and $\frac{Y}{N}$ must be orthogonal.[2] This evidence is consistent with the data. If the measure where near 100%, there would be no space for additional shocks and the model itself implies strong correlation between N and labour productivity. Actually, without moving costs the model generates 90% of actual output variability (generating a strong positive correlation between N and Y/N). However, this would have induced some perplexity on the measure since data does not provide this feature, and there is no way of generating it. Finally, suppose that at 70% both the model (with the additional shocks orthogonal to TFP) and the data would have suggested a high correlation between labour productivity and labour input. This would have been a basis for rejecting the model as a good instrument for measuring the importance of TFP shocks.

Principle 4: This is a corollary to Principle 3. The using statistical estimation theory to estimate models used to deduct scientific inference is bad practice. Estimating the magnitude of a measurement instrument, whether it is thermometer or a model economy

[2]The reason for it is that the factors other than technological shocks (e.g. public consumption) that give rise to variations in labour input result in productivity being low when hours are high (see, Christiano and Eichenbaum, 1992).

makes no sense.

Principle 5: A legitimate challenge to a finding is to introduce a feature into the model economy that is serving as the measurement instrument in a quantitatively reasonable way and show that the answer to the question changes. *Example:* Questioning the 70% result above on the basis that the model is not realistic is not legitimate since all models are abstractions and therefore unrealistic. But if you introduce a new feature in a convincing way, by motivating it with additional empirical evidence, your number is a legitimate challenge to the 70%. Indeed there are many such measurements in the RBC literature. They all are between 70% and 75%. Actually, we will see below a simple model that accounts for 77% of output variability. However, recall that the looking for a model that explains fluctuation is only one of the many quantitative questions one might ask.

10.5 The Basic RBC Model and its Performance

The basic RBC model departs from the simplest stochastic growth model only by the introduction of leisure. The planner, or representative consumer problem solves

$$V(x_0, k_0, z_0) = \max_{\{c_t, k_{t+1}\}} \mathbf{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{\left[c_t^{1-\phi} (L - n_t)^{\phi} \right]^{1-\sigma}}{1-\sigma} \right]$$

$$k_{t+1} + c_t = e^{z_t} k_t^{\alpha} \left(x_t n_t \right)^{1-\alpha} + (1-\delta) k_t$$

$$z_{t+1} = \rho z_t + \varepsilon_t;\ 0 < \rho < 1;\ \varepsilon_t \sim N(0, \sigma_{\varepsilon}^2)$$

$$\ln x_t = \ln \gamma + \ln x_{t-1}.$$

We now describe how to choose the different parameters. First of all notice that the functional specification of consumer's preferences allow steady state growth and roughly constant n, consistent with Kaldor facts. Some authors - especially in the early contributions - use micro evidence on asset pricing to argue in favor of a "log-log" formulation for preferences ($\sigma = 1$) of the form $\ln c_t + \theta \ln (L - n_t)$. We here do the same.[3] The calibration of the parameter θ is done at the end, as it depends on all other parameters.

[3] Soon below we will consider a model a bit more general than this.

First, we compute $\frac{c}{y}$. With a Cobb-Douglas production function, the steady state level of the capital per worker solves[4]

$$\frac{k}{n} = \left[\frac{\alpha}{r+\delta}\right]^{\frac{1}{1-\alpha}},$$

where $r = \beta^{-1} - 1$; which generates $\frac{k}{y} = \frac{k}{n}/\frac{y}{n}$. Notice that $\frac{y}{n}$ is known since $\frac{y}{n} = \left(\frac{k}{n}\right)^{\alpha}$. Hence from $\frac{i}{n} = (\gamma - 1 + \delta)\frac{k}{y}$ we get $\frac{c}{y} = 1 - \frac{i}{n}$ (notice that we also get the steady state wage rate $w = (1-\alpha)\left(\frac{k}{n}\right)^{\alpha} = (1-\alpha)\frac{y}{n}$).

Second, from the agent's first order conditions we have

$$w = \frac{u_2(c,n)}{u_1(c,n)} = \frac{\theta c}{L-n}.$$

If we aim at matching the average postwar level for n as a 20% of the available time we get[5]

$$1 - \alpha = \frac{wn}{y} = \frac{\theta n}{L-n}\frac{c}{y} = 0.2\theta\frac{c}{y}.$$

Hence, once α is calibrated as described below we get θ. Cobb-Douglas is very common practice since it has constant shares even though relative factor prices are different and seems to be the case that across countries labour share is roughly 70%. Hence $1-\alpha$ is typically chosen between .3 and .35,[6] here we set $\alpha = \frac{1}{3}$. This delivers $\theta = 3.48$.

In a deterministic economy, a good measure for the discount factor when one period is one quarter is $\beta = .996$. Some authors believe that for quarterly data a good value form β should be .99 (the annual deterministic level) in order to account for stochasticity. However, this number must be calibrated. One possibility is to choose β so that the

---

[4]From the Euler equation we get

$$c_t^{-\sigma} = \beta \mathbf{E}_t\left[c_{t+1}^{-\sigma}\left(1 - \delta + \alpha e^{z_{t+1}}\left(\frac{k_t}{n_{t+1}}\right)^{\alpha-1}\right)\right]$$

in the deterministic model with $z_{t+1} = 0$ we get

$$\left(\frac{c_{t+1}}{c_t}\right)^{\sigma} = \left(1 - \delta + \alpha\left(\frac{k_t}{n_{t+1}}\right)^{\alpha-1}\right),$$

which delivers the desired conditions for the steady state, i.e. when $c_t = c_{t+1}$.

[5]Perhaps a better number for this is now $\frac{n}{L-n} = 0.3$.

[6]King, Plosser and Rebelo (1988a,b) chose $1 - \alpha = .58$. This number is somewhat sensitive to th treatment of the government sector and of proprietor's income. See Cooley and Prescott (1995) for a discussion.

steady state real interest rate coincides with the average return to capital in the economy. The US data from 1948-1986 suggest $\beta = .984$.

The estimation of ρ on quarterly detrended data is between .95 and .98, here $\rho = .979$ and $\sigma_\varepsilon = .0072$. The mean of z_1 is normalized to zero since it only affects the scale of the economy. The same for $x_0 = 1$. The average rate of technical progress between 1948 and 1986 is 1.6%, suggesting $\gamma = 1.004$.[7] The conventional value for δ is 10% per annum, hence $\delta = 0.025$. We follow the convention, however Stokey and Rebelo (1995, Appendix B) estimate the US post war ratio of capital consumption allowances to the capital stock (excluding consumer durables and government capital) of the order of 6% per annum. It is believed that the dynamics near the steady state are not affected by γ and δ, although the average investment share is obviously affected.

The most widely used (especially in the early papers) numerical methodology is *log-linearization.* The first order conditions of the model are evaluated and log linearized around the deterministic steady state. Before presenting some figures let us revise the typical economic mechanism at work in these dynamic model.

10.5.1 Mechanism

In order to understand the fundamental mechanism at work in RBC model we now consider the following (deterministic) specification with labour:[8]

$$\max_{\substack{c_1,c_2\\ n_1,n_2}} \sum \beta^t \left[\frac{c_t^{1-\sigma}}{1-\sigma} - \theta\frac{n_t^{1-\upsilon}}{1-\upsilon}\right] + \beta\left[\frac{c_2^{1-\sigma}}{1-\sigma} - \theta\frac{n_2^{1-\upsilon}}{1-\upsilon}\right]$$

subject to

$$c_t + a_t = w_t n_t + (1+r)\, a_{t-1} \qquad ((\beta^t \lambda_t))$$

where a_t is the lever of asset in period t. Now solve it as a static problem and take the usual first order conditions:

$$c_t^{-\sigma} = \lambda_t \qquad (10.3a)$$

$$\phi n_t^{-\upsilon} = \lambda_t w_t \qquad (10.3b)$$

$$\lambda_t = \beta\,(1+r)\,\lambda_{t+1} \qquad (10.3c)$$

Interpretation: the multiplier λ_t is the value (price) of wealth.

---

[7] Most studies ignore population growth. Incorporating population growth would suggest $\gamma = 1.008$.

[8] These preferences can be seen as a generalized version of the 'log-log' specification we discussed above.

Rearranging (10.3a) and (10.3b) we get the usual intra-temporal condition, which defines the efficient (relative) allocation of consumption and labour give the real wage equating marginal returns to marginal costs:

$$c_t^{-\sigma} w_t = \theta n_t^{-\upsilon} \tag{10.4}$$

If we take logs of this we get

$$\ln c_t = \frac{\ln w_t + \upsilon \ln n_t - \ln \theta}{\sigma} \tag{10.5}$$

Rearranging (10.3a) and (10.3c) one gets the usual Euler equation. If we again take logs, we get:

$$\ln c_{t+1} - \ln c_t = \frac{1}{\sigma} \ln \beta (1+r). \tag{10.6}$$

Positive consumption growth if and only if $\beta(1+r) > 1$: The market return r is larger than the private cost of patience $\beta^{-1} - 1$. As usual, consumption growth is reduced if the intertemporal elasticity of substitution $\frac{1}{\sigma}$ is low. In other words, when the utility curvature σ is large there is a greater desire for consumption smoothing and consumption is relatively flat across time.[9]

Finally, if we substitutes (10.5) into (10.6) we get the following intertemporal relationship between labour

$$\ln n_{t+1} - \ln n_t = \frac{1}{\upsilon} \left[\ln \frac{w_{t+1}}{w_t} - \ln \beta (1+r) \right]. \tag{10.7}$$

Notice that the parameter θ does not appear in the above equation. This is because it is an intertemporal relationship. And, of course, it does not depend on σ either, since it is a relationship between labour only.

Fist, whenever $w_t > w_{t+1}$ there is a tendency to increase this period employment n_t compared to n_{t+1}. On the other hand, even though w_t increases by a lot, whenever both w_{t+1} and w_t increase by the same (proportional) amount there is no reason for n_t to adjust. This is the case of a fully permanent shock on wages for example. And the reason is a wealth effect. In fact, since consumption will increase, condition (**??**) suggests that we can even have a decrease in n_t as a consequence of a permanent increase in wages. This is so whenever the wealth effect dominates the today's wage effect.

Finally notice, that r enters into (10.7). This is one of the key a new insights emphasized by the RBC literature: *high interest rates discourage employment growth.* The only

[9] If you want to plug numbers into it. Empirically, σ seems to be between 1 and 2, if σ is estimated using the Euler equation.

way of bringing leisure into the next period is to save today. When r is large it make more sense to work today, save and enjoy it tomorrow. So leisure have less value today then tomorrow when r is large. Of course, similarly to the case of the Euler equation, all these arguments are mitigated by a low intertemporal substitution of employment $\frac{1}{v}$ (Frish elasticity).[10]

The basic mechanism of the RBC model can hence be summarized as follows.

First, Y_t tends to fluctuate with the technological shock z_t, and ρ gives some persistence to such a shocks. However, there are several other effect triggered by a technological shock.

One regards labour: When productivity (hence the wage) z_t is high it is better to work more today, save and work less tomorrow. This amplifies the cycle. As we saw above this mechanism is particularly strong only when ρ is relatively small. In other terms, when ρ is large employment N does not fluctuate much. Again, if the worker believes this to be a permanent increase, then why increasing hours today if tomorrow I will have roughly the same opportunities of increasing my (permanent) income? On the other hand, this mechanism cannot generate persistence if shocks are 'too' temporary.

Another mechanism regards investment: When productivity today z_t is large investment has been already made. But if ρ (persistence) is high then the entrepreneur should better invest today since it is expected to have high productivity tomorrow. Hence investment fluctuates a lot with a large ρ. A purely temporary shock generates positive investment for pure consumption smoothing purposes. High Y induce the agent to save for tomorrow. And the only storage technology available to the agent is investment (K). A purely temporary shock hence increases K, which then goes back to its steady state level via depreciation. This is the only persistence generated in Y by investment.

Let us now look at the main mechanism affecting intertemporal consumption. Recall that one of the key fact characterizing business cycles is that consumptions investment move together as they are both pro-cyclical. However, feasibility ($Y = C+I$) suggests that whenever I_t increases, then C_t tends to decrease. The RBC model can however generate co-movement between I_t and C_t since the signal inducing an increase in I_t is a positive productivity shock, which increases Y_t as well. As a consequence C_t and I_t can increase together since Y_t (hence feasibility) increases as well. Because of this tension between feasibility and investment C_t has typically a hump shaped behavior. After a positive persistent shock, it increase a bit and then even more the next period as investment is a bit lower in the second period than in the first period.

There is now a number of papers trying to generate business cycles using other source of

[10] Again, empirically $\frac{1}{v}$ is around 0 and 0.5 for men and less than 1 for women.

signals that increase investment. However, they naturally do not change Y_t. The obvious failure of these models is that C_t then tends to be countercyclical. A recent paper by Lorenzoni (2004) argues that a positive purely informational signal (demand) shocks might allow for co-movements between C and I since it might induce a substantial increase in N hence in $Y = F(z, N, K)$.[11]

Finally, let us have a look at the behavior of real wages. By definition of the productivity shock, real wages increase with Y. However, since N input increases as well, diminishing returns mitigate a bit this effect and also give more persistence in real wages.

10.5.2 Evaluating Quantitatively the model

What are the main finding of the basic RBC model and what the data tells us?[12]

Moments	US Data	Baseline RBC Model
σ_Y	1.81	1.34
σ_{e^z}	0.98	0.94
$corr(N, Y)$	.88	.97
$corr(I, Y)$	.80	.99

Moments	US Data	Baseline RBC Model
σ_C/σ_Y	.74	.44
σ_I/σ_Y	2.93	2.95
σ_N/σ_Y	.99	.48
σ_w/σ_Y	.38	.54
$corr(Y/N, Y)$	.55	.98
$corr(w, Y)$	.12	.98

Summary of the Performance evaluation:

1. The model generates right investment relative variability and roughly the right signs for autocorrelations, but the magnitudes are wrong

2. σ_C/σ_Y and σ_N/σ_Y are too low in the model.

3. σ_w/σ_Y is too high in the model.

4. Contemporaneous correlations are all too high. Contemporaneous correlation with C is also too high.

[11] Notice that purely temporal shocks generate small and persistent variations in C since u is concave.

[12] Most figures are from Tables 1 and 3 in King et al. (1999)

5. Since $\sigma_Y^{\text{data}}/\sigma_Y^{\text{model}} = \frac{1.81}{1.34} = .77$ we might argue that 77% of the output variability is explained by the model with only technological shocks. However, Eichenbaum (1991) criticizes this interpretation of the variance ratio. Most of the RBC studies find that technological shocks explain between 70% (Kydland and Prescott, 1991) to 75% of output variability. However, notice that this number has no meaning unless the model is not constructed precisely for this purpose. In this example, the model is constructed mainly with a didactical purpose. Actually, a good exercise would be to go and see the whole set of findings of King et al. (1999) and argue about this.

10.6 Extensions

We already discussed how time-to-build and indivisible labour increase σ_N/σ_Y and reduce autocorrelations with labour. We also saw that the introduction of government shocks help to reduce autocorrelations since induce orthogonal noise. The introduction of heterogeneous agents also improves on autocorrelations with labor and wages. Now the RBC literature is huge. For example, have a look at the web site: `http://dge.repec.org/index.html`. You will find many New-Keynesian articles there as well.

I found particularly stimulating a recent paper titled: "Business Cycle Accounting," by Chari et al. (2003). In my view this paper goes a step beyond Solow (1957). What Solow did in his article was to show that the deterministic growth model was unable to match the actual data. He hence rejected the deterministic model in favor of the stochastic model. He then computed the *Slow residual* performing an accounting exercise based on the production model. The whole variability of Y_t not explained by changes in N_t and K_t was defined as technological shock on A_t. They do a similar, much more complicated, exercise. They use household and firms maximization to determine other stochastic components that explain Y_t. They call them wage, investment, and efficiency wedges respectively. Basically the efficiency wedge is the new Solow residual. They then ask the question: Which one of these wedges is important in explaining the great depression or other recessions an how much? They find that the investment wedge is unimportant.

Obviously this approach to business cycles makes clear that the normative content other authors find in business cycle theorists disappeared. Now business cycles are not considered fluctuations around the potential level of output. It is not longer believed that government interventions have no role. Wedges come from imperfections. This both implies that institutions are important and that some of the problem can be solved - at least partially - by optimal policies.

International Business cycles Backus, Kehoe and Kydland (1992) study the implication of general equilibrium growth theory based models for international business cycles. They assume complete markets. They find important deviations of the data from the model.

1. Consumption is less correlated between countries than the theory would predict.

2. Investment and labour supply are too negatively correlated with output.

3. N and I co-move with domestic Y in the data but not in the model. The failure is particularly important for I.

4. N and I are positively correlated between countries in the data, but strongly negatively correlated in the model

Baxter and Crucini (1995), Kollmann (1996), and Heathcote and Perri (2002) restrict international asset markets to just borrowing and lending. This helps reduce consumption correlation between countries. But they generate basically no domestic co-movement between I and Y, as in the complete markets model. The same failure in generating the positive correlation of N and I between countries present in the data.

Kehoe and Perry (2001) use an endogenous debt model with default to endogenize liquidity constraints and insurance. Their model performs better in the co-movement of I and N than both the complete market or the exogenously incomplete market model. But has a new problem. In the data output and net exports are negatively correlated whereas in the model they are positively correlated. The default constraint reduce that international flow of investment when the country has a good shock.

Welfare cost of the Business Cycles Assume that log consumption is distributed normally with a trend:

$$\ln c_t \sim N\left(a + gt - \frac{1}{2}\sigma^2,\ \sigma^2\right).$$

If we now remove the variability in c we have an increase in expected lifetime utility by having only the deterministic trend. Say that in order to have the same welfare we must increase the level of consumption in the stochastic environment by 100λ percentage points in all dates and states; we get

$$\mathbf{E}_0\left[\sum_{t=0}^{\infty}\left(\frac{1}{1+\theta}\right)^t\frac{((1+\lambda)\,c_t)^{1-\gamma}}{1-\gamma}\right]=\sum_{t=0}^{\infty}\left(\frac{1}{1+\theta}\right)^t\frac{(Ae^{gt})^{1-\gamma}}{1-\gamma}.$$

with $A=e^a$. This comparison implies that the welfare gain λ is

$$\lambda\approx\frac{1}{2}\gamma\sigma^2.$$

First, we saw above that it is not clear how on should compute the risk aversion parameter. Tallarini (2000) computes welfare gains using Kreps-Porteus preferences. Lucas suggested that assuming $\gamma\in(1,4)$ and that we are able to eliminate the whole aggregate variance in consumption (which is $\sigma=0.032$) we would obtain a welfare increase in consumption equivalent of less than 0.2%. Only with a very large $\gamma\in(50,100)$ we could get a $\lambda\in(0.025,0.05)$, that is, a percentage increase of consumption of about $2-5\%$.

In a sequence of papers Krusell and Smith analyze the case with incomplete markets and heterogeneous agents, and investigate how the business cycle affects the welfare agents at different positions in the income distribution in the economy.

Identification of Monetary and Real Shocks (σ) In addition, in order to perform a meaningful exercise we must identify the variance σ of consumption *that can be eliminated* by some macroeconomic policy. We saw that Prescott (1986a,b) identifies real shock with *technological shocks* but there can be more. *Household technology shocks* that are typically identified as due to changes in labor supply preferences, consumption preferences, ...

Monetary or Nominal shocks are typically assumed to *die out in the long run*. Otherwise there is an identification problem. There are shocks that can be induced by the monetary policy or so. One example is the *information* that a change in the money aggregates has on people choices. Unfortunately they are difficult to identify. What is key according to Lucas (2003) is that if a money stock change does not have a one to one effect of the level of prices it must be the case that also *relative prices* are prevented from adjusting so that to create inefficiencies.

Let's say that the *70-75%* of the variance of consumption (and output, and hours) can be attributed to real shocks and *25-30%* to nominal shocks. That is the sort of accepted decomposition. Hence only the 30% variance in business cycles frequency can be reduced (by demand management policies)

Bibliography

[1] Cooley, T. F., and E. C. Prescott (1995) "Economic Growth and Business Cycles," in T. F. Cooley ed., *Frontiers of Business Cycle Research*, Princeton University Press: 1-38.

[2] King, R. G. (1995) "Quantitative Theory and Econometrics," *Federal Reserve Bank of Richmond Economic Quarterly*, 81: 53-105.

[3] King, R. G., and S. T. Rebelo (1999) "Resuscitating Real Business Cycles," in J. B. Taylor, and M. Woodford eds., *Handbook of Macroeconomics*, Vol. 1B, Elsevier: 927-1007.

[4] Mas-Colell, A., Michael D. Winston, and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press.

[5] Sargent, T. (1987) *Dynamic Macroeconomic Theory*, Harward University Press.

[6] Sargent, T. J. and L. Ljungqvist (2000), *Recursive Macroeconomic Theory*, MIT Press.

[7] Stokey, N., R. Lucas and E. Prescott (1991), *Recursive Methods for Economic Dynamics*, Harvard University Press.

10.7 Recursive General Equilibrium

10.8 General Equilibrium with Heterogeneous Agents

10.8.1 Finitely Many Agents and Ex-Ante Heterogeneity: The Negishi's Weights Approach

10.8.2 Decentralization

Bibliography

[1] Cugno, F. and L. Montrucchio (1998) *Scelte Intertemporali*, Carocci, Roma.

[2] Gale, D. (1967a), "On Optimal Development in a Multi-Sector Economy," *The Review of Economic Studies,* 34, 1-18.

[3] Gale, D. (1967b), "A Geometric Duality Theorem with Economic Applications," *The Review of Economic Studies,* 34, 19-24.

[4] Mas-Colell, A., Michael D. Winston and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press.

Chapter 11

Endogenously Incomplete Markets

11.1 A Model with Endogenous Liquidity Constraints (from Kehoe and Levine, Econometrica 2001)

We will now study different market arrangements.

1. Complete Markets → Full risk sharing (Benchmark I)

2. Self-Insurance → One risk free asset (Permanent Income Model)

3. Models with limited commitment (Default)

4. Models with asymmetric information (Moral Hazard)

11.1.0.1 Model (Physical Environment)

- Infinite discrete time periods $t = 0, 1, ...$

- Two type of consumers $i = 1, 2$

- A single consumption good each period c_t^i hence the lifetime consumption vector of type i consumer is $\mathbf{c}^i = \{c_t^i\}_{t=0}^{\infty} \in l_{\infty}^{++}$

- Preferences are the same across consumers:

$$(1 - \delta) \sum_t \delta^t u(c_t^i),$$

with $\delta \in (0, 1)$; $u \in C^2$ and $u' > 0$ and $u'' < 0$; with Inada: $\lim_{c \to 0} u' = \infty$.

- Two type of endowments.
  - Human capital (or wages, or idiosyncratic income) $w_t^i \in \{\omega^b,\ \omega^g\}$ with $w_t^i \neq w_t^{-i}$ and $w_t^i = w_{t+1}^i$. That is, the human capital endowments change deterministically over time switching from ω^g to ω^b every period, with $w_t^1 + w_t^2$ constant for any $t \geq 0$.
  - One unit of physical capital (or durable goods) with per period return $R > 0$ in consumption goods. We denote by θ_t^i the consumer of type i holding of physical capital in period t (shares).
- Obviously: $\theta_t^1 + \theta_t^2 \leq 1$. The other social feasibility constraint is

$$c_t^1 + c_t^2 \leq \omega^g + \omega^b + R \equiv \omega.$$

Market Arrangements:

1. Complete Markets
2. Liquidity Constrained Economy
3. Debt Constrained Economy

11.1.1 Complete Markets

Definition A Competitive AD-equilibrium is an allocation $\mathbf{c} = \{c_t^1, c_t^2\}_{t=0}^\infty$ and a set of prices $\mathbf{p} = \{p_t\}_{t=0}^\infty$ such that given $\mathbf{p}$, $\{c_t^i\}_{t=0}^\infty$ solves the following agent i problem

$$\max_{\mathbf{c}^i} \ (1-\delta)\sum_t \delta^t u(c_t^i) \quad \text{s.t.}$$

and

$$\sum_t p_t c_t^i \ \leq \ \sum_t p_t \left[w_t^i + d\theta_0^i\right], \text{ with } \theta_0^i \text{ given.}$$

and markets clear, that is, $c_t^1 + c_t^2 = \omega$ for any t.

Computation of the Equilibrium

- The first order conditions imply

$$p_t \lambda^i = \delta^t u'(c_t^i) \qquad \forall i, t,$$

we saw in the previous term that since there is no aggregate uncertainty, from market clearing we can show that for each agent $i = 1, 2$ we have $c_t^i = \bar{c}^i \ \forall t$. Using this and normalizing $p_0 = 1$, the first order condition imply $p_t = \delta^t$.

- The budget constraint for agent i becomes

$$(1-\delta)\left(\sum_{t=0}^{\infty}\delta^t(w_t^i+d\theta_0^i)\right)=\bar{c}^i.$$

- Hence, assume agent 1 starts with ω^g

$$\begin{aligned}\bar{c}^1 &= (1-\delta)\left[\left(1+\delta^2+\delta^4+...\right)\omega^g+\delta\left(1+\delta^2+\delta^4+...\right)\omega^b\right]+d\theta_0^1\\ &= \frac{1-\delta}{1-\delta^2}\omega^g+\delta\frac{1-\delta}{1-\delta^2}\omega^b+d\theta_0^1\\ &= \frac{1}{1+\delta}\omega^g+\delta\frac{1}{1+\delta}\omega^b+d\theta_0^1=\frac{\omega^g+\delta\omega^b}{1+\delta}+d\theta_0^1,\end{aligned}$$

and, similarly for agent 2 we have

$$\bar{c}^2=\frac{\delta\omega^g+\omega^b}{1+\delta}+d\theta_0^2.$$

- The consumer problem can be written as

$$\begin{aligned}&\max_{\{c_t^i,\theta_t^i\}}\ (1-\delta)\sum_t\delta^t u(c_t^i)\quad \text{s.t.}\\ c_t^i+\nu_t\theta_{t+1}^i &\le w_t^i+(\nu_t+d)\,\theta_t^i,\ \theta_t^i\ge-\Theta,\ \theta_0^i\text{ given.}\end{aligned}$$

where ν_t is the period t price of of capital and Θ is chosen high enough so that to rule out Ponzi Games but to not make the constraint $\theta_t^i\ge-\Theta$ otherwise binding.

- **Definition**. An equilibrium is an infinite sequence of consumption levels $\{c_t^1,c_t^2\}_{t=0}^{\infty}$, capital holdings $\{\theta_t^1,\theta_t^2\}_{t=0}^{\infty}$, and capital prices $\nu=\{\nu_t\}_{t=0}^{\infty}$ such that consumers solve the consumer maximization problem and such that the social feasibility conditions are satisfied.

11.1.2 Liquidity Constrained Economy

- The consumer's i problem is

$$\begin{aligned}&\max_{\{c_t^i,\theta_t^i\}}\ (1-\delta)\sum_t\delta^t u(c_t^i)\quad \text{s.t.}\\ c_t^i+\nu_t\theta_{t+1}^i &\le w_t^i+(\nu_t+d)\,\theta_t^i,\ \theta_t^i\ge-B,\ \theta_0^i\text{ given.}\end{aligned}$$

where $B\ge 0$ can be quite small (incomplete markets).

- An important example is when $B = 0$. In this case agents can only carry out intertemporal trade to smooth consumption by exchanging physical capital which can only be kept in positive amounts. This constraint can be enforced in a fully decentralized anonymous way.
- The key feature that is not very much evident in this environment is that there are no securities or other assets that can be traded besides physical capital. Obviously, this will become important when we will consider a stochastic environment.

Definition

An equilibrium is an infinite sequence of consumption levels $\mathbf{c}$, capital holdings $\theta = \{\theta_t^1, \theta_t^2\}_{t=0}^{\infty}$, and capital prices $\nu = \{\nu_t\}_{t=0}^{\infty}$ such that consumers maximize utility given their constraints, and such that the social feasibility conditions are satisfied.

- K & L focus on symmetric steady states, i.e. allocations that depend only on the today's endowment. The first order conditions:

$$\begin{aligned} u'\left(c^g\right)\nu &= \delta\left(\nu+d\right)u'\left(c^b\right) \\ u'\left(c^b\right)\nu &\geq \delta\left(\nu+d\right)u'\left(c^g\right) \end{aligned}$$

  the budget constraints

$$\begin{aligned} c^g + \nu\theta^b &= \omega^g + \left(\nu+d\right)\theta^g \\ c^b + \nu\theta^g &= \omega^b + \left(\nu+d\right)\theta^b \end{aligned}$$

  and market clearing

$$\begin{aligned} \theta^b + \theta^g &= 1 \\ c^b + c^g &= \omega. \end{aligned}$$

Definition

- In this deterministic environment the symmetric equilibrium always exists and in it can be of two types: Full insurance, and Partial insurance.
- If we have Full Insurance ($c^g = c^b$) then

$$u'(\omega - c^g)\nu = \delta\left(\nu+d\right)u'\left(c^g\right),$$

  $\Rightarrow c^g = \frac{\omega}{2}$ and $\frac{(\nu+d)}{\nu} = (1+r) = \frac{1}{\delta}$.
- If we have Partial insurance, $c^g \in \left(\frac{\omega}{2}, \omega^g\right]$ and $\theta^b = -B$. In this case, it must be that $\omega^g > \frac{\omega}{2} = \frac{\omega^g+\omega^b+d}{2}$ since $\theta = 0$ is always feasible.

11.1.3 Debt Constrained Economy

- Here agents have access to a full set of securities (i.e. they solve an Arrow-Debreu like problem)
- However, in each period t they face an individual rationality constraint of the following type
$$(1-\delta)\sum_{n=0}^{\infty}\delta^n u(c_{t+n}^i) \geq (1-\delta)\sum_{n=0}^{\infty}\delta^n u(w_{t+n}^i) \tag{IC}$$
this constraint can be either interpreted as a bankruptcy or opt out value. The agents can default on the debt. In this case they will lose the physical capital endowment and will be excluded from the market forever. The human capital is assumed to be inalienable.
- The model implicitly assumes the presence of a credit agency or government, who keep trace of who goes bankrupt.
- The consumers' problem is
$$\begin{aligned} & \max_{\mathbf{c}^i}\ (1-\delta)\sum_t \delta^t u(c_t^i) \quad \text{s.t.} \\ & IC, \text{ and} \\ \sum_t p_t c_t^i \quad \leq & \quad \sum_t p_t\left[w_t^i + R\theta_0^i\right], \text{ with } \theta_0^i \text{ given.} \end{aligned}$$
- Notice that the central authority must also control how much each agent borrows, so that the trade are incentive compatible.
- The Radner sequence formulation of the constraint is similar to the liquidity constrained model
$$c_t^i + q_t\theta_{t+1}^i \leq w_t^i + (\nu_t + R)\,\theta_t^i,\ \theta_t^i \geq -\Theta$$
where Θ is chosen high enough so that to rule our Ponzi Games but to not make the constraint $\theta_t^i \geq -\Theta$ otherwise binding.

11.1.4 Equilibrium and Efficiency

Definition 47 *An equilibrium in these economies is an infinite sequence of consumption levels, asset holding and prices both for assets and so that consumers maximize lifetime utility subject to he budget constraints, the restrictions for asset holdings, the eventual IC constraint (in the debt constrained economy), and the allocation is socially feasible.*

- They focus on symmetric steady states, i.e. allocations that depend only on the today's endowment:
$$c_t^i \in \left\{c^g, c^b\right\}.$$
- In this deterministic environment the symmetric equilibrium exists in both economies and in both cases can be of two types: Full insurance, where $c^g = \frac{\omega}{2}$ and Partial insurance, where $c^g \in \left(\frac{\omega}{2}, \omega^g\right]$. In the latter case, it must be that $\omega^g > \frac{\omega}{2} = \frac{\omega^g + \omega^b + R}{2}$.
- A standard argument also shows that the debt constrained equilibrium is Pareto efficient.
  - Assume it not the case, that is, there is a new feasible allocation c' so that both agents are made better off with respect to the equilibrium allocations. In this case, the new allocation cannot be feasible since it must be that at the new allocation c' the agent that stays better does not satisfy the budget constraint and that the other agent has the budget either violated or just satisfied. Hence
$$\sum_t p_t \left(c_t^{1\prime} + c_t^{2\prime}\right) > \sum_t p_t \omega.$$
  - **NB:** This is the firm's profits maximization violation in our First Welfare Theorem in the RBC model.
  - Notice that in the whole reasoning the IC constraint does not play a role since it will be even more satisfied by a Pareto dominating allocation.
- The interest rate in the steady state is
$$r = \frac{u'(c^g)}{\delta u'(c^b)} - 1.$$
When $c^g > c^b$ we have that $r > \frac{1}{\delta} - 1$, the subjective discount rate. This is so since the borrowers are constrained but the lender no. Hence, in order to satisfy market clearing we have to reduce the interest rate, otherwise the lenders (high endowment) would be willing to lend more than the equilibrium amount.

11.2 Stochastic Case

- Now, at each period the endowments to the agents are generated by a shock $z_t \in \{1, 2\}$ which denotes who gets the high endowment. The variable η_t is assumed to be Markov with transitions $\pi \in (0,1)$, the probability of reversal. When $\pi = 1$ we are in the deterministic case above. The case $\pi = 0$ is trivial.

- We will denote by $z^t = (z_0, ...z_t)$ the history of shocks up to period t, where z_0 is the initial state of the economy. Finally, we denote by

$$\mu(z^t) = pr\left(z_t \mid z_{t-1}\right) pr\left(z_{t-1} \mid z_{t-2}\right) \; ... \; pr\left(z_1 \mid z_0\right)$$

the probability of history z^t (given z_0).

- The problem in the liquidity constrained case is

$$\max_{\mathbf{c}^i} \; (1-\delta) \sum_{z^t} \delta^t \mu(z^t) u\left(c_t^i\left(z^t\right)\right) \quad \text{s.t.}$$

$$c_t^i\left(z^t\right) + \nu_t\left(z^t\right)\theta_{t+1}^i\left(z^t,\eta\right) \;\; \leq \;\; w_t^i + \left(\nu_t\left(z^t\right) + R\right)\theta_t^i\left(z^t\right), \; \theta_t^i\left(z^t\right) \geq -B, \; \theta_0^i \text{ fixed,}$$

In the debt constrained is

$$\max \; (1-\delta) \sum_{z^t} \delta^t \mu(z^t) u\left(c_t^i\left(z^t\right)\right)$$

s.t.

$$(1-\delta) \sum_{z^{t+n}/z^t} \delta^n \frac{\mu(z^{t+n})}{\mu(z^t)} u\left(c_{t+n}^i\left(z^{t+n}\right)\right) \geq (1-\delta) \sum_{z^{t+n}/z^t} \delta^n \frac{\mu(z^{t+n})}{\mu(z^t)} u\left(w_{t+n}^i\left(h^{t+n}\right)\right)$$

$$\begin{aligned} c_t^i\left(z^t\right) + q_t\left(z^t,1\right)\theta_{t+1}^i\left(z^t,1\right) + q_t\left(z^t,2\right)\theta_{t+1}^i\left(z^t,2\right) &\leq w_t^i + \left(\nu_t\left(z^t\right) + R\right)\theta_t^i\left(z^t\right), \\ \theta_t^i\left(z^t\right) &\geq -\Theta, \; \theta_0^i \text{ fixed.} \end{aligned}$$

- Notice that agents now decide state contingent plans, and $q_t\left(z^t, z_{t+1}\right)$ is the price of an Arrow-Debreu security traded in state z^t, period t that promises a unit of physical capital in state (z^t, z_{t+1}) next period.

- Standard arbitrage implies that

$$q_t\left(z^t,1\right) + q_t\left(z^t,2\right) = \nu_t\left(z^t\right).$$

- It turn out that the debt constrained economy has a unique symmetric steady state of the same form as before, whereas the liquidity constrained does not have it when $\pi \in (0,1)$. In addition

$$\frac{dc^g}{d\left(1-\pi\right)} > 0$$

i.e. more persistence in the shocks reduces trade, hence insurance.

- The first order conditions in the liquidity constrained model are

$$u'(c_t) \geq \delta \mathbf{E}_t \left[\frac{\nu_{t+1} + d}{\nu_t} u'(c_{t+1}) \right]$$

which is the first order condition for the situation where the agent is restricted to trades in which $\theta^i(z^t, 1) = \theta^i(z^t, 2)$.

- Alvarez and Jermann (Econometrica, 2000) describe how the equilibrium allocation in the default economy can be equivalently described by imposing state contingent (and possibly agent specific) liquidity constraints on the arrow securities:

$$\theta_{t+1}(z^t, z) \geq -B^i(z^t, z)$$

- Such constraints must be imposed so that the agents' following period IC constraint is satisfied.

Chapter 12

Recursive Contracts

12.1 Default Contracts with one sided commitment

12.1.1 Set up and contracts

Consider a consumer who lives two periods and faces an exogenous stochastic income process. Income realizations are discrete $y_t \in Y = \left\{y^1,\ y^2, ...,\ y^N\right\}$, and positive $y^i > 0$ with $y^{i+1} > y^i$. The distribution over Y is summarized by a vector $\pi_i > 0$, $i = 1,\ ...,\ N$, with $\sum_{i=1}^N \pi_i = 1$.

An insurance insurance contract in this environment is a collection of functions $\mathcal{W}_0 = \{c_t(h^t)\}_{t=1}^2$ that map each history of income shocks $h^1 = y_1$, and $h^2 = (y_1, y_2)$ into a consumption level $c_t\left(h^t\right)$. We can, for example, write $c_1^i = c_1\left(y_1^i\right)$ and $c_2^{i,j} = c_2\left(y_1^i, y_2^j\right)$.

The agent's expected discounted lifetime utility takes the following form

$$\mathbf{U}_2\left(\mathcal{W}_0\right) = \sum_{i=1}^N \pi_i \left[u\left(c_1^i\right) + \beta\left(\sum_{j=1}^N \pi_j u\left(c_2^{i,j}\right)\right)\right], \tag{12.1}$$

with u a concave and differentiable function. In class, we saw models where, although the agent faces a complete set of securities, she cannot attain the fully optimal allocation. The reason is an endogenous restriction on the trade of securities she faces, which are generated by the possibility of default. They are summarized by a set of *default constraints* of the form

$$u\left(c_1^i\right) + \beta\left[\sum_{j=1}^N \pi_j u\left(c_2^{i,j}\right)\right] \geq A^i - P, \qquad (IC_1^i)$$

for all $i = 1,\ 2,\ ...,\ N$; and

$$u\left(c_2^{i,j}\right) \geq u\left(y_2^j\right) - P \qquad (IC_2^{ij})$$

for all i, j; where $A^i = u\left(y_2^i\right) + \beta \sum_{j=1}^N \pi_j u\left(y_2^j\right)$, is the autarchy value and $P \geq 0$ is a default penalty.

The *constrained* Pareto optimal allocation can be computed by solving the problem of a risk neutral planner whose aim is to maximize his net revenues subject to a welfare constraint for the consumer. We assume that the planner discounts at the factor β. The planner problem is

$$V_2\left(U_0\right) = \max_{\mathcal{W}_0} \mathbf{V}_2(\mathcal{W}_0) = \max_{\mathcal{W}_0} \sum_{i=1}^N \pi_i \left[\left(y_1^i - c_1^i\right) + \beta \sum_{j=1}^N \pi_j \left(y_2^j - c_2^{i,j}\right)\right] \tag{12.2}$$

$$\text{s.t.:} \quad (IC_1^i) \text{ for all } i, \ (IC_2^{ij}) \text{ for all } i, j, \tag{12.3}$$

and the constraint

$$\mathbf{U}_2(\mathcal{W}_0) = \sum_{i=1}^N \pi_i \left[u\left(c_1^i\right) + \beta\left(\sum_{j=1}^N \pi_j u\left(c_2^{i,j}\right)\right)\right] \geq U_0 \tag{12.4}$$

which guarantees the agent gets at least a welfare value of U_0 from the contract.

Similarly to the period zero contract, we will denote by $\mathcal{W}_1^i = \left\{c_2^{i,j}\right\}_{j=1}^N$ the insurance contract *continuation* after the occurrence of income level y_1^i, and $\mathbf{V}_1(\mathcal{W}_1^i)$ and $\mathbf{U}_1(\mathcal{W}_1^i)$ will denote the values associated to any contract continuation $\mathcal{W}_1^i$, where:

$$\begin{aligned}
\mathbf{V}_1(\mathcal{W}_1^i) &= \sum_{j=1}^N \pi_j \left(y_2^j - c_2^{i,j}\right) \\
\mathbf{U}_1(\mathcal{W}_1^i) &= \sum_{j=1}^N \pi_j u\left(c_2^{i,j}\right).
\end{aligned}$$

Note that $\mathbf{V}_2(\mathcal{W}_0)$ differs from $V_2\left(U_0\right)$, which is the value associated to the optimal contract $\mathcal{W}_0^*$, the contract that solves problem (12.2)-(12.4). The subscripts in the values denote the time horizon.

12.1.2 The Recursive formulation

Step 1: a simple observation. With this notation in mind we can rewrite the period zero problem as follows:

$$V_2\left(U_0\right) = \max_{\mathcal{W}_0} \mathbf{V}_2(\mathcal{W}_0) = \max_{\left\{c_1^i, \mathcal{W}_1^i\right\}_i} \sum_{i=1}^N \pi_i \left[\left(y_1^i - c_1^i\right) + \beta \mathbf{V}_1(\mathcal{W}_1^i)\right]$$

s.t.

$$\mathbf{U}_2(\mathcal{W}_0) = \sum_{i=1}^{N} \pi_i \left[u\left(c_1^i\right) + \beta \mathbf{U}_1(\mathcal{W}_1^i) \right] \geq U_0,$$

for all $i = 1,\ 2,\ ...,\ N$

$$u\left(c_1^i\right) + \beta \mathbf{U}_1(\mathcal{W}_1^i) \geq A^i - P,$$

and for all i, j

$$u\left(c_2^{i,j}\right) \geq u\left(y_2^j\right) - P.$$

Notice that the continuations of the contract $\mathcal{W}_1^i$ affect the initial period default and participation constraints only through a particular transformation of it: the agent's value $\mathbf{U}_1(\mathcal{W}_1^i)$. All the other details are somehow irrelevant for period zero decisions; the specificities of $\mathcal{W}_1^i$ only matter for the second period default constraints $u\left(c_2^{i,j}\right) \geq u\left(y_2^j\right) - P$. This is a key observation, that in most cases gives a good hint on what are the relevant *state variables* of the problem.

Step 2: the proof of recursivity. We now want to give a form to the value $\mathbf{V}_1(\mathcal{W}_1^i)$, when $\mathcal{W}_1^i$ is chosen *optimally* at period zero. We are not interested in understanding how a generic contract maps into the planner values, we only care about the 'optimal' one.

Denote by $\mathcal{W}_1^{i*}$ the continuation of the optimal contract - the contract $\mathcal{W}_0^*$ that solves problem (12.2)-(12.4) - and by $U_1^i = \mathbf{U}_1(\mathcal{W}_1^{i*})$ the continuation lifetime utility it implies for the agent. Clearly, $\mathcal{W}_1^{i*}$ must be such that for all $j = 1,\ ...,\ N$

$$u\left(c_2^{*i,j}\right) \geq u\left(y_2^j\right) - P.$$

That is, the continuation of an optimal contract must be incentive compatible in period 2. As a consequence, its value for the planner $\mathbf{V}_1(\mathcal{W}_1^{i*})$ cannot be higher than that associated to the constrained optimal contract (chosen at $t = 1$) delivering U_1^i to the agent and satisfying the incentive constraints, i.e.

$$\mathbf{V}_1(\mathcal{W}_1^*) \leq V_1(U_1^i) = \left\{ \begin{array}{c} \max_{\mathcal{W}_1^i} \mathbf{V}_1(\mathcal{W}_1^i);\ \text{s.t.} \mathbf{U}_1(\mathcal{W}_1^i) \geq U_1^i, \\ \text{and for all } j\ u\left(c_2^{i,j}\right) \geq u\left(y_2^j\right) - P. \end{array} \right\}. \tag{12.5}$$

The key result of recursive contracts is that optimality requires that for all i the above inequality is satisfied with equality, i.e. $\mathbf{V}_1(\mathcal{W}_1^{i*}) = V_1(U_1^i)$. The implication of the Bellman Principle is that all that matters to reconstruct the optimal contract continuation is the particular 'statistic' $U_1^1 = \mathbf{U}_1(\mathcal{W}_1^{i*})$. This number in turn induces a constraint, which enters into the planner 're-maximization' problem in the following period.

Proposition 13 *For all i it must be that $\mathbf{V}_1(\mathcal{W}_1^{i*}) = V_1(U_1^i)$.*

Proof: By contradiction (assuming existence). Recall the previous notation and assume that for some $\bar{\imath}$

$$\mathbf{V}_1(\mathcal{W}_1^{\bar{\imath},*}) < V_1(U_1^{\bar{\imath}}).$$

Call $\mathcal{W}_1^{\bar{\imath}**}$ the continuation of the contract that satisfies the maximization problem (12.5), i.e. $\mathbf{V}_1(\mathcal{W}_1^{\bar{\imath}**}) = V_1(U_1^{\bar{\imath}})$. Clearly, under $\mathcal{W}_1^{\bar{\imath}**}$ the agent gets his utility $U_1^{\bar{\imath}}$. We are now going to show that the planner can gain by using $\mathcal{W}_1^{\bar{\imath}**}$ instead of $\mathcal{W}_1^{\bar{\imath}*}$ after $y_1^{\bar{\imath}}$. We will construct a new contract that delivers exactly the same return in any contingency but in one - when $y_1 = y^i$ - in which case the new contact delivers a higher return for the planner. Recall that by construction $\mathbf{U}_1(\mathcal{W}_1^{**\bar{\imath}}) \geq \mathbf{U}_1(\mathcal{W}_1^{*\bar{\imath}}) = U_1^{\bar{\imath}}$. This implies that

$$u\left(c_1^{\bar{\imath}*}\right) + \beta \mathbf{U}_1(\mathcal{W}_1^{**\bar{\imath}}) \geq A^{\bar{\imath}} - P,$$

and

$$\sum_{i \neq \bar{\imath}}^{N} \pi_i \left[u\left(c_1^{*i}\right) + \beta \mathbf{U}_1(\mathcal{W}_1^{*i})\right] + \pi_{\bar{\imath}}\left[u\left(c_1^{*\bar{\imath}}\right) + \beta \mathbf{U}_1(\mathcal{W}_1^{**\bar{\imath}})\right] \geq U_0.$$

But then the planner can change the original contract $\mathcal{W}_0^{**}$ as follows. He keeps all initial payments the same as those in the original contract $\mathcal{W}_0^{*}$, i.e., $c_1^{i**} = c_1^{i*}$ for all i. Moreover, he keeps the contract continuations mostly as the originals. In particular, for all $i \neq \bar{\imath}$ $\mathcal{W}_1^{**i} \equiv \mathcal{W}_1^{*i}$. That is, it does not change it at these nodes. Finally, only for $i = \bar{\imath}$ he change $\mathcal{W}_0^{*}$ by using the newly computed $\mathcal{W}_1^{**\bar{\imath}}$ instead of the old one $\mathcal{W}_1^{*\bar{\imath}}$. As announced, this newly formed contract $\mathcal{W}_0^{**}$ still satisfies both the initial participation constraint for U_0 and all the incentive constraints, and delivers a higher return for the planner. This contradicts the optimality of the original contract $\mathcal{W}_0^{*}$ as there is another contract $\mathcal{W}_0^{**}$ which dominates it. This contradiction proves the statement. There cannot be a realization y^i such that an optimal contract delivers $\mathbf{V}_1(\mathcal{W}_1^{i*}) < V_1(U_1^i)$. And we are done. **Q.E.D.**

Step 3: the 'handy' recursive formulation. We are now ready to enjoy the 'beauty' of being able to write the problem in recursive form. For any given level of welfare U_0 to be delivered to the agent in period zero, we have

$$V_2\left(U_0\right) = \max_{\{c^i, U^i\}_{i=1}^N} \sum_{i=1}^{N} \pi_i \left[\left(y^i - c^i\right) + \beta V_1(U^i)\right] \quad \text{s.t.}$$

s.t.

$$u\left(c^i\right) + \beta U^i \geq A^i - P,$$

for all $i = 1,\ 2,\ ...,\ N$ and

$$\sum_{i=1}^{N} \pi_i \left[u\left(c^i\right) + \beta U^i\right] \geq U_0.$$

Where for all numbers U^i we have

$$\begin{aligned} V_1(U^i) &= \max_{\{c^j\}_{j=1}^N} \sum_{j=1}^N \pi_j \left(y^j - c^j\right) \\ \text{s.t. for all } j &: \quad u\left(c^j\right) \geq u\left(y^j\right) - P, \\ &\quad \text{and} \\ \sum_{j=1}^N \pi_j u\left(c_1^j\right) &\geq U^i. \qquad (\lambda^i) \end{aligned}$$

In the recursive formulation, for any given level of promised utility U_0, at period zero the planner chooses consumption levels c_1^i and promises future utilities U_1^i for $i = 1, 2, ..., N$. For the future, the planner is not choosing complicated objects such a continuation contracts any more, he is choosing the numbers U^i.

Moreover, the planner knows exactly how to evaluate each U^i: $V_1(U^i)$ is the value of a relatively simple constrained maximization problem. The constraint (λ^i) is called the 'promise-keeping' constraint that requires the contract to deliver the promised level of utility U_1^i to the agent. It plays the role of the *law of motion* for the state variable U.

12.1.3 Characterization of the contract in the infinite horizon case

One of the key advantages of the recursive formulation is that the complexity of the problem does not change as we increase the time horizon. For example, in an infinite horizon environment a contract is a very complicated object: $\mathcal{W}_0 = \{c_t(h^t)\}_{t=1}^\infty$. It is a huge collection of functions mapping histories of income shocks $h^t = (y_1, ..., y_t)$ into consumption levels c_t.

The recursive formulation is however very similar to that described above for the two period model. When the time horizon is infinite, the value function $V = V_\infty$ becomes time invariant, and solves the following functional equation:

$$V(U) = \max_{\{c^i, U^i\}_{i=1}^N} \sum_{i=1}^N \pi_i \left[\left(y^i - c^i\right) + \beta V(U^i)\right]$$

s.t.: for all $i = 1, 2, ..., N$

$$u\left(c^i\right) + \beta U^i \geq A^i - P, \qquad (\pi_i \mu_i)$$

and

$$\sum_{i=1}^{N} \pi_i \left[u\left(c^i\right) + \beta U^i \right] \geq U. \quad (\lambda)$$

where the autarchy value can be written as $A^i = u\left(y^i\right) + \beta A$ where $A = \sum_i \pi_i A^i$. [Check it!].

Let us characterize the optimal contract. The first order conditions are

$$\begin{aligned} -V'(U) &= \lambda && \text{(the envelope)} \\ -V'\left(U^i\right) &= \lambda + \mu^i = \frac{1}{u'\left(c^i\right)} \end{aligned}$$

Since V is concave and $\mu_i \geq 0$ we have

$$U^i \geq U,$$

i.e., the utility of the agent increases through time.

Moreover, since A^i increases with y^i we have that if $\mu_i > 0$ then $\mu_{i+1} > 0$. Hence, for all U we would have a set of indexes where for all for all $i \leq \underline{i}(U)$ $\lambda = \frac{1}{u'(c^i)} = \frac{1}{u'(\underline{c})}$, and $V'(U) = V'\left(U^i\right)$ (hence $U^i = U$). And for $i > \underline{i}(U)$ consumption and utility will be higher as $-V'\left(U^i\right) = \frac{1}{u'(c^i)} > \frac{1}{u'(\underline{c})} = -V'(U)$. Since when $\mu_i > 0$ the incentive constraint must be binding, for such indexes we have $u\left(c^i\right) + \beta U^i = A^i - P$.

By applying repeatedly this argument we get that consumption and utilities increase through time: $c^i_{t+1} \geq c_t$ and $U^i_{t+1} \geq U_t$. This is a key property of the optimal contract under limited commitment. However, when payments are at the level that satisfy the incentive constraint for the maximal income y^N, i.e., when

$$u\left(c^N\right) + \beta U^N = A^N - P,$$

they will not change any more.

If we now assume there is continuum of identical agents, each agent living infinitely periods, and facing the same stochastic process of income. By the law of large numbers, in such economy π_i will also denote the fraction of agents getting shock y^i. It is possible to show that, since $\pi_N > 0$, in the long run each agent will eventually get this high income shock, and from there onward she will be fully insured. The steady state distribution of consumption in this economy will hence degenerate at one point: c^N.

Exercise 64 *Argue that the steady state distribution is not unique as it depends on the initial condition U_0.*

Exercise 65 *Assume that $P = 0$ and that the planner has limited commitment as well. The optimal contract must satisfy the following set of incentive constraints*

$$\text{for all } i: \quad y^i - c^i + \beta V(U^i) \geq 0.$$

Derive the optimal contract in this case. Would it be very different from that with one sided commitment? And the steady state distribution, how would it look like?