1 Infinite-horizon Models and Stochastic Processes

The life-cycle model is useful for studying lifetime patterns of behavior or things that are correlated with age. Modeling overlapping generations, however, introduces several complexities that we may wish to avoid by assuming that households are infinitely-lived. In particular, whereas life-cycle models carry the dimension of age, infinitely-lived models are age-independent and do not require extensive demographic accounting in the way that overlapping generations models do. This reduction in the state space allows for other complexities to be introduced into the model.

1.1 The Deterministic Dynastic Household Model

Infinitely-lived agent models are often developed from the perspective that individuals within a households may think past their own lifetimes and save for future generations of household members by acting as if they were infinitely-lived. We can specify the problem by generalizing the life-cycle model into the infinite horizon according to the following utility maximization problem:

$$\max_{\{n_t, a_{t+1}\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \beta^{t-1} u(c_t, n_t) \tag{1}$$

$$\text{s.t. } \; c_t = wn_t + Ra_t - a_{t+1}. \tag{2}$$

Notice that we have removed the labor productivity term z_t for now. We will introduce that again later as a parameter that varies randomly, but for now, we just want to understand the basics of solving an infinite-horizon model.

We will want to represent this dynamic optimization problem using a Bellman equation like we did in the life-cycle model. Recall that when we solved the life-cycle model using the Bellman equation, we used backwards induction. Such an approach was useful when we knew the terminal value at the end of life, but such information is no longer useful in the case of an infinite horizon. Instead, we rely on an important observation that can help us to choose an approach. That observation is that if, for example, we were to observe our assets today and make a savings decision, and we were to wake up tomorrow with the same amount of assets, then the optimal decision should look exactly the same. In other words, the decision is entirely independent of time, and depends only on non-temporal state space variables. Further, this implies that our continuation function today should be exactly the same as our value function today. This gives us an important goal in solving the dynamic optimization problem: $V_t(a) = V_{t+1}(a)$ for any period t and for any point a in the state space domain. Moreover, since the value function and continuation value should be the same function, we actually do not need the time subscript anymore. In fact, we can write the Bellman equation in any period by adopting the conventional *prime* notation as follows:

$$V(a) = \max_{n,a'} u(c,n) + \beta V(a') \tag{3}$$

$$\text{s.t. } \; c = wn + Ra - a'. \tag{4}$$

In order to solve the problem, know that if we have a function $V(a)$ such that $V_t(a) = V_{t+1}(a)$ for all a, then we have our target value function. But how do we solve that? Computationally, we rely on a combination of the *Contraction Mapping Theorem* and *Blackwell's Sufficient Conditions* to solve for $V(a)$ when the dynamic optimization problem satisfies certain properties.[1] This

[1] The first property requires monotonicity of the operator generating the sequence $\{V_n(a)\}_{n=0}^{\infty}$. The second condition requires that the discount factor is bounded strictly in between zero and one.

mathematical framework suggests that if we were to start of with a weakly concave function, for example $V_0(a) = 0$ for all a, then the following sequence of function evaluations will finally converge to the value function $V(a)$:

$$V_1(a) = \max_{n,a'} u(c, n) + \beta V_0(a') \tag{5}$$

$$\text{s.t. } c = wn + Ra - a', \tag{6}$$

where the solution to this problem is used in the next iteration:

$$V_2(a) = \max_{n,a'} u(c, n) + \beta V_1(a') \tag{7}$$

$$\text{s.t. } c = wn + Ra - a'. \tag{8}$$

Like we mentioned above, once we have a function such that $V_t(a) = V_{t+1}(a)$, then we have a solution to the problem. In fact, we can use this information as a convergence criteria for our computation:

$$\|V_{n+1}(a) - V_n(a)\| < \varepsilon. \tag{9}$$

Notice that because our infinite-horizon problem is not dependent on time, the only dimension we have to keep track of is our state space variable a. In other words, value functions and policy functions now only depend on assets at the beginning of the period. Also, since we do not have a terminal period, we can not directly institute our no-Ponzi condition. As a result, we will simply assume that assets can only take positive values for now and note that this issue is remedied by a so-called *transversality condition* in infinite-horizon models.

1.2 Stochastic Productivity in the Dynastic Model

When we studied the life-cycle model, we made the assumption that productivity rose deterministically early in life, then declined later in life. We showed this to be true on average by estimating the profile of log hourly wages by age. In addition to this deterministic component of productivity, individuals face significant productivity risk for several reasons, including job loss/gain, promotions, training, economic recessions/expansions, changing industries, and entrepreneurial opportunities. How individuals prepare for and respond to this uncertainty has important implications for public policy.

We begin by assuming that the productivity parameter z now has a corresponding stochastic process that is independently and identically distributed over time. In particular, suppose that individuals took productivity draws $z \in [\underline{z}, \overline{z}]$ from a distribution $F(z)$ in every period. We call the domain of the stochastic process ($[\underline{z}, \overline{z}]$) the *support* of the distribution $F(z)$. The function $F(z)$ tells us how likely it is to draw a value within the interval $[z_1, z_2] \subset [\underline{z}, \overline{z}]$. Specifically, we can mathematically specify the probability of drawing a particular value of $z \in [z_1, z_2]$ by the expression $\int_{z_1}^{z_2} z dF(z)$. Alternatively, this is sometimes specified as $\int_{z_1}^{z_2} z f(z) dz$, where $f(z) = F'(z)$.

Now that we have introduced some of the general language corresponding to stochastic processes, we consider a simple example. Suppose that productivity can only take two values: a low value z_l and a high value z_h. Further, suppose that the probability of a low draw is p_l, so that the probability of a high draw is $p_h = 1 - p_l$. The agent now wishes to maximize *expected utility* by observing the current productivity shock, and optimizing by choosing labor supply and savings as follows:

$$V(z_l, a) = \max_{n,a'} u(c, n) + \beta \left[p_l V(z_l, a') + (1 - p_l) V(z_h, a')\right] \tag{10}$$

$$\text{s.t. } c = wz_l n + Ra - a' \tag{11}$$

for a low current productivity draw and

$$V(z_h, a) = \max_{n,a'} u(c,n) + \beta\left[p_l V(z_l, a') + (1-p_l)V(z_h, a')\right] \tag{12}$$

$$\text{s.t. } c = wz_h n + Ra - a' \tag{13}$$

for a high productivity draw. The agent can not control draws of the random variable z, but the agent can control future values of a. Accordingly, it is useful to think of the state space as having exogenous variables, such as z, and endogenous variables, such as a. Notice that we simply have to iterate over the productivity types $z \in \{z_l, z_h\}$ in addition to our discretized points of the asset dimension to solve the problem. We can generalize the previous Bellman equations for an arbitrary number of productivity shocks z_j and corresponding probabilities $p(z_j)$ as follows:

$$V(z_i, a) = \max_{n,a'} u(c,n) + \beta \sum_{j=1}^{N} p(z_j) V(z_j, a') \tag{14}$$

$$\text{s.t. } c = wz_i n + Ra - a', \tag{15}$$

where $i, j \in \{1, \ldots, N\}$ and $\sum_{j=1}^{N} p(z_j) = 1$.

Up until now, we have only considered stochastic processes that are independent over time. In reality, labor productivity and many other economic variables are correlated over time. To capture this correlation over time, we will make use of very simple modeling tool known as a *Markov matrix*. For a given number of possible values that a stochastic process can take N, the Markov matrix is an $N \times N$ square matrix whose column elements represent the probability of a state being realized in the next period for a particular row representing the current state. Consider, as in the previous example, that productivity can only take two values: $z \in \{z_l, z_h\}$. Let $p(z'_l|z_l)$ represent the probability that productivity is low in the next period given that probability is low today, and let $p(z'_h|z_h)$ denote the probability of repeating a high productivity shock. Then, we could denote the Markov matrix as

$$P = \begin{bmatrix} p(z'_l|z_l) & 1 - p(z'_l|z_l) \\ 1 - p(z'_h|z_h) & p(z'_h|z_h) \end{bmatrix}. \tag{16}$$

In the Markov matrix P above, the rows represent the possible states in the current period, while the columns represent the possible states in the next period. In particular, the top row represents a low current productivity shock, while the bottom row represents a high current productivity shock. Similarly, the left column represents a low productivity shock in the next period, whereas the right column represents a high productivity shock in the next period. One important feature of every Markov matrix is that all elements along a particular row must sum to 1. This means that, for a given state today, row elements must give us the probability of going into every possible future state. We can generalize this concept by considering the probability of going from productivity state i today to a state j in the next period and summing across all possible values $j \in \{1, \ldots, N\}$:

$$\sum_{j=1}^{N} p(z_j|z_i) = 1. \tag{17}$$

Finally, we can write our generalized Bellman equation for an individual with current productivity shock z_i as:

$$V(z_i, a) = \max_{n,a'} u(c, n) + \beta \sum_{j=1}^{N} p(z_j|z_i)V(z_j, a') \tag{18}$$

$$\text{s.t. } \; c = wz_i n + Ra - a'. \tag{19}$$

1.3 Aggregation in Stochastic Infinite-horizon Models

Again we started our model development process by focusing on the dynamic optimization problem of an individual, categorizing our problem as a microeconomics issue. However, we would like aggregate over all productivity types to address macroeconomic policy issues. In order to do that, we need to need to generate a distribution over agents in the economy.

1.3.1 Generating the Stationary Distribution over the Stochastic Process

Consider first the case of time-independent draws, and assume again a two-state productivity process: $z \in \{z_l, z_h\}$. If $p(z_l) = \frac{1}{2}$ and $p(z_h) = \frac{1}{2}$, then the problem reduces to a coin-flipping exercise. In other words, we could assume that everybody in the model economy started off with low productivity. If everybody flipped the coin, approximately half would receive a high shock, and half would remain in the low state. We could then repeat the exercise in the following period. Half of all low productivity types would go to the high state, and half would remain in the low state. Similarly, half of all individuals in the high state would remain in the high state, whereas half would transition into the low state. *However, even though households are transitioning across states, the average number of low-state individuals and high-state individuals would not change!* This means that it only took one period for the entire distribution to arrive at its long-run value. In particular, the distributions $\{d_l, d_h\}$ in each period, where d_l and d_h are the probability measures over low types and high types, respectively, would have been $\{1, 0\}, \{\frac{1}{2}, \frac{1}{2}\}, \{\frac{1}{2}, \frac{1}{2}\}, \ldots$ This turns out to be a property of time-independent distributions and not of time-dependent distributions.

Suppose now that the following Markov matrix represented the productivity shock:

$$P = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}, \tag{20}$$

and suppose we tried the same experiment by starting all individuals at the low state. Now, $\frac{3}{4}$ would remain in the low state in the second period, while $\frac{1}{4}$ would transition into the high state. Consider one more iteration - the corresponding distribution over low types would be $\frac{3}{4} \times \frac{3}{4} + \frac{1}{4} \times \frac{1}{4} = \frac{5}{8}$ (i.e., low types remaining low plus high types transitioning to low), while the corresponding distribution over high types would be $1 - \frac{5}{8} = \frac{3}{8}$. To summarize, the sequence of distributions $\{d_l, d_h\}$ in the first three periods would be $\{1, 0\}, \{\frac{3}{4}, \frac{1}{4}\}$, and $\{\frac{5}{8}, \frac{3}{8}\}$. The distribution did not immediately converge to its long-run value! You might have guessed that if we kept going, the distribution would have converged to $\{\frac{1}{2}, \frac{1}{2}\}$, and indeed it would have. We call this limiting distribution the *stationary distribution*.

We can calculate the stationary distribution one of two ways. First, we can simply guess an initial probability measure over types and iterate as we did above. If we chose to do this, we would simply iterate until consecutive probability measures were sufficiently close to one another, i.e.,

$$||(d_{\{l,n\}}, d_{\{h,n\}}) - (d_{\{l,n+1\}}, d_{\{h,n+1\}})|| < \varepsilon \tag{21}$$

for some small ε. Alternatively, we could have taken the Monte Carlo approach of starting lots of individuals in a particular state, and simulated draws from a uniform distribution according to the Markov process.

1.3.2 Generating the Stationary Distribution over the State Space

Next, we need to consider the joint distribution over both the productivity shock and the endogenous variables of the state space. In this case, the distribution must be defined over both the random variable z and the agents' state variable a. To understand why we need to define the probability distribution over the state space, consider an individual who has experienced several consecutive low productivity shocks. This individual would have run their stock of savings down try to smooth consumption when times got tough. Now suppose that after a rough spell, the individual draws a high shock and begins to rebuild the stock of savings. This rebuilding happens slowly, as the agent experiences some positive and some negative shocks along the way. We learned from the previous section how to account for the distribution over productivity shocks, but how can account for the distribution over savings when individuals experience this constant ebb and flow across the asset space?

Let's begin by assuming that the productivity shock can take two possible values $z \in \{z_l, z_h\}$ and the asset grid is discretized into three possible values: $\{a_1, a_2, a_3\}$. We want to represent the distribution over productivity and assets using a matrix. Let the rows of the matrix represent the discrete points of the asset grid, and let the columns represent the states of the shock process. Then, we can fill in the initial matrix M_0 with any arbitrary set of masses $m_{\{(i,j\}}$, such that $\sum_{(i,j)} m_{\{i,j\}} = 1$:

$$M_0 = \begin{bmatrix} m_{\{1,1\}} & m_{\{1,2\}} \\ m_{\{2,1\}} & m_{\{2,2\}} \\ m_{\{3,1\}} & m_{\{3,2\}} \end{bmatrix}. \tag{22}$$

Let P be the 2×2 Markov matrix over the states of the productivity shock z denoted in (16). Now consider what happens to the mass of individuals in the entry $\{2, 1\}$ of M_0, $m_{\{2,1\}}$. These are individuals who have medium savings and a low current productivity shock. Our goal is to figure out where this mass of individuals will end up in the next period. In particular, we want to fill in the matrix M_1 knowing only the Markov process P and the policy function $a'(z_l, a)$ that solves (18) - (19). We know from (16) that percentage $p(z_l'|z_l)$ of the mass of individuals $m_{\{2,1\}}$ will remain in the low state, and percentage $1 - p(z_l'|z_l)$ will transition into the high state. That tells us what percentage of individuals will end up in each column of M_1, but how do we know which row they will go to?

The policy function tells us that every individual in the mass $m_{\{2,1\}}$ will save $a'(z_l, a_2)$. Since they are in the low productivity state, suppose that they dis-save so that $a_1 < a'(z_l, a_2) < a_2$. This creates a problem for us, since we do not have a row corresponding to such a value. Instead, we will assume that a portion of these individuals will randomly be assigned to the lower grid point a_1 and the rest will be assigned to the higher grid point a_2. We determine the measure of individuals going to the respective grid-point according to the proximity of $a'(z_l, a_2)$ to each bounding grid-point. In particular, percentage

$$\omega_2 = \frac{a'(z_l, a_2) - a_1}{a_2 - a_1} \tag{23}$$

will go to the row corresponding to grid point a_2, and the remaining percentage $1 - \omega_2$ will go to the row corresponding to grid point a_1. Finally, we can calculate the percentage of individuals that go from $m_{\{2,1\}}$ go to each element of M_1:

$$\Gamma(m_{\{2,1\}}) = \begin{bmatrix} m_{\{2,1\}}(1-\omega_2)p(z_l'|z_l) & m_{\{2,1\}}(1-\omega_2)(1-p(z_l'|z_l)) \\ m_{\{2,1\}}\omega_2 p(z_l'|z_l) & m_{\{2,1\}}\omega_2(1-p(z_l'|z_l)) \\ 0 & 0 \end{bmatrix}. \tag{24}$$

Notice that we have introduced a function Γ, which will tell us how much of each measure $m_{\{i,j\}}$ goes to each entry in the subsequent matrix.

We can repeat this process for each $\{i,j\}$ combination, and arrive at our next matrix, M_1. M_1 will share the property that all of its elements must sum to one: $\sum_{(i,j)} m_{\{i,j\}} = 1$. To keep track of masses as they correspond to their respective matrices, introduce the superscript $m^n_{\{i,j\}}$, so that:

$$M_n = \begin{bmatrix} m^n_{\{1,1\}} & m^n_{\{1,2\}} \\ m^n_{\{2,1\}} & m_{\{2,2\}} \\ m^n_{\{3,1\}} & m^n_{\{3,2\}} \end{bmatrix}. \tag{25}$$

Now that we know how to get M_{n+1} from the matrix M_n according to the *transition function* $\Gamma(m)$, we will generalize the convergence criteria from (21) to approximate the stationary distribution corresponding to the state space $\{z,a\}$:

$$\max_{i,j} |m^n_{\{i,j\}} - m^{n+1}_{\{i,j\}}| < \varepsilon. \tag{26}$$

Equation (26) says that we have successfully approximated the distribution over the state space once the maximum disparity between consecutive entries of the matrix is less than some predetermined tolerance.

1.3.3 Aggregating over the State Space

In this section, we will assume that we have successfully solved for the stationary distribution over the state space, which we denote as $M(z,a)$. This stationary distribution, along with the policy and value functions from the solution to the agent's problem give us the elements we need to calculate aggregate variables. In particular, let $\mu(z,a)$ denote the measure of individuals in state $\{z,a\}$. In practice, this measure would be an entry in the matrix corresponding to the approximated stationary distribution, but we can think of it more abstractly here. Let $a'(z,a)$ be the savings function, defined over the state space. Further, let $\mathcal{Z}$ and $\mathcal{A}$ be the domains of the productivity shock and assets. Then, we can write aggregate savings as:

$$A = \int_{z\in\mathcal{Z}} \int_{a\in\mathcal{A}} a'(z,a)\mu(z,a)\, da\, dz. \tag{27}$$

This abstract interpretation is, in practice, simply:

$$A = \sum_{z_j\in\mathcal{Z}} \sum_{a_i\in\mathcal{A}} a'(z_j,a_i) m_{\{i,j\}}, \tag{28}$$

where the mass $m_{\{i,j\}}$ corresponds to the $(i,j)^{th}$ entry in the distribution matrix M. In fact, any variable x defined over the state space $x(z,a)$, including the value function, consumption, and even the state space itself (in this case, productivity and assets) can be abstractly aggregated:

$$X = \int_{z\in\mathcal{Z}} \int_{a\in\mathcal{A}} x(z,a)\mu(z,a)\, da\, dz. \tag{29}$$

We conclude this section with two important observations. First, suppose that we arrive at a stationary distribution $M(z, a)$. It is important to note that aggregate variables will also be stationary (i.e., will not change over time), even though each individual's state is constantly changing. Economists in this field sometimes say that "idiosyncratic shocks wash out in the aggregate," which is another way of saying the aforementioned. It is also an application of the law of large numbers, which holds since we have infinitely-many agents in each mass $m_{\{i,j\}}$. Finally, note that even though we have been focused on the stationary distribution, we could have started from any initial distribution M_0, and the sequence of distributions generated by the function Γ: $\{M_t\}_{t=1}^{\infty}$ would have coincided with the actual time path of the distribution implied by the decision rule and the Markov matrix. In other words, we can use this sequence of distribution to calculate aggregates and determine how they change over time.