Chapter 1

Conditions Implied by Dynamic Optimization

1.1 Introduction

Two of the hallmarks of both macroeconomics and finance are the concern with time and with uncertainty. In the first half of this book, the focus is on time.

The key to much of economics is an understanding of choices made with an eye to the future. The consequences of such choices unfold over time, but it is the view of the future at the time of decision that governs such choices. Throughout the first half of this book, we will make the *certainty-equivalence approximation* of looking at the decisions agents would make if all uncertainty vanished and they were certain to face the expected values of future variables. Using the certainty-equivalence approximation, one proceeds as if all the agents in a model had perfect foresight.[1] As we proceed to study "perfect-foresight" models, it is important to keep sight of their purpose of providing a certainty-equivalence approximation to stochastic models.

[1] In the second half of the book, we will see how far optimal decisions under uncertainty depart from the certainty equivalence approximation. One simple generalization is that when only aggregate, economy-wide uncertainty is at issue, the certainty-equivalence approximation is typically quite a good approximation. The certainty-equivalence approximation is often not a very good approximation when the idiosyncratic risk faced by heterogeneous households and firms is at issue. (The law of averages helps to make aggregate uncertainties much smaller as a percentage of mean values than the risks faced by individual households and firms.) In any case, the certainty-equivalence approximation is the foundation on which higher-order approximations will be built.

In order to develop the intuition behind dynamic optimization, this chapter moves from a two-period model, to several related T-period models, and finally to a continuous-time model. For now, the focus is on deriving and interpreting the basic conditions that are required in order for a set of dynamic choices to be optimal.

1.2 A Two-Period Model of Production, Consumption, and Capital Accumulation

Many of the most important principles of dynamic optimization are apparent in a two-period model of production, consumption and capital accumulation, with a fixed amount of inelastically supplied labor in the background. Consider an agent facing the optimization problem

$$\max_{\substack{c_0, c_1, \\ k_0, k_1, k_2}} \quad u_0(c_0) + u_1(c_1)$$

$$\begin{aligned} \text{subject to} \qquad k_0 &= \bar{k} && (1.1) \\ k_1 - k_0 &= f(k_0) - c_0 && (1.2) \\ k_2 - k_1 &= f(k_1) - c_1 && (1.3) \\ k_2 &\geq 0. && (1.4) \end{aligned}$$

The subscripts are for the two active time periods 0 and 1 and any later periods used for accounting purposes. The letters are the traditional labels:

u : single-period utility function
f : production function (net of depreciation)
c : consumption
k : capital stock.

The utility function is additively time-separable with no explicit time discounting (though the possibility of time discounting is implicit in the distinction between u_0 and u_1). Equation (1.1), which says that the initial capital stock k_0 is fixed at $\bar{k}$ is the *initial condition.* Equation (1.4), which says that the capital stock must remain nonnegative after all the consumption is over, is the *terminal condition*. Equations (1.2) and (1.3), which say that the change in the capital stock from one period to the next is (net) output minus consumption, are the *accumulation equations*.[2]

In economic models, Lagrange multipliers represent shadow prices. Therefore, Lagrange multipliers are interesting in their own right, and it is often helpful to use them even where it is not absolutely necessary. In this problem, the Lagrange multipliers represent the *marginal values of capital* at various dates indicated by the subscripts. Using these Lagrange multipliers liberally (Λ_0, Λ_1, Λ_2 and Λ_3), the Lagrangian for this problem is:

$$\begin{aligned}\mathcal{L} &= u_0(c_0) + u_1(c_1) + \Lambda_0[\bar{k} - k_0] \\ &\quad +\Lambda_1[f(k_0) - c_0 + k_0 - k_1] \\ &\quad +\Lambda_2[f(k_1) - c_1 + k_1 - k_2] \\ &\quad +\Lambda_3 k_2\end{aligned} \tag{1.5}$$

Setting the derivatives of $\mathcal{L}$ with respect to c_0 and c_1 to zero yields two *first-order conditions*:

$$u_0'(c_0) = \Lambda_1 \tag{1.6}$$

$$u_1'(c_1) = \Lambda_2. \tag{1.7}$$

In words, for optimality, the marginal utility of consumption must equal the marginal value of capital in the next period, since each unit of consumption reduces the amount of capital at the beginning of the next period by one unit.

Setting the derivatives with respect to k_0, k_1 and k_2 equal to zero yields three *Euler equations* relating the Λ's at different dates[3]:

$$\Lambda_0 = [1 + f'(k_0)]\,\Lambda_1 \tag{1.8}$$

$$\Lambda_1 = [1 + f'(k_1)]\,\Lambda_2 \tag{1.9}$$

$$\Lambda_2 = \Lambda_3 \tag{1.10}$$

[2] [**?**] call an accumulation equation a *state equation* or a *transition equation*.

[3] [**?**] also call an Euler equation a *costate equation*.

Note from equation (1.10) that the marginal value of capital flattens out and becomes constant after the active periods are past. During those active periods, the marginal value of one unit of capital is equal to the value of its marginal product of $1 + f'(k)$ units of capital for the next period.

Finally, the Kuhn-Tucker conditions for (1.4) ($k_2 \geq 0$) yield the following additional two relations:

$$\Lambda_3 k_2 = 0 \tag{1.11}$$
$$\Lambda_3 \geq 0. \tag{1.12}$$

Equation (1.11) is called the *transversality condition.* In conjunction with the nonnegativity condition (1.12), the *transversality condition* says that either k_2 or Λ_3 must be equal to zero. In particular, the optimal value of the final capital stock k_2 can be strictly positive only if the final marginal value of capital Λ_3 is zero. The transversality condition guards the agent against sacrificing consumption for the sake of excess capital at the end.

In the T-period and continuous-time models to follow, the same types of equations and other relations recur. The statement of each dynamic optimization problem involves an utility function or other objective function, an initial condition, a terminal condition, and a set of accumulation equations indexed by time. Optimality for each problem requires a set of first-order conditions, a set of Euler equations, and a transversality condition.[4] Pay attention to these similarities as we examine the various models.

1.3 A T-Period Model of Production, Consumption, and Capital Accumulation

The T-period version of the two-period model of the previous section gives rise to the following optimization problem:

$$\max_{c,k} \sum_{t=0}^{T-1} u_t(c_t)$$

subject to
$$k_0 = \bar{k} \tag{1.13}$$
$$k_{t+1} - k_t = f(k_t) - c_t \qquad (t = 0, \ldots, T-1) \tag{1.14}$$
$$k_T \geq 0. \tag{1.15}$$

[4] There is also often a nonnegativity condition of some kind for Λ.

The Lagrangian for the T-period problem is

$$
\begin{aligned}
\mathcal{L} \quad = \quad & u_0(c_0) + u_1(c_1) + u_2(c_2) + \ldots + u_{T-1}(c_{T-1}) \\
& +\Lambda_0[\bar{k} - k_0] \\
& +\Lambda_1[f(k_0) - c_0 + k_0 - k_1] \\
& +\Lambda_2[f(k_1) - c_1 + k_1 - k_2] \\
& + \\
& \vdots \\
& +\Lambda_T[f(k_{T-1}) - c_{T-1} + k_{T-1} - k_T] \\
& +\Lambda_{T+1} k_T.
\end{aligned} \tag{1.16}
$$

The *first-order conditions* obtained by setting the derivative of $\mathcal{L}$ with respect to c_t equal to zero for $t = 0, \ldots, T-1$, are

$$u_t'(c_t) = \Lambda_{t+1} \qquad (t = 0, \ldots, T-1) \tag{1.17}$$

The *Euler equations* obtained by setting the derivative of $\mathcal{L}$ with respect to k_t equal to zero for $t = 0, \ldots, T-1$ are

$$\Lambda_t = [1 + f'(k_t)]\, \Lambda_{t+1} \qquad (t = 0, \ldots, T-1). \tag{1.18}$$

In addition, setting the derivative with respect to k_T equal to zero,

$$\Lambda_T = \Lambda_{T+1}. \tag{1.19}$$

Finally, the *transversality condition* and associated nonnegativity condition for Λ_{T+1} obtained from the Kuhn-Tucker conditions for (1.15) are

$$\Lambda_{T+1} k_T \quad = \quad 0 \tag{1.20}$$

$$\Lambda_{T+1} \quad \geq \quad 0. \tag{1.21}$$

As in the two-period model, the first-order conditions say that the marginal utility of consumption in each period is equal to the marginal value of capital anticipated in the next period. Thinking of the net marginal product of capital as the real interest rate, the Euler equations say that the ratio between the marginal value of capital now and in the next period is equal to one plus the real interest rate. The transversality condition says that the value of all capital left over, evaluated at its marginal value, must be zero.

1.4 The T-Period Model Viewed from Another Angle

What happens if we write the T-period model of production, consumption, and capital accumulation in a different, but equivalent form? Viewing the same model from various different angles can often give additional insights and strengthen one's intuition. Checking the consistency of the various solutions to the model stated in different ways can suggest important questions where there is a gap in one's understanding as well as helping one to detect simple, but potentially troublesome mistakes. Here, among other things, the alternative formulation allows us to see what happens when the capital stock is in the objective function. Consider the same T-period economic problem with (net) investment (i) taking center stage instead of consumption:

$$\max_{i,k} \sum_{t=0}^{T-1} u_t(f(k_t) - i_t)$$

$$\begin{aligned} \text{subject to} \qquad k_0 &= \bar{k} && && (1.22)\\ k_{t+1} - k_t &= i_t && (t = 0, \ldots, T-1) && (1.23)\\ k_T &\geq 0. && && (1.24) \end{aligned}$$

The Lagrangian is

$$\begin{aligned} \mathcal{L} = \; & u_0(f(k_0) - i_0) + u_1(f(k_1) - i_1) + \ldots + u_{T-1}(f(k_{T-1} - i_{T-1}) \\ & +\Lambda_0[\bar{k} - k_0] \\ & +\Lambda_1[i_0 + k_0 - k_1] \\ & +\Lambda_2[i_1 + k_1 - k_2] \qquad (1.25) \\ & + \\ & \vdots \\ & +\Lambda_T[i_{T-1} + k_{T-1} - k_T] \\ & +\Lambda_{T+1}k_T. \end{aligned}$$

The *first-order conditions* obtained by setting the derivative of $\mathcal{L}$ with respect to i_t equal to zero for $t = 0, \ldots, T-1$, are

$$u_t'(f(k_t) - i_t) = \Lambda_{t+1} \qquad (t = 0, \ldots, T-1) \qquad (1.26)$$

Since $f(k_t) - i_t = c_t$, this is equivalent to the first-order conditions (1.17) above.

The *Euler equations* obtained by setting the derivative of $\mathcal{L}$ with respect to k_t equal to zero for $t = 0, \ldots, T-1$ are

$$\Lambda_t = \Lambda_{t+1} + f'(k_t)u'(f(k_t) - i_t) \qquad (t = 0, \ldots, T-1). \tag{1.27}$$

Using (1.26) to substitute in Λ_{t+1} for $u'(f(k_t) - i_t)$ yields

$$\Lambda_t = [1 + f'(k_t]\Lambda_{t+1},$$

which is identical to (1.18). The remaining equations are identical to (1.19), (1.20), and (1.21).

1.5 A General T-Period Model of Dynamic Optimization

With the preceding models as background, it is time to look at a general T-period model that will yield general results that apply to a wide variety of more specific models. Consider the general problem

$$\max_{x,k} \sum_{t=0}^{T-1} U(k_t, x_t, t)$$

$$\begin{aligned}
\text{subject to} \qquad k_0 &= \bar{k} && && (1.28)\\
k_{t+1} - k_t &= A(k_t, x_t, t) && (t = 0, \ldots, T-1) && (1.29)\\
k_T &\geq 0. && && (1.30)
\end{aligned}$$

U is a general objective function. A is a general accumulation function. The dependence of U and A on time is shown within the parentheses rather than as a subscript in order to save subscripts on U and A for partial derivatives. The correspondence with the preceding sections is as follows:

Section:	(1.3)	(1.4)
$U(k_t, x_t, t)$:	$u_t(x_t)$	$u_t(f_t(k_t) - x_t)$
$A(k_t, x_t, t)$:	$f_t(k_t) - x_t$	x_t
x_t :	c_t	i_t

In the second and third columns, we make one change in the models above—making the production function f dependent on time, as it will be when technology is gradually evolving—to indicate one reason there might be a direct dependence of U or A on t. Many other models also fit within this general form, as the exercises at the end of the chapter illustrate.

The Lagrangian for the general T-period model is

$$\begin{aligned}\mathcal{L} \;=\; & U(k_0, x_0, 0) + U(k_1, x_1, 1) + \ldots + U(k_{T-1}, x_{T-1}, T-1) \\ & +\Lambda_0[\bar{k} - k_0] \\ & +\Lambda_1[A(k_0, x_0, 0) + k_0 - k_1] \\ & +\Lambda_2[A(k_1, x_1, 1) + k_1 - k_2] \\ & + \\ & \vdots \\ & +\Lambda_T[A(k_{T-1}, x_{T-1}, T-1) + k_{T-1} - k_T] \\ & +\Lambda_{T+1}k_T.\end{aligned} \tag{1.31}$$

The *first-order conditions* obtained by setting the derivative of $\mathcal{L}$ with respect to x_t equal to zero for $t = 0, \ldots, T-1$, are

$$U_x(k_t, x_t, t) + \Lambda_{t+1}A_x(k_t, x_t, t) = 0 \qquad (t = 0, \ldots, T-1) \tag{1.32}$$

The *Euler equations* obtained by setting the derivative of $\mathcal{L}$ with respect to k_t equal to zero for $t = 0, \ldots, T-1$ are

$$\Lambda_{t+1} - \Lambda_t = -[U_k(k_t, x_t, t) + \Lambda_{t+1}A_k(k_t, x_t, t)] \qquad (t = 0, \ldots, T-1). \tag{1.33}$$

Setting the derivative with respect to k_T equal to zero,

$$\Lambda_T = \Lambda_{T+1}. \tag{1.34}$$

Finally, as above, the *transversality condition* and associated nonnegativity condition for Λ_{T+1} obtained from the Kuhn-Tucker conditions for the constraint $k_T \geq 0$ are

$$\Lambda_{T+1}k_T = 0 \tag{1.35}$$

$$\Lambda_{T+1} \geq 0. \tag{1.36}$$

In view of (1.34), the transversality condition and associated nonnegativity condition can also be written

$$\Lambda_T k_T = 0 \tag{1.37}$$
$$\Lambda_T \geq 0. \tag{1.38}$$

1.6 A General Continuous-Time Model of Dynamic Optimization

The continuous-time equations for dynamic optimization can be obtained straightforwardly by taking a limit as the length of a period goes to zero. To see this, consider the following problem with a period of length $h = T/n$ for some integer n:

$$\max_{x,k} \sum_{t=0,h,2h,\ldots,T-h} hU(k_t, x_t, t)$$

$$\text{subject to} \quad k_0 = \bar{k} \tag{1.39}$$
$$k_{t+h} - k_t = hA(k_t, x_t, t) \qquad (t = 0, \ldots, T-h) \tag{1.40}$$
$$k_T \geq 0. \tag{1.41}$$

Following the same pattern as in the previous model shows that the necessary conditions for dynamic optimization are almost identical to those in the previous section, since the problems are essentially the same except for a renumbering of the time periods and the factor h multiplying the objective function and accumulation function. In place of (1.32), (1.33), (1.37) and (1.38), one obtains the following four equations:

$$h[U_x(k_t, x_t, t) + \Lambda_{t+h} A_x(k_t, x_t, t)] = 0 \qquad (t = 0, h, \ldots, T-h) \tag{1.42}$$

$$\Lambda_{t+h} - \Lambda_t = -h[U_k(k_t, x_t, t) + \Lambda_{t+h} A_k(k_t, x_t, t)] \qquad (t = 0, h, \ldots, T-h). \tag{1.43}$$

$$\Lambda_T k_T = 0 \tag{1.44}$$
$$\Lambda_T \geq 0. \tag{1.45}$$

As $h \to 0$, the Euler equation (1.43) guarantees that Λ_t will be continuously differentiable with respect to time. Thus, $\Lambda_{t+h} \to \Lambda_t$ as $h \to 0$. Dividing (1.42) and (1.43) by h and taking the limit as $h \to 0$, the continuous-time *first-order condition* is

$$U_x(k_t, x_t, t) + \Lambda_t A_x(k_t, x_t, t) = 0 \qquad (t \in [0, T)\,). \tag{1.46}$$

Using a dot for time-derivatives (*e.g.*, $\dot{\Lambda}_t = \frac{d\Lambda_t}{dt}$), the continuous-time *Euler equation* is

$$\dot{\Lambda}_t = -[U_k(k_t, x_t, t) + \Lambda_t A_k(k_t, x_t, t)] \qquad (t \in [0, T)\,). \tag{1.47}$$

1.7 Summary of the Continuous-Time Results

As $h \to 0$, the continuous-time model can be written (omitting t-subscripts) as

$$\max_x \int_0^T U(k, x, t)\, dt$$

$$\text{subject to} \qquad k_0 = \bar{k} \tag{1.48}$$

$$\dot{k} = A(k, x, t) \tag{1.49}$$

$$k_T \geq 0. \tag{1.50}$$

As is traditional, the maximization is here represented as being over only x alone because the accumulation equation (1.49) implies that the path of k_t is determined by the path of x_t. Following this line of thinking, in the dynamic optimization literature, variables like k that have rates of change constrained by differential equations like (1.49) are called *state variables*; while variables like x that the agent *could* vary discontinuously from one moment to the next at will (though it may not be optimal to do so) are called *control variables.* Multipliers like Λ are called *costate variables.* This is useful terminology.[5]

Omitting arguments as well as t subscripts for clarity, the four key results above that are implied by optimization are the *first-order condition*

[5]Even though k is not called a control variable, it is at least indirectly controllable. It is good to remember that the Euler equations above arise from considering changes in the path of k in conformance with the accumulation equation.

$$U_x + \Lambda A_x = 0, \tag{1.51}$$

the *Euler equation*

$$\dot{\Lambda} = -[U_k + \Lambda A_k], \tag{1.52}$$

the *transversality condition*

$$\Lambda_T k_T = 0 \tag{1.53}$$

and the nonnegativity condition for Λ_T

$$\Lambda_T \geq 0. \tag{1.54}$$

1.7.1 The Hamiltonian

A good way to remember the two most important necessary conditions for dynamic optimization—the first order condition and the Euler equation—is by the mnemonic device of the *Hamiltonian.* The Hamiltonian is analogous to, but not identical to the Lagrangian. The Hamiltonian $\mathcal{H}$ is defined as

$$\mathcal{H} = U(k, x, t) + \Lambda A(k, x, t). \tag{1.55}$$

In terms of the Hamiltonian $\mathcal{H}$, the first-order condition (1.51) can be recast as

$$\mathcal{H}_x = 0. \tag{1.56}$$

The Euler equation can be recast as

$$\dot{\Lambda} = -\mathcal{H}_k. \tag{1.57}$$

Interestingly enough, if one is willing to take a partial derivative with respect to Λ, even the accumulation equation (1.49) can be recast in terms of the Hamiltonian as

$$\dot{k} = \mathcal{H}_\Lambda. \tag{1.58}$$

The most important fact about the Hamiltonian is that a number of necessary conditions for dynamic optimization can be summarized in the single statement that for each k and t on an optimal path, the Hamiltonian must be maximized over x. In addition to the first order condition (1.56) implied by

$\mathcal{H}$ being maximized over x at each point in time, optimization requires, for example, that

$$\mathcal{H}_{xx} \leq 0. \tag{1.59}$$

We leave the proof of (1.59) to an exercise. A general proof that the Hamiltonian must be maximized at each point in time (the *Pontryagin maximum principle*) can be found in [**?**].

1.8 The Maximum Principle

The Pontryagin Maximum Principle states that given the value of the costate variable Λ at each point in time, the Hamiltonian must necessarily be maximized *globally* over all possible values of the control variables. It applies in the strict sense only in continuous time, though it will be approximately true in discrete time if the time intervals are short.

We will not prove it here, but it is not hard to explain the essence of the proof of the Maximum Principle. As far as the overall goal is concerned, only two things matter about what is being done this instant: (1) the value of the objective function this instant and (2) the accumulation of the state variable that occurs during this instant: $\mathcal{H} = U + \Lambda A$. Since the amount of accumulation that can occur in an instant is infinitesimal, there is no problem in valuing that accumulation at a fixed marginal value measured by Λ even if following an alternative policy for an instant involves a large change in the *rate* of accumulation for that instant. In particular, suppose the Hamiltonian was not at a global maximum at some point when following a proposed policy. Modifying that policy by switching to the global maximum for an instant at that point would improve the sum of the value of the objective function at that instant and the value of the accumulation at that instant.

In discrete time, switching to the global maximum of the Hamiltonian from some other local maximum has a noticeable effect on the amount of accumulated in the period, so the costate variable Λ that appears in the definition of the Hamiltonian may be noticeably altered, so that the global maximum given the original value of Λ might not be the global maximum given the new value of Λ. However, in discrete time the Hamiltonian still must be locally optimized, since any local change in the control variables causes only an infinitesimal change in the amount of accumulation. As the length of the period gets shorter, it becomes impossible for the local maximum achieved by the Hamiltonian in the optimal program to be very far below the global maximum for the Hamiltonian at that point.

Chapter 2

Conditions Sufficient for an Optimum

2.1 Introduction: Sufficient Conditions for Static Optimization

In the straightforward calculus problem

$$\max_x f(x)$$

with differentiable f, maximization implies as a minimal requirement the necessary condition $f'(x) = 0$. All of the necessary conditions for dynamic optimization discussed in the previous chapter are akin to this "flat top" condition. But finding where the derivative equals zero is not sufficient to guarantee that one has found a global maximum. Even if the second-order condition is satisfied strictly ($f''(x) < 0$) it guarantees only a local maximum—that one has found a hilltop, but not necessarily the *highest* hilltop. The simplest additional condition sufficient to guarantee that if $f'(x) = 0$, one has found a global maximum is for f to be globally concave.

A function f is globally concave if and only if its graph lies beneath all of its tangent lines (or all of its tangent hyperplanes).[1] Symbolically, if f is

[1] **if:** If the graph of f lies beneath all of its tangent lines, then the tangent line through any point on the curve—and therefore the point itself—must be above the secant line segment connecting any two points on either side of the first point. Points on the curve being above all relevant secant line segments is the primary definition of concavity.

only if: If f is concave, the separating hyperplane theorem guarantees that the graph of f lies below its tangent lines.

concave,

$$f(x) \leq f(x^*) + f'(x^*)[x - x^*], \tag{2.1}$$

where there is a tangency at x^* and x represents any other point on the curve. If x^* satisfies the first-order condition $f'(x^*) = 0$ with f concave, the inequality (2.1) simplifies to

$$f(x) \leq f(x^*)$$

—guaranteeing that $(x^*, f(x^*))$ is a global maximum.

2.2 Relating Dynamic Optimization to Static Optimization

Sufficient conditions for dynamic optimization are closely related to sufficient conditions for static optimization. Indeed, one of the most instructive sufficient conditions for dynamic optimization is static optimization at each instant of the *classic Lagrangian* $L(k, x, t)$ defined by

$$L(k, x, t) = H(k, x, t) + \dot{\Lambda}_t k = U(k, x, t) + \Lambda_t A(k, x, t) + \dot{\Lambda}_t k \tag{2.2}$$

over k and x.

More precisely, consider the problem

$$\max_{x,k} \int_0^T U(k, x, t)\, dt$$

$$\begin{aligned} \text{subject to} \qquad k_0 &= \bar{k} && (2.3) \\ \dot{k} &= A(k, x, t) && (2.4) \\ k_T &\geq 0. && (2.5) \end{aligned}$$

A program is *feasible* only if it satisfies (2.3), (2.4) and (2.5). If (k_t^*, x_t^*) is feasible, and there is any function of time Λ_t for which the maximum of

$$L(k, x, t) = U(k, x, t) + \Lambda_t A(k, x, t) + \dot{\Lambda}_t k$$

over k and x at each t is attained by (k_t^*, x_t^*), *and* for which

$$\Lambda_T k_T^* = 0 \quad (2.6)$$

$$\Lambda_T \geq 0, \quad (2.7)$$

then (k_t^*, x_t^*) achieves a global maximum for the dynamic optimization problem over all feasible programs.

Moreover, this result holds even if one modifies the dynamic optimization problem by adding the constraint $x \in X_t$ at each time t where X_t is a time-varying constraint set.

2.2.1 Proof

Integrating by parts,

$$\int_0^T \dot{\Lambda}_t k_t \, dt = \Lambda_T k_T - \Lambda_0 k_0 - \int_0^T \Lambda_t \dot{k}_t \, dt. \quad (2.8)$$

Therefore, for any function of time Λ_t, (2.4) and integration by parts together yield the identity

$$\begin{aligned} \int_0^T U(k_t, x_t, t) \, dt &= \int_0^T \{U(k_t, x_t, t) + \Lambda_t[A(k_t, x_t, t) - \dot{k}_t]\} \, dt \\ &= \int_0^T [U(k_t, x_t, t) + \Lambda_t A(k_t, x_t, t) + \dot{\Lambda}_t k_t] \, dt \\ &\quad - \Lambda_T k_T + \Lambda_0 k_0 \\ &= \int_0^T L(k_t, x_t, t) \, dt - \Lambda_T k_T + \Lambda_0 k_0. \end{aligned} \quad (2.9)$$

If the program (k_t^*, x_t^*) maximizes $L(k, x, t)$ for each t, then for any other feasible program (k_t, x_t),

$$\int_0^T L(k_t, x_t, t) \, dt \leq \int_0^T L(k_t^*, x_t^*, t) \, dt. \quad (2.10)$$

Both programs being feasible implies that

$$\Lambda_0 k_0 = \Lambda_0 \bar{k} = \Lambda_0 k_0^*, \quad (2.11)$$

Finally, by (2.7), (2.5) and the transversality condition (2.6) for the program (k_t^*, x_t^*),

$$-\Lambda_T k_T \leq 0 = -\Lambda_T k_T^*. \tag{2.12}$$

Therefore,

$$\begin{aligned}\int_0^T U(k_t, x_t, t)\, dt &= \int_0^T L(k_t, x_t, t)\, dt - \Lambda_T k_T + \Lambda_0 k_0 \qquad (2.13)\\ &\leq \int_0^T L(k_t^*, x_t^*, t)\, dt - \Lambda_T k_T^* + \Lambda_0 k_0^* \\ &= \int_0^T U(k_t^*, x_t^*, t)\, dt.\end{aligned}$$

In words, the program (k_t^*, x_t^*) achieves an equal or higher value of the objective as compared to any other path.

There is no need to modify the proof when there are constraints on x of the form $x_t \in X_t$ for a time-varying set X_t, except to read "global maximum over x and k" and "feasible" in this constrained sense.

2.3 Sufficient Conditions for Dynamic Optimization

For practical purposes, one would like to have simple conditions guaranteeing that when one finds a program that satisfies the necessary conditions for dynamic optimization, it will be a global optimum. The result of the previous section reduces this to the problem of guaranteeing that the program (k_t^*, x_t^*) maximizes $L(k, x, t)$ for each t, since the transversality condition (2.6) and associated nonnegativity condition (2.7) are necessary conditions for an dynamic optimization in any case.

2.3.1 Concavity of U and A in k and x (Mangasarian's Theorem)

Perhaps the simplest condition to add to the necessary conditions to make them sufficient, is making sure that U and A are each jointly concave in k and x, and checking that Λ_t is positive on the entire interval $[0, T]$. Then

$$L(k, x, t) = U(k, x, t) + \Lambda_t A(k, x, t)$$

must also be concave. Beyond the transversality condition (2.6) and associated nonnegativity condition (2.7), the two necessary conditions for dynamic optimization can be expressed as

$$L_k(k_t^*, x_t^*, t) = U_k(k_t^*, x_t^*, t) + \Lambda_t A_k(k_t^*, x_t^*, t) + \dot{\Lambda}_t = 0. \quad (2.14)$$
$$L_x(k_t^*, x_t^*, t) = U_x(k_t^*, x_t^*, t) + \Lambda_t A_x(k_t^*, x_t^*, t) = 0, \quad (2.15)$$

By the multivariate extension of (2.1), coupled with (2.14) and (2.15), if L is concave in k and x, then

$$\begin{aligned} L(k_t, x_t, t) &\leq L(k_t^*, x_t^*, t) + L_k(k_t^*, x_t^*, t)[k_t - k_t^*] + L_x(k_t^*, x_t^*, t)[x_t - x_t^*] \\ &= L(k_t^*, x_t^*, t). \end{aligned} \quad (2.16)$$

In words, if L is concave, then it must lie under its tangent planes. In particular, at any point where the tangent plane is horizontal in all directions, being below the tangent plane means being lower, period. Thus, where the tangent plane is horizontal, L must be at a maximum since it has to be lower everywhere else.

Since the term $\dot{\Lambda}_t x$ is linear in x, concavity of $L(k, x, t)$ in k and x is equivalent to concavity of $\mathcal{H}(k, x, \Lambda, t)$ in k and x.

Concavity of U—that is, increasing difficulty of raising the value of the objective function as one goes further in any one direction—often follows naturally from an assumption of decreasing marginal utility or a combination of decreasing marginal utility and decreasing returns of some type. Concavity of A—that is, increasing difficulty of raising the rate of capital accumulation as one goes further in any one direction—often follows naturally from some form of decreasing returns. In the examples of the previous chapter, concavity of u and f is enough to induce concavity (in the relevant variables for each case) of $u(c)$, $u(f(k) - i)$, $f(k) - c$ and i.

Exercises

1. Proof of sufficient conditions for minimization.

2. Proof of sufficient conditions for the general T-period model.

3. Show that the proof of the first Proposition works even when the constraint set X_t depends on k_t—that is, when x is constrained by $x_t \in X_t(k_t)$.

4. Arrow's Theorem: Define the function

$$L^{\star}(k,t) = \max_{x} L(k,x,t).$$

   - Show that if $L^{\star}$ is concave in k, then the necessary conditions for dynamic optimization are also sufficient.
   - Show that if U and A are each jointly concave in k and x and Λ_t is positive, then $L^{\star}$ will be concave in x.

5. Show that if $L(k,x,t) = g(Q(k,x,t))$ where Q is concave jointly in k and x and g is monotonically increasing, then the necessary conditions are sufficient.

Chapter 3

Discounting and an Infinite Horizon

3.1 Introduction

Assuming a preference for earlier rather than later consumption ("impatience") has a long history going back at least to Irving Fisher. At the individual level, the evidence for such a preference is equivocal at best,[1] but in a dynastic setting, it corresponds to the not-implausible assumption that people care about the average welfare of all of their children more than they care about their own welfare.[2]

Beyond a belief in impatience or imperfect altruism, there are important technical reasons for the interest economists have had in models with utility discounting. Dynamic control theory predicts that, with a finite horizon, as the end-time T approaches, an agent's behavior will be strongly influenced by that approaching end. In a saving and investment problem, this is reflected in increases in consumption toward the end as the end when the capital stock will become worthless approaches. It is often desirable to keep one's analysis from being dominated by such "endgame" considerations by letting the end-time T go to infinity. An infinite horizon does not typically present serious technical

[1] When asked to imagine themselves in a hypothetical situation, slightly more than half of a sample of older people in the Health and Retirement Survey and slightly less than half of a sample of undergraduates choose an upward-sloping consumption profile even in the face of a zero real interest rate. Other evidence also calls into question whether people in general prefer earlier to later consumption. See [?].

[2] How such a preference for one's own welfare over the average welfare of all one's children can persist in the fact of sociobiological pressures is unclear.

problems as long as the integral of the objective function $\int_0^\infty \mathcal{U}(k_t, x_t, t)\, dt$ converges. Convergence of this integral requires $\mathcal{U}$ to eventually decrease as t gets large. Ordinarily, the long-run decline in $\mathcal{U}$ results from a negative direct dependence on t in the long run—utility discounting.

3.1.1 Additive Time-Separability, Stationarity and Time Consistency

Economists in general, and macroeconomists in particular, have begun to take their utility functions for granted. When a particular form has become traditional, it is useful to step back and ask which aspects of the traditional form are solidly based and which aspects are arbitrary and dispensable. The familiar additively time-separable, exponentially discounted utility—utility of the form

$$\int_0^\infty e^{-\rho t} u(c_t)\, dt$$

—is actually less arbitrary than it might appear.

Let us examine some of the bases for choosing this form of utility function. To begin with, additive time-separability is an implication of ordinary separability holding between every two subsets of time periods whenever there are at least three time periods.[3]

Given additive time-separability, exponential discounting is the only form of discounting that is both stationary and allows time consistency.[4] Stationarity means that the preferences for consumption after some date in relation to that starting date does not depend on the starting date. In other words, a utility function is stationary if preferences between earlier and later consumption are the same when viewing the year 2011 from the year 2010 as when viewing the year 2001 from the year 2000. Time-consistency means that preferences between earlier and later consumption look the same in advance as they do when one gets there. In other words, a utility function is time-consistent if preferences between earlier and later consumption are the same when viewing the comparison of 2010 and 2011 from the vantage point of the year 2000 as when viewing the comparison of 2010 and 2011 from the nearer vantage point of 2010 itself. Much more can be said about these issues, but the only point here is to motivate an interest in the traditional exponential discounting.[5]

[3] See the first exercise for an idea why.

[4] See [?].

[5] Time-consistency without stationarity is possible if there is a time-varying utility dis-

3.2 Discrete Time

3.2.1 Discounting

It is not difficult to adapt the basic necessary conditions to a situation with explicit discounting. In the discrete-time model, let

$$\mathcal{U}(k, x, t) = \beta^t U(k, x, t). \tag{3.1}$$

so that the overall objective is given by

$$\sum_{t=0}^{T-1} \beta^t U(k_t, x_t, t).$$

In words, $\mathcal{U}(k, x, t)$ is the objective function expressed in a present value as viewed from time zero, while $U(k, x, t)$ is the objective function expressed as a current value as of time t. (Much, perhaps all, of the dependence of $\mathcal{U}(k, x, t)$ on time is in the β^t term multiplying $U(k, x, t)$, but we have allowed for the possibility of some other dependence on time that is better modeled separately from this term.)

In effect, $\mathcal{U}$ measures utility time-zero utils. U measures utility in current utils. In order to be consistent, the values of the Lagrange multipliers that give the marginal value of capital should be expressed in current utils (rather than time-zero utils) as well. That is, they should be expressed as current values as of time t rather than as present values as viewed from time zero. To do this, define the current value marginal value of capital, λ_t, by

$$\Lambda_t = \beta^t \lambda_t, \tag{3.2}$$

or equivalently

$$\lambda_t = \beta^{-t} \Lambda_t. \tag{3.3}$$

count rate attached to calendar dates. Stationarity without time-consistency can arise from the kind of hyperbolic discounting studied most recently by David Laibson, which makes the discount rate in the near future higher than the discount rate in the far future, with the meaning of "near future" and "far future" changing as the calendar location of "now" changes. This alteration of the discount rate between two calendar dates as "now" moves along is what gives rise to time-inconsistency rather than a non-constant discount rate per se, which can arise from nonstationarity. Note furthermore that one gets the combination of stationarity and time-consistency even when ρ is negative; the only problem is that integral is more likely to fail to converge.

Substituting the definitions (3.1) and (3.2) into (**??**) yields the first-order condition

$$\beta^t U_x(k_t, x_t, t) + \beta^{t+1}\lambda_{t+1} A_x(k_t, x_t, t) = 0, \tag{3.4}$$

or equivalently,

$$U_x(k_t, x_t, t) + \beta\lambda_{t+1} A_x(k_t, x_t, t) = 0. \tag{3.5}$$

Substituting the above definitions into (**??**) yields the Euler equation

$$\beta^{t+1}\lambda_{t+1} - \beta^t\lambda_t = -[\beta^t U_k(k_t, x_t, t) + \beta^{t+1}\lambda_{t+1} A_k(k_t, x_t, t)], \tag{3.6}$$

or equivalently,

$$\beta\lambda_{t+1} - \lambda_t = -[U_k(k_t, x_t, t) + \beta\lambda_{t+1} A_k(k_t, x_t, t)]. \tag{3.7}$$

Solving for λ_t yields a form that may look more familiar to those accustomed to working with discrete-time models:

$$\lambda_t = U_k(k_t, x_t, t) + \beta(1 + A_k(k_t, x_t, t))\lambda_{t+1}. \tag{3.8}$$

The transversality condition becomes

$$\beta^T \lambda_T k_T = 0 \tag{3.9}$$

which, when T is finite, is equivalent to

$$\lambda_T k_T = 0. \tag{3.10}$$

Finally, the associated nonnegativity condition still boils down to

$$\lambda_T \geq 0. \tag{3.11}$$

3.2.2 An Infinite Horizon

The only equation affected in form by letting T go to infinity is the transversality condition. The straightforward infinite counterpart to (3.9) is

$$\lim_{t\to\infty} \beta^t \lambda_t k_t = 0. \tag{3.12}$$

for practical purposes, this almost always suffices as a transversality condition. More accurately, the transversality condition should be stated in terms of the size of a feasible differential reduction in k, $-\Delta k$. For any feasible differential reduction in k, dynamic optimization requires

$$\lim_{t\to\infty} \beta^t \lambda_t[-\Delta k_t] = 0. \tag{3.13}$$

If, as is typically the case, the size of a feasible differential reduction in k grows in proportion to k itself as $t \to \infty$, then (3.12) is exactly right. See [**?**] for more details on the transversality condition in an infinite horizon setting.

3.3 Continuous Time

3.3.1 Discounting

One could readily derive results for the continuous-time model with discounting by taking a limit of the discrete-time model as the length of a time period goes to zero,[6] but it is easier to start with the continuous-time necessary conditions and add discounting.

In the continuous-time model, let

$$\mathcal{U}(k, x, t) = e^{-\rho t} U(k, x, t), \tag{3.14}$$

so that the overall objective is given by

$$\int_0^T e^{-\rho t} U(k_t, x_t, t)\, dt.$$

As in the discrete-time case, $\mathcal{U}(k, x, t)$ is the objective function expressed in a present value as viewed from time zero, while $U(k, x, t)$ is the objective function expressed as a current value as of time t. (Much, perhaps all, of the dependence of $\mathcal{U}(k, x, t)$ on time is in the $e^{-\rho t}$ term multiplying $U(k, x, t)$, but we have allowed for the possibility of some other dependence on time that is better modeled separately from this term.)

In order to be consistent and have a form of the costate variable that expresses the marginal value of capital as a current value as of time t (in current utils) rather than as a present value viewed from time zero (in time-zero utils), let us also define the current value marginal value of capital, λ_t, by

[6]See exercise 2.

$$\Lambda_t = e^{-\rho t}\lambda_t, \tag{3.15}$$

or equivalently

$$\lambda_t = e^{\rho t}\Lambda_t. \tag{3.16}$$

Substituting the definitions (3.14) and (3.15) into (**??**) yields the first-order condition

$$e^{-\rho t}U_x(k_t, x_t, t) + e^{-\rho t}\lambda_t A_x(k_t, x_t, t) = 0 \tag{3.17}$$

or equivalently,

$$U_x(k_t, x_t, t) + \lambda_t A_x(k_t, x_t, t) = 0. \tag{3.18}$$

Substituting the above definitions into (**??**) yields the Euler equation

$$e^{-\rho t}\dot{\lambda} - \rho e^{-\rho t}\lambda = -[e^{-\rho t}U_k(k_t, x_t, t) + e^{-\rho t}\lambda_t A_k(k_t, x_t, t)] \tag{3.19}$$

or equivalently,

$$\dot{\lambda}_t = \rho\lambda_t - [U_k(k_t, x_t, t) + \lambda_t A_k(k_t, x_t, t)]. \tag{3.20}$$

Alternatively,

$$\dot{\lambda}_t = (\rho - A_k(k_t, x_t, t))\lambda_t - U_k(k_t, x_t, t). \tag{3.21}$$

The transversality condition becomes

$$e^{-\rho T}\lambda_T k_T = 0 \tag{3.22}$$

which, when T is finite, is equivalent to

$$\lambda_T k_T = 0. \tag{3.23}$$

Finally, the associated nonnegativity condition comes to

$$\lambda_T \geq 0. \tag{3.24}$$

3.3.2 An Infinite Horizon

Again, the only equation affected in form by letting T go to infinity is the transversality condition. The straightforward infinite counterpart to (3.22) is

$$\lim_{t\to\infty} e^{-\rho t}\lambda_t k_t = 0. \tag{3.25}$$

for practical purposes, this almost always suffices as a transversality condition. More accurately, the transversality condition should be stated in terms of the size of a feasible differential reduction in k, $-\Delta k$. For any feasible differential reduction in k, dynamic optimization requires

$$\lim_{t\to\infty} e^{-\rho t}\lambda_t[-\Delta k_t] = 0. \tag{3.26}$$

If, as is typically the case, the size of a feasible differential reduction in k grows in proportion to k itself as $t \to \infty$, then (3.25) is exactly right.[7]

3.4 Summary of Results

3.4.1 Discrete Time

With $\mathcal{U} = \beta^t U$ and $\Lambda_t = \beta^t \lambda_t$ the *first order condition* is

$$U_x + \beta\lambda_{t+1}A_x = 0. \tag{3.27}$$

(Here and below, all of the arguments of U and A are at time t.) The Euler equation is

$$\lambda_t = U_k + \beta[1 + A_k]\lambda_{t+1}. \tag{3.28}$$

The transversality condition, in a form that remains essentially valid if $T \to \infty$, is

$$\beta^T \lambda_T k_T = 0. \tag{3.29}$$

3.4.2 Continuous Time

Omitting the time t subscripts, with $\mathcal{U} = e^{-\rho t}U$ and $\Lambda = e^{-\rho t}\lambda$ the *first order condition* is

[7] See [?] for more details on the transversality condition in an infinite horizon setting.

$$U_x + \lambda A_x = 0. \tag{3.30}$$

The Euler equation is

$$\dot{\lambda} = [\rho - A_k]\lambda - U_k. \tag{3.31}$$

The transversality condition, in a form that remains essentially valid if $T \to \infty$, is

$$e^{-\rho T}\lambda_T k_T = 0. \tag{3.32}$$

3.4.3 The Current Value Hamiltonian

In continuous time, define the current-value Hamiltonian H by

$$H(k, x, \lambda, t) = U(k, x, t) + \lambda A(k, x, t). \tag{3.33}$$

Then the *first-order condition* can be written

$$H_x = 0, \tag{3.34}$$

while the *Euler equation* is

$$\dot{\lambda} = \rho\lambda - H_k. \tag{3.35}$$

The Pontryagin maximum principle holds for both Hamiltonians equally well, since the current-value Hamiltonian is related to the present-value Hamiltonian $\mathcal{H}$ by

$$\mathcal{H}(k, x, \Lambda, t) = e^{-\rho t} H(k, x, \lambda, t). \tag{3.36}$$

Thus, maximizing the current-value Hamiltonian over x is essentially the same problem as maximizing the present-value Hamiltonian over x.

3.4.4 The Current Value Classic Lagrangian

Similarly, define the current-value classic Lagrangian L by

$$L(k, x, \lambda, t) = U(k, x, t) + \lambda[A(k, x, t) - \rho k] + \dot{\lambda} k. \tag{3.37}$$

Then the *first-order condition* can be written

$$L_x = 0, \tag{3.38}$$

while the *Euler equation* is

$$L_k = 0. \tag{3.39}$$

The Pontryagin maximum principle and the sufficiency result hold for both classic Lagrangians equally well, since given

$$\dot{\Lambda} = \frac{d}{dt} e^{-\rho t} \lambda = e^{-\rho t} [\dot{\lambda} - \rho \lambda],$$

the current-value classic Lagrangian is related to the present-value classic Lagrangian $\mathcal{L}$ by

$$\mathcal{L}(k, x, \Lambda, t) = e^{-\rho t} H(k, x, \lambda, t). \tag{3.40}$$

Thus, maximizing the current-value classic Lagrangian over x (for the maximum principle) or over x and k (for the sufficiency result) is essentially the same problem as maximizing the present-value classic Lagrangian x or over x and k.

At a steady state, $\dot{\lambda} = 0$, and the current-value classic Lagrangian becomes the classic Lagrangian for the discounted case discussed in chapter 18.

3.5 An Example: The Ramsey-Cass Model

The earliest use of the tools of dynamic optimization in economics is by Ramsey (1928) [**?**], addressing the question of the optimal path of capital accumulation for an economy using the calculus of variations and a zero discount rate. Cass [**?**] revisited the question of the optimal path of capital accumulation allowing for a positive utility discount rate and using the optimal control techniques here which, in effect, add the multiplier λ to the calculus of variations.

The social planner's problem Cass examines is

$$\max_c \quad \int_0^\infty e^{-\rho t} u(c) \, dt \tag{3.41}$$

$$\text{subject to} \qquad \dot{k} = f(k) - \delta k - c. \tag{3.42}$$

Where δ is rate at which capital depreciates. (From here on, we will list the maximization as being over the control variables rather than also over the state variables implied by the policies for the control variables. For brevity, we will also typically omit the initial condition $k_0 = \bar{k}$ in the statement of a problem since this condition seldom varies in form.)

The current-value Hamiltonian for (3.41) is

$$H = u(c) + \lambda[f(k) - \delta k - c].$$

The first order condition is

$$H_c = u'(c) - \lambda = 0,$$

or

$$u'(c) = \lambda.$$

Thus, the marginal utility of consumption equals the marginal value of capital.

The Euler equation is

$$\dot{\lambda} = \rho\lambda - H_k = [\rho + \delta - f'(k)]\lambda,$$

or

$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - f'(k).$$

Thus, the proportional growth rate of the marginal value of capital is equal to the difference between the utility discount rate and the net marginal product of capital $f'(k) - \delta$. (In comparing with earlier results, pay particular attention to the special case when the depreciation rate δ is zero.)

3.6 Extension: A Variable Discount Rate

Especially in dealing with the behavior of profit-maximizing firms, for whom the relevant discount rate is the interest rate r, it is important to know how to deal with variable discount rates.

Consider a general problem with a variable discount rate r:

$$\max_{x} \quad \int_0^T e^{-\int_0^t r_{t'}\,dt'} U(k_t, x_t, t)\,dt \tag{3.43}$$

$$\text{subject to} \qquad \dot{k} = A(k_t, x_t, t), \tag{3.44}$$

where t' is just a dummy variable. For this problem, where we want to discount by the variable interest rate r, define the current-value marginal value of capital λ by

$$\lambda_t = e^{\int_0^t r_{t'}\, dt'} \Lambda_t. \tag{3.45}$$

In order to differentiate λ, we need the fundamental theorem of calculus, which says that[8]

$$\frac{d}{dt}\int_0^t r\, dt' = r,$$

so that by the chain rule,

$$\frac{d}{dt} e^{\int_0^t r\, dt'} = r e^{\int_0^t r\, dt'}$$

and the product rule give us

$$\begin{aligned} \dot{\lambda} &= r e^{\int_0^t r\, dt'} \Lambda + e^{\int_0^t r\, dt'} \dot{\Lambda} \\ &= r\lambda + e^{\int_0^t r\, dt'} \dot{\Lambda}. \end{aligned}$$

Together with the general Euler equation (**??**), this implies

$$\begin{aligned} \dot{\lambda} &= r\lambda - [U_k(k, x, t) + \lambda A_k(k, x, t)] \\ &= r\lambda - H_k, \end{aligned} \tag{3.46}$$

where the current value Hamiltonian H is defined as

$$H = U + \lambda A.$$

Thus, the formula for the Euler equation has essentially the same form even when the discount rate varies with time! It is also easy to show that the first-order condition

$$H_x = 0$$

and the more general principle of maximization of the current-value Hamiltonian is still valid in terms of the current-value Hamiltonian with a variable discount rate.

[8]For clarity, we will drop the time and time-dummy subscripts from here on in this section.

Exercises

Additive Time-Separability.

This exercise is meant to clarify why separability every which way implies additive separability as long as there are at least three periods. To see the logic of this result, consider the case of exactly three time periods and objective function $\Omega(c_1, c_2, c_3)$. Separability every which way implies that each ratio of marginal utilities is a function of only the two immediately relevant variables; that is,

$$\begin{aligned}
\frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_1} \div \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_2} &= \phi(c_1,c_2) \\
\frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_2} \div \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_3} &= \chi(c_2,c_3) \\
\frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_1} \div \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_3} &= \psi(c_1,c_3)
\end{aligned}$$

Note that this implies

$$\psi(c_1,c_3) = \chi(c_2,c_3)\phi(c_1,c_2)$$

for *any* c_2. Starting from an arbitrary reference point $(\bar{c}_1, \bar{c}_2, \bar{c}_3)$, define

$$\begin{aligned}
u_1'(c_1) &= \frac{\phi(c_1,\bar{c}_2)}{\phi(\bar{c}_1,\bar{c}_2)} \\
u_2'(c_2) &= \frac{1}{\phi(\bar{c}_1,c_2)} \\
u_3'(c_3) &= \frac{1}{\psi(\bar{c}_1,c_3)}.
\end{aligned}$$

Exercise: Show that, given these definitions,

$$\begin{aligned}
\frac{u_1'(c_1)}{u_2'(c_2)} &= \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_1} \div \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_2}, \\
\frac{u_2'(c_1)}{u_3'(c_2)} &= \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_2} \div \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_3}, \\
\frac{u_1'(c_1)}{u_3'(c_2)} &= \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_1} \div \frac{\partial\Omega(c_1,c_2,c_3)}{\partial c_3},
\end{aligned}$$

and, therefore, that the proposed marginal utilities $u_1'(c_1)$, $u_2'(c_2)$ and $u_3'(c_3)$ accurately represent the slopes of the indifference surfaces of $\Omega(c_1, c_2, c_3)$ and so represent the same preferences. (Note that given u_1', u_2' and u_3', there is no problem in obtaining the period utility functions u_1, u_2 and u_3 by integration.)

Answer:

$$\begin{aligned}\frac{u_1'(c_1)}{u_2'(c_2)} &= \phi(c_1, c_2) \\ &= \frac{\partial\Omega(c_1, c_2, c_3)}{\partial c_1} \div \frac{\partial\Omega(c_1, c_2, c_3)}{\partial c_2},\end{aligned}$$

$$\begin{aligned}\frac{u_2'(c_1)}{u_3'(c_2)} &= \frac{\psi(\bar{c}_1, c_3)}{\phi(\bar{c}_1, c_2)} \\ &= \chi(c_2, c_3) \\ &= \frac{\partial\Omega(c_1, c_2, c_3)}{\partial c_2} \div \frac{\partial\Omega(c_1, c_2, c_3)}{\partial c_3}\end{aligned}$$

and

$$\begin{aligned}\frac{u_1'(c_1)}{u_3'(c_2)} &= \frac{\phi(c_1, \bar{c}_2)\psi(\bar{c}_1, c_3)}{\phi(\bar{c}_1, \bar{c}_2)} \\ &= \frac{\phi(c_1, \bar{c}_2)\phi(\bar{c}_1, \bar{c}_2)\chi(\bar{c}_2, c_3)}{\phi(\bar{c}_1, \bar{c}_2)} \\ &= \phi(c_1, \bar{c}_2)\chi(\bar{c}_2, c_3) \\ &= \psi(c_1, c_3) \\ &= \frac{\partial\Omega(c_1, c_2, c_3)}{\partial c_1} \div \frac{\partial\Omega(c_1, c_2, c_3)}{\partial c_3}.\end{aligned}$$

Taking a limit as $h \to 0$

Show that with $\beta = e^{-\rho h}$, the equations (3.5) and (3.7) limit into (3.18) and (3.20) as $h \to 0$. What happens to the transversality condition and associated nonnegativity condition as $h \to 0$?

The Discrete-Time Current Value Hamiltonian

Find the definition of a discrete-time current value Hamiltonian that duplicates equations (3.34) and (3.35) as closely as possible (but with $1-\beta$ in place of ρ).

Interpreting λ

Derive an integral equation for λ from the Euler equation, using the transversality condition as an endpoint condition. Do this first for a constant discount rate, then for a variable discount rate. (Hint: there are actually several forms of such an integral equation that are actually equivalent but look quite different. Also, you may need to use Leibniz' rule for differentiating integrals—the rule that if θ is a scalar parameter,

$$\frac{\partial}{\partial\theta}\int_{a(\theta)}^{b(\theta)} f(t,\theta)\,dt = b'(\theta)f(b(\theta),\theta) - a'(\theta)f(a(\theta),\theta) + \int_{a(\theta)}^{b(\theta)} f_\theta(t,\theta)\,dt,$$

where $f_\theta = \frac{\partial f}{\partial\theta}$.)

Consider a firm deciding how much to invest in the face of investment adjustment costs. We can write such a problem as

Solved Exercise: Investment Adjustment Costs

$$\max_{n,i} \quad \int_0^\infty e^{-\int_0^t r\,dt'}[pF(k,n) - wn - p_i i]\,dt \tag{3.47}$$

$$\text{subject to} \qquad \dot{k} = k\phi(i/k). \tag{3.48}$$

The variable n is labor employed and $F(k,n)$ is the production function. The real interest rate r, relative output price p, relative investment goods price p_i and real wage w are all stated in terms of the same price index, but it is not important which price index as long as it is used consistently. The function ϕ satisfies $\phi'(\cdot) > 0$ and $\phi''(\cdot) < 0$. Investment adjustment costs which make it more costly to add to the capital stock when investment is all ready happening fast are modeled by the concavity of ϕ.[9]

Denoting the costate variable by q here instead of the usual λ, the current-value Hamiltonian is

[9] There is an implicit assumption that investment adjustment costs are in terms of wastage of the investment good.

$$H = pF(k,n) - wn - p_i i + q[k\phi(i/k)].$$

The first-order conditions for n and i are

$$\begin{aligned} H_n &= pF_n - w = 0 \\ H_i &= -p_i + q\phi'(i/k) = 0, \end{aligned}$$

or

$$pF_n = w \tag{3.49}$$

$$q = \frac{p_i}{\phi'(i/k)}. \tag{3.50}$$

The Euler equation is

$$\dot{q} = rq - H_k = [r + (i/k)\phi'(i/k) - \phi(i/k)]q - pF_k.$$

Chapter 4

Multiple Control Variables and Multiple State Variables

4.1 Introduction

Nothing in the derivation of the general first-order conditions and Euler equations in the previous chapters prevents the control x and the state k from being vectors. The costate λ will be a vector of the same size as the state k. Looking at the components of these vectors, (1) to each control variable there is a first-order condition, (2) each state variable comes endowed with an accumulation equation and (3) each state-variable/costate-variable pair has an Euler equation. Thus, accumulation equations and Euler equations come paired according to the state variable they relate to. All of the equations have the same form for each variable as if that variable were the only control variable or the only state variable.

Although state and costate variables are linked together in pairs so that there are exactly as many costate variables as there are state variables, the number of control variables can be very different from the number of state variables. Indeed, it is surprisingly easy to work with models with one state variable and many control variables—allowing one to address a wide range of phenomena.

In later chapters we will find that additional state variables add much more to the difficulty of a complete analysis of a model than do additional control variables, but in this chapter we are only trying to get the basic equations of each model. Additional state variables do not cause any trouble in getting the basic equations of a model, so we will look at both models with one state variable and more than one control variable and some models with more than one state variable.

4.2 The Cass Model with Two Types of Consumption

Because the components of the relevant vector equations look the same as the scalar equations one gets with just one variable, in a sense there is nothing truly new about the basic conditions for optimization with more than one control variable or more than one state variable. We will illustrate by a model with two types of nondurable consumption.

4.2.1 Two Types of Nondurable Consumption

Consider a retired household's problem of deciding how much to buy of two different consumption goods:

$$\max_{c_1,c_2} \quad \int_0^\infty e^{-\rho t} u(c_1, c_2)\, dt \tag{4.1}$$

$$\text{subject to} \qquad \dot{a} = ra + y - p_1 c_1 - p_2 c_2, \tag{4.2}$$

where y is non-capital income (such as social security checks) and p_1 and p_2 are the prices of the two types of consumption goods c_1 and c_2. (If p_1 and p_2 are nominal prices, then r needs to be the nominal interest rate for the accumulation equation to make sense economically.) The current-value Hamiltonian for (4.1) is

$$H = u(c_1, c_2) + \lambda[ra + y - p_1 c_1 - p_2 c_2]$$

The first-order conditions are

$$H_{c_1} = u_{c_1}(c_1, c_2) - p_1 \lambda = 0$$

$$H_{c_2} = u_{c_2}(c_1, c_2) - p_2 \lambda = 0$$

or

$$\begin{aligned} u_{c_1} &= p_1 \lambda \\ u_{c_2} &= p_2 \lambda. \end{aligned}$$

The Euler equation is

$$\dot{\lambda} = \rho\lambda - H_a = (\rho - r)\lambda.$$

This control problem can be used to address the case of a working household deciding on consumption and leisure by a simple reinterpretation. Let c_1 be ordinary consumption and c_2 leisure. Let y be the amount earned with the maximum possible amount of work. Leisure is measured by the amount by which actual labor is lower than that maximum. Taking the price of ordinary consumption as the numeraire $p_1 = 1$ and $p_2 = w$, where w is the real wage. Then, writing c for ordinary consumption c_1 and ℓ for leisure c_2, the budget constraint becomes

$$\dot{a} = ra + y - c - w\ell,$$

the Hamiltonian becomes

$$H = u(c, \ell) + \lambda[ra + y - c - w\ell]$$

and the first-order conditions are

$$\begin{aligned} u_c &= \lambda \\ u_\ell &= w\lambda. \end{aligned}$$

The Euler equation is unchanged.

Usually, we will find it more convenient to think in terms of consumption and labor rather than consumption and leisure. There is no real difficulty to having labor enter the utility function directly and allow it to often have a negative effect on utility. In this case, $p_1 = 1$, $p_2 = -w$ and y should be interpreted as non-capital, non-labor income (often zero). Writing n for the amount of labor, the budget constraint becomes

$$\dot{a} = ra + y - c + wn,$$

the Hamiltonian becomes

$$H = u(c, n) + \lambda[ra + y - c + wn]$$

and the first-order conditions are

$$\begin{aligned} u_c &= \lambda \\ u_n &= -w\lambda. \end{aligned}$$

The Euler equation is again unchanged. We revisit this problem in more detail in the next chapter.

4.3 Constraints on the Control Variables

In this section, we will focus on dealing with constraints on the control variables. The Pontryagin Maximum Principle states that the Hamiltonian must be maximized at each point in time (globally in the case of continuous time, at least locally in discrete time). The easiest way to maximize the Hamiltonian subject to constraints is to augment the Hamiltonian with these constraints and their associated multipliers and then get first-order conditions and sometimes Kuhn-Tucker conditions in the familiar way. The multipliers associated with these constraints must be carefully distinguished from the costate variables. In particular, there are no Euler equations attached to the multipliers for the constraints on the control variables.

4.3.1 The Cass Model Revisited

We can make both consumption and investment explicit in the Cass model if we also make an explicit constraint that consumption plus investment equals output. That is,

$$\max_{c,i} \quad \int_0^\infty e^{-\rho t} u(c)\, dt \tag{4.3}$$

$$\begin{aligned} \text{subject to} \quad \dot{k} &= i - \delta k & (4.4) \\ f(k) &= c + i. & (4.5) \end{aligned}$$

The augmented current-value Hamiltonian is

$$H = u(c) + \lambda[i - \delta k] + \mu[f(k) - c - i].$$

The costate variable is λ, since λ is the multiplier for the accumulation equation (4.4). By constrast, μ is an ordinary Lagrange multiplier associated with the maximization of the stripped-down Hamiltonian $u(c) + \lambda(i - \delta k)$. (Without the constraint, maximization of this stripped down Hamiltonian would be easy: infinitely high consumption c and infinitely high investment i.) Using the augmented Hamiltonian, the first-order conditions are

$$\begin{aligned} H_c &= u'(c) - \mu = 0 \\ H_i &= \lambda - \mu = 0. & (4.6) \end{aligned}$$

Not surprisingly, μ can be eliminated, yielding the familiar equation

$$u'(c) = \lambda.$$

There is no Euler equation for μ, since it is not a costate variable. The Euler equation for λ is still

$$\dot{\lambda} = \rho\lambda - H_k = (\rho + \delta)\lambda - f'(k)\mu = [\rho + \delta - f'(k)]\lambda,$$

where the last step requires the fact that $\mu = \lambda$ due to (4.6).

It does not matter whether one calculates the H_k in the Euler equation from the stripped-down Hamiltonian or from the augmented Hamiltonian as here. For one thing, the augmented and stripped-down Hamiltonians are identically equal since, for each constraint, either the constraint or the associated multiplier is always equal to zero making the product identically zero. (In this particular case, $f(k) - c - i$ is identically equal to zero.) Furthermore, by the Envelope Theorem, the change in the maximized Hamiltonian is equal to the partial derivative of either version of the Hamiltonian with respect to k. (See Section (**??**) for more on the Envelope Theorem. The Envelope Theorem states that the derivative of a maximized objective with respect to a parameter—that is, with respect to any variable not maximized over—is equal to the partial derivative with respect to that parameter.)

4.3.2 A Model with Consumer Durables

Consumer Durables as Another State Variable

Consider a model with both nondurable consumption c and a stock of a durable consumer good d. The quantity of the durable good is another state variable, in addition to k. Let us put these two types of consumption in the context of a growth model. Thus, the social planner's problem is

$$\max_{c, i_k, i_d} \int_0^\infty e^{-\rho t} u(c, d)\, dt \tag{4.7}$$

$$\begin{aligned} \text{subject to} \quad \dot{k} &= i_k - \delta_k k \\ \dot{d} &= i_d - \delta_d d \\ f(k) &= c + i_k + i_d. \end{aligned} \tag{4.8}$$

The augmented current-value Hamiltonian is

$$H = u(c,d) + \lambda_d[i_d - \delta_d d] + \lambda_k[i_k - \delta_k k] + \mu[f(k) - c - i_d - i_k].$$

The first order conditions are

$$\begin{aligned} H_c &= u_c(c,d) - \mu = 0 \qquad (4.9) \\ H_{i_d} &= \lambda_d - \mu = 0 \\ H_{i_k} &= \lambda_k - \mu = 0. \qquad (4.10) \end{aligned}$$

(Subscript represent partial derivatives on the utility function u but not otherwise.) There are two Euler equations, for λ_d and λ_k:

$$\begin{aligned} \dot{\lambda}_d &= \rho\lambda_d - H_d = (\rho + \delta_d)\lambda_d - u_d(c,d) \qquad (4.11) \\ \dot{\lambda}_k &= \rho\lambda_k - H_k = (\rho + \delta_k)\lambda_k - f'(k)\mu. \qquad (4.12) \end{aligned}$$

The first-order conditions imply

$$u_c(c,d) = \mu = \lambda_d = \lambda_k.$$

Substituting λ_k for every occurence of λ_d and μ in the Euler equations, this means that

$$[\rho + \delta_k - f'(k)]\lambda_k = \dot{\lambda}_k = [\rho + \delta_d]\lambda_k - u_d(c,d) \qquad (4.13)$$

Equation (**??**) implies, in turn, that

$$u_d(c,d) = [f'(k) - \delta_k + \delta_d]\lambda_k, \qquad (4.14)$$

and in view of (4.9) that

$$\frac{u_d(c,d)}{u_c(c,d)} = f'(k) - \delta_k + \delta_d. \qquad (4.15)$$

In words, the ratio of the marginal utility of the durable to the marginal utility of nondurable consumption is proportional to the effective price ratio: the net marginal product of capital plus the depreciation rate for consumer durables.

Equivalence to a One-State-Variable Model

In model (4.7), the marginal value of capital and of the consumer durable are equal because one can always decrease investment in one and increase investment in the other by a corresponding amount. Because these marginal values are equal, one can actually model the same problem with just one state variable. To demonstrate this, following Mankiw (), let $w = d + k$ and let $\lambda = \lambda_d = \lambda_k$ be the costate variable for w. Treat d and k as *control variables*. That is, consider

$$\max_{c,i,d,k} \quad \int_0^\infty e^{-\rho t} u(c,d)\, dt \tag{4.16}$$

$$\begin{aligned} \text{subject to} \qquad \dot{w} &= i - \delta_k k - \delta_d d \\ f(k) &= c + i \\ w &= d + k. \end{aligned} \tag{4.17}$$

The augmented current-value Hamiltonian is

$$H = u(c,d) + \lambda[i - \delta_d d - \delta_k k] + \mu[f(k) - c - i] + \nu[w - d - k].$$

The first-order conditions are

$$H_c = u_c(c,d) - \mu = 0 \tag{4.18}$$

$$H_i = \lambda - \mu = 0 \tag{4.19}$$

$$H_d = u_d(c,d) - \delta_d \lambda - \nu = 0 \tag{4.20}$$

$$H_k = f'(k)\mu - \delta_k \lambda - \nu = 0. \tag{4.21}$$

The only Euler equation is

$$\dot{\lambda} = \rho\lambda - H_w = \rho\lambda - \nu. \tag{4.22}$$

Equation (4.19) allows one to replace μ by λ. By (4.20) and (4.21),

$$\nu = [f'(k) - \delta_k]\lambda = u_d(c,d) - \delta_d \lambda.$$

Since $\lambda = \lambda_d = \lambda_k$, substituting these expressions in for ν in (4.22) yields the Euler equations (4.11) and (**??**). Careful inspection shows that the whole set of equations is equivalent.

Irreversible Investment

So far we have imposed no requirement that net investment in capital or consumer durables be positive. Adding a requirement that investment in capital and in consumer durables be positive has no effect as long as both types of investment would be positive anyway, but when the social planner would like to have one or the other type of investment be negative, the ability to trade off capital for consumer durables by decreasing investment in one and increasing investment in the other breaks down. Adding the constraints

$$\begin{aligned} i_d &\geq 0 \\ i_k &\geq 0 \end{aligned}$$

to the two-state variable problem (4.7) above, the augmented current-value Hamiltonian is

$$H = u(c,d) + \lambda_d[i_d - \delta_d d] + \lambda_k[i_k - \delta_k k] + \mu[f(k) - c - i_d - i_k] + \xi_d i_d + \xi_k i_k.$$

The Kuhn-Tucker conditions state that $\xi_d \geq 0$ and $\xi_k \geq 0$, with either one equal to zero whenever the corresponding level of gross investment is strictly positive. The first order conditions are

$$\begin{aligned} H_c &= u_c(c,d) - \mu = 0 \\ H_{i_d} &= \lambda_d - \mu + \xi_d = 0 \\ H_{i_k} &= \lambda_k - \mu + \xi_k = 0. \end{aligned} \tag{4.23}$$

The two Euler equations are formally the same as (4.11) and (4.12), but they cannot be simplified in the same way when ξ_d or ξ_k is positive.

The first-order conditions and other Kuhn Tucker conditions imply that

$$\lambda_d = \lambda_k = \mu = u_c(c,d)$$

when both levels of gross investment are strictly positive;

$$\lambda_d \leq \lambda_k = \mu = u_c(c,d)$$

when $i_d = 0$ but $i_k > 0$; and

$$\lambda_k \leq \lambda_d = \mu = u_c(c,d)$$

when $i_k = 0$ but $i_d > 0$. When both levels of gross investment are zero, both λ_d and λ_k can be less than μ and u_c and it is unclear which of λ_d or λ_k is smaller.

Intuitively, if the marginal value of consumer durables is especially low, it can shut off investment in consumer durables entirely. Because investment must be nonnegative, investment in consumer durables may not be able to go low enough to allow equalization of this low marginal value of consumer durables with the marginal value of capital or the marginal utility of consumption. If the marginal value of investment in capital is especially low, it can shut off investment in capital entirely.

Chapter 5

Characterizing Phase Diagrams in Fully Optimizing Concave Models with One State Variable

5.1 Introduction

There is more structure to the phase diagrams that arise from dynamic optimization than many economists realize. This chapter derives general characteristics of the phase diagram for fully optimizing models with a concave objective function and a concave accumulation function. In this chapter, we discuss models with one state variable, for which the phase diagram can be graphed in two dimensions. In the next chapter we extend certain key principles laid out in this chapter to the case of many state variables, and point out which principles cannot be generalized to the many-state-variable case.

For the sake of concreteness, we will continue to call the state variable k "capital" and the costate variable λ the "marginal value of capital." Keep in mind that the following results, because they are purely mathematical, apply to *any* fully optimizing dynamic optimization model with one state variable and concave objective and accumulation functions. We will standardize by always putting the state variable k on the horizontal axis and the costate variable λ (rather than one of the control variables) on the vertical axis.

Formally, consider the following general dynamic optimization problem:

$$\max_x \int_0^\infty e^{-\rho t} U(k, x; z) dt,$$

s.t.

$$\dot{k} = A(k, x; z) \tag{5.1}$$

and

$$x \in X(k, z),$$

where the constraint set $X(k, z)$ forms a convex region in (k, x) space for the relevant z. As above (in the previous chapter), the only direct dependence on time is in the discounting of the objective function. Changes over time in the vector of exogenous variables z can be represented and analyzed as changes in the dynamics indicated by the phase diagram. However, the results of this section are for a fixed value of z. Assume further that the objective function U and the accumulation function A are (twice-differentiable and) concave and that free disposal of capital guarantees a nonnegative marginal value of capital $(\lambda \geq 0)$.

For this case, we will show that (1) a higher marginal value of capital λ always raises $\dot{k}$, the rate of increase in capital, (2) a higher capital stock k always raises $\dot{\lambda}$, the rate of increase in the marginal value of capital; (3) the two eigenvalues for the linearized dynamics around the steady-state are real and add up to the discount rate ρ, with the corollary of (4) no spiralling. Furthermore, (5) except in one case involving a large utility discount rate ρ, the steady state has local saddle-point stability, with a downward-sloping saddle-path.

5.2 The Maximized Hamiltonian $H^\star$

As shown in chapter **??**, the dynamics can be expressed in terms of the current-value Hamiltonian

$$H(k, x, \lambda) = U(k, x) + \lambda A(k, x).$$

by

$$\dot{k} = A(k, x) = H_\lambda(k, x, \lambda)$$

and

$$\dot{\lambda} = \rho\lambda - H_k(k, x, \lambda),$$

with x at a value that maximizes the Hamiltonian H. Defining the *maximized* Hamiltonian $H^\star$ by

$$H^\star(k, \lambda) = \max_x H(k, x, \lambda),$$

the envelope theorem guarantees that at optimal values of x, $H^\star_k = H_k$ and $H^\star_\lambda = H_\lambda$. Thus, the dynamics are also described by

$$\dot{k} = H^\star_\lambda(k, \lambda) \tag{5.2}$$

$$\dot{\lambda} = \rho\lambda - H^\star_k(k, \lambda). \tag{5.3}$$

The behavior of $H^\star_k$ and $H^\star_\lambda$ can be characterized by the following result:

Lemma 1 *$H^\star_k(k, \lambda)$ is (at least weakly) decreasing in k as long as it is well-defined and $H(k, x, \lambda)$ is jointly concave in k and x, while $H^\star_\lambda(k, \lambda)$ is (at least weakly) increasing in λ as long as it is well defined. If $H^\star$ is twice differentiable, this means that $H^\star_{kk} \leq 0$, while $H^\star_{\lambda\lambda} \geq 0$.*

Before diving into the proof of the lemma, think about what Lemma 1 means for dynamics on the phase diagram.

By (5.3), the fact that $H^\star_k$ is decreasing in k means that higher k always raises $\dot{\lambda}$. Thus, quite generally, the marginal value of capital is rising ($\dot{\lambda} > 0$) to the right of the $\dot{\lambda} = 0$ locus, while the marginal value of capital is falling ($\dot{\lambda} < 0$) to the left of the $\dot{\lambda} = 0$ locus; in any one-state variable, concave model of pure dynamic optimization, the dynamic arrows should point *up* to the right of the $\dot{\lambda} = 0$ locus and *down* to the left of the $\dot{\lambda} = 0$ locus.

By (5.2) the fact that $H^\star_\lambda$ is increasing in λ means that higher λ always raises $\dot{k}$. Thus, quite generally, capital is accumulating ($\dot{k} > 0$) above the $\dot{k} = 0$ locus, while capital is decumulating ($\dot{k} < 0$) below the $\dot{k} = 0$ locus; in any one-state variable, model of pure dynamic optimization (regardless of whether or not it is concave), the dynamic arrows should point *to the right* above the $\dot{k} = 0$ locus and *to the left* below the $\dot{k} = 0$ locus.

Both of these results are global. They apply everywhere, not just in the neighborhood of the steady state. Regardless of the slopes of the $\dot{\lambda} = 0$ locus and the $\dot{k} = 0$ locus, to get the visual sense of these results, it is helpful to see the $\dot{\lambda} = 0$ locus as separating the right from left and to see the $\dot{k} = 0$ separating above from below. Then the general rules can be stated simply as (1) to the right, λ is rising; (2) to the left, λ is falling; (3) up high, k is accumulating; (4) down low, k is decumulating. Schematically, Figure 1 gives the right idea.[1] [Figure 1 has a shaded area in the shape of a Maltese cross in the middle to indicate a vaguely vertical $\dot{\lambda} = 0$ locus of indefinite slope, and a vaguely horizontal $\dot{k}$ locus of indefinite slope. The four quadrants have arrows showing the dynamics in each quadrant.]

5.2.1 Proof that $H^\star_k$ is decreasing in k.

$H^\star_k$ must be decreasing in k (when it is well-defined) if $H^\star$ is concave in k, as it will be if H is jointly concave in k and x. To see this consider that if $x^\star_0$ maximizes H at (k_0, λ) while $x^\star_1$ maximizes H at (k_1, λ), and H is jointly concave in k and x, then for any $a \in [0, 1]$,

$$\begin{aligned} H^\star(ak_0 + (1-a)k_1, \lambda) &= \max_x H(ak_0 + (1-a)k_1, x, \lambda) \\ &\geq H(ak_0 + (1-a)k_1, ax^\star_0 + (1-a)x^\star_1, \lambda) \\ &\geq aH(k_0, x^\star_0, \lambda) + (1-a)H(k_1, x^\star_1, \lambda) \\ &= aH^\star(k_0, \lambda) + (1-a)H^\star(k_1, \lambda). \end{aligned} \tag{5.4}$$

The first and last equalities in (5.4) follow from the definition of $H^\star$. The first inequality follows from the nature of maximization, with the convexity of the $k - x$ region formed by $X(k, z)$ guaranteeing that the indicated point is in the domain of H. The second inequality follows from the joint concavity of H in k and x.

Intuitively, joint concavity of U and A in k and x guarantees that the marginal product of capital (the combined value of its contribution to utility and to accumulation) must fall as k increases, even with x adjusting optimally along the way. (The adjustment in the vector of control variables x makes the resulting fall in the marginal product of capital as small as possible, but cannot stop the decline.) In many applications, constant returns to scale in production will mean that A is only weakly concave, but strict concavity of U will typically induce strict concavity of $H^\star$ in k. Any one of various types of increasing returns to scale in production may overturn the concavity of

[1] Figure 1 is appropriate for all three of the stable cases discussed below. But it is not appropriate for the unstable case discussed below, since Figure 1 assumes that the $\dot{\lambda} = 0$ locus is more nearly vertical than the $\dot{k} = 0$ locus.

$H^\star$, leading to a qualitatively different phase diagram. (See for example Paul Romer's first paper on endogenous growth, 19??.)

5.2.2 Proof that $H^\star_\lambda$ is increasing in λ.

If $x_0^\star$ maximizes the Hamiltonian H at (k, λ_0), while $x_1^\star$ maximizes the Hamiltonian H at (k, λ_1), then by definition,

$$U(k, x_0^\star) + \lambda_0 A(k, x_0^\star) \geq U(k, x_1^\star) + \lambda_0 A(k, x_1^\star) \tag{5.5}$$

and

$$U(k, x_1^\star) + \lambda_1 A(k, x_1^\star) \geq U(k, x_0^\star) + \lambda_1 A(k, x_0^\star). \tag{5.6}$$

Therefore,

$$\begin{aligned}(\lambda_1 - \lambda_0)[H^\star_\lambda(k, \lambda_1) - H^\star_\lambda(k, \lambda_0)] &= (\lambda_1 - \lambda_0)[A(k, x_1^\star) - A(k, x_0^\star)] \\ &= \{U(k, x_1^\star) + \lambda_1 A(k, x_1^\star) - U(k, x_0^\star) - \lambda_1 A(k, x_0^\star)\} \\ &\quad + \{U(k, x_0^\star) + \lambda_0 A(k, x_0^\star) - U(k, x_1^\star) - \lambda_0 A(k, x_1^\star)\} \\ &\geq 0.\end{aligned}$$

(The same result can be obtained by subtracting (5.5) from (5.6) and rearranging.) In words, $H^\star_\lambda$ moves in the same direction as λ.

Intuitively, a higher marginal value of capital λ makes capital accumulation more valuable and therefore leads to more accumulation by means of adjustments in the control variables. Paul Milgrom and John Roberts, in "Comparing Equilibria" AER June 1994, 441–449, indicate the generality of this type of result. In particular, $H^\star_\lambda$ being increasing in λ in no way depends on concavity.

5.3 Dynamics in the Neighborhood of the Steady State

To go further in our analysis, we need to zoom in on the neighborhood of the steady state (or of *a* steady state). Making a first order Taylor approximation of (5.2) and (5.3) around the steady state (k^*, λ^*) on the assumption that $H^\star$ is twice differentiable, [2]

[2] Note the distincition between a five-pointed star * for maximization and a regular asterisk * for a steady-state value.

$$\begin{aligned}\dot{k} &\approx 0 + H^{\star}_{\lambda k}(k^*, \lambda^*)[k - k^*] + H^{\star}_{\lambda\lambda}(k^*, \lambda^*)[\lambda - \lambda^*] \\ \dot{\lambda} &\approx 0 - H^{\star}_{kk}(k^*, \lambda^*)[k - k^*] + (\rho - H^{\star}_{k\lambda}(k^*, \lambda^*))[\lambda - \lambda^*].\end{aligned}$$

Using a tilde ($\tilde{\ }$) to represent a small deviation from a steady-state value that can be treated as a differential for calculus, these equations can be rewritten in matrix form as

$$\begin{bmatrix} \dot{\tilde{k}} \\ \dot{\tilde{\lambda}} \end{bmatrix} = \begin{bmatrix} H^{\star}_{\lambda k} & H^{\star}_{\lambda\lambda} \\ -H^{\star}_{kk} & \rho - H^{\star}_{k\lambda} \end{bmatrix} \begin{bmatrix} \tilde{k} \\ \tilde{\lambda} \end{bmatrix}, \tag{5.7}$$

where all of the derivatives of $H^{\star}$ are evaluated at (k^*, λ^*) and

$$\dot{\tilde{k}} = \dot{k} - (\dot{k^*}) = \dot{k}$$

$$\dot{\tilde{\lambda}} = \dot{\lambda} - (\dot{\lambda^*}) = \dot{\lambda},$$

both regarded as small.

The key characteristics of the 2x2 *dynamic matrix* in (5.7) are (1) by Lemma 1 above, both off-diagonal elements ($H^{\star}_{\lambda\lambda}$ and $-H^{\star}_{kk}$) are positive; and (2) by the symmetric equality of mixed partial derivatives, $H^{\star}_{\lambda k} = H^{\star}_{k\lambda}$, making the trace (the sum of the elements on the main diagonal) equal to ρ.

Using ϑ to denote an eigenvalue, the characteristic equation of a 2x2 matrix is

$$\vartheta^2 - \text{trace}\ \vartheta + \det = 0,$$

where det is the determinant. The determinant of the dynamic matrix is

$$\det = \rho H^{\star}_{k\lambda} - (H^{\star}_{\lambda k})^2 + H^{\star}_{kk} H^{\star}_{\lambda\lambda}.$$

If $H^{\star}_{k\lambda}$ is negative or ρ is small, the determinant will be negative, since both $-(H^{\star}_{k\lambda})^2$ and $H^{\star}_{kk} H^{\star}_{\lambda\lambda}$ are negative. (See Lemma 1.) But if $H^{\star}_{k\lambda}$ is positive and ρ is large, the determinant may be positive. The sign of the determinant matters because, as a consequence of the characteristic equation, the determinant is equal to the product of the two eigenvalues. The only way there can be one positive and one negative root for saddlepoint stability is if the determinant is negative.

To make the same point about the sign of the determinant in another way, use the quadratic formula to solve the characteristic equation as follows:

$$\vartheta = \frac{\text{trace}}{2} \pm \sqrt{\frac{\text{trace}^2}{4} - \text{det}}. \tag{5.8}$$

With a positive trace, the only way there can be a negative root is if $(\text{trace}^2/4) - \text{det} > (\text{trace}/2)^2$. If the determinant is negative, this is guaranteed. If the determinant is positive, there will be two positive roots, indicating an unstable steady state.

Graphically, a negative determinant is assured if one of the $\dot{k} = 0$ and $\dot{\lambda} = 0$ loci slopes up and the other one down. If they both slope in the same direction, a negative determinant corresponds to the $\dot{\lambda} = 0$ locus being more nearly vertical than the $\dot{k} = 0$ locus.

Can the eigenvalues be complex? When the determinant is negative it is easy to see that the quantity under the square root sign is positive, making both eigenvalues real. If the determinant is positive, it looks at first as if there could be two complex roots with a positive real part, indicating an outward spiral, but in a pure optimization model, substitution from (5.7) into (5.8) gives the following formula for the eigenvalues

$$\begin{aligned} \vartheta &= \frac{\rho}{2} \pm \sqrt{\frac{\rho^2}{4} - \rho H^\star_{k\lambda} + (H^\star_{k\lambda})^2 - H^\star_{kk} H^\star_{\lambda\lambda}} \\ &= \frac{\rho}{2} \pm \sqrt{\left[\frac{\rho}{2} - H^\star_{k\lambda}\right]^2 - H^\star_{kk} H^\star_{\lambda\lambda}}. \end{aligned} \tag{5.9}$$

The quantity under the radical sign is clearly positive. Therefore, both roots are real in a pure one-state variable dynamic optimization problem.

5.4 A Typology

Focusing on the neighborhood of a steady state, in a fully optimizing, one-state variable dynamic control model, there are only four qualitatively distinct patterns of dynamics, plus three borderline cases. Since the off-diagonal elements of the dynamic matrix and (in order for the sum to equal ρ) at least one of the diagonal elements are positive, there are three possible sign patterns for the dynamic matrix, aside from borderline cases: (I) $\begin{bmatrix} + & + \\ + & - \end{bmatrix}$, (II) $\begin{bmatrix} + & + \\ + & + \end{bmatrix}$, and

(III) $\begin{bmatrix} - & + \\ + & + \end{bmatrix}$. Type II can be split into (A) a stable case, with a negative determinant, and (B) an unstable case, with a positive determinant. The borderline cases are (I/II) $\begin{bmatrix} + & + \\ + & 0 \end{bmatrix}$, (II/III) $\begin{bmatrix} 0 & + \\ + & + \end{bmatrix}$ and the borderline stable case of type III. The four main types are shown in Figure 2. The remainder of Section 5.4 is an economic interpretation of the differences between the four different types of dynamics. Among other things, the remainder of the section justifies the name of each type given in Figure 2. [Figure 2 is arranged counterclockwise, starting from the upper left, $\begin{matrix} \text{I} & \text{III} \\ \text{IIA} & \text{IIB} \end{matrix}$. This arrangement evokes the counterclockwise swing of the loci as one passes from I to II to III, and puts the most important cases, I and III on top. The cases are labeled as follows— I: Thrift-k Complementarity Dominant; III: Thrift-k Substitutability Dominant; II A: Discount Rate Dominant, Stable Case, II B: Discount Rate Dominant, Unstable Case.]

As can be seen from Figure 2, in all three of the stable cases, the saddle path is downward sloping, while the explosive path is upward sloping. (See Exercise 4. In addition, in chapter **??** there is a proof that the saddle-path is downward-sloping using the techniques of dynamic programming, in which concavity of the value function in k, $V_{kk} < 0$ implies that the saddle path, described by the function V_k, is downward-sloping.) In the unstable case, there is one upward-sloping explosive path and one downward-sloping explosive path.

5.4.1 The Derivatives of $H^\star$

From (5.7), it is clear that the second derivatives of $H^\star$ are the key to the dynamics in the neighborhood of a steady state. It is useful to express these derivatives in terms of H, U and A.

In order to get a foothold in interpreting the derivatives of $H^\star$, let us begin by limiting ourselves to the case in which the constraint set X does not depend on k. Then the envelope theorem allows us to ignore induced movements in the vector of control variables x.[3] Only in taking a second derivative, will it be

[3]Even when the constraint set X depends on k, since it does not depend on λ, the envelope theorem still applies in a very straightforward way to a first derivative with respect to λ. As for the derivative with respect to k, when the constraint set X depends on k, the envelope theorem applies to the augmented Hamiltonian rather than the Hamiltonian itself. (When the constraint set X is *not* a function of k, the additional terms in the augmented Hamiltonian are not a function of k, and so do not affect the envelope theorem for derivatives with respect to k.) These extra terms causes no great difficulties in calculating things for a particular model, but the augmented Hamiltonian can take so many different forms that it

necessary to take these induced movements into account. Using $\frac{d}{dk}$ and $\frac{d}{d\lambda}$ to denote full derivatives including this induced variation in x, we can calculate

$$\begin{aligned}
\begin{bmatrix} H^\star_{kk} & H^\star_{k\lambda} \\ H^\star_{\lambda k} & H^\star_{\lambda\lambda} \end{bmatrix} &= \begin{bmatrix} H_{kk} & H_{k\lambda} \\ H_{\lambda k} & H_{\lambda\lambda} \end{bmatrix} \\
&\quad + \begin{bmatrix} H_{kx} \\ H_{\lambda x} \end{bmatrix} \left[\tfrac{dx}{dk} \quad \tfrac{dx}{d\lambda} \right] \\
&= \begin{bmatrix} U_{kk} + \lambda A_{kk} & A_k \\ A_k & 0 \end{bmatrix} + \begin{bmatrix} U_{kx} + \lambda A_{kx} \\ (A_x)^T \end{bmatrix} \left[\tfrac{dx}{dk} \quad \tfrac{dx}{d\lambda} \right] \\
&= \begin{bmatrix} U_{kk} + \lambda A_{kk} + (U_{kx} + \lambda A_{kx})\frac{dx}{dk} & A_k + (U_{kx} + \lambda A_{kx})\frac{dx}{d\lambda} \\ A_k + (A_x)^T \frac{dx}{dk} & (A_x)^T \frac{dx}{d\lambda} \end{bmatrix},
\end{aligned} \tag{5.10}$$

where $(A_x)^T$ is the transpose of A_x. Because the off-diagonal elements of the dynamic matrix are always positive, the qualitative distinctions all have to do with the sign and magnitude of $H^\star_{k\lambda} = H^\star_{\lambda k}$ at the steady state. There are two equivalent ways to express this cross derivative, depending on the order of differentiation. In order to make the equivalence of the two expressions

$$A_k + A_x \frac{dx}{dk} = H^\star_{\lambda k} = H^\star_{k\lambda} = A_k + [U_{kx} + \lambda A_{kx}] \frac{dx}{d\lambda} \tag{5.11}$$

less mysterious, consider the case in which the control variables are all at an interior solution, so that $\frac{dx}{dk}$ and $\frac{dx}{d\lambda}$ can be obtained by differentiating $H_x = 0$:

$$H_{xk} + H_{xx} \frac{dx}{dk} = 0,$$

so that

$$\frac{dx}{dk} = -[H_{xx}]^{-1} H_{xk} = -[U_{xx} + \lambda A_{xx}]^{-1} [U_{xk} + \lambda A_{xk}],$$

and

$$H_{x\lambda} + H_{xx} \frac{dx}{d\lambda} = 0,$$

so that

$$\frac{dx}{d\lambda} = -[H_{xx}]^{-1} H_{x\lambda} = -[U_{xx} + \lambda A_{xx}]^{-1} A_x.$$

---

is inconvenient for a general discussion.

Thus, using either expression for $H^{\star}_{k\lambda}$, (plus the fact that a scalar is its own transpose),

$$H^{\star}_{k\lambda} = A_k - [U_{kx} + \lambda A_{kx}][U_{xx} + \lambda A_{xx}]^{-1} A_x. \tag{5.12}$$

However, the symmetry equation (5.11) remains true even when one of the control variables is at a corner solution, as long as the constraint set X does not depend on k.

5.4.2 Thrift

To aid in interpreting the sign and magnitude of $H^{\star}_{k\lambda}$ we will call the change in behavior indicated by $\frac{dx}{d\lambda}$ an increase in *thrift*. That is, increased *thrift* is the set of actions an agent uses in order to accumulate faster when λ, the marginal value of the stock to be accumulated, increases. Since by (5.10),

$$(A_x)^T \frac{dx}{d\lambda} = H^{\star}_{\lambda\lambda} \geq 0,$$

an increase in thrift does in fact raise the rate of accumulation.

In a particular model, the increased thrift induced by an increase in λ will be a particular pattern of change in behavior. Identifying the behavior associated with thrift and getting an intuitive grasp on the nature of thrift in any particular model is one of the keys to giving an intuitive interpretation of the dynamics of a model. Identifying thrift clearly may require rewriting the optimization problem to eliminate k from all but the accumulation function A and the objective function U. For example, in the Basic Real Business Cycle Model described in the Introduction (Chapter **??**), greater thrift is a combination of more labor and less consumption. Eliminating k from all but the accumulation equation and the objective eliminates the control variables i and y. In the Cass model, greater thrift is just a matter of less consumption. In Mankiw's model, greater thrift is a combination of reduced consumption of nondurables and reduced consumption of the services of durables.

5.4.3 Divide and Conquer

Already, by recasting a model so that the constraint set does not depend on k, we have concentrated all dependence on k into the objective and accumulation functions. Now, in order to further the economic interpretation of the different types of phase diagram dynamics, we will further concentrate the dependence on k into either the objective function alone or the accumulation function alone.

Looking at cases in which either U or A is not a function of k is more general than it seems. The first-order condition $U_x + \lambda A_x = 0$ is valid in all of the directions in which the control variables can be freely varied, or more formally, if the differential dx is consistent with the constraint set, then $[U_x + \lambda A_x]^T dx = 0$. Unless both terms are individually zero, $[U_x]^T dx = 0$ and $[A_x]^T dx = 0$, in all of the directions in which the control variables can be freely varied, it is always possible (at least in the neighborhood of the steady state) to eliminate k from either the objective function U or the accumulation function A, whichever is chosen.[4]

When k is only in the Objective Function ($A_k \equiv 0$).

When $A_k \equiv 0$, by (5.11),

$$\begin{aligned} H^{\star}_{k\lambda} &= A_k + (U_{kx} + \lambda A_{kx})\frac{dx}{d\lambda} \\ &= U_{kx}\frac{dx}{d\lambda}. \end{aligned}$$

In words, when the dependence on k is concentrated into the objective function U, $H^{\star}_{k\lambda}$ is positive when k and *thrift* are complements in the objective function, but negative when k and *thrift* are substitutes in the objective function. Since the dynamic matrix for this case looks like

$$\begin{bmatrix} U_{kx}\frac{dx}{d\lambda} & + \\ + & \rho - U_{kx}\frac{dx}{d\lambda} \end{bmatrix},$$

the slope of the $\dot{k} = 0$ locus depends on whether k and *thrift* are complements or substitutes in U. The $\dot{k} = 0$ locus is downward-sloping if k and *thrift* are complements, and upward-sloping if k and *thrift* are substitutes.

[4]Since in the neighborhood of the steady state, x can be freely varied in this direction, any dependence of the constraint set on k introduced by this recasting of the model will be irrelevant for dynamics in the neighborhood of the steady state. If both $[A_x]^T dx$ and $[U_x]^T dx$ *are* zero in all of the directions dx in which the control variables can be freely varied, then by concavity, U and A are each separately maximized in those dimensions of x. Except when a constraint on the vector x in some other dimension is exactly at the critical point between binding and not binding (where the balance could be tipped by a change in λ) this means that a change in λ will lead to no change in the value of x needed to maximize $H = U + \lambda A$, so that $\frac{dx}{d\lambda} = 0$. (In other words, there is no appropriate way for the agent to alter actions for increased thrift.) That makes this a very special case. (See Exercise **??** for more details.)

The $\dot{\lambda} = 0$ locus is downward-sloping if k and *thrift* are substitutes, and upward-sloping if k and *thrift* are powerful enough complements to overcome ρ.

When k is only in the Accumulation Function ($U_k \equiv 0$).

In what is often the more natural case when $U_k \equiv 0$,

$$\begin{aligned} H^{\star}_{k\lambda} &= A_k + (U_{kx} + \lambda A_{kx}\frac{dx}{d\lambda} \\ &= A_k + A_{kx}\frac{dx}{d\lambda}. \end{aligned}$$

Since at the steady state, $\dot{\lambda} = 0$ implies $\rho = \frac{U_k + \lambda^* A_k}{\lambda^*} = A_k$,

$$\rho - H^{\star}_{k\lambda} = -A_{kx}\frac{dx}{d\lambda}$$

and the dynamic matrix looks like

$$\begin{bmatrix} \rho + A_{kx}\frac{dx}{d\lambda} & + \\ + & -A_{kx}\frac{dx}{d\lambda} \end{bmatrix}.$$

Thus, the slope of the $\dot{\lambda} = 0$ locus depends on whether k and *thrift* are complements or substitutes in the accumulation function A. The $\dot{\lambda} = 0$ locus is upward-sloping if k and *thrift* are complements in A, and downward-sloping if k and *thrift* are substitutes in A.

The $\dot{k} = 0$ locus is downward-sloping if k and *thrift* are complements in A, and upward-sloping if k and *thrift* are powerful enough substitutes in A to overcome ρ.

5.4.4 Summary of the Dynamic Types

Looking again at Figure 2, dynamic type I can be thought of either as a case of complementarity between k and *thrift* in accumulation or of strong enough complementarity between k and *thrift* in the objective function to overcome the discount rate ρ. To encompass both views, we label case I "Thrift-k Complementarity Dominant." Dynamic type II can be thought of either as a case of moderate substitutability between k and *thrift* in accumulation or moderate complementarity between k and *thrift* in the objective function U overwhelmed by ρ. To encompass both views, we label case II "Discount

Rate Dominant," divided into the "Stable" and "Unstable" case. Dynamic type III can be thought of as a case of substitutability between k and *thrift* in the objective function U or strong enough substitutability between k and *thrift* in the accumulation function A to overcome the discount rate ρ. To encompass both views, we label case III "Thrift-k Substitutability Dominant." The borderline case (I/II) can be thought of as a case of thrift-k independence in the accumulation function. The borderline case (II/III) can be thought of as a case of thrift-k independence in the objective function.

5.5 Examples of the Dynamic Types

5.5.1 Cass, Prescott and Mankiw

The Cass Model, the Basic Real Business Cycle Model and Mankiw's model with consumer durables form a trio of models with k only in the accumulation function.

In the Cass Model, increased *thrift* consists only of a reduction in consumption. This reduction in consumption has a linear effect on accumulation and so is neither a complement nor a substitute to k. Thus, the Cass model manifests thrift-k independence and the $\dot{\lambda} = 0$ locus is vertical (I/II).

In the Basic Real Business Cycle Model, *thrift* consists of a combination of reduced consumption and increased labor. As in the Cass model, the reduced consumption has no interaction with k in the accumulation function, but the increased labor is complementary with k in the accumulation function. Thus, the Basic RBC Model manifests thrift-k complementarity and the $\dot{\lambda} = 0$ locus is upward-sloping (I).

In Mankiw's Model with consumer durables, *thrift* consists of a reduction in both durable and nondurable consumption. The reduction in nondurable consumption has only a linear effect on accumulation, no interaction with the stock of machines k (Mankiw's W). However, the reduction in consumption of the services of consumer durables allows them to be (costlessly) reconverted to use as capital. This has the same effect on the accumulation function as if the total stock of machines (the sum of consumer durables and those in use as capital) increased. Thus, thrift and an increase in the stock of machines k are substitutes. Mankiw's Model exhibits thrift-k substitutability in the accumulation function and so the $\dot{\lambda} = 0$ locus is backward sloping. If the discount rate ρ is large relative to this substitutability, it will be the ρ-dominant dynamic type II. If the discount rate ρ is small relative to this substitutability, Mankiw's model will be of dynamic type III.

5.5.2 A Model with Unstable Dynamics

A good example of a model with unstable dynamics (IIB) is a modified version of Mankiw's Model. First, for simplicity, let the agent consume *only* durables. Second, allow for the possibility that k, instead of being a stock of machines that depreciates, is a biological stock with a positive natural rate of increase. The rate of increase can be enhanced by procedures that require labor (which is in inelastic supply) and which interfere with use as a consumer durable. Use of the stock by the government for its purposes actually uses up the stock. Formally, the model is

$$\max_x \int_0^\infty e^{-\rho t} u(k-x)\,dt$$

s.t.

$$\dot{k} = \mu k + f(x) - g.$$

Both u and f are increasing and concave and $\mu > 0$. (Note that if μ were negative, this would be a special case of Mankiw's model, since we could set $-\delta = \mu$.) $u'(0) = +\infty$ insures that we need not worry about the implicit constraint $x \le k$, but let us make the constraint $x \ge 0$ explicit in case $f'(0)$ is finite. The current-value Hamiltonian is

$$H = u(k-x) + \lambda[\mu k + f(x) - g].$$

Rewriting the model in terms of the part of the stock used for consumer durables $d = k - x$ would allow one to eliminate x and concentrate all of the dependence on k into the accumulation function, demonstrating the thrift-k substitutability in the accumulation function that guarantees that the $\dot{\lambda} = 0$ locus is downward-sloping. (The recast Hamiltonian would be $H = u(d) + \lambda[\mu k + f(k-d) - g]$.) But for the rest of the analysis, it is best to focus on the behavior of the part of the stock involved in enhanced production x. The first order condition for x can be written

$$u'(k-x) = \lambda f'(x)$$

when $k > 0$, with $u'(k-x) \ge \lambda f'(x)$ when $x = 0$. It is clear from the first order condition that an increase in λ leads to an increase in x: $\frac{dx}{d\lambda} \ge 0$. The Euler equation is

$$\dot{\lambda} = (\rho - \mu)\lambda - u'(k-x) = \lambda[\rho - \mu - f'(x)],$$

where the last equality holds as long as $x > 0$. Exercise 9 gives an example showing that this model can have a steady state for at least some functional forms and parameter values.

The cross-derivative of $H^\star$ is easily calculated as

$$\begin{aligned} H^\star_{k\lambda} &= A_k + (U_{kx} + \lambda A_{kx})\frac{dx}{d\lambda} \\ &= \mu - u''(k-x)\frac{dx}{d\lambda} \\ &> 0. \end{aligned}$$

Thus, the $\dot{k} = 0$ locus is also downward-sloping, putting the dynamics squarely in type II.

The remaining issue is whether the dynamics are stable or unstable. The dynamics will be unstable if and only if the $\dot{k} = 0$ locus is steeper than the $\dot{\lambda} = 0$ locus. In the stable case, with the $\dot{\lambda} = 0$ locus steeper than the $\dot{k} = 0$ locus, points on the $\dot{k} = 0$ locus to the right of the steady state will have $\dot{\lambda} > 0$. However, in the unstable case, with both loci downward sloping, the $\dot{k} = 0$ locus to the right of the steady state is to the left of the $\dot{\lambda} = 0$ locus and so has $\dot{\lambda} < 0$. Which case are we dealing with here? In order to hold $\dot{k} = \mu k + f(x) - g$ constant as k increases, x must fall and

$$\frac{\dot{\lambda}}{\lambda} = \rho - \mu - f'(x)$$

must also fall. Since $\dot{\lambda} = 0$ at the steady state, in that neighborhood a fall in $\frac{\dot{\lambda}}{\lambda}$ requires a fall in $\dot{\lambda}$ itself. Thus, the dynamics are unstable. (See Figure 3.)

Interpreting Unstable Dynamics

With unstable dynamics, it is clear that one will not end up at the steady state if one does not have the steady state capital stock to begin with. To make sense of the model's dynamics, it is best to think of the corresponding finite horizon problem with end-time T and then take the limit as $T \to \infty$. With a finite end-time T, the transversality condition $\lambda(T)k(T) = 0$ must be satisfied. Looking at the dynamic paths shown in Figure 3, if initially, $k > k^*$ the way to satisfy this condition when T is large is have λ start very slightly below the downward-sloping explosive path, follow along very close to the downward-sloping explosive path for a long time, then veer off down to the $\lambda = 0$ axis near T. Thus, as $T \to \infty$, the appropriate dynamics are to follow the downward-sloping explosive path (which is also the slower explosive path) more and more closely. In some ways this downward-sloping explosive path acts like a saddle-path even though it is not a saddle path. If, initially, $k < k^*$, as $T \to \infty$ it may be impossible to stay away from both $k = 0$ and $\lambda = 0$ for more than a finite length of time. If the axes are far away, it will still be appropriate to follow the downward-sloping saddle-path for some time before veering down toward $\lambda = 0$.[5]

Exercises

1. Using the proof of Lemma 1 as a guide, prove that maximization of the Hamiltonian $U(k, x) + \lambda A(k, x)$ implies that $U(k, x)$ falls with λ (given k).

2. If you had a model in which more of the state variable was always undesirable, how could you recast the model in order to fit it into the framework for characterizing phase diagrams in this chapter? What does this mean for the original way of looking at the model?

3. Draw the three borderline cases for dynamics: (I/II), (II/III) and the borderline stable case. What are the dynamics like in the borderline stable case?

4. The slope of the saddle-path and of the explosive path.

    (a) Verify that the explosive path is upward sloping for each case by graphing the intersection of each "quadrant" divided up by the $\dot{k} = 0$ and $\dot{\lambda} = 0$ loci with the right-angled area made by the arrow-angle for that quadrant with the vertex put at the steady state. Explain why the explosive path must lie in this region. Show that in each of the stable cases, every point in this region is above and to the right or below and to the left of the steady state, implying that the explosive path is upward sloping.

    (b) Verify that the saddle-path must be downward sloping for each stable case by graphing the intersection of each quadrant with the right-angled area made by the opposite of the arrow angle for that

[5] One can conjecture that when T is large enough, the optimal thing will be to have λ jump to the path that leads to the origin and follow that path, as the way to spend the most possible time before reaching $\lambda = 0$.

quadrant with the vertex of the anti-arrow angle put at the steady state. Explain why the saddle path must lie in this region.

(c) Verify that the saddle-path is downward-sloping and the explosive path upward-sloping using matrix algebra.

5. Graph the effect of shifting the $\dot{k} = 0$ locus up in each of the three stable cases (I, IIA and III). In each case, examine the effects of a permanent, of a temporary and of an anticipated upward shift in the $\dot{k} = 0$ locus. Think about the borderline cases as well.

6. Graph the effect of shifting the $\dot{\lambda} = 0$ locus *out* in each of the three stable cases (I, IIA and III). In each case, examine the effects of a permanent, of a temporary and of an anticipated outward shift in the $\dot{\lambda} = 0$ locus. Think about the borderline cases as well.

7. Looking at $H^\star_{kk}$ and $H^\star_{\lambda\lambda}$ with matrix algebra.

   (a) Show that when all the control variables in x are at an interior solution, $H^\star_{kk}$ and $H^\star_{\lambda\lambda}$ can be expressed as

$$H^\star_{kk} = H_{kk} - H_{kx}[H_{xx}]^{-1}H_{xk} \leq 0 \tag{5.13}$$

$$\begin{aligned} H^\star_{\lambda\lambda} &= -H_{\lambda x}[H_{xx}]^{-1}H_{x\lambda} \\ &= -[A_x]^T[U_{xx} + \lambda A_{xx}]^{-1}A_x \\ &\geq 0, \end{aligned} \tag{5.14}$$

   where $[A_x]^T$ is the transpose of A_x.

   (b) Use matrix algebra to confirm that if

$$\begin{bmatrix} H_{kk} & H_{kx} \\ H_{xk} & H_{xx} \end{bmatrix}$$

   is negative definite as required by concavity of H in k and x, then the expression in (5.13) for $H^\star_{kk}$ is negative, and the expression for $H^\star_{\lambda\lambda}$ in (5.14) is positive. Note that though it helps in making this proof using matrix algebra, the concavity of H is not necessary to establish that $H^\star_{\lambda\lambda}$ is positive.

8. Suppose that both $[A_x]^T dx$ and $[U_x]^T dx$ are zero vectors at the steady state.

(a) Write down a model in which this would happen automatically and generically, not just by chance. (It is possible to do the rest of this exercise even if you cannot answer this part.)

(b) Show that U and A are separately maximized over x and explain why, starting from the steady state, a change in λ will lead to no change in the value of x needed to maximize $H = U + \lambda A$, so that $\frac{dx}{d\lambda} = 0$.

(c) Show that the dynamic matrix for the neighborhood of the steady state is

$$\begin{bmatrix} A_k & 0 \\ -H^{\star}_{kk} & -\frac{U_k}{\lambda^*} \end{bmatrix}.$$

(d) Classify, graph and describe the different types of dynamics possible in this case. Are the dynamics for this case always stable, or is it possible to get unstable dynamics?

9. Show that the model

$$\max_x \int_0^\infty e^{-1.5t}\sqrt{k - x}\, dt$$

s.t.

$$\dot{k} = k + \sqrt{x} - 2$$

has a steady state at $(k^*, \lambda^*) = (2, 1)$.

Chapter 6

More on Characterizing Phase Diagrams

6.1 N State Variables

Consider the general dynamic optimization problem discussed in the previous chapter,

$$\max_x \int_0^\infty e^{-\rho t} U(k, x; z) dt,$$

s.t.

$$\dot{k} = A(k, x; z) \tag{6.1}$$

and

$$x \in X(k, z),$$

but with k and A interpreted as vectors. Again, assume that the objective function U and the accumulation function A are (twice-differentiable and) concave and that each component of k has nonnegative value. The current-value Hamiltonian is

$$H(k, \lambda, x) = U(k, x) + \lambda^T A(k, x),$$

where λ is the vector of costate variables, the marginal values for each component of k, and λ^T is its tranpose. The term $\lambda^T A(k, x)$ can be thought of also as the vector dot product $\lambda \cdot A(k, x)$. Defining the *maximized* Hamiltonian $H^\star$ as before by

$$H^\star(k, \lambda) = \max_x \; H(k, x, \lambda),$$

the dynamics of the system are described by the pair of vector equations

$$\begin{aligned} \dot{k} &= H_\lambda(k, x, \lambda) = H^\star_\lambda(k, \lambda) && (6.2) \\ \dot{\lambda} &= \rho\lambda - H_k(k, x, \lambda) = \rho\lambda - H^\star_k(k, \lambda). && (6.3) \end{aligned}$$

6.1.1 The Saddle Shape of $H^\star$

The dependence of $H^\star$ on k and λ can be characterized as follows:

Lemma 2 *$H^\star(k, \lambda)$ is concave in k as long as $H(k, x, \lambda)$ is jointly concave in k and x; $H^\star(k, \lambda)$ is convex in λ even without this assumption.*

This result is akin to the concavity of a firm's profit function in the vector of quasi-fixed factors and convexity in factor and output prices. In this comparison the vector k corresponds to the quasi-fixed factors, the vector λ corresponds to the prices and the vector x corresponds to the variable factors.

Proof.

The proofs of the one-state variable lemma (1) can be followed line by line to prove the multi-state-variable version of the lemma (as you are asked to do in Exercise **??**). This approach directly establishes the concavity of $H^\star$ in k. For λ one obtains the result

$$\begin{aligned} [\lambda_1 - \lambda_0]^T [H^\star_\lambda(k, \lambda_1) - H^\star_\lambda(k, \lambda_0)] &= [\lambda_1 - \lambda_0]^T [A(k, \lambda_1) - A(k, \lambda_0)] \\ &\geq 0, \end{aligned}$$

for any pair of costate vectors λ_0 and λ_1, which is equivalent to convexity of $H^\star$ in λ.

Interpretation

By the same token, the concavity of $H^\star$ in k is equivalent to

$$[k_1 - k_0]^T [H_k^\star(k_1, \lambda) - H_k^\star(k_0, \lambda)] \leq 0.$$

Thus, one can say that a change in the costate vector λ changes $\dot{k}$ in a direction with a positive dot product to the change in λ, while a change in the vector k changes $\dot{\lambda}$ in a direction with a positive dot product to the change in k. In particular, this means that an increase in one scalar component of λ by itself will raise the rate of increase of the corresponding state variable, while an increase in any scalar component of k by itself will raise the rate of increase in the corresponding costate variable.

In a two-dimensional slice of the 2N-dimensional phase diagram with the component state variable k_i on the horizontal axis and the corresponding component costate variable λ_i on the vertical axis, rules like those stated in the previous chapter apply: above the $\dot{k}_i = 0$ locus, $\dot{k}_i > 0$ while $\dot{k}_i < 0$ below; to the right of the $\dot{\lambda}_i = 0$ locus, $\dot{\lambda}_i > 0$ while $\dot{\lambda}_i < 0$ to the left. Moreover, the same can be said for any linear combination of component state variables and the same linear combination of component costate variables.

6.1.2 The Dynamic Matrix

The dynamics in the neighborhood of the steady state for the N-state-variable case are given by

$$\begin{bmatrix} \dot{\tilde{k}} \\ \dot{\tilde{\lambda}} \end{bmatrix} = \begin{bmatrix} H_{\lambda k}^\star & H_{\lambda\lambda}^\star \\ -H_{kk}^\star & \rho I - H_{k\lambda}^\star \end{bmatrix} \begin{bmatrix} \tilde{k} \\ \tilde{\lambda} \end{bmatrix}. \tag{6.4}$$

$$H_{k\lambda} = (H_{\lambda k})^T.$$

The key characteristics of the 2Nx2N *dynamic matrix* in (6.4) are (1) by Lemma 2 above, both off-diagonal elements ($H_{\lambda\lambda}^\star$ and $-H_{kk}^\star$) are positive semidefinite; and (2) the following lemma:

Lemma 3 *All of the eigenvalues of the dynamic matrix come in pairs adding up to ρ.*

Lemma 3 has at least four important consequences.

First, it means that the trace of the dynamic matrix is Nρ, since there are N pairs.

Second, Lemma 3 means that if ρ is small enough, there will be N positive roots (or roots with positive real parts) and N negative roots (or roots with negative real parts). As one considers larger values of ρ, the chance of having more positive roots than negative roots increases.

Third, the fact that the characteristic polynomial is real means that complex roots must come in conjugate pairs. A conjugate pair can only add up to ρ if the two roots are of the form $\frac{\rho}{2} \pm bi$. Otherwise, a complex root can only occur as a member of a quartet of the form $a \pm bi$, $\rho - a \pm bi$. If $a > \rho$, one pair of roots indicates a path spiralling in while the other pair represents a path spiralling out.

Fourth, as is apparent from the proof, the characteristic polynomial is a function of $(\vartheta - \frac{\rho}{2})^2$. In particular, in the two-state-variable case, the characteristic equation is a quadratic in $(\vartheta - \frac{\rho}{2})^2$ and so can be solved explicitly by means of the quadratic formula (as opposed to the *much* more difficult quartic formula necessary to explicitly solve an arbitrary fourth order polynomial equation).

Proof.

Writing $\mathcal{M}$ for the dynamic matrix, it is clear that the eigenvalues of $\mathcal{M}$ come in pairs adding up to ρ if and only if the eigenvalues of $\mathcal{M} - \frac{\rho}{2}I$ come in pairs adding up to zero. That is, what we need to show is that the characteristic polynomial of $\mathcal{M} - \frac{\rho}{2}I$ is symmetric around zero.

We can use the result from matrix algebra that the determinant of a partitioned matrix

$$\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}$$

is equal to the determinant of

$$\begin{bmatrix} I & 0 \\ -\mathcal{D}\mathcal{B}^{-1} & I \end{bmatrix} \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\mathcal{B}^{-1}\mathcal{A} & I \end{bmatrix} = \begin{bmatrix} 0 & \mathcal{B} \\ \mathcal{C} - \mathcal{D}\mathcal{B}^{-1}\mathcal{A} & 0 \end{bmatrix}.$$

When each of the four submatrices is NxN, this determinant is in turn equal to

$$(-1)^{\mathrm{N}} |\mathcal{B}|\, |\mathcal{C} - \mathcal{D}\mathcal{B}^{-1}\mathcal{A}|.$$

Even when $\mathcal{B}$ is singular, since a determinant is a continuous functions of a matrix,

$$\begin{vmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{vmatrix} = \lim_{\epsilon \to zero} (-1)^{\mathrm{N}} |\mathcal{B} + \epsilon I|\, |\mathcal{C} - \mathcal{D}(\mathcal{B} + \epsilon I)^{-1} \mathcal{A}|.$$

Using χ as the variable for the characteristic polynomial of $\mathcal{M} - \frac{\rho}{2} I$ to keep the roots clearly distinguished from the roots of $\mathcal{M}$ itself, the characteristic polynomial of $\mathcal{M} - \frac{\rho}{2} I$ is therefore

$$\begin{aligned} |\mathcal{M} - (\frac{\rho}{2} + \chi) I| &= \begin{vmatrix} H^\star_{\lambda k} - \frac{\rho}{2} I - \chi I & H^\star_{\lambda\lambda} \\ -H^\star_{kk} & \frac{\rho}{2} I - \chi I - H^\star_{k\lambda} \end{vmatrix} \\ &= (-1)^{\mathrm{N}} |H^\star_{\lambda\lambda}|\, | - H^\star_{kk} + (H^\star_{k\lambda} - \frac{\rho}{2} I + \chi I)[H^\star_{\lambda\lambda}]^{-1} (H^\star_{\lambda k} - \frac{\rho}{2} I - \chi I)| \end{aligned}$$

Then since a matrix and its tranpose have the same determinant, $H^\star_{kk}$ and $H^\star_{\lambda\lambda}$ are symmetric, and $H^\star_{k\lambda} = (H^\star_{\lambda k})^T$,

$$|-H^\star_{kk} + (H^\star_{k\lambda} - \frac{\rho}{2} I + \chi I) H^\star_{\lambda\lambda} (H^\star_{\lambda k} - \frac{\rho}{2} I - \chi I)| = |-H^\star_{kk} + (H^\star_{k\lambda} - \frac{\rho}{2} I - \chi I) H^\star_{\lambda\lambda} (H^\star_{\lambda k} - \frac{\rho}{2} I + \chi I)|,$$

implying that the characteristic polynomial of $\mathcal{M} - \frac{\rho}{2} I$ is unchanged when χ is replaced by $-\chi$.

Note that the characteristic polynomial of $\mathcal{M} - \frac{\rho}{2} I$ is therefore a function of χ^2. The eigenvalues of $\mathcal{M}$ itself are related by $\vartheta = \frac{\rho}{2} + \chi$, so that the characteristic polynomial of $\mathcal{M}$ itself is a function of $(\vartheta - \frac{\rho}{2})^2$.

6.2 More on the One-State-Variable Case

There are a number of other useful facts about one-state-variable phase diagrams that hold in great generality. We will state these facts here so that they can at least be referred to later.

6.2.1 Comparative Discount Rate Analysis

It is often reasonable to think that the utility discount rate ρ is quite small in relation to the other forces in a model. Therefore it is valuable to study the limiting case as $\rho \to 0$. With few exceptions one can even study the case when $\rho = 0$ (or in a model with growth in the background, when $\hat{\rho} = 0$) directly using an overtaking criterion to deal with the fact that the integral of the objective function does not converge, and the dynamics in this case are identical to the limit as $\rho \to 0$ from above.

To understand the role of the discount rate ρ in the model, it is important to see that ρ affects only the dynamics. Given k and λ, the solution to

$$H^\star(k, \lambda) = \max_x \; U(k, x) + \lambda A(k, x)$$

is the same regardless of the value of ρ. The optimizing value of x, and the values of $H^\star$, U, and A are all fixed by a knowledge of k and λ alone, without any reference to ρ. Since $\dot{k} = A$, this means that the $\dot{k} = 0$ locus is unaffected by changes in ρ. On the other hand, since

$$\dot{\lambda} = \rho\lambda - H_k^\star(k, \lambda),$$

an increase in ρ raises $\dot{\lambda}$ at a given point (k, λ) on the phase diagram. Points that were on the old $\dot{\lambda} = 0$ locus will have $\dot{\lambda} > 0$; the new $\dot{\lambda} = 0$ locus will be to the left of the old one, since a reduction in k lowers $\dot{\lambda}$.

As long as the dynamics are stable, the leftward shift in the $\dot{\lambda} = 0$ accompanying an increase in ρ always reduces the steady state capital stock k^*. The effect of the increase in ρ on λ^* depends on the slope of the unmoving $\dot{k} = 0$ locus.

If the dynamics are unstable, the leftward shift of the $\dot{\lambda} = 0$ locus accompanying an increase in ρ *raises* the steady-state capital stock k^*. This is more intuitive than it may seem at first. In the unstable case, greater impatience as measured by ρ raises the critical level of capital at which the economy can "take off" on a rightward path of increasing k—making such a take-off more difficult and making a gradual collapse of the economy on a leftward path of shrinking k more likely.

6.2.2 The Isograms of the Hamiltonian

"Isogram" is the general word for any curve on which some quantity is constant. Isoquants, indifference curves, and the $\dot{k} = 0$ and $\dot{\lambda} = 0$ loci are all examples of isograms. There is a remarkable connection between the dynamics of a dynamic control model with $\rho = 0$ and the isograms of the Hamiltonian, or iso-Hamiltonian curves. In general,

$$\dot{k} = H_\lambda^\star(k, \lambda)$$

and

$$\dot{\lambda} = \rho\lambda - H_k^\star(k, \lambda)$$

imply that

$$\begin{aligned}\dot{H} &= \dot{H}^\star \\ &= H_k^\star(k,\lambda)\dot{k} + H_\lambda^\star(k,\lambda)\dot{\lambda} \\ &= H_k^\star(k,\lambda)H_\lambda^\star(k,\lambda) + H_\lambda^\star(k,\lambda)[\rho\lambda - H_k^\star(k,\lambda)] \\ &= \rho\lambda H_\lambda^\star(k,\lambda) \\ &= \rho\lambda\dot{k}.\end{aligned} \tag{6.5}$$

When $\rho = 0$, $\dot{H} = 0$—the Hamiltonian is constant. In other words, the dynamic paths ("flow lines") followed by the model when $\rho = 0$ are the isograms for the Hamiltonian (the "iso-Hamiltonian curves"). As iso-Hamiltonian curves, these curves are unchanged as curves in $k - \lambda$ space if ρ is increased to a positive number, although the dynamics change with ρ so that the iso-Hamiltonian curves are then no longer flow lines.

When $\rho > 0$, the value of the Hamiltonian rises and falls with k when following a flow line. On the saddle path this means that H is rising when going down (and right) along the saddle path. H is falling when going up (and left) along the saddle path. Other flow lines either have H continually increasing or continually decreasing along with k or have a maximum or minimum value of H where the flow line crosses the $\dot{k} = 0$ locus.

There is one more intriguing fact about $H^\star$. Since $H_{kk}^\star \leq 0$ and $H_k^\star = \rho\lambda - \dot{\lambda}$, the $\dot{\lambda} = 0$ locus for $\rho = 0$ is the solution to

$$\max_k \; H^\star(k, \lambda)$$

for any given λ. Similarly, since $H_{\lambda\lambda}^\star \geq 0$ and $H_\lambda^\star = \dot{k}$, the $\dot{k} = 0$ locus is the solution to

$$\min_\lambda \; H^\star(k, \lambda)$$

for any given k. Thus, the steady state when $\rho = 0$ can be viewed as the equilibrium in a zero-sum game with payoff $H^\star$ with one player controlling k and the other player controlling λ.

6.2.3 The Isograms of the Objective Function U

It is often useful to know the behavior of U. We can state several quite general results.

The Saddle Path and U

First, going down along the saddle path (to the right), the Hamiltonian $U+\lambda A$ is increasing, since with $\rho > 0$, the Hamiltonian moves in the same direction as k along any flow line, including the saddle path. But going down the saddle path (to the right), λA is decreasing both to the left of the steady state and for at least some distance to the right of the steady state (since λA must decline from 0 to something negative on the saddle path immediately to the right of the steady state). Therefore, in the neighborhood of the steady state, the objective function U must be increasing when going down (to the right) along the saddle path to allow the Hamiltonian to be increasing, and the objective function U is decreasing when going up (to the left) along the saddle path in the neighborhood of the steady state.

λ and U

Second, if we define the stripped Hamiltonian G and its maximized counterpart $G^\star$ by

$$G(k,x,\ell) = G(k,x,\frac{1}{\lambda}) = \frac{U(k,x)}{\lambda} + A(k,x) = \ell U(k,x) + A(k,x)$$

($\ell = \frac{1}{\lambda}$) and

$$G^\star(k,\ell) = \max_x \ \ell U(k,x) + A(k,x),$$

then the same value of x maximizes the stripped Hamiltonian G as maximizes the regular Hamiltonian H. Writing the value of the objective function associated with maximizing H and G as $U^\star(k,\lambda)$, that is

$$U^\star(k,\lambda) = U(k, x^\star(k,\lambda)),$$

the envelope theorem tells us that

$$U^\star(k,\lambda) = G^\star_\ell(k,\ell) = G^\star_\ell\left(k, \frac{1}{\lambda}\right).$$

Thus, $G^\star_\ell$ tell us the value of the objective function at each point on the phase diagram. The set of curves on which U is constant ("iso-U curves") is identical to the set of curves on which $G^\star_\ell(k,\frac{1}{\lambda})$ is constant.

The $\frac{\dot\lambda}{\lambda}$ = constant Curves and U

Third, the envelope theorem also tells us that

$$\frac{\dot\lambda}{\lambda} = \rho - \frac{H_k^\star(k,\lambda)}{\lambda} = \rho - G_k^\star\left(k, \frac{1}{\lambda}\right).$$

Thus, the set of curves on which $\frac{\dot\lambda}{\lambda}$ is constant is identical to the set of curves on which $G_k^\star(k, \frac{1}{\lambda})$ is constant. This set of curves—which we will call the set of $\frac{\dot\lambda}{\lambda}$ isograms—is the same regardless of ρ; a change in ρ changes only the values of $\frac{\dot\lambda}{\lambda}$ labeling each isogram, not its shape.

Clearly, by the same proof we used for Lemma 1, $G_{\ell\ell}(k,\ell) \geq 0$ and

$$U_\lambda^\star(k,\lambda) = -\frac{1}{\lambda^2} G_{\ell\ell}^\star\left(k, \frac{1}{\lambda}\right) \leq 0.$$

Thus, U falls when λ increases. (See also Exercise 1.)

Moreover,

$$\begin{aligned} U_k^\star(k,\lambda) &= G_{\ell k}^\star\left(k, \frac{1}{\lambda}\right) \\ &= G_{k\ell}^\star\left(k, \frac{1}{\lambda}\right) \\ &= \frac{d}{d\ell}[G_k^\star - \rho] \\ &= \frac{d\lambda}{d\ell}\frac{d}{d\lambda}\left[-\frac{\dot\lambda}{\lambda}\right] \\ &= -\frac{1}{\ell^2}\frac{d}{d\lambda}\left[-\frac{\dot\lambda}{\lambda}\right] \\ &= \lambda^2 \frac{d}{d\lambda}\left[\frac{\dot\lambda}{\lambda}\right]. \end{aligned} \tag{6.6}$$

The symmetry relationship (6.6) implies a relationship between $\frac{\dot\lambda}{\lambda} =$ constant curves on which $G_k^\star(k, \frac{1}{\lambda})$ is constant and $U =$ constant curves on which $G_\ell^\star(k, \frac{1}{\lambda})$ is constant. Since $d\ell = -\frac{d\lambda}{\lambda^2}$, the slope of a $\frac{\dot\lambda}{\lambda}$ isogram can be found by setting

$$G_{kk}^\star dk + G_{k\ell}^\star\left(\frac{-d\lambda}{\lambda^2}\right) = 0,$$

implying

$$\frac{d\lambda}{dk} = \lambda^2 \frac{G^\star_{kk}}{G^\star_{k\ell}}$$

along a $\frac{\dot{\lambda}}{\lambda}$ isogram. Similarly, the slope of an iso-U curve can be found by setting

$$G^\star_{\ell k} dk + G^\star_{\ell\ell} \left(\frac{-d\lambda}{\lambda^2} \right) = 0,$$

implying

$$\frac{d\lambda}{dk} = \lambda^2 \frac{G^\star_{\ell k}}{G^\star_{\ell\ell}}$$

along an iso-U curve. Since $G^\star_{kk}$ is negative, while $G^\star_{\ell\ell}$ is positive, an iso-U curve must have a slope of opposite sign to the $\frac{\dot{\lambda}}{\lambda}$ it is crossing.

Where the $\frac{\dot{\lambda}}{\lambda}$ isograms are vertical, the iso-U curves must be horizontal. The best example is the Cass model, where $\frac{\dot{\lambda}}{\lambda} = \rho + \delta - f'(k)$ is constant if k is constant, while consumption and therefore $u(c)$ is constant if λ is constant.

In the neighborhood of the steady state, iso-U curves must slope down if the $\dot{\lambda} = 0$ (and $\frac{\dot{\lambda}}{\lambda} = 0$ locus) slopes up; while iso-U curves in the neighborhood of the steady state must slope up if the $\dot{\lambda} = 0$ locus slopes down. If the $\dot{\lambda} = 0$ locus is vertical, then the iso-U curve through the steady state must be horizontal.

The $\dot{k} = 0$ Locus and U

Fourth, along the $\dot{k} = 0$ locus, $A = 0$, so that the Hamiltonian $U + \lambda A$ is equal to the objective function U. Since $H^\star_\lambda = A = 0$ along the $\dot{k} = 0$ locus, and $H^\star_k = \rho\lambda - \dot{\lambda}$, both H and U are increasing as one goes right along the $\dot{k} = 0$ locus until one reaches the steady-state for $\rho = 0$. Beyond that "dynamic efficiency boundary," $H = U$ decreases as one goes further to the right along the $\dot{k} = 0$ locus. The $\rho = 0$ steady state, is the "Golden Rule" steady state that maximizes steady state U. Higher values of ρ yield steady states on the $\dot{k} = 0$ locus to the left of the $\rho = 0$ steady state with lower values of U.

Clearly, the iso-U curve through the steady state for $\rho = 0$ must be tangent to the $\dot{k} = 0$ locus at that point. In the neighborhood of any steady state for $\rho > 0$, the iso-U curve must slope up more than the $\dot{k} = 0$ locus in order for an increase in λ to bring U down enough to counteract the increase in U that one gets going right along the $\dot{k} = 0$ locus.

The Road to Oblivion City and U

In dynamic type I, in which the $\dot{\lambda} = 0$ locus is upward sloping, the iso-U curve must be downward-sloping by the third principle, but slope down less than the $\dot{k} = 0$ locus by the fourth principle. This sandwiching of the iso-U curve through the steady state between a horizontal line and the downward-sloping $\dot{k} = 0$ locus pins things down pretty well.

What of those cases (II and III) in which the $\dot{\lambda} = 0$ locus is downward-sloping? The iso-U curve through the steady state must be upward-sloping and steeper than the $\dot{k} = 0$ curve (a fact that bites only in dynamic type III in which the $\dot{k} = 0$ locus slopes upward). But is there any limit to how steep the upward slope of the iso-U curve through the steady state can be? Indeed there is.

The fifth principle about iso-U curves is that, in the stable case, starting from the steady state, the iso-U curve must be flatter or slope up less than the explosive eigenvector. This sandwiches the iso-U curve through the steady state either between a horizontal line and the explosive path or an upward-sloping $\dot{k} = 0$ locus and the explosive path.

To begin with, the explosive path is a flow line. By (6.5), along a flow line,

$$\dot{H} = \dot{U} + \dot{\lambda}A + \lambda\dot{A} = \rho\lambda\dot{k}.$$

Therefore, using the fact that $\dot{k} = A$,

$$\dot{U} = (\rho\lambda - \dot{\lambda})A - \lambda\dot{A}.$$

Considering the stable case first, and writing $-\kappa$ for the negative eigenvalue, the positive eigenvalue must be $\rho + \kappa$ since the sum of the eigenvalues equals the trace ρ. Therefore, on the explosive path in the neighborhood of the steady state, $\dot{A} = (\rho + \kappa)A$, so that

$$\dot{U} = -(\kappa\lambda + \dot{\lambda})A.$$

Since $\dot{\lambda}$ is small in the neighborhood of the steady state, $\dot{U}$ has the same sign as $-\kappa A$. In the stable case, this means that U declines when the explosive path is followed to the right, while U rises when the explosive path is followed to the left. In either direction, U is negatively related to k when going along the explosive path. The iso-U curve must be flatter so that the fall in U one would get by going along the explosive path to the right is counteracted by a reduction in λ and corresponding shot in the arm for U.

In the unstable case, we can use the smaller positive eigenvalue $-\kappa$ and the equations above will still be valid. But in this case, U *rises* when following the explosive path to the right. Therefore, in the unstable case, the iso-U curve through the steady state must be steeper than the explosive path.

Exercises

1. Spell out in detail the proof of Lemma 2, following the pattern of the proof of Lemma 1.

2. Show that if $\begin{bmatrix} \upsilon_k \\ \upsilon_\lambda \end{bmatrix}$ is a right eigenvector of the dynamic matrix with eigenvalue ϑ, then $[\, \upsilon_\lambda^T \quad -\upsilon_k^T \,]$ is a left eigenvector of the dynamic matrix with eigenvalue $\rho - \vartheta$.

Real Business Cycle Theory: A Semiparametric Approach, by Miles S. Kimball

September 7, 2004 at 16:22

Chapter 7

Why Steady State Growth Requires the King-Plosser-Rebelo Form of the Utility Function

What we have done so far makes it clear that the King-Plosser-Rebelo utility function works well with steady-state growth. In this section, we would like to give a rough argument for why the King-Plosser-Rebelo form of the utility function is necessary—why there isn't a more general form of the utility function that would do the trick.[1]

7.1 Labor, Consumption and the Real Wage

Over the course of the last century or two, real wages and per capita consumption have risen by at least an order of magnitude (that is, by a factor of 10) [some data would be good here]. Yet the average workweek N has changed little in comparison with the dramatic increases in real wages and in consumption. Also, the ratio of labor income to consumption $\frac{WN}{C}$ has remained relatively constant despite those dramatic changes. We need to ask what kind of utility function can reproduce these stylized facts.[2]

[1] A more formal proof can be found in King, Plosser and Rebelo's separate appendix.

[2] If the per capita quantity of labor N did have a secular trend, we could not use steady-state analysis at all. N is bounded below by zero and bounded above by the amount of time physically available, and so cannot have an unending trend. In a sense, any drift in N would

The only way for both N and $\frac{WN}{C}$ to be constant is for $\frac{W}{C}$ to be constant as well. For $\frac{W}{C}$ to be constant, the W must be proportional to C and both must have the same growth rate. What kind of utility function is consistent with the real wage W remaining proportional to consumption C, without any change in N?

To formalize the idea that for a given value of N, the real wage W must be proportional to consumption C, we can write

$$W = Cv'(N),$$

where at this point $v'(N)$ can represent an arbitrary positive function of N, though the name we have chosen for this arbitrary function hints at where we are headed. Given a general time-separable utility function $u(C, N)$, under perfect competition, the real wage is

$$W = -\frac{u_N(C, N)}{u_C(C, N)}.$$

Drawing the indifference curves on a graph with N on the horizontal axis and C on the vertical axis as in Figure **??**, for any C and N, the real wage W at that point is equal to the slope of the indifference curve.

On the graph of indifference curves, looking up and down any vertical line with $N =$ constant, the slopes of the indifference curves must be proportional to C. This means that the indifference curves must all be vertical expansions or vertical contractions of each other, as shown in Figure **??**. Formally, let

$$C = \psi(N)$$

be the equation of a typical indifference curve. Then since the slope $\psi'(N)$ equals the real wage, and $W = Cv'(N)$ along this indifference as it does everywhere, $\psi(N)$ must satisfy

$$\frac{\psi'(N)}{\psi(N)} = \frac{W}{C} = v'(N),$$

or

$$\frac{d}{dN} \ln(\psi(N)) = v'(N).$$

Given the arbitrary function $v'(N)$, the solution is

have to be transitional, even if it is a transition that lasts a long time.

$$\ln(\psi(N)) = v(N) + \text{constant}, \tag{7.1}$$

or

$$\psi(N) = \text{constant} \cdot e^{v(N)}.$$

The arbitrary constant allows one to range over all of the indifference curves. Thus, in the vertical dimension, all of the indifference curves are multiples of each other.

The equation (7.1) actually contains the key to representing preferences of this type. Since $\psi(N) = C$ on the typical indifference curve, we can describe the typical indifference curve by

$$\ln(C) - v(N) = \text{constant}.$$

In order to have indifference curves of this type, the utility function must be of the form

$$u(C, N) = U(\ln(C) - v(N)), \tag{7.2}$$

where outer function $U(\cdot)$ must be increasing for consumption to be a good and labor to be a bad.

This is all that is required for labor to be constant in the face of an upward trend in both the real wage and consumption that leaves $\frac{W}{C}$ constant. The exact form of U governs the household's intertemporal choices. Since the relationship between labor, consumption and the real wage is an *intra*temporal (within-period) relationship, it does not depend on the exact form of U.

7.2 Consumption Growth and the Real Interest Rate

The other requirement that narrows the utility function down to the King-Plosser-Rebelo form is the requirement that a constant proportional rate of growth in consumption (with no change in labor) lead to a constant proportional rate of decline in the marginal utility of consumption λ, so that the real interest rate r, which is given by

$$r = \rho - \frac{\dot{\lambda}}{\lambda}$$

and the proportional growth rate of consumption $\frac{\dot{c}}{c}$ can both be constant.

Given $u(c, n) = U(\ln(c) - v(n))$, the marginal utility of consumption λ is

$$\lambda = \frac{1}{c} U'(\ln(c) - v(n)). \tag{7.3}$$

In words, this equation says among other things that, *moving along an indifference curve*—which means holding $\ln(c) - v(n)$ constant—the marginal utility of consumption is inversely proportional to consumption c. Graphically, this is clear from the fact that the vertical distance Δc between any two indifference curves is proportional to c, so that with Δu between the two indifference curves is constant (by definition), $\frac{\Delta u}{\Delta c}$ is proportional to $\frac{1}{c}$.

Since $\dot{n} = 0$ along a steady state growth path, taking logarithms of both sides of (7.3) and differentiating with respect to time yields the following relationship:

$$\begin{aligned} \frac{\dot{\lambda}}{\lambda} &= \frac{d}{dt} \ln(\lambda) \\ &= \frac{d}{dt} [\ln(U'(\ln(c) - v(n))) - \ln(c)] \\ &= \left[\frac{U''(\ln(c) - v(n))}{U'(\ln(c) - v(n))} - 1 \right] \frac{\dot{c}}{c}. \end{aligned}$$

Therefore,

$$r = \rho + \left[1 - \frac{U''(\ln(c) - v(n))}{U'(\ln(c) - v(n))} \right] \frac{\dot{c}}{c} \tag{7.4}$$

along a steady state growth path on which $\dot{n} = 0$.

Given (7.4), the only way both the proportional growth rate of consumption $\frac{\dot{c}}{c}$ and the real interest rate r can be constant along this path is if

$$\frac{U''(x)}{U'(x)}$$

is a constant. Let us call that constant $1 - \beta$, so that the coefficient of $\frac{\dot{c}}{c}$ in (7.4) is equal to β. Then solving the implied differential equation, $U(\cdot)$ must be of the form

$$U(x) = \frac{e^{(1-\beta)x}}{1 - \beta},$$

or

$$U(x) = x$$

if $\beta = 1$ so that $\frac{U''(x)}{U'(x)} = 0$. (The two additional arbitrary constants that arise in solving the differential equation transform the utility function linearly and so do not affect the implied preferences.)

Finally, substituting in $x = \ln(c) - v(n)$, we must have

$$u(c,n) = U(\ln(c) - v(n)) = \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}$$

or

$$u(c,n) = \ln(c) - v(n)$$

if $\beta = 1$. This is the King-Plosser-Rebelo utility function.

Chapter 8

The Ramsey Model via Phase Diagrams

The function of economic models is to provide metaphors that are useful in discussing the economy. As metaphors, economic models are unusual in having a quantitative edge—the subject of part II. But surprisingly often, economists focus on the qualitative behavior of economic models.

Metaphors draw their power from vividness and memorability, aptness and subtlety. When the focus is on the qualitative behavior of a model, graphs are indispensible in making economic models into powerful metaphors. Therefore, an important part of the art of economics is the design and study of graphs. For dynamic economic models, a key type of graph is the phase diagram. This chapter introduces the use of phase diagrams by applying them to the Ramsey model.

8.1 The Ramsey Model Revisited

The undistorted Ramsey model, with government purchases financed by lump-sum taxes, can be concisely represented by the social planner's problem

$$\max_c \int_0^\infty e^{-\rho t} u(c) dt \tag{8.1}$$

s.t.

$$\dot{k} = f(k) - \delta k - c - g. \tag{8.2}$$

As before, c is consumption, ρ is the level of impatience, k is the capital stock, δ is the depreciation rate and g is the exogenous level of government purchases (which have a utility which is assumed to be additively separable from the utility from c). The utility function u has $u'(\cdot) > 0$ and $u''(\cdot) < 0$ and the production function f has $f'(\cdot) > 0$ and $f''(\cdot) < 0$. Labor has been pushed to the background of the model by assuming that the quantity of labor is totally inelastic (and for convenience, the fixed quantity of labor has been normalized to 1).

Using the current-value Hamiltonian

$$H = u(c) + \lambda[f(k) - \delta k - c - g]$$

the first-order condition

$$u'(c) = \lambda \tag{8.3}$$

and Euler equation

$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - f'(k) \tag{8.4}$$

can be readily derived as in Chapter 1.

Figure (8.1) shows the negative relationship between consumption c and the marginal value of capital and marginal utility of consumption λ implied by the function $u'(\cdot)$. Its companion, Figure 8.1 shows the inverse function $u'^{-1}(\cdot)$, which is obtained by reflecting **??** through the 45° line. Both $u'(\cdot)$ and $u'^{-1}(\cdot)$ are decreasing functions. Combining the first-order condition (8.3) the accumulation equation (8.2) yields

$$\dot{k} = f(k) - \delta k - g - u'^{-1}(c). \tag{8.5}$$

If goods are thought of as a flow, a useful image is to think of λ as the tightness of the spigot for the part of the flow which is directed toward consumption. Higher λ tightens the spigot, reducing outflow to consumption and allowing more goods to flow on into investment. Lower λ loosens and opens up the spigot, increasing outflow to consumption and leaving a reduced flow on into investment.

8.2 Constructing the Phase Diagram

A phase diagram is a graphical representation of a differential equation (or difference equation). It shows the evolution of variables over time. In order to construct a two-dimensional phase diagram, it is necessary to choose which two variables to represent on the axes. For the Ramsey model itself, there are two equally good choices—putting k and c or putting k and λ on the axes. (In the Ramsey model, the qualitative effect of switching between $k - \lambda$ space and $k - c$ space is just to flip everything in the graph upside down.) The great advantage of the choice of k and λ is that it generalizes to other models in a way that the choice of k and c does not. State variables and the associated costate variables such as k and λ are intrinsic to the structure of any dynamic optimization problem, while the centrality of consumption as a preeminent control variable is peculiar to the Ramsey model. Being able to compare and contrast different models is crucial to developing the economic intuition. Thus, whenever possible, we will follow a standard convention of putting the state variable (k) on the horizontal axis and the costate variable (λ) on the vertical axis of phase diagrams throughout this book.

The key to the dynamics of the phase diagram are the locus of points at which $\dot{k} = 0$ and the locus of points at which $\dot{\lambda} = 0$. These are sometimes called the k-isocline and the λ-isocline. The intersection of these two curves is the steady state.

8.2.1 The $\dot{\lambda} = 0$ Locus

Setting $\dot{\lambda}$ to zero in (8.4) yields the equation for the $\dot{\lambda} = 0$ locus (the λ-isocline):

$$f'(k) = \rho + \delta. \tag{8.6}$$

Since ρ and δ are constants, and $f'(k)$ is decreasing in k, this equation determines a particular value of k. In particular, if (for use throughout the book) we define

$$\kappa = f'^{-1}(\rho + \delta), \tag{8.7}$$

then the $\dot{\lambda} = 0$ locus is given by

$$k = \kappa.$$

This is a vertical line on the phase diagram, as shown in Figure 8.2.1.

Because $f'(\cdot)$ is decreasing, (8.4) implies that to the right of this isocline ($k > \kappa$), $f'(k) < f'(\kappa) = \rho + \delta$, so that $\dot{\lambda} > 0$. To the left of this isocline ($k < \kappa$), $\dot{\lambda} < 0$.

Indeed, Chapter 5 demonstrates the general result that in any optimal control problem with one endogenous state variable and concave objective and accumulation functions (thereby satisfying the key sufficiency conditions of Chapter 2), $\dot{\lambda} > 0$ to the right of the $\dot{\lambda} = 0$ locus and $\dot{\lambda} < 0$ to the left of the $\dot{\lambda} = 0$ locus. Being able to state general principles like this is an important advantage of standardizing the phase diagram by putting the costate variable on the vertical axis and the state variable on the horizontal axis.

8.2.2 The $\dot{k} = 0$ Locus

Setting $\dot{k}$ to zero in (8.2) yields the equation

$$f(k) - \delta k - g = c.$$

By (8.3), λ is a monotonically decreasing function of c, which makes c a monotonically decreasing function of λ:

$$c = u'^{-1}(\lambda).$$

The $\dot{k} = 0$ locus can be written as

$$f(k) - \delta k - g = u'^{-1}(\lambda) \tag{8.8}$$

and is shaped as shown in Figure 8.2.2. For comparison, Figure cass c k isocline shows what the $\dot{k} = 0$ isocline would look like in $k - c$ space. As k increases, the left-hand-side of (8.2.2)—which gives consumption when $\dot{k} = 0$—increases as long as $f'(k) > \delta$—which is always true to the left of the $\dot{\lambda} = 0$ locus where $f'(k) = \rho + \delta$ and for some distance to the right of the $\dot{\lambda} = 0$ locus. Beyond the point where $f'(k) = \rho$, a higher level of capital would reduce the level of consumption when $\dot{k} = 0$ in a reflection of dynamic inefficiency. Consumption and the marginal utility of consumption λ are inversely related. Therefore, in (8.2.2), the k-isocline is downward sloping to the left of the λ-isocline and for some distance to the right of it, eventually becoming upward sloping.

As indicated by (8.5), at higher values of λ and lower values of c than the $\dot{k} = 0$ locus, net investment will be positive, making $\dot{k} > 0$. At lower values of λ and higher values of c, net investment will be negative, making $\dot{k} < 0$. Thus, in the standardized phase diagram with λ on the vertical axis and k on the horizontal axis, $\dot{k} > 0$ above the $\dot{k} = 0$ locus and $\dot{k} < 0$ below the $\dot{k} = 0$ locus.

Indeed, Chapter 5 demonstrates the general result that in any optimal control problem with one endogenous state variable and concave objective and accumulation functions that $\dot{k} > 0$ above the $\dot{k} = 0$ locus and $\dot{k} < 0$ below the $\dot{k} = 0$ locus. Standardizing the phase diagram by putting the costate variable on the vertical axis and the state variable on the horizontal axis makes this common thread stand out.

8.2.3 The Saddle Path

Combining the dynamics for λ shown by (8.2.1) and the dynamics for k shown by (8.2.2) yields Figure 8.2.3. Note that the dynamic paths are precisely horizontal when crossing the $\dot{\lambda} = 0$ locus, and are precisely vertical when crossing the $\dot{k} = 0$ locus.

The intersection of the $\dot{k} = 0$ locus and the $\dot{\lambda} = 0$ locus is the steady state. We will denote steady-state values of all variables by asterisks. Thus, on the phase diagram, the steady state is the point (k^*, λ^*). At the steady state there is no tendency to move toward anywhere else.

Most of the dynamic paths shown lead off to infinity or off the edges of the graph. The exception is the pair of dynamic paths shown in bold, which lead in to the steady state. These are the saddle paths. Often, both together are referred to in the singular as *"the saddle path."*

The importance of the saddle path arises from the fact that while the initial value of k is given by history, the initial value of λ is not. While k is pointed toward the past, λ is pointed toward the future. The initial value of λ is determined by the need to have a future that is (1) consistent with rational expectations, (2) consistent with the transversality condition and (3) otherwise sensible. In practice, the paths that go off to infinity or off the edges of the graph are all eliminated by one of these three criteria.[1] Therefore, the focus is on paths that lead eventually to a steady state. When no further movements of exogenous shifters can be foreseen, that leads to a focus on the saddle path. Given an initial value of k by history, the initial value of λ is given by the saddle path.

Based on the saddle-path, the prediction of the model is that if the capital stock k begins below its long-run steady-state value, the marginal value of

[1] See exercises ....

capital λ will begin above its long-run steady-state value. The capital stock will gradually increase toward its steady-state value while the marginal value of capital gradually falls to *its* steady-state value.

If the capital stocks begins *above* its steady state value, the marginal value of capital will begin below its steady state value. The capital stock will gradually fall, while the marginal value of capital gradually rises as the economy approaches the steady state.

8.3 The Economic Geography of the Phase Diagram

The standardized phase diagram shows k and λ by its axes. By means of the isoclines and the dynamic arrows, the phase diagram also shows $\dot{k}$ and $\dot{\lambda}$. Yet there are many important economic variables that are not shown directly on the phase diagram, such as consumption, output, investment, and the real wage, real rental rate, and real interest rate—c, y, R, r, w and i. If the phase diagram is thought of as a map of an abstract empire, the map shows the location of the capital city and the direction of the key roads, but not what is going on at the various places along the way. What is missing is the economic geography of the phase diagram. In place of questions like "In which region do they grow the grapes?" or "Where do they mine the copper?" for an ordinary map we have questions like "In which region is output high?" or "Where is the real interest rate low?"

Fortunately, in the Ramsey model, as in most other optimal control models with one state variable, all of the other key variables can be determined from k, λ and the exogenous variables. In general, the state variable (k) is a sufficient statistic for the past, the costate variable (λ) is a sufficient statistic for the future, and the exogenous variables (such as government purchases g in the next section) are a sufficient statistic for everything else about the present needed to determine what is going on at a particular location on the phase diagram. The formal name for what is going on at a particular location on the phase diagram—that is, what the economy will do for given values of k and λ—is *the contemporaneous equilibrium*. The *economic geography* of the phase diagram is a description of how the contemporaneous equilibrium varies with k and λ.

The economic geography of the Ramsey model is particularly simple. Since $u'(c) = \lambda$, consumption depends only on λ, to which it is negatively related (decreasing when λ increases and increasing when λ decreases). Since $y = f(k)$, output depends only on k, to which it is positively related. Since $R =$

$f'(k)$, the rental rate depends only on k, to which it is negatively related. Since $r = R - \delta = f'(k) - \delta$, the real interest rate also depends only on k, to which it is negatively related. Since $w = f(k) - kf'(k)$, the real wage depends only on k. The real wage is positively related to k since $\frac{d}{dk}[f(k) - kf'(k)] = -kf''(k) > 0$.

Gross investment i is a little more difficult, since i depends on both k and λ. To be specific,

$$i = y - c - g = f(k) - u'^{-1}(\lambda) - g.$$

Thus, investment is positively related to both k and λ. (When λ increases, c falls, leaving more resources for investment.) Just as one draws indifference curves for utility as a function of various types of consumption and isoquants for output as a function of inputs, one can draw curves on the phase diagram for a fixed value of i. Such constant-i curves are called *isograms* for investment.[2] Since investment depends positively on both k and λ, the investment isograms are downward sloping; formally,

$$\left.\frac{d\lambda}{dk}\right|_{i=\text{constant}} = -\frac{\frac{\partial i}{\partial k}}{\frac{\partial i}{\partial \lambda}} < 0, \tag{8.9}$$

where the partial derivatives are for gross investment i once i has been reduced to a function of k, λ and the current values of the exogenous variables. Since $i = \dot{k} + \delta k$,

$$\frac{\partial i}{\partial k} > \frac{\partial(\dot{k})}{\partial k},$$

which by (8.9) implies that the investment isograms have a steeper downward slope than the $\dot{k}$ = constant curves. In particular, the investment isograms cut the $\dot{k} = 0$ locus from above, as shown in Figure 8.3.

8.4 The Economic Geography of the Saddle Path

Suppose the initial capital stock is below k^*, so that the economy begins high up on the saddle path to the left of the steady state. Moving down along the saddle path to the right, the capital stock k rises while the marginal value of

[2] "Isogram" is the generic word for curves such as the contours on a topographical map; isoquants and indifference curves in economics; isoglosses in linguistics; and isobars, isotherms, isohels and isohyets in meteorology.

capital λ falls. In view of the results above, the gradual increase in the capital stock leads to a gradual rise in output y and the real wage w and a gradual fall in the real rental rate R and the real interest rate r. The gradual fall in the marginal value of capital leads to a gradual rise in consumption c. By the nature of the saddle path, $\dot{k}$ gradually falls from its positive value toward zero, while $\dot{\lambda}$ gradually rises from its negative value toward zero (at least in the neighborhood of the steady state).

Finally, since investment i is positively related to λ, investment will gradually fall if the saddle path cuts the investment isograms from above, but investment will gradually rise if the saddle path cuts the investment isograms from below. Either case is possible. If $\delta = 0$, then gross and net investment are the same, and the investment isograms parallel the $\dot{k} = 0$ locus. Since the saddle path is steeper than the $\dot{k} = 0$ locus, in this case the saddle path cuts the investment isograms from above (at least in the neighborhood of the steady state). If, on the other hand, δ becomes large enough, the investment isograms will become close to vertical and the saddle path will cut the investment isograms from below.[3]

If the initial capital stock is greater than k^*, the gradual movement back up the saddle causes the opposite pattern for all of these variables.

8.5 Shocks to the Rate of Time Preference and to Government Purchases

In the absence of any changes in the exogenous variables, convergence along the saddle path is the only story to be told. It is time now to look at such exogenous changes. We will look first at shocks to the rate of time preference ρ and then at shocks to the level of government purchases g.

For now, in studying the effects of shocks, we will make the *certainty-equivalence approximation*: looking at the decisions agents would make if all uncertainty vanished and they were certain to face the expected values of future variables. In other words, using the certainty-equivalence approximation, one proceeds as if all the agents in a model had perfect foresight. As we proceed to study "perfect-foresight" models, it is important to keep sight of their purpose of providing a certainty-equivalence approximation to stochastic models.[4]

[3]This is more difficult to show. Indeed, it is easiest to show using techniques from later on in the book.

[4]See Chapter **??** for a more detailed discussion of certainty equivalence and the certainty-

When looking at the effects of a particular shock, think of it as a realization from a distribution of possible shocks that are equally likely to be positive or negative. Given the certainty-equivalence approximation, the dynamic effects or "impulse response functions" predicted by the perfect foresight model yield the impulse response functions for the stochastic model as well. As shocks repeatedly buffet the economy, the impulse responses from the various shocks all add on top of one other, algebraically. Thus, the dynamics of perfect foresight models are the key to understanding the corresponding stochastic models.

When thinking about shocks, it is best to think of shocks as packets of genuine news—of genuinely new information about the future. When using the certainty-equivalence approximation, that new information takes the form of a change in the future expected time path of one or more exogenous variables (and maybe a change in the current value of one or more of the exogenous variables). Sometimes, the expected time path for exogenous variables will be unchanged for some time into the future. In this case, there is a difference between when the shock is unsheathed and when it actually cuts, between when it is unveiled and when it actually hits, between when it is revealed and when it actually materializes. As we will see, a shock begins affecting the economy as soon as it is revealed. An economy with forward-looking agents does not wait until the shock actually materializes to respond. Moreover, even when some change in an exogenous variable happens immediately, what is expected to happen thereafter has a very important influence on the effect of a shock.

8.5.1 Shocks to the Rate of Time Preference (ρ)

One broad category of shocks is the category of "preference shocks." In the Ramsey model, the simplest example of a preference shock that affects behavior is a change to ρ (the "degree of impatience," "utility discount rate," or "rate of time preference"). Note that a simple multiplication of the felicity

equivalence approximation. Chapter **??** investigates how good the certainty equivalence approximation is. One simple generalization is that when only aggregate, economy-wide uncertainty is at issue, the certainty-equivalence approximation is typically quite a good approximation. The certainty-equivalence approximation is often not a very good approximation when the idiosyncratic risk faced by heterogeneous households and firms is at issue. (The law of averages helps to make aggregate uncertainties much smaller as a percentage of mean values than the risks faced by individual households and firms.) In any case, the certainty-equivalence approximation is the foundation on which higher-order approximations will be built, so it is the right place to start.

function at every instant by the same positive constant makes no difference at all to optimal choices. For choices to be affected, the balance between felicity at two different dates must be altered. An increase in ρ alters that balance by reducing the importance of future felicity in comparison to present felicity.

Just as the k- and λ-isoclines are the key to understanding the dynamics of the model in the absence of shocks, shifts in the k- and λ- isoclines are the key to understanding the response of the model to shocks. Equation (8.5) indicates that $\dot{k}$ is not affected by ρ. In particular, the $\dot{k} = 0$ locus is unaffected. But equation (8.4) indicates that , given k and λ (that is, given a specific location on the phase diagram), the increase in ρ causes an increase in $\dot{\lambda}$. In particular, a point on the *old* $\dot{\lambda} = 0$ locus must have $\dot{\lambda} > 0$. Therefore, the *old* $\dot{\lambda} = 0$ locus must be to the right of the *new* $\dot{\lambda} = 0$ locus; in other words, an increase in ρ shifts the $\dot{\lambda} = 0$ locus to the left. (The leftward shift in the $\dot{\lambda} = 0$ locus can also be seen from (8.6). An increase in ρ requires the $\dot{\lambda} = 0$ locus to have a higher marginal product of capital. Capital per worker must be smaller in order to get this higher marginal product of capital.)

To begin with, consider an immediate permanent increase in impatience ρ. In a dynamic economic model, any shock has at least two aspects: (a) the arrival of the information that the shock will happen and (b) the alterations in the paths of exogenous variables indicated by that information.[5] For convenience, we will call the moment at which the information arrives *time zero.*

In the long run, as shown in Figure 8.5.1, a permanent increase in ρ leads to a fall in the steady-state capital stock k^* and an increase in the steady-state marginal value of capital λ^*. However, at time zero, of k and λ, only the marginal value of capital λ—which is a forward-looking expectation—can jump. The capital stock k—which is a historical variable—cannot jump. Backward induction is the key to deducing what should happen. Since the

[5]When there is a delay between the arrival of the information and the indicated alteration in the paths of exogenous variables, it is called an *anticipated* shock. When there is no such delay, it is called an *immediate*shock. (This will be our default case, whenever we do not explicitly state that a shock is anticipated.) When the alterations in the paths of exogenous variables are expected to eventually die away to nothing, it is called a *temporary* shock. If there is any alteration that does not ultimately die away, it is called a *permanent* shock. When we talk about temporary and permanent shocks without further modification, we will be focusing on the simple case when the alterations in the exogenous variables can be described by simple step functions that move from zero to some other value, remaining constant at that value—and then in the case of temporary shocks, eventually going back to zero.

only way to get to the new steady state following the new dynamics is on the saddle path, the marginal value of capital λ must jump down to the new saddle path on impact at time zero. After the initial jump to the new saddle path, the economy follows that saddle path up to the new steady state.

The phase diagram shows directly only what happens to k and λ. It is important to find what happens to the other key economic variables as well. The effects of the permanent impatience shock on the various economic variables can be summarized neatly by a dynamic effect table, which shows for each variable (1) which way it jumps on impact, (2) which way it evolves as the economy moves along the saddle path and (3) the direction of the overall effect from its initial value to its value in the new steady state. Here the economic geography of the phase diagram comes in handy. Note that ρ does not appear in any of the first-order conditions; it appears only in a dynamic equation—specifically the Euler equation, which determines $\dot{\lambda}$ after time zero). Conditional on k and λ—the location on the phase diagram—ρ has no effect on any of the control variables or on the prices associated with those control variables. To be specific, given λ and k, consumption is determined by $c = u'^{-1}(\lambda)$, with labor n exogenous in any case, output is determined by $y = f(k)$. The real wage is determined by $w = f(k) - kf'(k)$. The real rental rate is determined by $R = f'(k)$; the real interest rate by $r = R - \delta = f'(k) - \delta$. Finally, investment is determined by $i = y - c - g = f(k) - u'^{-1}(\lambda) - g$.

Investment is the one variable that depends on both k and λ. The initial jump downward in λ increases consumption, without affecting output, which depends only on k—thus leading to a fall in investment i. As mentioned previously, what happens to investment when moving Northwest along the saddle path is ambiguous, since the effects of the movement in k and the movement in λ are in opposite directions.[6] It might seem that the overall effect from initial value to the value in the new steady state would also be ambiguous, since it is also a movement to the Northwest, but we know that at both the old and the new steady-state, $i = \delta k$ (since $\dot{k}$ must be zero). Thus, the lower capital stock in the new steady state implies a lower level of steady state investment as well.[7]

[6] To make a solid claim of ambiguity, one must actually produce an example of each case, which we omit here since establishing these examples is more advanced than the level of this chapter. One pair of such examples involves the limiting cases when $\delta = 0$ and when $\delta \to \infty$.

[7] One could also get to the same conclusion about the overall steady state effect on investment from the fact worked out previously that the investment isograms are steeper than the (here unmoving) $\dot{k} = 0$ locus, implying that the movement in k dominates in its effect on i when moving along the $\dot{k} = 0$ locus.

Given these results for investment, it is straightforward to fill in the dynamic effects table:

Variable	Impact Effect	Saddle Path Effect	Overall Steady State Effect
c	$+$	$-$	$-$
y	0	$-$	$-$
i	$-$	?	$-$
w	0	$-$	$-$
R	0	$+$	$+$
r	0	$+$	$+$
$\dot{k}$	$+$	$-$	0
$\dot{\lambda}$	$-$	$+$	0

(8.10)

Note that the third column is the sum of the first two columns, although not everything that might seem ambiguous from simple addition is really ambiguous. (On this point, note that the overall steady state effects on both consumption and investment are determinate, even though addition of the first two columns looks ambiguous.)

8.5.2 Shocks to Government Purchases (g)

The logical counterpoint to the category of preference shocks is the category of shocks to constraints. The simplest type of constraint shock is a shock to extraneous government purchases. By the word *purchases*, we mean to exclude government transfers, which are often spoken of as part of government spending, but which should be viewed economically as negative taxes. By the word *extraneous*, we mean government purchases that use up resources but have no other effect on the production function and enter utility only in an additively separable way. (In practical terms, military spending may be the type of government spending closest to being extraneous government purchases.)

If we include the effect of extraneous government purchases on utility *explicitly* by an additive time-dependent function ϕ_t, the social planner's problem becomes

$$\max_{c,g} \int_0^\infty e^{-\rho t}[u(c_t) + \phi_t(g_t)]dt \tag{8.11}$$

s.t.

$$\dot{k} = f(k_t) - \delta k_t - c_t - g_t. \tag{8.12}$$

(For clarity, we have also shown the dependence of c, g and k on time explicitly.) A necessary condition for solving social planner's problem (8.11) is that for whatever path of extraneous government purchases g that is chosen, the social planner be solving the subproblem

$$\max_c \int_0^\infty e^{-\rho t} u(c_t) dt \tag{8.13}$$

s.t.

$$\dot{k}_t = f(k_t) - \delta k_t - c_t - g_t. \tag{8.14}$$

Why is this necessary? If $\int_0^\infty e^{-\rho t} u(c_t) dt$ can be increased without changing the path of g, then

$$\int_0^\infty e^{-\rho t} u(c_t) dt + \int_0 e^{-\rho t} phi_t(g_t)$$

must also be increased since the second term $\int_0 e^{-\rho t} \phi_t(g_t)$ has been left unchanged. Thus, anything that increases $\int_0^\infty e^{-\rho t} u(c_t) dt$ without changing the path of g also moves the economy closer to a solution of (8.11). This straightforward result means that optimization of (8.11) implies optimization of the simpler problem (8.1) or (8.13)we have been emphasizing, conditional on some path of extraneous government purchases g.

One can learn something about optimization conditional on a path of extraneous government purchases by treating the path of government purchases as if it were exogenous, then optimizing. That is the approach we will take here. The actual paths of g that we will condition on will be most relevant if at any point in time ϕ_t is tightly curved as a function of g, indicating a strong necessity for a certain level of purchases, after which more is not so important. This "necessary" level of government purchases can vary over time because of the dependence of phi_t on time.

As with changes in ρ, the key to understanding the response of the model to changes in extraneous government purchases g is to understand the effects of g on the k- and λ-isoclines and on the rest of the economic geography of the phase diagram. All of these effects on the economic geography of the phase diagram can be summarized and compared for both ρ and g by the following direct effect table: each entry in the column for g shows the direct effect of g on that variable *for given values of k and λ.* The first column shows each variable as a function of k, λ and the exogenous g and ρ.

Variable	Function	Direct Effect of ρ	Direct Effect of g
c :	$u'^{-1}(\lambda)$	0	0
y :	$f(k)$	0	0
i :	$f(k) - u'^{-1}(\lambda)$	0	–
w :	$f(k) - kf'(k)$	0	0
R :	$f'(k)$	0	0
r :	$f'(k) - \delta$	0	0
$\dot{k}$:	$f(k) - \delta k - u'^{-1}(\lambda)$	0	–
$\dot{\lambda}$:	$\rho + \delta - f'(k)$	+	0

Holding k and λ fixed—in other words, staring at a given point on the phase diagram—an increase in extraneous government purchases g has no effect on consumption, output, the real wage, the real rental rate or the real interest rate. The direct effect of extraneous government purchases is limited to a reduction in investment i, with a consequent reduction in capital accumulation $\dot{k}$.

A Permanent Increase in Government Purchases

The direct effects on $\dot{k}$ and $\dot{\lambda}$ indicate the way the k- and λ-isoclines shift. In the case of an increase in extraneous government purchases g, capital accumulation $\dot{k}$ is reduced at a given value of k and λ. A point on the old $\dot{k} = 0$ locus will have $\dot{k} < 0$, and so must be below the new $\dot{k} = 0$ locus. Therefore, the $\dot{k} = 0$ locus must shift up. Since $\dot{\lambda}$ is unchanged at a given value of k and λ, there is no shift of the $\dot{\lambda} = 0$ locus.

As shown by Figure 8.5.2, a permanent increase in extraneous government purchases g—financed by lump-sum taxes—leads to a new steady state that is due north of the old steady state. Since no capital accumulation is required, the initial jump in λ upon receipt of the new information can take the economy instantly to the new steady state.

With no saddle-path effects, the impact effects and overall steady state effects of a permanent increase in g are identical.

Variable	Impact Effect	Saddle Path Effect	Overall Steady State Effect
c	$-$	0	$-$
y	0	0	0
i	0	0	0
w	0	0	0
R	0	0	0
r	0	0	0
$\dot{k}$	0	0	0
$\dot{\lambda}$	0	0	0

(8.15)

It is easy to summarize this table: in the Ramsey model, a permanent increase in government purchases immediately crowds out consumption. It has no other effect on any key variable. Investment i is affected by both λ and directly by g, but these effects must cancel out since $i^* = \delta k^*$ in both steady states, with the same value of k^*.

A Temporary Increase in Government Purchases

If extraneous government purchases g increase by a certain amount, and everyone knows that they will return to the old value at time T, the isoclines and dynamic arrows on the phase diagram shift for an interim period $(0, T)$, then return to normal when g returns to its previous value. It is important to realize that, since it is foreseen in advance, the shift back at time T is a different kind of shift than the shift at time zero. The marginal value of capital λ depends on expectations of the future. It can jump when those expectations are changed by new information, but not when events are playing out as expected.

With g returning to normal, the economy should eventually return to the old steady state. Backward induction is the easiest way to deduce the dynamic general equilibrium that will eventually return the economy to the old steady state. From time T on, the economy follows the old dynamics, and the only way to get back to the once and future steady state will be along the old saddle path. Although λ can jump at time 0 when there is new information, λ *cannot* jump at time T, since there is no new information then—only the economy playing out its prearranged script. Therefore, the economy must use the initial jump in λ and the dynamics during the interim period to arrive at the old saddle path at time T.

After the initial jump in λ, it is easy to mechanically trace out where the interim dynamics will take the economy. It is instructive to trace out the consequences of various jumps in λ. A jump down or a jump up beyond the interim steady-state would take the economy of on paths that would never lead back to the old saddle path. But if λ jumps part way up toward the new steady state, the interim dynamics take the economy on a path that crosses the old steady state. A small jump upward in λ would lead to a quick arrival back at the old saddle path, and so is called for when T is small. A jump upward in λ almost to the interim steady state leads to a lot of time spent in the neighborhood of the interim steady state—a region in which movement is very slow. Indeed, jumping up close enough to the interim steady state, the economy can spend an arbitrarily long time in the neighborhood of the interim steady state. This is an example of the Turnpike Theorem. A large value of T calls for such a jump up nearly to the interim steady state.

The dynamic effects on the key variables can be summarized by the following table, giving the impact effects, the effects during the time of higher extraneous government purchases g, the reaction at time T to the return of g to normal, and the effects after T, when the economy follows the old saddle path. The overall effect is always zero, since the economy returns to the old steady state. Thus, the rows of Table 8.17 must add to zero.

Variable	Impact $(t = 0)$	During $(t \in (0, T))$	At $(t = T)$	After$(t \in (T, \infty))$
c	$-$	$+$	0	$+$
y	0	$-$	0	$+$
i	$-$	$-$	$+$	?
w	0	$-$	0	$+$
R	0	$+$	0	$-$
r	0	$+$	0	$-$
$\dot{k}$	$-$	$-$	$+$	$-$
$\dot{\lambda}$	0	$-$	0	$+$

(8.16)

It is a straightforward matter to draw graphs of the key variables against time from the information in Table 8.17. These graphs are graphs of the *impulse response functions* of the variables to a temporary government purchase shock.

An Anticipated Increase in Government Purchases

Suppose now that government purchases have not yet changed, but that everyone knows that they will increase at time T to a higher level that they will then remain at from then on. Working backward, the economy must end up at the new steady state and therefore must arrive at the new saddle path at time T. The economy must use its initial jump in λ and the old dynamics to get to the new saddle path.

In order to use the old dynamics to get to the new saddle path, the initial jump in λ must go partway up toward the new steady state. The larger the initial upward jump, the shorter the time needed to get to the new saddle path using the old dynamics. Therefore, if T is large, the initial jump upward in λ will be small, while if T is small, the initial jump upward in λ will take the economy almost to the new steady state.

The following dynamic effect table gives the effects on key variables of the initial impact; of the time before the actual change in g at time T; after time T when moving along the new saddle path; and the overall effect from the old steady state to the new steady state.

Variable	Impact($t = 0$)	Before($t \in (0, T)$)	At ($t = T$)	After($t \in (T, \infty)$)	Overall
c	$-$	$-$	0	$-$	$-$
y	0	$+$	0	$-$	0
i	$+$	$+$	$-$	?	0
w	0	$+$	0		0
R	0	$-$	0	$+$	0
r	0	$-$	0	$+$	0
$\dot{k}$	$+$	$+$	$-$	$+$	0
$\dot{\lambda}$	0	$+$	0	$-$	0

(8.17)

It is not hard to draw impulse response functions from this table. An important relationship becomes apparent if one compares the impulse response functions for permanent, temporary and anticipated government purchase shocks. Figure 8.5.2 shows the three impulse response graphs for consumption. What should be apparent is that—to a good approximation that will be made explicit later on as linearization—the impulse response to a permanent shock is the sum of the impulse responses to a temporary and to an anticipated shock. Linearization makes effects additive. Thus, since in terms of g

itself, TEMPORARY + ANTICIPATED = PERMANENT, this is true for the effects as well.

Exercises

1. Show that if output is given by $y = nf(k/n)$, that at $n = 1$, $w = \frac{\partial y}{\partial n} = f(k) - kf'(k)$.

2. Analyze the effects of a temporary increase in ρ.

3. Analyze the effects of an anticipated future increase in ρ.

4. Show that the dynamic after-effects of a sudden, permanent increase in population (with no change in the capital stock) are similar to the dynamic after-effects of the sudden destruction of part of the capital stock.

Chapter 9

Technology Shocks and Taxes in the Ramsey Model

When looking at the effects of technology shocks and taxes, it is useful to make a modest extension of the Ramsey model to two sectors—an investment-goods-producing sector and a consumption-goods-producing sector. Such an extension can be made at surprisingly low cost as long as the two sectors have a production function of the same shape, yielding an equal degree of capital intensity in both sectors. This chapter will hold to that simplifying assumption.

9.1 Reducing the Two-Sector Model to a Modified One-Sector Model

The maximum principle says that, whatever else is happening, the Hamiltonian should be maximized. By identifying and solving a static optimization subproblem that arises in maximizing the Hamiltonian, the two sector model can be reduced to a one sector model.

9.1.1 The Original Social Planner's Problem

The two-sector model corresponds to the following social planner's problem:

$$\max_{c} \int_{0}^{\infty} e^{-\rho t} u(c) dt \tag{9.1}$$

s.t.

$$\begin{aligned}
\dot{k} &= i - \delta k \\
i &= z n_i f(\frac{k_i}{z n_i}) - g \qquad (9.2) \\
c &= \xi z n_c f(\frac{k_c}{z n_c}) \qquad (9.3) \\
n_i + n_c &= 1 \\
k_i + k_c &= k.
\end{aligned}$$

In addition to the constraints above, all quantities must be nonnegative. This will be discussed below. As before, i is gross investment, c is consumption, ρ is the level of impatience, k is the capital stock, δ is the depreciation rate and g is the exogenous level of government purchases (which have a utility which is assumed to be additively separable from the utility from c). The quantities n_i and n_c are the amounts of labor used in the consumption and the investment sector respectively, while k_i and k_c are the amounts of capital used in the consumption and investment sectors. The parameter z is the labor-augmenting technology shifter, which applies to both sectors, while ξ is a multiplicative technology shifter *applying only to the consumption sector*. For simplicity, government purchases are from the investment sector.

9.1.2 The Static Subproblem

Rather than going directly to Hamiltonian, the first-order conditions and Euler equations, consider the subproblem of finding the production possibility frontier. Defining p_c as the price of consumption goods relative to investment goods, which varies from 0 to ∞ to trace out the production possibility frontier, the subproblem of finding the PPF can be expressed as

$$\max i + p_c c \qquad (9.4)$$

s.t.

$$\begin{aligned} i &= zn_i f(\frac{k_i}{zn_i}) - g \\ c &= [\xi zn_c f(\frac{k_c}{zn_c}) \\ n &= n_i + n_c \\ k &= k_i + k_c \\ i &\geq 0 \\ c &\geq 0, \end{aligned}$$

where $n = 1$ for the particular problem at hand, but for later use we will solve the problem for a more general value of n. We are explicitly including the nonnegativity constraints on investment and consumption because the production possibility frontier between investment and consumption will turn out to be linear; for most values of p_c, the optimal solution to (9.4) is a corner solution.

Labeling the multiplier for the adding-up constraint on labor as w, since it will be the real wage (in terms of investment goods) and the multiplier on the adding-up constraint on capital as R, since it will be the real rental rate (in terms of investment goods), the Lagrangian for this static problem is

$$\mathcal{L} = i + p_c c + \mu_i[zn_i f(\frac{k_i}{zn_i}) - g - i] + \mu_c[\xi zn_c f(\frac{k_c}{zn_c}) - c] + w[n - n_i - n_c] + R[k - k_i - k_c] + \nu_i i + \nu_c c.$$

The first order condition for optimal i and c are

$$1 + \nu_i = \mu_i$$

and

$$p_c + \nu_c = \mu_c.$$

Thus, at an interior solution, $\mu_i = 1$ and $\mu_c = p_c$. The first-order conditions for k_i, k_c, n_i, and n_c are

$$\mu_i z \left[f(\frac{k_i}{zn_i}) - (\frac{k_i}{zn_i}) f'(\frac{k_i}{zn_i}) \right] = w \tag{9.5}$$

$$\mu_c \xi z \left[f(\frac{k_c}{zn_c}) - (\frac{k_c}{zn_c}) f'(\frac{k_c}{zn_c}) \right] = w \tag{9.6}$$

$$\mu_i f'(\frac{k_i}{zn_i}) = R \tag{9.7}$$

$$\mu_c \xi f'(\frac{k_c}{zn_c}) = R \tag{9.8}$$

Dividing (9.5) by (9.7) and (9.6) by (9.8) yields

$$\frac{f(\frac{k_i}{zn_i}) - (\frac{k_i}{zn_i})f'(\frac{k_i}{zn_i})}{f'(\frac{k_i}{zn_i})} = \frac{w}{zR} = \frac{[f(\frac{k_c}{zn_c}) - (\frac{k_c}{zn_c})f'(\frac{k_c}{zn_c})}{f'(\frac{k_c}{zn_c})} \tag{9.9}$$

It is not difficult to show that if $f'(\cdot) > 0$ and $f''(\cdot) < 0$, then $\frac{f(x)-xf'(x)}{f'(x)}$ is a monotonically increasing function of x. Thus, the effective capital/labor ratio $\frac{k}{zn}$ in each sector is the *same* monotonic function of the effective wage/rental rate ratio $\frac{w}{zR}$. Since the social planner or perfect competition in the factor markets equalizes the wage/rental rate across sectors, the two sectors must have the same effective capital/labor ratio:

$$\frac{k_i}{zn_i} = \frac{k_c}{zn_c}. \tag{9.10}$$

Since z is the same in both sectors, this implies $k_i/n_i = k_c/n_c$.

In other words, having a production function of the same shape makes it optimal for the two sectors to have the same capital intensity at all times. An Edgeworth box makes clearer what is going on.

With isoquants of the same shape for both goods, the tangencies must all lie along the straight line connecting the two origins, O_i and O_c. Not only do the two sectors have the same capital labor ratio, this capital/labor ratio is equal to the overall capital/labor ratio in the economy:

$$k_i/n_i = k_c/n_c = k/n.$$

Substituting k/n into (9.2) and (9.3), one finds that

$$i + g = zn_i f(\frac{k}{zn})$$
$$c = \xi zn_c f(\frac{k}{zn}),$$

implying

$$\frac{c}{\xi} + i + g = z(n_i + n_c)f(\frac{k}{zn}) = znf(\frac{k}{zn}). \tag{9.11}$$

Graphically, (9.11) is a linear production possibility frontier.

Finally, given equal effective capital/labor ratios, the combination of (9.5) and (9.6) or of (9.7) and (9.8) yields an equation for how the relative price of consumption goods is determined by technology:

$$\frac{\mu_c}{\mu_i}\xi = \frac{p_c + \nu_c}{1 + \nu_i}\xi = 1. \tag{9.12}$$

Since the multipliers on the nonnegativity constraints are zero when the constraints are slack and themselves nonnegative when the constraints are binding, (9.12) implies that

$$p_c = \frac{1}{\xi} \tag{9.13}$$

at an interior solution—that is, when both $i > 0$ and $c > 0$. When $c = 0$ and $i > 0$, $p_c \leq \frac{1}{\xi}$. When $i = 0$ and $c > 0$, $p_c \geq \frac{1}{\xi}$. In terms of Figure 9.1.2, the slope of the production possibility frontier is $\frac{1}{\xi}$. When the $p_c \leq \frac{1}{\xi}$, the upper-left-hand corner is optimal. When $p_c \geq \frac{1}{\xi}$, the lower-left-hand corner is optimal.

9.1.3 The Reduced Social Planner's Problem

Consider the reduced social planner's problem

$$\max_c \int_0^\infty e^{-\rho t} u(c) dt \tag{9.14}$$

s.t.

$$\dot{k} = i - \delta k \tag{9.15}$$

$$i + \frac{c}{\xi} = z f(\frac{k}{z}). \tag{9.16}$$

(In addition to these constraints, $i \geq 0$ and $c \geq 0$.) The reduced social planner's problem is well defined. Given a solution to the reduced social planner's problem, the static optimization problem with $n = 1$ can deliver the appropriate point on the production possibility frontier between investment and consumption at every instant. To be specific,

$$
\begin{aligned}
n_i &= \left(\frac{i+g}{c+i+g}\right) n = \frac{i+g}{c+i+g} \\
n_c &= \left(\frac{c}{c+i+g}\right) n = \frac{c}{c+i+g} \\
k_i &= \left(\frac{i+g}{c+i+g}\right) k \\
k_c &= \left(\frac{c}{c+i+g}\right) k.
\end{aligned}
$$

And as long as the reduced social planner's problem delivers $i > 0$ and $c > 0$, we must have $p_c = \frac{1}{\xi}$.

The important condition to check is that the static optimization problem does not impose contraints beyond what have been written in to the reduced social planner's problem. Since any combination of i and c on the production possibility frontier described by (9.16) can be achieved by an appropriate choice of n_i, n_c, k_i and k_c, this condition is satisfied.

9.2 Analyzing the Reduced Two-Sector Model

The augmented current-value Hamiltonian for the social planner's problem 9.14 is

$$H = u(c) + \lambda[i - \delta k] + \mu[zf(k/z) - (c/\xi) - g - i].$$

(For simplicity, the nonnegative constraints have been left out.) The first-order conditions are

$$
\begin{aligned}
\mu &= \lambda && (9.17) \\
u'(c) &= \frac{\mu}{\xi}. && (9.18)
\end{aligned}
$$

Equation (9.17) says that since (in the absence of investment adjustment costs in the model) one unit of investment goods always adds one unit to the capital stock, the marginal value of investment (μ) is equal to the marginal value of capital (λ). Combining (9.17) and (9.18) yields

$$u'(c) = \frac{\lambda}{\xi} \tag{9.19}$$

The Euler equation is

$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - f'(k/z) \tag{9.20}$$

9.3 Measurement with Two Different Goods

In this model, investment and government purchases are treated as perfect substitutes in production, but investment and consumption are clearly two different goods. When there is a differential change in the technology for producing consumption goods and producing investment goods, the conversion rate between consumption and investment goods changes. Indeed, the conversion rate is ξ units of consumption for every one unit of investment. Clearly, this changes whenever ξ changes. It is not immediately apparent what to use as the numeraire or yardstick for the model. Because it is an entirely real model, one can choose any numeraire that is convenient. Initially, it is tempting to use consumption goods as the numeraire, but the structure of the model will be more deeply illuminated by taking *investment goods* as the numeraire. Standard national income accounting would construct a GDP deflator, which is in the spirit of using a basket with both consumption and investment goods as the yardstick. A serious treatment of national income accounting must wait until the section on *quantitative* analytics. For qualitative analysis, we will consistently use investment goods as the numeraire (or yardstick). This gives rise to several strange-sounding results. An improvement in consumption technology, while it increases the amount of consumption goods, also makes them count less in terms of investment goods and so need have no effect on output *as measured in investment good equivalents.* Similarly, an improvement in consumption technology can raise the real *consumption* wage (how much consumption an hour of work can buy) without raising the real wage in terms of investment goods and can raise the marginal product of capital measured in terms of consumption goods without raising the marginal product of capital measured in terms of investment goods.

As a reminder, we will often refer to this concept of output as “output with rescaled consumption.” Giving output with rescaled consumption the letter y,

$$y = zf(\frac{k}{z}).$$

Output with rescaled consumption depends only on k and z. It does not depend on ξ. The wage *in terms of investment goods* w, is given by

$$w = z[f(\frac{k}{z}) - (\frac{k}{z}f'(\frac{k}{z})]$$

The rental rate *in terms of investment goods* R is given by

$$R = f'(\frac{k}{z}).$$

Both w and R are determined by k and z with no dependence on ξ.

Note that r is the real interest rate in terms of investment goods. The usual consumption Euler equation involves the real interest rate *in terms of consumption goods*. Thus, anticipated changes in consumption can depend on changes in the relative price of consumption and investment goods as well as on r.

9.4 Comparative Steady State Analysis

There are two complementary approaches to figuring out the dynamic response to shocks. In the previous chapter we took a bottom-up approach, asking first about the behavior at each point on the phase diagram. Here illustrated a top-down approach, starting with comparative steady-state analysis. Knowing the comparative steady-state analysis can given an anchor to some of the dynamic analysis. However, in the end, the top-down approach cannot give all the needed results and must be paired with a bottom-up approach.

Define as above

$$\kappa = f'^{-1}(\rho + \delta).$$

As long as ρ does not change, κ is a constant. Then the steady state condition $\dot{\lambda} = 0$ or

$$f'(k/z) = \rho + \delta$$

implies

$$k^* = z\kappa \tag{9.21}$$

The steady-state condition $\dot{k} = 0$ yields

$$i^* = \delta k^*, \tag{9.22}$$

which, in connection with the material balance condition (9.16) implies

$$\frac{c^*}{\xi} + g = zf(k^*/z) - \delta k^* = z[f(\kappa) - \delta\kappa]. \tag{9.23}$$

That is, given a fixed value of ρ and κ, the amount of output available for consumption and government purchases in the steady state is proportional to z.

These equations make it possible to do comparative steady state analysis. The long-run steady-state effects of permanent improvements in z (across-the-board labor-augmenting technology) and ξ (consumption goods technology) are given by the following table:

Variable	Steady-State Effect of z	Steady-State Effect of ξ
k	+	0
λ	−	?
c	+	+
y	+	0
i	+	0
w	+	0
R	0	0
r	0	0

(9.24)

Consider first a permanent across-the board improvement in technology z. Because z does not affect the form of the Euler equation $\frac{\dot{\lambda}}{\lambda} = \rho - r$ and the associated steady-state condition $\rho = r$, the steady-state real interest rate r and the steady-state rental rate $R = r + \delta$ are unchanged. With R unchanged, steady-state k/z must also remain unchanged, which implies that steady-state k, $y = zf(k/z)$, and $w = z[f(k/z) - (k/z)f'(k/z)]$ rise in proportion to z. In the steady state $i = \delta k$ and therefore $(c/\xi) + g = y - i = z[f(k/z) - \delta(k/z)]$ also rise in proportion to k. The consequent rise in c implies a fall in λ.

Now consider a permanent improvement in the consumption goods technology ξ. The table indicates that a permanent change in ξ affects only c and λ in steady state. Again there is no change in steady-state r, R or k/z. But with no change in z, this means that the steady-state capital stock k does not change. Because steady-state k does not change, steady-state i does not change and there is no change in steady-state y or w: the improvement in consumption goods technology has no effect on steady-state output or the real wage as measured in terms of investment goods. With no change in y

or i, $(c/\xi) + g = y - i$ must be unchanged, which implies that "consumption measured in terms of investment goods" c/ξ must stay the same—implying that steady-state consumption c must increase in proportion to ξ.

The effect of ξ on λ in the steady state depends on the felicity function u. If c_1 is what steady-state consumption would be with $\xi = 1$, then $c^* = \xi c_1$. By (9.19),

$$\lambda^* = \xi u'(c^*) = \xi u'(\xi c_1).$$

Taking the derivative with respect to ξ,

$$\begin{aligned} \frac{\partial \lambda^*}{\partial \xi} &= u'(\xi c_1) + \xi c_1 u''(\xi c_1) \qquad (9.25) \\ &= u'(c^*)\left[1 - \left(\frac{-c^* u''(c^*)}{u'(c^*)}\right)\right] \end{aligned}$$

The ratio $\frac{-c^* u''(c^*)}{u'(c^*)}$ is the *resistance to intertemporal substitution*. It is often inappropriately called "relative risk aversion," even in models with perfect foresight. In Chapter **??**, we will show that whether or not this ratio is relative risk aversion depends on how the utility function is extended to deal with uncertainty. In the familiar functional form

$$u(c) = \frac{c^{1-\beta}}{1-\beta},$$

the resistance to intertemporal substitution is constant at β and when

$$u(c) = \ln(c),$$

the resistance to intertemporal substitution is constant at 1.

The resistance to intertemporal substitution is equal to the reciprocal of the *elasticity of intertemporal substitution* s:

$$s(c^*) = \frac{-u'(c^*)}{c^* u''(c^*)}.$$

In terms of the elasticity of intertemporal substitution,

$$\frac{\partial \lambda^*}{\partial \xi} = u'(c^*)\left[1 - \frac{1}{s(c^*)}\right]$$

Thus, the consumption goods technology raises the marginal value of capital in the steady state when $s(c^*) > 1$ (*e.g.,* when $u(c) = 2\sqrt{c}$), but lowers the marginal value when $s(c^*) < 1$ (*e.g.,* when $u(c) = -1/c$). When $s(c^*) = 1$ (*e.g.,* when $u(c) = \ln(c)$), the consumption goods technology does not affect the steady-state marginal value of capital.

9.5 The Dynamic Response to Consumption-Goods Technology Shocks

On the phase diagram, the fact that a permanent improvement in consumption-goods technology ξ has no affect on k^* means that an increase in ξ must leaving the $\dot{\lambda} = 0$ unchanged. The $\dot{k} = 0$ locus must shift in the same direction as λ^*.

The dynamic response to a permanent increase in ξ is simple: λ can jump immediately to the new steady state. Thus, the comparative steady-state analysis essentially gives the whole story of the response to an increase in ξ. In general, there are no complex dynamic responses to permanent shocks that have no long-run effect on the state variable.

The dynamic arrows on the phase diagram depend only the (k, λ) location and the current values of exogenous variables. In particular, the position of the isoclines depends only on the current values of the exogenous variables. Therefore, for as long as ξ is altered, the isoclines must shift in the same way in response to a temporary shock as to a permanent shock. As shown in Figures 9.5, 9.5, and 9.5, a temporary improvement in consumption technology ξ will cause the capital stock to run down while the improvement lasts if $s > 1$, and will cause the capital stock to accumulate while the improvement lasts if $s < 1$. When $u(c) = \ln(c)$, so that $s = 1$, changes in ξ have no effects on the isoclines and so induce no dynamic effects on the phase diagram regardless of the time-path of the changes in ξ.

Intuitively, there are two effects. On one hand, consumption smoothing suggests that much of a temporary windfall such as a temporary improvement in consumption goods technology should be saved. On the other hand, the temporary improvement in consumption goods technology means that now is a particularly good time to consume—what can be called the *carpe diem* ("seize the day") effect. If the elasticity of intertemporal substitution is below 1, the consumption smoothing effect dominates. If the elasticity of intertemporal substitution is above 1, the *carpe diem* effect dominates. The two effects cancel each other out if the elasticity of intertemporal substitution is equal to 1.

The story evident on the phase diagram is reinforced by telling what happens to variables other than k and λ. The following tables show these effects for $s < 1$ and $s = 1$. The case when $s > 1$ is left as an exercise.

The Effects of a Temporary Improvement in Consumption Technology
When the Elasticity of Intertemporal Substitution ¡ 1

Variable	Impact ($t = 0$)	During ($t \in (0, T)$)	At ($t = T$)	After($t \in (T, \infty)$)
c	+	−	−	−
$\frac{c}{\xi}$	−	−	+	−
y	0	+	0	−
i	+	+	−	?
w	0	+	0	−
R	0	−	0	+
r	0	−	0	+
$\dot{k}$	+	+	−	+
$\dot{\lambda}$	0	+	0	−

(9.26)

The Effects of a Temporary Improvement in Consumption Technology
When the Elasticity of Intertemporal Substitution = 1

Variable	Impact ($t = 0$)	During ($t \in (0, T)$)	At ($t = T$)	After($t \in (T, \infty)$)
c	+	0	−	0
$\frac{c}{\xi}$	0	0	0	0
y	0	0	0	0
i	0	0	0	0
w	0	0	0	0
R	0	0	0	0
r	0	0	0	0
$\dot{k}$	0	0	0	0
$\dot{\lambda}$	0	0	0	0

(9.27)

Table 9.27 shows the strong form of consumption technology neutrality exhibited when $s = 1$ (logarithmic utility)—other than consumption itself, and of course utility, none of the key economic variables is affected by even a complex consumption technology shock. Even when $s \neq 1$, the same can be said for the reaction to an immediate, permanent consumption technology shock: none of the other key economic variables is affected by the shock.

9.6 The Dynamic Response to Across-the-Board Technology Shocks

In Section 9.4, we found that a permanent across-the-board technology improvement raised the steady state capital stock k^* and lowered the steady state marginal value of capital λ^*. The fact that k^* increases implies that the $\dot{\lambda} = 0$ locus must shift to the right. But it is not clear from such reasoning which way the $\dot{k} = 0$ locus shifts. For that, it is best to analyze the contemporaneous general equilibrium of the model at a given value of k and λ. Table 9.28 shows the direct effects of z on the other important variables. For completeness, the table gives the direct effects of ξ as well.

Variable	Function	Direct Effect of z	Direct Effect of ξ		
			$s < 1$	$s = 1$	$s > 1$
c :	$u'^{-1}(\lambda/\xi)$	0	+	+	+
$\frac{c}{\xi}$:	$\frac{u'^{-1}(\lambda/\xi)}{\xi}$	0	−	0	+
y :	$zf(k/z)$	+	0	0	0
i :	$zf(k/z) - \frac{u'^{-1}(\lambda/\xi)}{\xi}$	+	+	0	−
w :	$z\left[f\frac{k}{z} - \frac{k}{z}f'\frac{k}{z}\right]$	+	0	0	0
R :	$f'(k/z)$	+	0	0	0
r :	$f'(k/z) - \delta$	+	0	0	0
$\dot{k}$:	$zf(k/z) - \delta k - \frac{u'^{-1}(\lambda)}{\xi}$	+	+	0	−
$\dot{\lambda}$:	$\rho + \delta - f'(k/z)$	−	0	0	0

(9.28)

Looking at the column for z, the positive direct effect of z on $\dot{k}$ implies that an increase in z must shift the $\dot{k} = 0$ locus down, into the region that had $\dot{k} < 0$ at the lower value of z. (See Figure 9.6. In other words, the location of the $\dot{k} = 0$ locus must shift to counteract the direct effect of z. By the same kind of reasoning, the negative direct effect of z on $\dot{\lambda}$ implies that an increase in z must shift the $\dot{\lambda} = 0$ locus to the right, into the region that had $\dot{\lambda} > 0$ at the lower value of z. In its effect on $\dot{\lambda}$, the higher value of k counteracts the higher value of z.

9.6.1 The Dynamic Response to a Permanent Across-the-Board Technology Shock

As with other permanent shocks, the dynamic response to a permanent across-the-board improvement in technology is to jump to the new saddle path and follow it to the new steady state. But with the new steady state to the southeast of the old steady state, it is not clear whether the new saddle path lies above or below the old steady state. If the new saddle path lies above the old steady state, then λ initially jumps up. (See Figure 9.6.1.) If the new saddle path lies below the old steady state, then λ initially jumps down. (See Figure 9.6.1.) In the borderline case in which the new saddle path passes through the old steady state, there is no initial jump λ at all. (See Figure 9.6.1.)

How should one interpret this ambiguity? The level of λ is determined by two factors: (1) overall wealth, in the broad sense of abundance versus scarcity, and (2) expected future real interest rates. A permanent, across-the-board improvement in technology raises both wealth and real interest rates. The higher wealth tends to pull λ down. The higher expected real interest rates tends to pull λ up.

There is more than one way to divide a reaction into a wealth effect and an interest rate effect. A useful way of slicing comes from integrating the Euler equation $\frac{\dot{\lambda}}{\lambda} = \rho - r$ to the equation

$$\ln(\lambda_t) = \ln(\lambda_\infty) + \int_t^\infty [r_\tau - \rho]\, d\tau. \tag{9.29}$$

The term $\ln(\lambda_\infty)$ is the marginal value of capital in the long-run. The term $\int_t^\infty [r_\tau - \rho]\, d\tau$ is the total amount by which the interest rate is expected to exceed ρ in the future. It is akin to a very long-term real interest rate. The long-run effect of a shock on $\ln(\lambda_\infty)$ can be identified as the wealth effect. The effect of a shock on $\int_t^\infty [r_\tau - \rho]\, d\tau$ can be identified as the interest rate effect.

The impact effect of a shock is the sum of the direct effect and the effect of the initial jump in λ. In the borderline case in which λ does not jump, the impact effect of a permanent across-the-board technology shock is just the direct effect given in Table 9.28.

Before adding the effect of the initial jump in λ to the direct effect of z

to find the overall impact effect of a permanent increase in z, it is useful to see what can be deduced from other directions in ways that do not depend on whether the wealth effect or the interest rate effect dominates. To begin with, the impact effect plus the saddle-path effect must add up to the steady-state effect of z. In the new steady state, it must still be true that interest rate $r = \rho$ because of the Euler equation $\frac{\dot{\lambda}}{\lambda} = \rho - r$. Thus it is still true that the rental rate $R = \rho + \delta$ and the ratio k/z must be unchanged. With k/z unchanged, the capital stock k, output y and the real wage w rise in proportion to z. Since $i = \delta k$ at the steady state, investment i must rise in proportion to z as well. Finally, with both y and i increasing in proportion to z, $y - i$ also increases in proportion to z, which implies an increase in c in view of the material balance condition $y - i = c + g$. The saddle path effects are in accordance with the discussion in Chapter 8. The adding up constraint determines a number of the impact effects, as indicated by (9.31). In addition, these impact effects, plus the fact that i increases in impact can be deduced from the fact evident from the saddle-path that $\dot{k}$ must jump from zero to a positive value on impact, while $\dot{\lambda}$ must jump from zero to a negative value on impact.

Variable	Impact Effect	Saddle Path Effect	Overall Steady State Effect
c	$*$	$+$	$+$
y	$*$	$+$	$+$
i	$+$	?	$+$
w	$*$	$+$	$+$
R	$+$	$-$	0
r	$+$	$-$	0
$\dot{k}$	$+$	$-$	0
$\dot{\lambda}$	$-$	$+$	0

(9.30)

We have put asterisks rather than question marks for the impact effects that we have not yet determined in order to emphasize that not knowing the answer yet is not the same as knowing that an effect is ambiguous. Indeed, the impact effects on y and w are determinate, and only the effect on c is genuinely ambiguous, depending on whether the interest rate or the wealth effect dominates. To see this, turn to the approach of adding the direct effect of z to the effect of the jump in λ on impact. In the last three columns, the results in Table 9.31 have been used to determine many signs.

Variable	Direct Effect of z	Effect of λ		Impact Effect	
		Dominant Effect:	Interest Rate	Equal	Wealth
c	0	$-$	$-$	0	$+$
y	$+$	0	$+$	$+$	$+$
i	$+$	$+$	$+$	$+$	$+$
w	$+$	0	$+$	$+$	$+$
R	$+$	0	$+$	$+$	$+$
r	$+$	0	$+$	$+$	$+$
$\dot{k}$	$+$	$+$	$+$	$+$	$+$
$\dot{\lambda}$	$-$	0	$-$	$-$	$-$

(9.31)

Since output y and the real wage w depend only on k plus exogenous variables, not λ, the jump in λ has no effect on them and the direct effect equals the impact effect for these two variables. As for the impact effect of a permanent across-the-board technology shock on consumption, consumption depends only on λ and the unchanging ξ. Thus, c moves in the opposite direction to the initial jump in λ. When λ jumps up because the interest rate effect is dominant, c jumps down on impact. When λ jumps down because the wealth effect is dominant, c jumps up on impact. And when λ does not jump because the interest rate effect and wealth effect exactly counterbalance each other, c also does not jump.

9.6.2 The Dynamic Response to Temporary or Anticipated Across-the-Board Technology Shocks

The details of the response to temporary or anticipated across-the-board technology shocks are left to exercises. The most important point is that temporary and anticipated shocks cause λ to make its initial jump in the same direction as the corresponding permanent shock. For anticipated shocks, this requires no further assumption. For temporary shocks, this requires the reasonable assumption that the old and new saddle paths do not cross (or at least, do not cross at values of k between the old and new steady state values). [1] In the borderline case in which the new saddle path passes through the old steady state, an anticipated shock requires no movement until the moment

[1]In analyzing dynamic paths on the phase diagram, one useful principle is that a dynamic path never crosses the saddle path or divergent path while the dynamics for that saddle path and divergent path are operative. This principle traps the dynamic path for a temporary shock to z between the two saddle paths when they do not cross.

when z changes. A temporary shock in the borderline case will also typically involve very little initial jump in λ.

9.7 Taxes in the Ramsey Model

A more complete treatment of taxation must wait until Chapter 17. However, it is useful to look at a few aspects of taxation in the context of the Ramsey model, arguing for some results that will not be fully proved until later.

Consider the optimizing household's problem when there is a payroll tax τ_w, a sales tax on consumption goods τ_c, a tax on interest income τ_r, an asset tax τ_a and a lump-sum tax τ_0:

$$\max_c \int_0^\infty e^{-\rho t} u(c) dt \tag{9.32}$$

s.t.

$$\dot{a} = [(1-\tau_r)r - \tau_a]a + (1-\tau_w)w - (1+\tau_c)c - \tau_0 \tag{9.33}$$

Using $\bar{\lambda}$ for the costate variable of the household's problem, the current value Hamiltonian is

$$H = u(c) + \bar{\lambda}\{[(1-\tau_r)r - \tau_a]a + (1-\tau_w)w - (1+\tau_c)c - \tau_0\}$$

The first-order condition (for c) is

$$u'(c) = \bar{\lambda}(1+\tau_c). \tag{9.34}$$

The Euler equation is

$$\frac{\dot{\lambda}}{\lambda} = \rho + \tau_a - (1-\tau_r)r. \tag{9.35}$$

The degree to which taxes affect household behavior is indicated by how they enter into these two equations. Let us consider the various taxes from the least distortionary in this model to the most distortionary. The lump-sum tax, of course, does not directly affect any marginal conditions. Therefore, as discussed in Chapter 17, one can find the effect of lump-sum taxes on household behavior in general equilibrium through the effect of government purchases on the material balance condition without any other consideration of the effects of lump-sum taxes. With the quantity of labor fixed, it is not surprising that the payroll tax τ_w also has no effect on either equation ((9.34) and (9.35)). It acts just like a lump-sum tax. The before-tax wage w can be determined in exactly the same way it otherwise would be by firm labor demand with no change in household behavior. Because of the totally inelastic labor supply, the household bears the full cost of the tax.

Next in order is the consumption tax. As long as it is definitely kept at a constant value τ_c, its effects can be fully counteracted by a change in $\bar{\lambda}$ inversely proportion to $1 + \tau_c$. Since $\bar{\lambda}$ is a shadow price that belongs to the household alone, there is nothing to prevent this adjustment, which will cause the household to behave in general equilibrium exactly the same as in the absence of the consumption tax. Thus, a constant consumption tax is not distortionary in the Ramsey model! What is going on? Distortions must arise from doing one thing rather than another. With labor supply totally inelastic, the only thing a household can do with its intertemporal budget other than consuming now is consuming later. A constant consumption tax has the same effect on consuming later as on consuming now; therefore it does not distort the choice between consuming now and consuming later.

The important limitation on $\bar{\lambda}$ is that it cannot jump unless the household receives new information. If events are unfolding as expected, $\bar{\lambda}$ must evolve according to the Euler equation (9.35). Thus, movements in the consumption tax that can be anticipated in advance can have very important effects. Because $\bar{\lambda}$ cannot jump to counteract it, an anticipated jump in τ_c causes the product $\bar{\lambda}(1+\tau_c)$ to jump in the same direction and c to jump in the opposite direction.

The proportional asset tax τ_a is clearly distortionary. Indeed, it effects the pair of equations (9.34) and (9.35) in exactly the same way as ρ—that is, the tax τ_a has the same effect as changing ρ to $\rho + \tau_a$ would have. The asset tax is in the category of *distortionary taxes that have effects equivalent to changing a parameter in the social planner's problem.*

Finally, consider the interest rate tax τ_r. In steady state, where r is constant, a constant value of τ_r would have the same effect as τ_a or a higher value of ρ. But out of steady state, even a constant value of τ_r has an effect different from any parameter in the social planner's problem. The interest tax tends to suppress the adjustments in economic behavior that are called forth by variations in the interest rate away from the steady state.

Exercises

1. Analyze the effects of a temporary increase in ξ on c, y, i, w, R, r, $\dot{k}$ and $\dot{\lambda}$ when the elasticity of intertemporal substitution is greater than

1. (Report your answer in an appropriate table.)

2. Analyze the effects of an anticipated future increase in ξ on the phase diagram and on c, y, i, w, R, and r.

3. Analyze the effects of a temporary increase in z on the phase diagram and on c, y, i, w, R, and r.

4. Analyze the effects of an anticipated future increase in z on the phase diagram and on c, y, i, w, R, and r.

Chapter 10

Appendix: Log-Linearization Techniques

This appendix can be read in two ways. In its entirety, it provides simple proofs for various log-linearization relationships. It also serves as a list of helpful rules if one ignores the intermediate steps in the various equations, looking instead only at the first and last quantities that are equated in each case.

10.1 Definitions

10.1.1 Absolute Deviations and Logarithmic Deviations

For any variable x, the (absolute) deviation $\tilde{x}$ is a total differential with respect to all variables except for time, evaluated at the steady state:

$$\tilde{x} = dx|_*, \tag{10.1}$$

where $|_*$ means "evaluated at the steady state values of all the variables." The deviation $\tilde{x}$ can be calculated as a function of the arguments of x in exactly the same way as any other total differential.

The logarithmic deviation $\check{x}$ is the total differential of $\ln(x)$ with respect to all variables except for time, evaluated at the steady state:

$$\check{x} = d\ \ln(x)|_* = \frac{dx}{x}|_* = \frac{dx|_*}{x^*} = \frac{\tilde{x}}{x^*}. \tag{10.2}$$

Multiplied through by x^*, this also implies that

$$x^* \check{x} = \tilde{x}. \tag{10.3}$$

In any of the expressions below, if $x^* = 0$, the expression $x^* \check{x}$ should be replaced by $\tilde{x}$, and likewise if any other variable is zero at the steady state. There are also a few cases (such as for the real interest rate) in which it is preferable to use the absolute deviation $\tilde{x}$ because it has a more interesting interpretation than the logarithmic deviation $\check{x}$. But for most variables, the logarithmic deviation is more interesting, since it can be interpreted in terms of percentage deviations from its steady-state value.

10.1.2 Time Derivatives of Deviations and Logarithmic Deviations

The time derivative of a deviation is

$$\begin{aligned} \dot{\tilde{x}} &= \frac{d}{dt}\tilde{x} \\ &= \frac{d}{dt}(dx|_*). \end{aligned} \tag{10.4}$$

The time derivative of a logarithmic deviation is

$$\begin{aligned} \dot{\check{x}} &= \frac{d}{dt}\check{x} \\ &= \frac{d}{dt}(d\ln(x)|_*) \\ &= \frac{d}{dt}\frac{dx|_*}{x^*} \\ &= \frac{1}{x^*}\frac{d}{dt}(dx|_*) \\ &= \frac{\dot{\tilde{x}}}{x^*}. \end{aligned} \tag{10.5}$$

10.2 Rules for Log-Linearizing Atemporal Equations

subsectionMultiplying, Dividing and Taking to a Power

One nice aspect of log-linearization is the simplicity of the product rule. Using an overbrace to group factors,

$$\overset{\vee}{\overbrace{xy}} = d\,\ln(xy)|_* = d\,\ln(x)|_* + d\,\ln(y)|_* = \check{x} + \check{y}. \tag{10.6}$$

Similarly, for quotients,

$$\overset{\vee}{\overbrace{x/y}} = d\,\ln(x/y)|_* = d\,\ln(x)|_* - d\,\ln(y)|_* = \check{x} - \check{y}. \tag{10.7}$$

In accordance with these rules, if a is a constant,

$$\overset{\vee}{\overbrace{ax}} = \check{a} + \check{x} = \check{x}, \tag{10.8}$$

and

$$\overset{\vee}{\overbrace{a/x}} = \check{a} - \check{x} = -\check{x}. \tag{10.9}$$

Also simple is the rule for dealing with powers. If a is a constant,

$$\overset{\vee}{\overbrace{x^a}} = d\,\ln(x^a)|_* = d\,a\ln(x)|_* = a\check{x}. \tag{10.10}$$

10.2.1 Adding and Subtracting

The logarithmic deviation of a sum involves the share of each term added together in the entire sum. Assuming that neither x^* nor y^* is zero,

$$\begin{aligned}
\overbrace{x+y}^{\vee} &= d\,\ln(x+y)|_* && (10.11)\\
&= \frac{dx}{x+y}|_* + \frac{dy}{x+y}|_*\\
&= \frac{x}{x+y}\frac{dx}{x}|_* + \frac{y}{x+y}\frac{dy}{y}|_*\\
&= \frac{x^*}{x^*+y^*}\check{x} + \frac{y^*}{x^*+y^*}\check{y}.
\end{aligned}$$

and the logarithmic deviation of a difference is

$$\begin{aligned}
\overbrace{x-y}^{\vee} &= d\,\ln(x-y)|_* && (10.12)\\
&= \frac{dx}{x-y}|_* - \frac{dy}{x-y}|_*\\
&= \frac{x}{x-y}\frac{dx}{x}|_* - \frac{y}{x-y}\frac{dy}{y}|_*\\
&= \frac{x^*}{x^*-y^*}\check{x} - \frac{y^*}{x^*-y^*}\check{y}.
\end{aligned}$$

Similarly, the logarithmic deviation of the sum of several terms is

$$\overbrace{\sum_{i=1}^{n} x_i}^{\vee} = \frac{x_1^*}{\sum_{i=1}^n x_i^*}\check{x}_1 + \frac{x_2^*}{\sum_{i=1}^n x_i^*}\check{x}_2 + \ldots + \frac{x_n^*}{\sum_{i=1}^n x_i^*}\check{x}_n. \qquad (10.13)$$

Each of the coefficients of $\check{x}_i$ is equal to the share of x_i^* in the sum $\sum_{i=1}^n x_i^*$. If $x_i^* = 0$ for any i, the expression must be modified by putting $\tilde{x}_i$ in place of $x_i^*\check{x}_i$.

10.2.2 Other Functions

For an exponential function,

$$\begin{aligned}
\overbrace{e^x}^{\vee} &= d\ln(e^(x))|_* && (10.14)\\
&= dx|_*\\
&= \tilde{x}\\
&= x^*\check{x}.
\end{aligned}$$

In general, for any differentiable function f of x,

$$\begin{aligned}\overbrace{f(x)}^{\vee} &= d\ln(f(x))|_* \\ &= \frac{f'(x)}{f(x)}dx|_* \\ &= \frac{f'(x^*)}{f(x^*)}\tilde{x} \\ &= \frac{x^* f'(x^*)}{f(x^*)}\breve{x}.\end{aligned} \tag{10.15}$$

The coefficient $\frac{x^* f'(x^*)}{f(x^*)}$ is the classic form of an elasticity, evaluated at the steady state.

For composite functions, an interesting version of the chain rule applies:

$$\begin{aligned}\overbrace{f(g(x))}^{\vee} &= d\ln(f(g(x)))|_* \\ &= \frac{f'(g(x))g'(x)}{f(g(x))}dx|_* \\ &= \frac{g(x^*)f'(g(x^*))}{f(g(x^*))}\frac{x^* g'(x^*)}{g(x^*)}\breve{x}.\end{aligned} \tag{10.16}$$

In words, the elasticity of $f(g(x))$ with respect to x is equal to the product of the elasticity of f (with respect to $g(x)$) and the elasticity of g (with respect to x).

The logarithmic deviation of multivariate functions adds up terms that accord with the dependence on each variable:

$$\begin{aligned}\overbrace{f(x,y)}^{\vee} &= d\ln(f(x,y))|_* \\ &= (\frac{f_x(x,y)}{f(x,y)}dx + \frac{f_y(x,y)}{f(x,y)}dy)|_* \\ &= \frac{f_x(x^*,y^*)}{f(x^*,y^*)}\tilde{x} + \frac{f_y(x^*,y^*)}{f(x^*,y^*)}\tilde{y} \\ &= \frac{x^* f_x(x^*,y^*)}{f(x^*,y^*)}\breve{x} + \frac{y^* f_y(x^*,y^*)}{f(x^*,y^*)}\breve{y}\end{aligned} \tag{10.17}$$

This formula can be extended to any number of variables.

10.2.3 Functional Relationships

If

$$\ln(z) = f(x)$$

then

$$\begin{aligned} \check{z} &= d\ln(z)|_* && (10.18) \\ &= df(x)|_* \\ &= f'(x^*)\tilde{x} \\ &= x^* f'(x^*)\check{x}, \end{aligned}$$

and if

$$\ln(z) = f(\ln(x)),$$

then

$$\begin{aligned} \check{z} &= d\ln(z)|_* && (10.19) \\ &= df(\ln(x))|_* \\ &= f'(\ln(x^*))\frac{1}{x^*}\tilde{x} \\ &= f'(\ln(x^*))\check{x}. \end{aligned}$$

In the multivariate case, if

$$\ln(z) = f(x, y)$$

then

$$\begin{aligned} \check{z} &= d\ln(z)|_* && (10.20) \\ &= df(x, y)|_* \\ &= f_x(x^*, y^*)\tilde{x} + f_y(x^*, y^*)\tilde{y} \\ &= x^* f_x(x^*, y^*)\check{x} + y^* f_y(x^*, y^*)\check{y}. \end{aligned}$$

Finally, if y is given implicitly as a function of x by

$$0 = f(x, y),$$

then total differentiation reveals that

$$0 = f_x(x, y)dx + f_y(x, y)dy = xf_x(x, y)\frac{dx}{x} + yf_y(x, y)\frac{dy}{y}.$$

Evaluating this at the steady state reveals that

$$0 = x^* f_x(x^*, y^*)\check{x} + y^* f_y(x^*, y^*)\check{y},$$

or

$$\check{y} = -\frac{x^* f_x(x^*, y^*)}{y^* f_y(x^*, y^*)}\check{x}. \tag{10.21}$$

10.3 Rules for Log-Linearizing Dynamic Equations

10.3.1 Interchanging the Time Derivative with the Total Differential with Respect to All Other Variables

Since the time derivative can be interchanged with the total differential with respect to all other variables, one can also think of the time derivative of a deviation as the deviation of a time derivative:

$$\begin{aligned} \dot{\tilde{x}} &= \frac{d}{dt}\tilde{x} \qquad (10.22) \\ &= \frac{d}{dt}(dx|_*) \\ &= (d[\frac{dx}{dt}])|_* \\ &= (d[\dot{x}])|_* \\ &= \widetilde{\dot{x}}. \end{aligned}$$

Similarly, the time derivative of a logarithmic deviation is the deviation of a logarithmic time derivative:

$$\begin{aligned}
\dot{\check{x}} &= \frac{d}{dt}\check{x} \\
&= (\frac{d}{dt}[d\ln(x)])|_* \\
&= (d[\frac{d\ln(x)}{dt}])|_* \\
&= (d[\frac{\dot{x}}{x}])|_* \\
&= \frac{(d[\dot{x}])|_*}{x^*} - (\frac{dx}{x^2}\dot{x})|_* \\
&= \frac{(d[\dot{x}])|_*}{x^*},
\end{aligned} \tag{10.23}$$

since by the definition of a steady state, $(\dot{x})|_* = 0$.

Note however, that

$$\dot{\check{x}} \neq \overbrace{\dot{x}}^{\vee},$$

since in general,

$$\begin{aligned}
\overbrace{\dot{x}}^{\vee} &= d\ln(\dot{x})|_* \\
&= \frac{d(\dot{x})}{\dot{x}}|_* \\
&= \frac{(d[\dot{x}])|_*}{(\dot{x})|_*} \\
&= \frac{(d[\dot{x}])|_*}{0} \\
&\neq \frac{(d[\dot{x}])|_*}{x^*}.
\end{aligned}$$

10.3.2 The Basic Rule

If

$$\dot{y} = x,$$

then

$$\begin{aligned} \dot{\tilde{y}} &= \widetilde{\dot{y}} \\ &= \tilde{x}, \end{aligned} \tag{10.24}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \\ &= \frac{\tilde{x}}{y^*}. \end{aligned} \tag{10.25}$$

In both cases, it is tempting to replace $\tilde{x}$ with $x^*\check{x}$, but if $\dot{y} = x$, then $x^* = \dot{y}^* = 0$, by the definition of a steady state.

10.3.3 Sums and Differences

If

$$\dot{y} = x_1 + x_2 + \ldots + x_n,$$

then

$$\begin{aligned} \dot{\tilde{y}} &= \widetilde{\dot{y}} \\ &= x_1 + x_2 \widetilde{+ \ldots} + x_n \\ &= \tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n \\ &= x_1^*\check{x}_1 + x_2^*\check{x}_2 + \ldots + x_n^*\check{x}_n, \end{aligned} \tag{10.26}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \\ &= \frac{\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n}{y^*} \\ &= \left(\frac{x_1^*}{y^*}\right)\check{x}_1 + \left(\frac{x_2^*}{y^*}\right)\check{x}_2 + \ldots + \left(\frac{x_n^*}{y^*}\right)\check{x}_n. \end{aligned} \tag{10.27}$$

If, on the other hand,

$$\dot{y} = x_1 - x_2,$$

then

$$\begin{aligned} \dot{\tilde{y}} &= \widetilde{\dot{y}} \\ &= \tilde{x}_1 - \tilde{x}_2 \\ &= x_1^* \check{x}_1 - x_2^* \check{x}_2 \end{aligned} \tag{10.28}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \\ &= \left(\frac{x_1^*}{y^*}\right) \check{x}_1 - \left(\frac{x_2^*}{y^*}\right) \check{x}_2. \end{aligned} \tag{10.29}$$

10.3.4 Products and Quotients

Using the basic rule, if

$$\dot{y} = x_1 x_2 \ldots x_n,$$

then at the steady state, at least one of the x_i^* must be zero. Let us say that $x_1^* = 0$. Then

$$\begin{aligned} \dot{\tilde{y}} &= x_1 \widetilde{x_2 \ldots x_n} \\ &= x_1^* \widetilde{x_2 \ldots x_n} + (x_2^* \ldots x_n^*) \tilde{x}_1 \\ &= (x_2^* \ldots x_n^*) \tilde{x}_1, \end{aligned} \tag{10.30}$$

and

$$\dot{\check{y}} = \frac{x_2^* \ldots x_n^*}{y^*} \tilde{x}_1. \tag{10.31}$$

If

$$\dot{y} = x_1 / x_2$$

then $x_1^* = 0$ and

$$\dot{\tilde{y}} = \frac{\tilde{x}_1}{x_2^*} \tag{10.32}$$

and

$$\dot{\check{y}} = \frac{\tilde{x}_1}{y^* x_2^*}. \tag{10.33}$$

If a product or a quotient is only one term in a larger sum or difference, the formulas of the previous subsection can be combined with those of the section on static log-linearization.

10.3.5 Univariate Functional Relationships

If

$$\dot{y} = f(x),$$

then

$$\begin{aligned} \dot{\tilde{y}} &= \widetilde{\dot{y}} \\ &= \widetilde{f(x)} \\ &= f'(x^*)\tilde{x} \\ &= f'(x^*)x^*\check{x}. \end{aligned} \tag{10.34}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \\ &= \frac{x^*}{y^*} f'(x^*)\check{x}. \end{aligned} \tag{10.35}$$

Sums are easy in this context, since if

$$\dot{y} = f(x) + g(x),$$

then

$$\begin{aligned} \dot{\check{y}} &= \frac{x^*}{y^*}[f(x)+g(x)]'\check{x} \quad (10.36) \\ &= \frac{x^*}{y^*}f'(x)\check{x} + \frac{x^*}{y^*}g'(x)\check{x}. \end{aligned}$$

10.3.6 Multivariate Functional Relationships

If

$$\dot{y} = f(x, y),$$

then using subscripts for partial derivatives,

$$\begin{aligned} \dot{\tilde{y}} &= \widetilde{\dot{y}} \quad (10.37) \\ &= f_x(x^*, y^*)\tilde{x} + f_y(x^*, y^*)\tilde{y} \\ &= f_x(x^*, y^*)x^*\check{x} + f_y(x^*, y^*)y^*\check{y}, \end{aligned}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \quad (10.38) \\ &= \frac{x^*}{y^*}f_x(x^*, y^*)\check{x} + f_y(x^*, y^*)\check{y}. \end{aligned}$$

Similarly, if

$$\dot{y} = f(x_1, x_2, \ldots, x_n, y),$$

then, abbreviating $f^*_{x_i} = \frac{\partial f(x_1, x_2, \ldots, x_n, y)}{\partial x_i}|_*$ and so on,

$$\begin{aligned} \dot{\tilde{y}} &= \widetilde{\dot{y}} \quad (10.39) \\ &= f^*_{x_1}\tilde{x}_1 + f^*_{x_2}\tilde{x}_2 + \ldots + f^*_{x_n}\tilde{x}_n + f^*_y\tilde{y} \\ &= f^*_{x_1}x^*_1\check{x}_1 + f^*_{x_2}x^*_2\check{x}_2 + \ldots + f^*_{x_n}x^*_n\check{x}_n + f^*_y y^*\check{y}, \end{aligned}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \quad (10.40) \\ &= \frac{x^*_1}{y^*}f^*_{x_1}\check{x}_1 + \frac{x^*_2}{y^*}f^*_{x_2}\check{x}_2 + \ldots + \frac{x^*_n}{y^*}f^*_{x_n}\check{x}_n + f^*_y\check{y}. \end{aligned}$$

10.3.7 Relationships in Logarithmic Growth Rates

If

$$\frac{\dot{y}}{y} = x,$$

then $x^* = 0$. Applying the results of the previous subsection to the relationship

$$\dot{y} = yx,$$

$$\begin{aligned} \dot{\tilde{y}} &= x^*\tilde{y} + y^*\tilde{x} \\ &= y^*\tilde{x} \end{aligned} \tag{10.41}$$

and

$$\begin{aligned} \dot{\check{y}} &= \frac{\dot{\tilde{y}}}{y^*} \\ &= \tilde{x}. \end{aligned} \tag{10.42}$$

Putting $f(x)$ in place of x, if

$$\frac{\dot{y}}{y} = f(x),$$

then

$$\begin{aligned} \dot{\check{y}} &= \widetilde{f(x)} \\ &= f'(x^*)\tilde{x} \\ &= f'(x^*)x^*\check{x}. \end{aligned} \tag{10.43}$$

Similarly, if

$$\frac{\dot{y}}{y} = f(x_1, x_2, \ldots, x_n, y),$$

then

$$\begin{aligned} \dot{\check{y}} &= \widetilde{f(x_1, x_2, \ldots, x_n, y)} \\ &= f^*_{x_1}\tilde{x}_1 + f^*_{x_2}\tilde{x}_2 + \ldots + f^*_{x_n}\tilde{x}_1 + f^*_y\tilde{y} \\ &= f^*_{x_1}x^*_1\check{x}_1 + f^*_{x_2}x^*_2\check{x}_2 + \ldots + f^*_{x_n}x^*_n\check{x}_1 + f^*_y y^*\check{y} \end{aligned} \tag{10.44}$$

Chapter 11

Production and the Solow Residual

11.1 Introduction: Log-Linearizing

In economics there is often a trade-off between the advantages of generality and the advantages of specificity. On one hand, using general functions to represent preferences and production possibilities allows one to establish general results. On the other hand, specific functional forms often make it easier to do quantitative calculations.

Fortunately, in any model where the key issues depend on the *local* behavior of preferences and production possibilities, it is possible to use general functional forms, yet still quantify effects in detail. In particular, one can use Taylor expansions that apply to very general functions, then pay attention to the specific magnitudes of the various coefficients in that Taylor expansion. Because this approach allows one to work with general functions, it is sometimes called *nonparametric*. However, in this kind of nonparametric approach, the coefficients in the Taylor expansion (or ratios of these coefficients) can be treated as if they were parameters, and we will often call them "parameters," since these coefficients describe the shape of the function at least locally.

In order to make the various coefficients easy to interpret, it is often better to make a Taylor expansion in logarithms, so that the coefficients will be like elasticities. Forming such a logarithmic Taylor expansion is called *log-linearizing*.

In this chapter, we will introduce the (nonparametric) log-linearizing approach by looking at the firm's production decision at a point in time. Although we will not look at any explicit dynamics in this chapter, these techniques will be applied to a dynamic problem in the next chapter.

The point one log-linearizes around is often the steady state, but not always. For a few lines, in order to introduce some new notation, let us use the steady-state asterisk to denote the point around which the log-linearizing takes place. If x^* is the point around which log-linearizing is taking place, it is helpful to have a way to denote $\ln(x) - \ln(x^*)$. We will use the notation "check" for such a logarithmic deviation; that is, we will write

$$\check{x} = \ln(x) - \ln(x^*).$$

For the first half of the book, we will assume that this deviation is small enough that a first-order Taylor expansion is reasonably accurate. In the second half of the book we will look at second-order Taylor expansions. (Formally, we will consider $\check{x}$, like $\tilde{x}$, to be a differential of calculus, so that we can write a first-order Taylor expansion as an exact equation in these differentials.)

In log-linearizing a variable y, the coefficient $\check{x}$ in the expression for $\check{y}$ is equal to

$$\frac{\partial \ln(y)}{\partial \ln(x)} = x\frac{\partial \ln(y)}{\partial x} = \frac{x}{y}\frac{dy}{dx},$$

evaluated at $x = x^*$, $y = y^*$, etc. Note also that by an ordinary Taylor expansion of $\ln(x)$ around x^*,

$$\check{x} = \widetilde{\ln(x)} = \frac{d\ln(x)}{dx}|_{x=x^*} \cdot \tilde{x} = \frac{\tilde{x}}{x^*},$$

where $\tilde{x}$ represents $x - x^*$ as in previous chapters.

11.2 The Production Relation

Consider the production function

$$y = \xi F(k, zn),$$

where ξ is the level of multiplicative technology and z is the level of labor-augmenting technology. Since

$$\ln(y) = \ln(\xi) + \ln(F(k, zn)),$$

one can calculate

$$\frac{\partial \ln(y)}{\partial \ln(\xi)} = 1,$$

$$\frac{\partial \ln(y)}{\partial \ln(k)} = \frac{kF_1(k, zn)}{F(k, zn)},$$

$$\frac{\partial \ln(y)}{\partial \ln(z)} = \frac{znF_2(k, zn)}{F(k, zn)}$$

and

$$\frac{\partial \ln(y)}{\partial \ln(n)} = \frac{znF_2(k, zn)}{F(k, zn)},$$

where F_1 is the derivative of F with respect to its first argument k and F_2 is the derivative of F with respect to its second argument zn. Therefore, if we omit asterisks (*) in order to emphasize that this log-linearizing is valid around any point,

$$\check{y} = \check{\xi} + \frac{kF_1(k, zn)}{F(k, zn)}\check{k} + \frac{znF_2(k, zn)}{F(k, zn)}[\check{z} + \check{n}]. \tag{11.1}$$

11.2.1 Constant, Increasing or Decreasing Returns to Scale

The degree of returns to scale is the elasticity of output with respect to a balanced expansion in all inputs. In a balanced expansion of k and n, $\check{k} = \check{n} = \check{x}$ where x is an overall measure of inputs, $\check{\xi} = \check{z} = 0$ and

$$\check{y} = [\frac{kF_1(k, zn)}{F(k, zn)} + \frac{znF_2(k, zn)}{F(k, zn)}]\check{x},$$

or

$$\check{y} = \gamma\check{x},$$

where

$$\gamma = \frac{kF_1(k, zn)}{F(k, zn)} + \frac{znF_2(k, zn)}{F(k, zn)}$$

is the degree of returns to scale. At least locally, $\gamma = 1$ represents constant returns to scale, $\gamma > 1$ represents increasing returns to scale and $\gamma < 1$ represents decreasing returns to scale.

11.2.2 The Marginal Productivity Shares of Capital and Labor

In general, when $\check{k} \neq \check{n}$, define the capital's marginal productivity share θ by

$$\theta = \frac{\frac{kF_1(k,zn)}{F(k,zn)}}{\frac{kF_1(k,zn)}{F(k,zn)} + \frac{znF_2(k,zn)}{F(k,zn)}} = \frac{\frac{kF_1(k,zn)}{F(k,zn)}}{\gamma}.$$

Then Labor's marginal productivity share is

$$1 - \theta = \frac{\frac{znF_2(k,zn)}{F(k,zn)}}{\frac{kF_1(k,zn)}{F(k,zn)} + \frac{znF_2(k,zn)}{F(k,zn)}} = \frac{\frac{znF_2(k,zn)}{F(k,zn)}}{\gamma},$$

and (11.1) can be rewritten as

$$\check{y} = \check{\xi} + \gamma[\theta\check{k} + (1-\theta)(\check{z} + \check{n})] = \gamma[\check{x} + (1-\theta)\check{z}] + \check{\xi}, \tag{11.2}$$

where the logarithmic change in overall input $\check{x}$ is given by

$$\check{x} = \theta\check{k} + (1-\theta)\check{n},$$

a weighted average of logarithmic changes in the two inputs, weighted by their marginal productivity shares.

By comparing (11.1) and (11.2), it is apparent that

$$\frac{\partial \ln(y)}{\partial \ln(k)} = \frac{kF_1(k,zn)}{F(k,zn)} = \gamma\theta \tag{11.3}$$

and

$$\frac{\partial \ln(y)}{\partial \ln(n)} = \frac{nF_2(k,zn)}{F(k,zn)} = \gamma(1-\theta) \tag{11.4}$$

11.3 The Solow Residual

The overall effect of shocks to both kinds of technology on output is

$$\check{\xi} + \gamma(1-\theta)\check{z}.$$

By (11.2), this overall effect of technology is

$$\check{\xi} + \gamma(1-\theta)\check{z} = \check{y} - \gamma\check{x} = \check{y} - \gamma[\theta\check{k} + (1-\theta)\check{n}].$$

If we write TFP for total factor productivity, then "true log total factor productivity" is

$$\ln(\text{TFP}) = \breve{y} - \gamma[\theta\breve{k} + (1-\theta)\breve{n}].$$

and

$$\frac{d}{dt}[\breve{y} - \gamma[\theta\breve{k} + (1-\theta)\breve{n}] = \dot{\breve{y}} - \gamma[\theta\dot{\breve{k}} + (1-\theta)\dot{\breve{n}}]$$

is "true total factor productivity growth"—abbreviated "true TFPG."

By contrast, the "standard Solow residual" arises from a particular way of trying to measure true TFPG. In brief, the standard Solow residual is

$$\frac{d}{dt}\{\breve{y} - [1 - \frac{wn}{py}]\breve{k} - \frac{wn}{py}\breve{n}, \}$$

where w is the wage for labor and p is the price of output. If the firm's output is taken as the numeraire, p is normalized to 1 and w will be the real product wage.

We can show that if returns to scale are constant ($\gamma = 1$) and there is perfect competition, true TFPG is equal to the standard Solow residual. Perfect competition implies that the real product wage must equal the marginal product of labor:

$$\frac{w}{p} = \frac{\partial y}{\partial n}.$$

so that the coefficient of $\breve{n}$ in the expression for $\breve{y}$ is

$$\gamma(1-\theta) = \frac{\partial \ln(y)}{\partial \ln(n)} = \frac{\partial y}{\partial n}\frac{n}{y} = \frac{wn}{py}. \tag{11.5}$$

Constant returns to scale makes the left side of (**??**) equal to $1 - \theta$, so that $1 - \theta = wn/py$ and $\theta = 1 - (wn/py)$, while under constant returns to scale, true TFPG is

$$\frac{d}{dt}\{\breve{y} - \theta\breve{k} - (1-\theta)\breve{n}\}.$$

11.3.1 Mismeasurement of True Total Factor Productivity

Integrating (or in discrete-time, summing up) the Solow residual over time yields a quantity that can be compared closely with (log) total factor productivity in some circumstances. To be more specific, the integral of the standard Solow residual is

$$\begin{aligned} \check{y} - [1 - \frac{wn}{py}]\check{k} - \frac{wn}{py}\check{n} &= \check{y} - \gamma[\theta\check{k} + (1-\theta)\check{n}] \\ &\quad +(\gamma - 1)[\theta\check{k} + (1-\theta)\check{n}] \\ &\quad +[(1-\theta) - \frac{wn}{py}][\check{n} - \check{k}]. \qquad (11.6) \end{aligned}$$

Robert Hall (1988) ["The Relation between Price and Marginal Cost in U.S. Industry," JPE 96, 921–947.] argues that if there are constant returns to scale ($\gamma = 1$), imperfect competition makes marginal revenue lower than price and so depresses the real product wage $\frac{w}{p}$ below the marginal physical product of labor. We will discuss imperfect competition more in Chapter **??**. For now, the only fact we need about imperfect competition is that if the labor market has perfect competition, but *the market for the firm's output* has imperfect competition, the physical marginal product of labor is equal to the markup ratio μ times the real wage:

$$\frac{\partial y}{\partial n} = \mu \frac{w}{p},$$

with $\mu > 1$. (In contrast, a markup ratio μ equal to one represents perfect competition.) Solving for $\frac{w}{p}$ and using (11.4),

$$\frac{w}{p} = \frac{1}{\mu}\frac{\partial y}{\partial n}$$

Using (11.3), this implies that[1]

$$\frac{wn}{py} = \frac{1}{\mu}\frac{n}{y}\frac{\partial y}{\partial n} = \frac{\gamma}{\mu}(1-\theta). \qquad (11.7)$$

[1] By similar reasoning, if one can observe the rental rate for capital R

$$\frac{Rk}{py} = \frac{1}{\mu}\frac{k}{y}\frac{\partial y}{\partial k} = \frac{\gamma}{\mu}\theta.$$

In Hall (1988), however, the rental rate R is treated as unobservable.

Thus, (11.6) can be rewritten

$$\begin{aligned} \check{y} - [1 - \frac{wn}{py}]\check{k} - \frac{wn}{py}\check{n} &= \check{y} - \gamma[\theta\check{k} + (1-\theta)\check{n}] \\ &\quad +(\gamma - 1)[\theta\check{k} + (1-\theta)\check{n}] \\ &\quad +[1 - \frac{\gamma}{\mu}](1-\theta)[\check{n} - \check{k}]. \end{aligned} \tag{11.8}$$

By (11.7), if $\gamma = 1$ (constant returns to scale), then $\frac{wn}{py} = \frac{\gamma}{\mu}(1-\theta)$ underestimates labor's share $1-\theta$. Why would it matter if labor's true share is underestimated? Take a look at (11.8). With $\gamma = 1$, the term $(\gamma-1)[\theta\check{k} + (1-\theta)\check{n}]$ drops out of (11.6), leaving only the last bias term

$$\begin{aligned} [(1-\theta) - \frac{wn}{py}][\check{n} - \check{k}] &= (1 - \frac{1}{\mu})(1-\theta)[\check{n} - \check{k}]. \\ &= (\mu - 1)\frac{wn}{py}[\check{n} - \check{k}]. \end{aligned}$$

Since labor moves proportionately more than capital over the course of a business cycle—in large part because of the stock-flow distinction—the difference $\check{n} - \check{k}$ is procyclical in the data, making

$$(1 - \frac{1}{\mu})(1-\theta)[\check{n} - \check{k}]$$

procyclical and making the standard Solow residual more procyclical than the true Solow residual.

Maintaining for a moment more the assumption of constant returns to scale (as well as the assumption of perfect competition in the labor market), there is a clear bound on the magnitude of the mismeasurement Hall (1988) points to, since labor's true marginal productivity share $1-\theta$ is always less than 1 (100%), and the observed value of $\frac{wn}{py}$ is on average something like .6 (60 %) or .7 (70%). For an industry in which $\frac{wn}{py}$ is .7, the maximum possible underestimate of labor's true share is only .3 (30%), even if labor's true share is 1. If in this example, one were willing to insist that capital's true share θ should be at least .15 (15%), the maximum possible underestimate of labor's true share would be only .15. To put things another way, since under constant returns, $1-\theta = \mu\frac{wn}{py}$, an observed value of .7 for $\frac{wn}{py}$ would be inconsistent with a markup any greater than $1/.7 = 1.43$, even if one is willing to contemplate a value of labor's share all the way up to 1.

When there are increasing returns to scale ($\gamma > 1$) the second bias term

$$[1 - \frac{\gamma}{\mu}](1 - \theta)[\check{n} - \check{k}]$$

may be close to zero if γ is close to μ (even if both γ and μ are large), but the first bias term

$$(\gamma - 1)[\theta\check{k} + (1 - \theta)\check{n}]$$

will tend to make the standard Solow residual look more procyclical. [Exercise: Show that if $\check{k} > 0$ and $\check{n} > 0$ as would be likely in a boom, then the total bias in the standard Solow residual is increasing in γ and increasing in μ.]

11.3.2 Other Kinds of Mismeasurement of the Solow Residual

The literature on mismeasurement of the true Solow residual and efforts to correct that mismeasurement is quite extensive. Even if one has dealt with the possibility of imperfect competition and increasing returns to scale to solve the first two problems of getting appropriate values for γ and θ, other potential problems remain.

The three other general types of mismeasurement problems one may run up against are apparent from looking at the expression

$$\check{y} - \gamma[\theta\check{k} + (1 - \theta)\check{n}],$$

which one can construct once one has appropriate values for γ and θ. They are (1) mismeasurement of $\check{y}$ (fluctuations in y), (2) mismeasurement of $\check{k}$ (fluctuations in k), and (3) mismeasurment of $\check{n}$, (fluctuations in n).

Mismeasurement of fluctuations in output can occur if certain types of output are unreported. For example, suppose otherwise idle workers repaint the factory during a recession and that this painting service is unreported. Then true output will fall less in the recession than reported output does. Observed output will look more procyclical than true output, biasing the observed Solow residual toward procyclicality.

As another example, as workers rush to get orders out the door, the physical quality of output may fall during a boom in ways that are unreported. In another quality dimension, delivery lags typically lengthen in boom times, reducing the overall quality of output as delivered. Countercyclical quality of output means again that observed output will be more procyclical than true output, biasing the procyclicality of the observed Solow residual upwards.

Mismeasurement of fluctuations in effective capital input can occur when data on the workweek of capital are unavailable or unused or when data on any other dimension of the intensity of capital utilization is missing. Typically, one would expect capital to be used more intensively in booms, making the fluctuations in observed capital less procyclical than the fluctuations in effective capital input, and biasing the observed Solow residual toward procyclicality.

Similarly, mismeasurement of fluctuations in effective labor input can occur when the intensity of labor utilization fluctuates in ways that are not all observed. In particular, suppose the amount of effort the average worker exerts per hour and the number of actual hours on the job per reported hour increase when output increases. Then observed labor hours will be less procyclical than effective labor input, biasing the observed Solow residual toward procyclicality.

On the other hand, consider the fact that typical data on labor hours does not weight workers by their quality. Since average labor quality tends to be countercyclical, total unweighted worker hours is more procyclical than an appropriately quality adjusted measure of labor input. Using unweighted worker hours thus biases the calculated Solow residual toward *countercyclicality.* This is the one major mechanism that biases the Solow residual in this direction.

Finally, fluctuations in the workweek of capital, which we mentioned above, may be associated with fluctuations in another dimension of labor quality—namely the time of day or week at which the work is being done. Work on the night shift may be more valuable because of the ready availability of capital to work with at that hour, and in that sense may be "higher quality" labor even if there is some diminution of a workers mental acuteness at night. (There may also be certain jobs for which the diminished mental acuteness at night outweighs the extra value from ready availability of capital to work with.) Of course, "higher quality" labor may cause more disutility to the typical worker, making it more costly. In principle, the relative quality of labor at night should be something one can gauge by looking at the shift-premium that relates the night wage to the day wage, much as any other dimension of worker quality can be gauged by relative wages.[2]

When one combines the shift premium with all the other dimensions of labor quality to get an omnibus measure of quality weighted worker hours, one gets a net effect from fluctuations in both kinds of worker quality. Since

[2]There is, however, some danger of double counting, if one corrects fully for the fluctuations of the workweek of capital in some other way *and* treats work at night as higher quality than work during the day. It is important to have a clearly specified model to sort everything out and avoid such double counting.

the additional workers in a boom tend to have less education and experience (implying lower quality) but also are more likely to work on a late shift (implying higher quality), the variations in these two different dimensions of quality tend to cancel each other out.

11.4 Summary

Efforts to adjust the Solow residual in order to measure true technology shocks are likely to continue for a long time, along with debates about the strengths and weaknesses of various approaches. This will remain an important issue in macroeconomics for the foreseeable future.

One of the most direct approaches is trying to improve one's measurement of true technology shocks is to construct better data. Another type of approach that we have more to say about is to model the variation in unobserved variables in a dynamic optimization framework. For example, in Chapter **??** we will look at a model in which the intensity of capital and labor utilization decision variables in a dynamic optimization problem.

Real Business Cycle Theory: A Semiparametric Approach, by Miles S. Kimball

September 7, 2004 at 16:22

Chapter 12

The Solow Growth Model

12.1 Log-Linearizing Rates of Change

The only difference between $\check{x}$ and $\ln(x) - \ln(x^*)$ is that $\check{x}$ carries a built-in reminder that one is dealing with an approximation in the neighborhood of the steady state. If one is willing to ignore constant terms in an equation, one can think of $\check{x}$ as if it were simply $\ln(x)$. In any case, since $\ln(x^*)$ is constant, if one takes a time derivative, then to a first-order approximation,

$$\dot{\check{x}} = \dot{\ln(x)} - \dot{\ln(x^*)} = \frac{\dot{x}}{x},$$

Though formally speaking, $\check{x}$ is a logarithmic differential, the only practical difference between $\dot{\check{x}}$ and $\frac{\dot{x}}{x}$ is that $\dot{\check{x}}$ carries a built-in reminder that one is dealing with an approximation in the neighborhood of the steady state.

Similarly, with $\tilde{x} = x - x^*$,

$$\dot{\tilde{x}} = \dot{x} - \dot{x}^* = \dot{x},$$

plus a reminder that one is dealing with an approximation in the neighborhood of the steady steady state.

Systematically pursuing a linear or log-linear approximation does lead one to ignore second-order terms in a way that may be hard to get used to at first. Any product of two differentials can be ignored, including the product of a time derivative and another differential. For example, to a linear approximation, around the steady-state,

$$\begin{aligned}\frac{\dot{x}}{x} &= \frac{\dot{x}}{x^*} + \dot{x}[\frac{1}{x} - \frac{1}{x^*}] \\ &= \frac{\dot{x}}{x^*}.\end{aligned}$$

In terms of the differentials, which make this fact formal and exact,

$$\dot{\check{x}} = \frac{\dot{\tilde{x}}}{x^*}.$$

In this chapter, we will illustrate the technique of log-linearizing rates of change and using the resulting equations to analyze the dynamics of a model by taking a quantitative look at the Solow Growth Model. Some of the economic substance that follows may be old hat, allowing you to concentrate on the techniques at hand, but you may also learn that there is more to the Solow Growth Model than you thought.

12.2 The Basic Equations of the Solow Growth Model

Despite its non-optimizing dynamics, the Solow Growth Model is useful for thinking about the economy. The one *ad hoc* feature of the Solow Growth Model is the assumption of an exogenous gross saving rate. The only thing that the Cass model adds to the Solow Growth Model is that it endogenizes the gross saving rate.

The Solow Growth Model is especially good at clarifying the more mechanical relationships in the economy. Clarifying the more mechanical relationships is valuable because the more mechanical an economic relationship, the more robust it is likely to be to different assumptions about economic behavior.

With the price of investment goods taken as the numeraire, and government purchases included as part of either consumption c or investment i, the basic equations of the Solow Growth Model are per capita output y as a constant-returns-to-scale function of the per capita capital stock k and effective labor zn:

$$y = F(k, zn); \tag{12.1}$$

inelastically supplied labor, normalized to one unit per capita:

$$n = 1; \tag{12.2}$$

output divided between consumption and investment, with a technology ratio of ξ in favor of producing consumption goods:

$$y = \frac{c}{\xi} + i; \tag{12.3}$$

gross saving, which equals gross investment, as a given fraction of output:

$$i = \mathcal{S}y; \tag{12.4}$$

and the rate of increase in the capital stock equal to gross investment minus depreciation and the dilution of the capital stock by population growth and the trend in labor-augmenting technology:

$$\dot{k} = i - \hat{\delta}k \tag{12.5}$$

where

$$\begin{aligned}\hat{\delta} \quad = \quad & \text{depreciation rate} + \text{population growth rate} \\ & +\text{trend rate of improvement in the investment goods technology.}\end{aligned}$$

In the steady-state, (12.5) becomes

$$\frac{i^*}{k^*} = \hat{\delta}. \tag{12.6}$$

In conjunction with (12.4), this means that

$$y^* = \frac{i^*}{\mathcal{S}} = \frac{\hat{\delta}k^*}{\mathcal{S}}. \tag{12.7}$$

12.3 Log-Linearizing the Solow Growth Model

In accordance with the results of the previous chapter, (12.1), in conjunction with (**??**) (which implies that $\check{n} = 0$) can be log-linearized to

$$\begin{aligned}\check{y} &= \theta\check{k} + (1-\theta)[\check{z} + \check{n}] \qquad (12.8) \\ &= \theta\check{k} + (1-\theta)\check{z}.\end{aligned}$$

Equation (12.3) can be log-linearized to

$$y^* \check{y} = \frac{c^*}{\xi^*}[\check{c} - \check{\xi}] + i^* \check{\imath}. \tag{12.9}$$

To interpret this a little more, note that with the prices of investment and consumption equal to $p_i = 1$ and $p_c = \frac{1}{\xi}$ we can define the output shares

$$\zeta_i = \frac{p_i i^*}{y^*} = \frac{i^*}{y^*} = \mathcal{S}^*$$

(using the steady-state version of (12.4)) and

$$\zeta_c = \frac{p_c^* c^*}{y^*} = \frac{c^*}{\xi^* y^*} = 1 - \mathcal{S}^*.$$

Then, after dividing through by y^*, (12.9) can be rewritten

$$\begin{aligned} \check{y} &= \zeta_c[\check{c} - \check{\xi}] + \zeta_i \check{\imath} \\ &= (1 - \mathcal{S}^*)[\check{c} - \check{\xi}] + \mathcal{S}^* \check{\imath}. \end{aligned} \tag{12.10}$$

Equation (12.4) itself can be log-linearized to

$$\check{\imath} = \check{\mathcal{S}} + \check{y} \tag{12.11}$$

Finally, (12.5) can be log-linearized as

$$k^* \dot{\check{k}} = i^* \check{\imath} - \hat{\delta} k^* \check{k}. \tag{12.12}$$

But at the steady state, in order to have $\dot{k} = 0$, (12.5) implies

$$i^* = \hat{\delta} k^*.$$

Therefore, dividing (12.12) by k^* yields

$$\dot{\check{k}} = \hat{\delta}[\check{\imath} - \check{k}]. \tag{12.13}$$

12.4 The Dynamics of the Solow Growth Model

Combining (12.11) and (12.8) yields

$$\check{\imath} = \check{\mathcal{S}} + \theta \check{k} + (1 - \theta)\check{z}.$$

Using (12.13), one can then forge the dynamics equation

$$\dot{\check{k}} = -\hat{\delta}(1 - \theta)\check{k} + \hat{\delta}[\check{\mathcal{S}} + (1 - \theta)\check{z}]. \tag{12.14}$$

12.4.1 The Convergence Rate of the Solow Growth Model

If there is no disturbance to any of the exogenous variables, but only an initial capital stock away from the steady-state value, only the coefficient of $\check{k}$ is relevant. This coefficient says that if the capital stock is x% above the steady-state value, then the capital stock will fall at the rate of $\hat{\delta}(1-\theta)x\%$ per year. Similarly, if the capital stock is $x\%$ below the steady-state value, then it will rise at the rate of $\hat{\delta}(1-\theta)x\%$ per year. In other words, whatever the gap between the capital stock and its steady-state value, the size of the gap will shrink each year by $\hat{\delta}(1-\theta)$ of the total size of the gap. Thus, $\hat{\delta}(1-\theta)$ (the absolute value of the coefficient of $\check{k}$ in (12.14)) is the convergence rate, which can be labelled κ:

$$\kappa = \hat{\delta}(1-\theta).$$

In words, the convergence rate is equal to labor's share $(1-\theta)$ times the steady-state investment/capital stock ratio $\hat{\delta} = \frac{i^*}{k^*}$. If, for example, labor's share is .7, and the sum of depreciation plus the steady-state growth rate of the capital stock is equal to .08/year = 8%/year, then the convergence rate κ is .056/year = 5.6%/year.

Intuitively, the convergence rate reflects the balance between saving and investment on one hand and depreciation plus dilution on the other hand that exists in the steady state—a balance which is disturbed when the capital stock is away from the steady state. The force of depreciation and dilution, by itself, would cause the capital stock to shrink at the proportional rate $\hat{\delta}$; while at the steady state, the balancing force of saving and investment, by itself, would cause the capital stock to grow at the proportional rate $\hat{\delta}$. How fast do these two forces get out of balance away from the steady state? A 1% increase in the capital stock causes a 1% increase in the total amount of depreciation and dilution $\hat{\delta}k$, but only a $\theta\%$ increase in the total amount of saving—because it causes only a $\theta\%$ increase in output. The total amount of depreciation and dilution rises more quickly with the capital stock than does saving. Quantitatively, the two forces, which individually would change the capital stock at the proportional rate $\hat{\delta}$, get out of balance with an elasticity of $(1-\theta)$ with respect to the capital stock. The product of δ—the base rate of change in k that would come from each of the two forces—and the elasticity with which they get out of balance yields the convergence rate.

12.4.2 The Exogenous Driving Forces of the Per Capita Capital Stock

The other coefficients in (12.14) tell one that movements in the capital stock are driven by the gross saving rate and the labor-augmenting technology. Indeed, the (approximate) differential equation (12.14) can be integrated (using the integrating factor $e^{-\hat{\delta}(1-\theta)t}$) to get

$$\check{k}_t = e^{-\hat{\delta}(1-\theta)t}\check{k}_0 + \int_0^{\infty} e^{-\hat{\delta}(1-\theta)(t-t')}\hat{\delta}[\check{s} + (1-\theta)\check{z}]\,dt'. \tag{12.15}$$

In words, the logarithmic deviation of the capital stock is equal to a distributed lag of the logarithmic deviation of the gross saving rate and the labor-augmenting technology.

Notice what is missing in (12.14) and (12.15): the output-augmenting consumption technology ξ. In words, the capital stock is independent of ξ! To see why, think of what happens when it becomes easier to produce consumption goods, with no change in how difficult it is to produce investment goods. In equilibrium, the relative price of consumption goods will fall, until the marginal revenue products of capital and labor in the two sectors are brought back into equality. At these new prices, saving a fraction s of all output leads to using a fraction s of capital and labor to produce investment goods. But with no improvement in the investment goods producing technology, and no immediate change in the total quantities of labor and capital, there is no change in the amount of investment. Thus, the improvement in the technology for producing consumption goods will have no affect on the capital stock.

With ξ out of the picture, the investment-goods-producing technology and the gross saving rate alone determine the capital stock. To clarify this statement, it is helpful to interpret the two technology shifters, z and ξ in the light of the discussion of total factor productivity in the previous chapter.

Normalizing the value in the initial steady state to zero, log total factor productivity in the investment goods sector is

$$\ln(\mathrm{TFP}_i) = (1-\theta)\check{z}.$$

Log total factor productivity in the consumption goods sector is

$$\ln(\mathrm{TFP}_c) = (1-\theta)\check{z} + \check{\xi}.$$

Thus,

$$\check{z} = \frac{\ln(\mathrm{TFP}_i)}{(1-\theta)}; \tag{12.16}$$

while

$$\check{\xi} = \ln(\mathrm{TFP}_c) - \ln(\mathrm{TFP}_i). \tag{12.17}$$

In words, $\check{z}$ can be calculated by looking at total factor productivity in the investment goods sector alone, while $\check{\xi}$ is the difference between log total factor productivity in the consumption goods sector and log total factor productivity in the investment goods sector.

12.5 The Long-Run and Short-Run Effects of Permanent Changes in the Saving Rate or Technology

12.5.1 A Graph

Figure **??** shows the short-run and long-run responses of output and the capital stock for the Cobb-Douglas case in which the log-linearized production function is exact. With $\ln(k)$ on the horizontal axis and $\ln(y)$ on the vertical axis, the log-linearized production function (12.8) has slope equal to capital's share θ, and an intercept equal to $(1-\theta)\ln(z) = \ln(TFP_i)$. Essentially the same graph can be applied to a general constant returns production function by replacing $\ln(k)$ and $\ln(y)$ with the logarithmic differentials $\check{k}$ and $\check{y}$. [Is "logarithmic differential" a better term to use consistently than "logarithmic deviation," which we are using most frequently now?]

By taking logarithms of (12.7), the $\dot{k} = 0$ locus is given by

$$\ln(y) = \ln(k) + \ln(\hat{\delta}) - \ln(\mathcal{S}).$$

Graphically, this a line with a slope of 1. The intercept on the vertical axis is $\ln(\hat{\delta}) - \ln(\mathcal{S})$. However, the intercept on the horizontal axis, $\ln(\mathcal{S}) - \ln(\hat{\delta})$, shows the direction of effects in a more intuitive way. The intercept on the horizontal axis shows the steady-state capital stock that could be supported by an output of 1 given the gross saving rate $\mathcal{S}$.

If the investment goods technology improves by 1%, the production function in Figure **??** shifts up by .01. This implies an immediate 1% increase in output at the same level of capital, which cannot jump, followed by a gradual movement up along the production function to the $\dot{k} = 0$ line and the new steady-state at the intersection of the new production function and the $\dot{k} = 0$ line. Graphically, one can see from the familiar multiplier calculation that the long-run increase in k and y is $\frac{1}{1-\theta}$%. In other words, the elasticity of $\ln(k)$ and $\ln(y)$ with respect to $\ln(TFP_i)$ is $\frac{1}{1-\theta}$, while the elasticity with respect to the general labor-augmenting technology z is equal to 1.

If the gross saving rate increases by 1% (say from 20% of output to 20.2% of output), the $\dot{k} = 0$ line shifts out by .01. There is no immediate change in output, since at the initial capital stock, the economy will be at the same point on the same production function. Then the economy moves up the production function to the new steady state at the intersection of the new $\dot{k} = 0$ line and the unmoved production function. The long-run elasticity of k with respect to $\mathcal{S}$ is $\frac{1}{1-\theta}$. Since the elasticity of y with respect k on the unmoved production function is θ, this yields an elasticity for y with respect to $\mathcal{S}$ of $\frac{\theta}{1-\theta}$.

12.5.2 Algebra

The Short Run

In the short-run, we would like to know how each variable depends on the capital stock and the exogenous variables z, ξ and $\mathcal{S}$. For output y, this is clear from (12.8). For investment i, the dependence on k and the exogenous variables follows easily from $\check{\imath} = \check{\mathcal{S}} + \check{y}$—a relation one can substitute into (12.10) to get

$$\check{y} = (1 - \mathcal{S}^*)(\check{c} - \check{\xi}) + \mathcal{S}^*[\check{\mathcal{S}} + \check{y}].$$

Solving for $\check{c}$,

$$\begin{aligned} \check{c} &= \check{y} + \check{\xi} - \frac{\mathcal{S}^*}{1 - \mathcal{S}^*}\check{\mathcal{S}}. \\ &= \theta\check{k} + (1 - \theta)\check{z} + \check{\xi} - \frac{\mathcal{S}^*}{1 - \mathcal{S}^*}\check{\mathcal{S}}. \end{aligned}$$

Finally, although using the price of the investment good as the numeraire is the simplest approach analytically, deflating GDP in the standard way,

$$\check{\text{GDP}} = \check{y} + \zeta_c\check{\xi} = \check{y} + (1 - \mathcal{S}^*)\check{\xi}.$$

This is also an interesting quantity to watch. Summarizing all of the relationships to k and the exogenous variables in matrix form, and using (12.16) and (12.17),

$$\begin{bmatrix} \check{y} \\ \check{\imath} \\ \check{c} \\ \text{G}\check{\text{D}}\text{P} \end{bmatrix} = \begin{bmatrix} \theta & 1-\theta & 0 & 0 \\ \theta & 1-\theta & 0 & 1 \\ \theta & 1-\theta & 1 & -\frac{\mathcal{S}^*}{1-\mathcal{S}^*} \\ \theta & 1-\theta & 1-\mathcal{S}^* & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{z} \\ \check{\xi} \\ \check{s} \end{bmatrix} \tag{12.18}$$

$$= \begin{bmatrix} \theta & 1 & 0 & 0 \\ \theta & 1 & 0 & 1 \\ \theta & 0 & 1 & -\frac{\mathcal{S}^*}{1-\mathcal{S}^*} \\ \theta & \mathcal{S}^* & 1-\mathcal{S}^* & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \ln(\text{TFP}_i) \\ \ln(\text{TFP}_c) \\ \check{s} \end{bmatrix}.$$

The Long Run

Bending notation to write $\check{k}^*$ for the long-run change in the steady-state capital stock from a permanent change in the exogenous variables, one finds from either (12.14) with $\dot{\check{k}}$ set to zero—or from (12.15)—that

$$\check{k}^* = \check{z} + \frac{\check{s}}{1-\theta} = \frac{\ln(\text{TFP}_i) + \check{s}}{1-\theta}.$$

Substituting this expression into (12.18) yields the equation for the long-run changes in y, i, c and GDP:

$$\begin{bmatrix} \check{y}^* \\ \check{\imath}^* \\ \check{c}^* \\ \text{G}\check{\text{D}}\text{P}^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{\theta}{1-\theta} \\ 1 & 0 & \frac{1}{1-\theta} \\ 1 & 1 & \frac{\theta}{1-\theta} - \frac{\mathcal{S}^*}{1-\mathcal{S}^*} \\ 1 & 1-\mathcal{S}^* & \frac{\theta}{1-\theta} \end{bmatrix} \begin{bmatrix} \check{z} \\ \check{\xi} \\ \check{s} \end{bmatrix} \tag{12.19}$$

$$= \begin{bmatrix} \frac{1}{1-\theta} & 0 & \frac{\theta}{1-\theta} \\ \frac{1}{1-\theta} & 0 & \frac{1}{1-\theta} \\ \frac{\theta}{1-\theta} & 1 & \frac{\theta}{1-\theta} - \frac{\mathcal{S}^*}{1-\mathcal{S}^*} \\ \frac{1}{1-\theta} & 1-\mathcal{S}^* & \frac{\theta}{1-\theta} \end{bmatrix} \begin{bmatrix} \ln(\text{TFP}_i) \\ \ln(\text{TFP}_c) \\ \check{s} \end{bmatrix}.$$

The Golden Rule

The Golden Rule steady state is the steady-state with the maximum level of consumption. The elasticity of steady-state consumption with respect to the gross saving rate, $\frac{\theta}{1-\theta} - \frac{\mathcal{S}^*}{1-\mathcal{S}^*}$, is useful in thinking about the Golden Rule. This elasticity is positive when $\mathcal{S}^* < \theta$, negative when $\mathcal{S}^* > \theta$ and zero when $\mathcal{S}^* = \theta$. The condition $\mathcal{S}^* = \theta$ is a convenient way to characterize the gross saving rate that achieves the Golden Rule (See Phelps ()). The comparison of the gross saving rate $\mathcal{S}^*$ and capital's share θ is also particularly easy to take to the data. Abel, Mankiw, Summers and Zeckhauser () find that $\frac{i}{y} < \theta$ for every year for which data is available, strongly indicating that $\mathcal{S}^* < \theta$—putting the economy below the Golden Rule, in the dynamically efficient region.

12.5.3 A Numerical Example

Suppose capital's share θ is $.333 = 33.3\%$, while gross saving rate is $.2 = 20\%$. Then in the short run,

$$\begin{bmatrix} \check{y} \\ \check{\imath} \\ \check{c} \\ \mathrm{G\check{D}P} \end{bmatrix} = \begin{bmatrix} .333 & 1 & 0 & 0 \\ .333 & 1 & 0 & 1 \\ .333 & 0 & 1 & -.25 \\ .333 & 1 & .8 & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \ln(\mathrm{TFP}_i) \\ \ln(\mathrm{TFP}_c) \\ \check{\mathcal{S}} \end{bmatrix}.$$

In the long run,

$$\begin{bmatrix} \check{y}^* \\ \check{\imath}^* \\ \check{c}^* \\ \mathrm{G\check{D}P}^* \end{bmatrix} = \begin{bmatrix} 1.5 & 0 & .5 \\ 1.5 & 0 & 1.5 \\ .5 & 1 & .25 \\ 1.5 & .8 & .5 \end{bmatrix} \begin{bmatrix} \ln(\mathrm{TFP}_i) \\ \ln(\mathrm{TFP}_c) \\ \check{\mathcal{S}} \end{bmatrix}.$$

The difference between the short-run and long-run effects of the exogenous variables is due to the induced capital accumulation.

The steady-state investment/capital ratio $\hat{\delta}$ has no effect on any of these elasticities, but it does affect the convergence rate. If $\hat{\delta} = .09/\text{year}$, the convergence rate is $\kappa = .06/\text{year}$.

Another Numerical Example Taking Account of Human Capital

Suppose we include human capital in the Solow Growth Model, but just lump it together with regular capital. That might make the combined share of both kinds of capital something like $\theta = .667$. Since education counts as investment in a model with human capital, it is appropriate to use a higher value of the gross saving rate as well, say $\mathcal{S}^* = .4$. Also, with education a form of investment, improvements in educational technology will raise TFP_i.

With these parameter values, in the short run,

$$\begin{bmatrix} \check{y} \\ \check{\imath} \\ \check{c} \\ \mathrm{G\check{D}P} \end{bmatrix} = \begin{bmatrix} .667 & 1 & 0 & 0 \\ .667 & 1 & 0 & 1 \\ .667 & 0 & 1 & -.667 \\ .667 & 1 & .6 & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \ln(\mathrm{TFP}_i) \\ \ln(\mathrm{TFP}_c) \\ \check{\mathcal{S}} \end{bmatrix}.$$

In the long run,

$$\begin{bmatrix} \check{y}^* \\ \check{\imath}^* \\ \check{c}^* \\ \text{G\v{D}P}^* \end{bmatrix} = \begin{bmatrix} 3 & 0 & 2 \\ 3 & 0 & 3 \\ 2 & 1 & 1.333 \\ 3 & .6 & 2 \end{bmatrix} \begin{bmatrix} \ln(\text{TFP}_i) \\ \ln(\text{TFP}_c) \\ \check{s} \end{bmatrix}.$$

Because of the greater importance of capital, and the lesser degree of decreasing returns to a broadly defined version of capital, induced capital accumulation has powerful effects, making the long-run elasticities much different from the short-run elasticities. Overall, in the long run, changes in technology and the gross saving rate matter more with these parameter values.

As for the convergence rate, even if the adjusted depreciation rate $\hat{\delta}$ is left at .09/year (despite the likelihood that human capital added to the model tends to have a lower depreciation rate than physical capital), the convergence is now only .03/year. Though the long-run effects are larger, they take a longer time to work themselves out.

Chapter 13

Q-Theory in Partial Equilibrium

In the Basic Real Business Cycle Model, investment demand is perfectly elastic. Thus, saving supply is simply accomodated. Also, the relative price of capital is always equal to 1 and the real interest rate moves one-for-one with the real rental rate of capital.

As is clear from Chapter **??**, it also interesting to consider the Basic Q-Theory Real Business Cycle Model, in which the investment demand curve is downward-sloping, the relative price of capital can vary and the real interest rate is not enslaved to the real rental rate. In this chapter, we solve the investment problem for a firm facing investment adjustment costs to derive the basic equations of the Q-theory of investment and log-linearize those equations.

13.1 Representing Investment Adjustment Costs

There are two ways of formalizing investment adjustment costs, corresponding to two ways of measuring investment. Abel () measures investment by the overall contribution to the capital stock and has investment adjustment costs show up in an extra cost of investment. We follow Hayashi (1982) in measuring investment by the cost of investment and representing investment adjustment costs by a variable efficiency of a given amount of investment in adding to the capital stock. Formally, let

$$\dot{k} = kJ(i/k),$$

where i is investment and k is capital. Assuming that $J'(\cdot) > 0$ and $J''(\cdot) < 0$

ensures that additional investment does indeed add to the capital stock, but that the *marginal* efficiency of investment in adding to the capital stock falls when the firm invests at a quicker pace. The way in which the capital stock k enters this relationship ensures that there is constant returns to scale in a firm's investment. If the initial capital stock and investment are both twice as large, the rate of addition to the capital stock will be twice as large and the growth rate of the capital stock will be the same, since

$$\frac{\dot{k}}{k} = J(i/k).$$

A useful way to define the depreciation rate δ is as the rate of investment in relation to the capital stock (i/k) that maintains the capital stock at a constant level. That is, δ is defined by

$$\frac{\dot{k}}{k} = J(\delta) = 0.$$

Clearly, in a steady state,

$$\frac{i^*}{k^*} = \delta.$$

It makes sense to measure capital so that *at the "normal" rate of investment* i^*/k^* one unit of investment makes one unit of capital. Measuring capital in this sensible way corresponds to scaling or normalizing J so that

$$J'(i^*/k^*) = J'(\delta) = 1.$$

For comparison, The model of capital accumulation without investment adjustment costs is a limiting case of the model with adjustment costs. In particular, if $J(x) = x - \delta$, then $kJ(i/k) = i - \delta k$. The function $J(x) = x - \delta$ satisfies $J'(\cdot) > 0$, $J(\delta) = 0$ and $J'(\delta) = 1$, but has $J''(\cdot) = 0$.

13.2 Log-Linearizing the Capital Accumulation Equation When Investment Adjustment Costs are Present

Given $(i^*/k^*) = \delta$ and $J'(\delta) = 1$, the accumulation equation (13.2) can be log-linearized to

$$\dot{\check{k}} = \delta[\check{\imath} - \check{k}]. \tag{13.1}$$

Note that (13.1) is identical to what it would be in the absence of investment adjustment costs. Investment adjustment costs play only a second-order role in the accumulation equation itself.

13.3 The Firm's Investment Problem

We will get the most insight into basic workings of the Q Theory by starting with a very simple partial equilibrium model. Therefore, consider a firm that does nothing but invest in capital and then rent that capital out at the real rental rate R. With the price of investment i as the numeraire, the present value of this activity, which the investing firm maximizes is

$$\max_i \int_0^\infty e^{\int_0^{t'} r\, dt''} [Rk - i]\, dt',$$

s.t.

$$\dot{k} = kJ(i/k). \tag{13.2}$$

The current-value Hamiltonian is

$$H = Rk - i + qkJ(i/k).$$

The first-order condition for investment is

$$1 = qJ'(i/k),$$

Since $i^*/k^* = \delta$, in steady-state the first-order condition implies $q^* = 1$. The first-order condition can be log-linearized as

$$\check{q} = j[\check{\imath} - \check{k}], \tag{13.3}$$

where

$$j = -\frac{(i^*/k^*)J''(i^*/k^*)}{J'(i^*/k^*)}.$$

The Euler equation is

$$\begin{aligned} \dot{q} &= rq - R - J(i/k)q + (i/k)J'(i/k)q \\ &= rq - R - J(i/k)q + (i/k), \end{aligned} \tag{13.4}$$

using the first-order condition (13.3). Dividing (13.4) through by q,

$$\frac{\dot{q}}{q} = r - J(i/k) + \frac{(i/k) - R}{q}. \tag{13.5}$$

Let us take two tacks toward interpreting this equation before we go on to log-linearize. First, rearranging (13.5) yields

$$r = \frac{R - (i/k)}{q} + J(i/k) + \frac{\dot{q}}{q}. \tag{13.6}$$

In words, the required return r is equal to the dividend/price ratio plus the growth rate of the total value of the firm's capital. The dividend is equal to the rental rate of capital R minus the amount of investment spending per unit of capital (i/k). The growth rate of the total value of the firm's capital is equal to the sum of the growth rate of the quantity of capital $J(i/k) = \frac{\dot{k}}{k}$ and the growth rate of the price of capital $\frac{\dot{q}}{q}$. Integrating (13.4) with the transversality condition

$$\lim_{t\to\infty} e^{\int_0^t r dt'} q_t k_t = 0,$$

yields an equation that gives the same story in integral form:

$$q_t = \int_t^\infty e^{\int_t^{t'} [r - J(i/k)]dt''}[R - \frac{i}{k}]dt'.$$

The Q-Theory Euler equation is easier to interpret after log-linearization around the steady state. Return to (13.5). In the steady state, $i^*/k^* = \delta$, $q^* = 1$ and $J(i^*/k^*) = J(\delta) = 0$ imply $r^* + \delta = R^*$ as in the absence of investment adjustment costs. In log-linearizing (13.5), since $J'(i^*/k^*) = (1/q^*) = 1$, the derivative of the second line of (13.5) with respect to i/k is zero, and

$$\begin{aligned} \dot{\breve{q}} &= \tilde{r} - \tilde{R} + r^*\breve{q} \\ &= \tilde{r} - R^*\breve{R} + r^*\breve{q}. \end{aligned} \tag{13.7}$$

In integral form, using the steady-state as a terminal condition,

$$\breve{q}_t = \int_t^\infty e^{-r^*(t'-t)}[\tilde{R} - \tilde{r}]. \tag{13.8}$$

Thus, to a first-order approximation, the movements in q are a present discounted value of the movements in the spread between the rental rate of capital and the real interest rate.

The most suprising feature of this equation is that the movements in the spread between the rental rate of capital and the real interest rate are discounted only at the rate r^* rather than at, say, the rate $r^* + \delta$. This low rate of discounting of the movements in the spread between the rental rate and the real interest rate means that distant future events have an important impact on q and therefore on investment.

13.4 Estimating j

By (**??**), the elasticity of i/k with respect to q is $\frac{1}{j}$. Given a good measure of q, the adjustment cost parameter j could be estimated. One approach is to use the ratio of the firm's value to the replacement cost of its capital to gauge q. In Part III of this book, we discuss the assumptions needed to get the scale symmetry that should make this measure of the average price of capital ("average Q") equal to the marginal price of capital ("marginal q"). This approach tends to yield very low estimates of the elasticity of (i/k) with respect to q and therefore very high estimates of j. In the next section, we will discuss the sense in which these are "very high" estimates of j.

Other approaches sidestep the use of highly volatile stock and bond prices, which may measures q with substantial errors, biasing the estimates of $\frac{1}{j}$ downward. Cummins, Hassett and Hubbard (1994) identify movements in a tax-adjusted version of q by looking at what happens around the time of statutory tax changes, and find much larger elasticities of I/K with respect to q—on the order of 5, implying $j = .2$.

The equation (13.10) above can itself serve as the basis of an equation for estimating $\frac{1}{j}$ that sidesteps stock and bond market prices. First, combine either (13.9) or (13.10) with (13.3) and write $\tilde{R}$ as $R^*\check{R}$ to get

$$[\dot{\check{\imath}} - \dot{\check{k}}] - r^*[\check{\imath} - \check{k}] = \frac{1}{j}[\tilde{r} - R^*\check{R}]. \tag{13.9}$$

or

$$\check{\imath} - \check{k} = \frac{1}{j}\int_t^\infty e^{-r^*(t'-t)}[R^*\check{R} - \tilde{r}]. \tag{13.10}$$

In order to implement either of these equations empirically, one needs a measure of variations in the real rental rate of capital, $\check{R}$. In the Cobb-Douglas case where

$$y = \xi k^\theta (zn)^{1-\theta},$$

the marginal product of capital is

$$R = \theta \xi k^{\theta-1}(zn)^{1-\theta} = \theta \frac{y}{k}.$$

Log-linearizing around the steady state,

$$\check{R} = \check{y} - \check{k}.$$

(Exercise 1 addresses the non-Cobb-Douglas case.)

To use available discrete-time data, one must do (approximate) time-aggregation. Finally, in implementing an Euler equation, one must remember that we have been using perfect-foresight as a way of representing the certainty-equivalence approximation to rational expectations. The Euler equation implicitly carries a rational expectations expectation error that is uncorrelated with information known at the time the expectation was formed.

Putting everything together, the kind of equation one would actually look at, based on (13.9), would be something like

$$\ln(i_{t+1}/k_{t+1}) - (1+r^*)\ln(i_t/k_t) = \text{constant} + \frac{1}{j}[r_t - R^*(\check{y}_t - \check{k}_t)] + \epsilon_{t+1}. \quad (13.11)$$

Here, one can estimate the equation by instrumental variables using as instruments things known before the beginning of the window of time from which the various lags of the data in the equation itself come from. One should allow for at least an MA(1) structure to the error. For the real interest rate r, one can use the ex post real rate, since it has the same kind of expectations error under rational expectations, which can be combined with the other rational expectations error without any problem. (Make sure that the interest rate r_t and steady-state rental rate R^* are what one would earn over one time period, not 100 times that amount or 400 times that amount.)

If one wished to estimate instead an equation based on (13.10), it would be something like

$$\ln(i_t/k_t) = \text{constant} + \frac{1}{j} E_t \sum_{\ell=0}^{L} e^{-\ell r^*}[r_{t+\ell} - R^* \ln(y_{t+\ell}/k_{t+\ell})].$$

Econometrically, estimating this equation is similar to estimating present value relationships like the relationship that should exist between stock prices and expected subsequent dividends.

13.5 Calibrating j by Considering the Partial Equilibrium Convergence Rate

In this section we will illustrate another way to gauge the size of the adjustment cost parameter j by looking at a partial equilibrium model simple enough that we can get some intuition about the likely speed of adjustment of a firm's capital stock in response to shocks to product demand or to technology.

13.5.1 The Model

Think of a single firm trying to minimize costs in the face of investment adjustment costs for capital that is attached to that particular firm, so there is no literal rental market for the firm's capital. It must produce output y (this demanded output being given exogenously) with a Cobb-Douglass technology. It faces a competitive labor market, hiring labor n at the real wage w, and an exogenously given real interest rate. The investment goods the firm purchases are the numeraire. Putting everything together as a maximization problem, the firm solves

$$\max_{n,i} - \int_0^\infty e^{-\int_0^t r\, dt'} [wn + i] dt$$

s.t.

$$\dot{k} = kJ(i/k) \tag{13.12}$$

$$y = \xi k^\theta n^{1-\theta}. \tag{13.13}$$

13.5.2 Basic Equations

The current-value Hamiltonian is

$$H = -(wn + i) + qkJ(i/k) + \mu[\xi k^\theta n^{1-\theta} - y].$$

The two first-order conditions are $1 = qJ'(i/k)$ and

$$w = \mu(1-\theta)\xi k^\theta n^{-\theta} \tag{13.14}$$

In addition, one can define the shadow rental rate R by the marginal revenue product of capital:

$$R = \mu\theta\xi k^{\theta-1} n^{1-\theta} \tag{13.15}$$

With this definition, the Euler equation is identical to (13.4).

13.5.3 Finding the Convergence Rate

In order to find the dynamics of this model, it is desirable to find equations for $\dot{\check{k}}$ and $\dot{\check{q}}$ in terms of $\check{k}$, $\dot{q}$ and exogenous variables alone.

First, since the equations earlier in this chapter are valid, including (13.3), it is easy to express $\dot{\check{k}}$ as

$$\dot{\check{k}} = \delta[\check{\imath} - \check{k}] = \frac{\delta}{j}\check{q}. \tag{13.16}$$

Turning to the Euler equation, the only term that is not already in such a reduced form is $\check{R}$. What we need is an expression for R in terms of the dynamic variables $\check{k}$, $\check{q}$ and exogenous variables. From (13.14) and (13.15),

$$R = \frac{\theta}{1-\theta}\frac{wn}{k}.$$

Log-linearizing,

$$\check{R} = \check{w} + \check{n} - \check{k}. \tag{13.17}$$

To eliminate $\check{n}$ from this equation, first log-linearize the production function (13.13) to

$$\check{y} = \check{\xi} + \theta\check{k} + (1-\theta)\check{n}$$

and solve for $\check{n}$:

$$\check{n} = \frac{\check{y} - \check{\xi} - \theta\check{k}}{1-\theta}. \tag{13.18}$$

Substituting (13.18) into (13.17),

$$\check{R} = \check{w} + \frac{\check{y} - \check{\xi} - \check{k}}{1-\theta}. \tag{13.19}$$

Then, substituting (13.3) and (13.19) into the log-linearized Euler equation qpartial log q Euler yields

$$\dot{\check{q}} = \tilde{r} + r^*\check{q} - R^*[\check{w} + \frac{\check{y} - \check{\xi} - \check{k}}{1-\theta}] \tag{13.20}$$

In matrix form, the reduced form accumulation equation and reduced form Euler equation make the reduced form dynamic equation of the model:

$$\begin{bmatrix} \dot{\check{k}} \\ \dot{\check{q}} \end{bmatrix} = \begin{bmatrix} 0 & \frac{\delta}{j} \\ \frac{R^*}{(1-\theta)} & r^* \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{q} \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{r} - R^* \left(\check{w} + \frac{\check{y} - \check{\xi} - \check{k}}{1-\theta} \right) \end{bmatrix} \tag{13.21}$$

The convergence rate κ is the absolute value of the negative eigenvalue of the 2x2 dynamic matrix in (13.21), which is

$$\begin{aligned} \kappa &= \sqrt{\frac{(\text{trace})^2}{4} - \det} - \frac{\text{trace}}{2} \\ &= \sqrt{\frac{(r^*)^2}{4} + \frac{R^*\delta}{j(1-\theta)}} - \frac{r^*}{2}. \end{aligned} \tag{13.22}$$

Note that this convergence rates goes to infinity as $j \to \infty$; the firm would adjust instantly except for the investment adjustment costs. In general equilibrium, in response to a macroeconomic shock, the endogenous response of the real interest rate and the real wage would slow down convergence. Holding the real interest rate fixed has the same effect on the convergence rate as an infinite elasticity of intertemporal substitution. Holding the real wage fixed has the same effect on the convergence rate as an infinite labor supply elasticity.

13.5.4 A Numerical Example

Using $r^* = .02$, $R^* = .10$, $\delta = .08$ and $\theta = .3$ as in Chapter **??**, Table **??** gives the partial equilibrium convergence rate of capital as a function of the adjustment cost parameter j.

Despite the fact that in this partial equilibrium model, only the positive investment adjustment cost keeps the convergence rate from being infinite, a value of j even as high as 1 makes the convergence rate an implausibly low 9.7% per year. A value of j equal to .2 gives a more plausible 22.9% per year partial equilibrium adjustment rate. And a value of j even as low as .05 implies a not-implausible partial equilibrium adjustment rate of 46.8% per year.

Exercises

1. Show that when the production function is not Cobb-Douglas, but the firm faces perfect competition and has constant returns to scale, then

$$\check{R} = \check{y} - \check{k} + (1-\theta)(\frac{1}{\sigma} - 1)[\check{n} - \check{k}]. \tag{13.23}$$

Answer

When the production function is not Cobb-Douglas, one can use the courterpart to (**??**) when labor n can vary. Start with the fact that when there is perfect competition and constant returns to scale, factor payments exhaust the value of the output:

$$py = Rk + wn.$$

Of course, for the numeraire, $p = 1$, so this reduces to

$$y = Rk + wn.$$

Log-linearizing this factor exhaustion result around the steady-state,

$$\check{y} = \theta[\check{R} + \check{k}] + (1 - \theta)[\check{w} + \check{n}]. \tag{13.24}$$

Then use the definition of the elasticity of capital labor substitution—

$$\check{k} - \check{n} = \sigma[\check{w} - \check{R}] \tag{13.25}$$

—to eliminate $\check{w}$ from (13.24). That is, substitute

$$\check{w} = \check{R} + \frac{\check{k} - \check{n}}{\sigma}. \tag{13.26}$$

into (13.24) and then solve for $\check{R}$.

2. Calculate the partial equilibrium adjustment rate for capital for a firm that has a constant-returns-to-scale production function that is not Cobb-Douglas, but has an elasticity of capital-labor substitution of σ in the neighborhood of the steady state.

Chapter 14

Introduction to Quantitative Analytics: the Cass Model

14.1 Introduction

The literal meaning of "analysis" is to take something apart. Qualitative analysis is taking models apart and studying the qualitative relationships of the pieces. It involves answering the questions "Up or down?" "Positive or negative effect?" when the answers are unambiguous.

To this point, we have emphasized qualitative analysis. Qualitative analysis will continue to be important throughout the book. But it is time to add more of a quantitative edge to our analysis.

Quantitative analysis is taking models apart and studying the quantitative relationships of the pieces. It answers the questions "How much?" "How fast?" "How strong?" "How dependent on parameter values?" *and* the questions "Up or down?" "Positive or Negative Effect?" *when the answers depend on parameter values.*

What we will call *quantitative analytics*, is a particular approach to quantitative analysis. It is a set of tools and methods for doing quantitative analysis in a way that emphasizes understanding as well as the numbers on the bottom line. Quantitative analytics is an approach that emphasizes taking a model apart and studying each piece and subsystem carefully—and then, while putting the model back together, trying to understand how the behavior of the model as a whole arises from the behavior of the pieces *and* from the interaction of those pieces.

We hope to persuade readers that quantitative analytics pays important dividends in understanding in comparison to relying *solely* on approaches that treat models as black boxes. This chapter begins to introduce some of the principles of quantitative analytics.

14.2 An Example: Analyzing the Convergence Rate in the Cass Model the Hard Way

One bit of advice we have for quantitative analytics is to log-linearize equations wherever possible, as soon as the basic equations of a model have been derived as first order conditions, Euler equations, accumulation equations or as formalizations of constraints. To motivate this piece of advice, let us show what might happen when one sets out to analyze the rate of convergence in the Cass Model, while ignoring this advice.

14.2.1 Calculating the Convergence Rate

By now, the Cass model should be relatively familiar. As a dynamic optimization problem, it is

$$\max_{c} \int_0^{\infty} e^{-\rho t} u(c)\, dt$$

s.t.,

$$\dot{k} = f(k) - \delta k - c - g.$$

The current value Hamiltonian is

$$H = u(c) + \lambda[f(k) - \delta k - c - g].$$

The first order condition for c can be written as

$$u'(c) = \lambda.$$

The Euler equation can be written as

$$\dot{\lambda} = [\rho + \delta - f'(k)]\lambda.$$

At the steady state, the first order condition, accumulation equation and Euler equation become

$$\begin{aligned} u'(c^*) &= \lambda^* \\ f(k^*) &= \delta k^* + c^* + g \\ f'(k^*) &= \rho + \delta. \end{aligned}$$

Linearizing around the steady-state (with a constant level of government purchases g since we don't need to have g vary to study the convergence rate)

$$u''(c^*)\tilde{c} = \tilde{\lambda},$$

and so

$$\dot{\tilde{k}} = [f'(k^*) - \delta]\tilde{k} - \tilde{c} = \rho\tilde{k} - \frac{1}{u''(c^*)}\tilde{\lambda}$$

and

$$\dot{\tilde{\lambda}} = [\rho + \delta - f'(k^*)]\tilde{\lambda} - \lambda^* f''(k^*)\tilde{k} = -\lambda^* f''(k^*)\tilde{k}.$$

In matrix form,

$$\begin{bmatrix} \dot{\tilde{k}} \\ \dot{\tilde{\lambda}} \end{bmatrix} = \begin{bmatrix} \rho & -\frac{1}{u''(c^*)} \\ -f''(k^*)\lambda^* & 0 \end{bmatrix} \begin{bmatrix} \tilde{k} \\ \tilde{\lambda} \end{bmatrix}.$$

The eigenvalues are

$$\vartheta = \frac{\text{trace}}{2} \pm \sqrt{\frac{\text{trace}^2}{4} - \text{det}}.$$

The convergence rate κ is the absolute value of the negative eigenvalues:

$$\begin{aligned} \kappa &= \sqrt{\frac{\text{trace}^2}{4} - \text{det}} - \frac{\text{trace}}{2} \\ &= \sqrt{\frac{\rho^2}{4} + \frac{f''(k^*)\lambda^*}{u''(k^*)}} - \frac{\rho}{2}. \end{aligned}$$

14.2.2 Interpreting the Convergence Rate

So far so good. It is clear at least that curvature of the production function f tends to raise the convergence rate, while curvature of the utility function u tends to lower the convergence rate. But having done things by linearizing, rather than log-linearizing, we are left with a lot more processing before we can interpret the rate of convergence quantitatively.

One difficulty is that as (ultimately) empirical quantities, $f''(k^*)$ and $u''(c^*)$ are quite exotic (unlike the discount rate ρ, which is a pure rate, measured by the unit $\frac{1}{\text{year}}$). With the capital stock measured in (real) dollars and output f measured in (real) $\frac{\text{dollar(s)}}{\text{year}}$, the unit for measuring $f''(k^*)$ is $\frac{1}{\text{dollar}\cdot\text{year}}$. Since consumption is measured in $\frac{\text{dollar(s)}}{\text{year}}$, while u is measured in $\frac{\text{util(s)}}{\text{year}}$, $u''(c^*)$ is measured by $\frac{\text{util}\cdot\text{year}}{\text{dollar}^2}$. Finally, since λ^* is measured in $\frac{\text{util(s)}}{\text{dollar}}$, the key quantity $\frac{f''(k^*)\lambda^*}{u''(c^*)}$ is measured by

$$\frac{\frac{1}{\text{dollar}\cdot\text{year}}\frac{\text{util}}{\text{dollar}}}{\frac{\text{util}\cdot\text{year}}{\text{dollar}^2}} = \frac{1}{\text{year}^2}.$$

Thus, $\frac{f''(k^*)\lambda^*}{u''(c^*)}$ the square of a pure rate measured by $1/\text{year}^2$ and it can be added to $\frac{\rho^2}{4}$ which is also the square of a pure rate. The dimensions of the units all work out in the end, but the individual factors $f''(k^*)$ or $u''(c^*)$ are very difficult to interpret.

One way to try to escape the problem of exotic dimensions is multiply and divide by quantities that will help to make elasticities out of $f''(k^*)$ and $u''(c^*)$. By multiplying and dividing by one,

$$\begin{aligned}\frac{f''(k^*)\lambda^*}{u''(c^*)} &= \left(\frac{-k^*f''(k^*)}{f'(k^*)}\right)\left(\frac{-u'(c^*)}{c^*u''(c^*)}\right)\frac{\lambda^*}{u'(c^*)}\frac{f'(k^*)c^*}{k^*}\\ &= \left(\frac{-k^*f''(k^*)}{f'(k^*)}\right)\left(\frac{-u'(c^*)}{c^*u''(c^*)}\right)(\rho+\delta)\frac{c^*}{k^*}.\end{aligned}$$

At this point, one would be fortunate to recognize $\frac{-u'(c^*)}{c^*u''(c^*)}$ as the elasticity of intertemporal substitution at c^*, which we can denote s.

Interpreting $\frac{-k^*f''(k^*)}{f'(k^*)}$ is more difficult. We will not even try to establish its interpretation at this point. But using the approach we will present below, we will find that

$$\frac{-k^* f''(k^*)}{f'(k^*)} = \frac{(1-\theta)}{\sigma},$$

where θ is capital's share at the steady state, $1-\theta$ is labor's share, and σ is the elasticity of substitution between capital and labor capital at the steady state. But this fact is far from obvious. Given this fact, the convergence rate is

$$\kappa = \sqrt{\frac{\rho^2}{4} + \frac{s(1-\theta)}{\sigma}(\rho+\delta)\frac{c^*}{k^*}} - \frac{\rho}{2}. \tag{14.1}$$

If one could get this far, the convergence rate can be interpreted quite well. The elasticity of intertemporal substitution, the elasticity of capital/labor substitution, labor's share, the utility discount rate and the steady-state ratio of consumption to the capital stock are all individually interpretable in a straightforward way. One can readily develop a sense for how reasonable or unreasonable various possible values of these numbers are. That is not to say that economists will all agree about these numbers, but at least most economists will understand what is being argued over.

14.3 A Better Way: Analyzing the Model into Small Pieces and Using Log-Linearization

The key whether an expression for the model's convergence rate is interpretable or not is whether or not it is expressed in terms of individually interpretable numbers such as rates, elasticities and share parameters, or in terms of essentially uninterpretable magnitudes such as $f''(k^*)$ and $u''(c^*)$. In the previous section, we managed to make the transition from a quantitatively uninterpretable expression to a readily interpretable one by dint of significant effort. There is an easier way. By log-linearizing first order conditions, constraints, accumulation equations and Euler equations at the very beginning, we guarantee that all coefficients and parameters will be expressed from the start in terms of pure numbers such as elasticities and shares and in terms of rates such as discount rates, interest rates, rental rates, depreciation rates etc. Since each individual equation is expressed in those terms, any algebraic combination of them will automatically be expressed in a corresponding, readily interpretable way. Good form concentrates the recasting of the model in terms of elasticities at the beginning instead of the end—with the advantages of greater transparency and more carryover from one model to another, since different models often share the same partial equilibrium pieces.

The smaller the pieces a model is analyzed into, the better log-linearization works. It is especially easy to interpret the parameters in an equation with only a few variables in it. Moreover, small pieces of a model are more likely to be familiar components shared with other models. Log-linearized equations combine well, with the combinations automatically remaining in interpretable form, so reassembly is not difficult. The additional costs incurred by breaking the model down into small parts are minimal, while the benefits are often substantial.

In practice, breaking the model down into small pieces often involves freely giving out the appropriate economic labels to quantities that have economic interpretations. Thus, as usual, we can define the rental rate

$$R = f'(k),$$

the real interest rate

$$r = R - \delta,$$

the real wage [Question: Where in the book do we explain this equation and where it comes from?]

$$w = f(k) - kf'(k),$$

output

$$y = f(k),$$

and gross investment

$$i = f(k) - c - g = y - c - g.$$

Quantities like investment and output could have been written into the problem from the beginning, but prices like the rental rate and the real wage arise out of the logic of the problem.

The steady-state versions of the originial three equations and the above definitions are

$$
\begin{aligned}
f(k^*) &= \delta k^* + c^* + g \\
f'(k^*) &= \rho + \delta. \\
u'(c^*) &= \lambda^* \\
R^* &= f'(k^*) \\
r^* &= R^* - \delta = \rho \\
w^* &= f(k^*) - k^* f'(k^*) \\
y^* &= f(k^*) \\
i^* &= y^* - c^* - g.
\end{aligned}
$$

Two simple implications are

$$i^* = \delta k^*$$

or

$$\frac{i^*}{k^*} = \delta,$$

and

$$R^* = \rho + \delta.$$

In the steady state, the investment/capital ratio must equal the depreciation rate and the rental rate must equal the sum of the depreciation rate and the discount rate.

Let's see how things would have gone using log-linearization. We need one item of new notation; for any variable x, we need some way to denote $\ln(x) - \ln(x^*)$ when this is considered a small enough quantity that the rules of calculus for differentials can be applied to it. We will use the notation "check" for such a logarithmic deviation; that is, we will write

$$\check{x} = \ln(x) - \ln(x^*).$$

By the rules of differentiation,

$$\check{x} = \widetilde{\ln(x)} = \frac{d\ln(x)}{dx}|_{x=x^*} \cdot \tilde{x} = \frac{\tilde{x}}{x^*}.$$

Analyzing the model into components as small as possible, there are seven components to the model: First, the determination of consumption c by the marginal value of capital λ. Second, the relation of the marginal value of capital λ to the interest rate r. Third, the accumulation equation. Fourth, the material balance condition (disposition of output identity). Fifth, overall output as determined by the production function. Sixth, factor prices as determined by the production function. Seventh, the capital market. Each of these is individually easy to study in an economically meaningful way and they are easy to put together. For concreteness, we will give more or less plausible numerical examples for each equation.

14.3.1 Consumption and the Marginal Value of Investment

It is easiest to start by log-linearizing first-order conditions and constraints. Taking the derivative of the logarithm of each side of the equation

$$u'(c) = \lambda$$

with respect to the logarithm of the arguments c and λ and multiplying these derivatives by the logarithmic differentials $\check{c}$ and $\check{\lambda}$ yields

$$\frac{c^* u''(c^*)}{u'(c^*)} \check{c} = \check{\lambda}. \tag{14.2}$$

(Mechanically, an equation can be log-linearized just by taking the total differential of the logarithm, which yields an expression in the straight deviations, then substituting in $\tilde{x} = x^* \check{x}$.)

Upon log-linearizing an equation that does not involve a time derivative, every coefficient must be a pure number—such as an elasticity or a share parameter. Why? Logarithmic deviations like $\check{c}$ can be thought of as percentage deviations and are themselves pure numbers or equivalently, pure percentages (the word "percent" is really just a synonym for .01). Thus the coefficients that relate the percentage deviations must be pure numbers like elasticities and shares. In a sense, the issue reduces to one of naming each coefficient in a suitable way, since each coefficient is guaranteed to be meaningful and interpretable.

For the sake of standardization, two useful guidelines to follow unless there is a good reason to do otherwise are (1) to make all parameter labels represent positive numbers whenever they have an unambiguous sign and (2) where there is a distinction between prices and quantities, to favor elasticities of quantities with respect to prices rather than elasticities of prices with respect to a quantity which are their reciprocals. These two guidelines suggest writing (14.2) as

$$\check{c} = -s\check{\lambda}, \tag{14.3}$$

where s is defined as

$$s = \frac{-u'(c^*)}{c^* u''(c^*)} > 0.$$

(In order to avoid too great a profusion of asterisks, we do not put asterisks on these parameters, but it is good to think of an asterisk as being there implicitly on most such parameters.) If we did not already know our way around, s would at this point simply be the elasticity of consumption with respect to the marginal utility or util price of consumption. The identity of s as the elasticity of intertemporal substitution might emerge only later.

There is much debate over the value of the elasticity of intertemporal substitution s. A serious discussion of this debate must wait until Chapter **??**. For now, as numerical examples, consider the two possible values $s = 1$, implying

$$\check{c} = -\check{\lambda}$$

and $s = .15$, implying

$$\check{c} = -.15\check{\lambda}.$$

14.3.2 The Marginal Value of Investment and the Real Interest Rate

The Euler equation can be written

$$\frac{\dot{\lambda}}{\lambda} = \frac{d\ln(\lambda)}{dt} = \rho - r.$$

If

$$\lim_{t\to\infty} \lambda_t = \lambda^*,$$

this differential equation has the unique solution

$$\ln(\lambda_t) = \int_t^{\infty} (r_{t'} - \rho) dt' = \int_t^{\infty} (r_{t'} - r^*)\, dt'.$$

In words, the logarithm of the marginal value of investment is equal to an integral from time t to the infinite future of the excess of the real interest rate over its steady-state value.

To log-linearize the Euler equation and the accumulation equation we need to clarify first that in the neighborhood of the steady state,

$$\dot{\check{\lambda}} = \frac{d}{dt}[\ln(\lambda) - \ln(\lambda^*)] = \frac{\dot{\lambda}}{\lambda} = \frac{\dot{\tilde{\lambda}}}{\lambda^*},$$

Similarly, in the neighborhood of the steady state,

$$\dot{\check{k}} = \frac{d}{dt}[\ln(k) - \ln(k^*)] = \frac{\dot{k}}{k} = \frac{\dot{\tilde{k}}}{k^*}.$$

When log-linearizing the dynamic equations, the coefficients will all be rates, since a rate of growth is being related to percentage deviations that are pure numbers. Alternatively, one can allow the variation in some rates remain in the equation as straight deviations rather than logarithmic deviations. Considering first the Euler equation, since

$$\frac{\dot{\lambda}}{\lambda} = \rho - r,$$

in the neighborhood of the steady state,

$$\dot{\check{\lambda}} = -\tilde{r}.$$

In integral form,

$$\check{\lambda}_t = \int_t^\infty \tilde{r}_{t'}\, dt'.$$

Combined with (14.3), this implies

$$\check{c} = -s \int_t^\infty \tilde{r}_{t'}\, dt'.$$

For example, this says that a real interest rate expected to be 1% above normal for 1 year implies consumption s% below normal (*e.g.*, 1% or .15 % below normal). Similarly, a real interest rate expected to be 2% above normal for 3 years implies consumption $6s$% below normal (*e.g.*, 6% or .9% below normal). On the other hand, a real interest rate expected to be 6% above normal for one month implies consumption only $.5s$% below normal (*e.g.*, .5% or .075% below normal).

14.3.3 Capital Accumulation

As for the accumulation equation, since

$$\frac{\dot{k}}{k} = \frac{i - \delta k}{k} = \frac{i}{k} - \delta,$$

and δ is a constant, and $\ln(i/k) = \ln(i) - \ln(k)$,

$$\begin{aligned}\dot{\check{k}} &= \widetilde{\left(\frac{i}{k}\right)} \\ &= \frac{i^*}{k^*}[\check{\imath} - \check{k}] \\ &= \delta[\check{\imath} - \check{k}].\end{aligned}$$

As a numerical example, take $\delta = .08/\text{year}$. Then

$$\dot{\check{k}} = \frac{.08}{\text{year}}[\check{\imath} - \check{k}].$$

This means for example that an investment/capital stock ratio 10% above normal for one year would make the capital stock at the end of that time approximately $\frac{.08}{\text{year}} \cdot 1\text{year} \cdot 10\% = .8\%$ larger than it would otherwise be.

14.3.4 Material Balance

The equation

$$y = c + i + g,$$

can be log-linearized as

$$\begin{aligned}\check{y} &= \frac{c^*}{y^*}\check{c} + \frac{i^*}{y^*}\check{\imath} + \frac{g^*}{y^*}\check{g} \\ &= \zeta_c\check{c} + \zeta_i\check{\imath} + \zeta_g\check{g},\end{aligned}$$

where ζ_c, ζ_i and ζ_g are respectively the shares of consumption, investment, and government purchases in output. Solving for $\check{\imath}$,

$$\check{\imath} = \frac{\check{y} - \zeta_c\check{c} - \zeta_g\check{g}}{\zeta_i}.$$

The Output Shares

It is still unclear how the output shares ζ_c and ζ_i relate to other steady-state relationships. We do know that

$$\frac{i^*}{k^*} = \delta,$$

so that

$$\zeta_i = \frac{i^*}{y^*} = \frac{i^*}{k^*}\frac{k^*}{y^*} = \delta\frac{k^*}{y^*}. \tag{14.4}$$

Since by definition,

$$\zeta_c + \zeta_i + \zeta_g = \frac{c^* + i^* + g^*}{y^*} = 1,$$

we can write

$$\zeta_c = 1 - \zeta_i - \zeta_g = 1 - \delta\frac{k^*}{y^*} - \zeta_g,$$

and the share of government purchases in output ζ_g might fairly be taken as a parameter.

The Capital/Output Ratio

If we define θ as capital's share,

$$\theta = \frac{R^* k^*}{y^*},$$

then the capital/output ratio $\frac{k^*}{y^*}$ can be expressed also as

$$\frac{k^*}{y^*} = \frac{\theta}{R^*} = \frac{\theta}{\rho + \delta}.$$

Substituting this expression for the capital/output ratio into (14.4), we find that

$$\zeta_i = \frac{\delta}{R^*}\theta = \frac{\delta}{\rho + \delta}\theta. \tag{14.5}$$

For what it is worth, this also means that

$$\zeta_c = 1 - \frac{\delta}{\rho + \delta}\theta - \zeta_g.$$

Numerical Example

If we set capital's share at .3 and the utility discount rate at .02/year (and $\delta = .08$/year as above), it implies a capital/output ratio of

$$\frac{k^*}{y^*} = \frac{.3}{(.02 + .08)}\text{years} = 3\text{years}$$

and an investment share of

$$\zeta_i = \frac{\delta}{\rho + \delta}\theta = \frac{.08}{.02 + .08}.3 = .24.$$

The appropriate government purchases share depends on what "government purchases" represent in the model. They represent only those government purchases—such as military spending—that are not good substitutes for private consumption and investment. In this light, consider $\zeta_g = .06$. Then $\zeta_c = 1 - .24 - .06 = .7$ and the log-linearized material balance equation becomes

$$\check{y} = .7\check{c} + .24\check{\imath} + .06\check{g}.$$

Solved for $\check{\imath}$, this becomes

$$\check{\imath} = 4.17\check{y} - 2.92\check{c} - .25\check{g}.$$

14.3.5 The Production Function and Output

Log-linearizing $y = f(k)$ reveals that

$$\check{y} = \frac{f'(k^*)k^*}{f(k^*)}\check{k} = \frac{R^*k^*}{y^*}\check{k} = \theta\check{k}$$

where, as above,

$$\theta = R^*k^*/y^* = \frac{f'(k^*)k^*}{f(k^*)}$$

is capital's (steady-state) share in output. Note that since the quantity of labor is normalized to 1, labor's (steady-state) share in output is

$$\frac{w^*}{y^*} = \frac{f(k^*) - k^*f'(k^*)}{f(k^*)} = 1 - \frac{k^*f'(k^*)}{f(k^*)} = 1 - \theta.$$

14.3.6 The Production Function and Factor Prices

The Cobb-Douglas Case

In the Cobb-Douglas case, with θ constant, the general form of the production function is

$$F(k,n) = \Xi k^{\theta} n^{1-\theta} = \Xi k^{\theta} = f(k),$$

since $n = 1$. The marginal physical products, which here are the factor prices, are

$$R = F_k(k,n) = \theta \Xi k^{-(1-\theta)} n^{(1-\theta)} = \theta \Xi k^{-(1-\theta)} = f'(k) \tag{14.6}$$

and

$$w = F_n(k,n) = (1-\theta)\Xi k^{\theta} n^{-\theta} = (1-\theta)\Xi k^{\theta} = f(k) - kf'(k). \tag{14.7}$$

Log-linearizing (14.6) and (14.7), with Ξ and n (and θ) constant,

$$\check{R} = (1-\theta)[\check{n} - \check{k}] + \check{\Xi} = -(1-\theta)\check{k} \tag{14.8}$$

and

$$\check{w} = \theta[\check{k} - \check{n}] + \check{\Xi} = \theta \check{k}. \tag{14.9}$$

The limitation of Cobb-Douglas production functions is that the elasticity of substitution between capital and labor σ is equal to 1. That is, from (14.8) and (14.9),

$$\check{k} - \check{n} = 1 \cdot [\check{w} - \check{R}].$$

In order to see what the effect values of σ other than 1 have, we must deal with a more general case.

In General: The Factor Price Possibility Frontier

It is helpful to consider both factor prices together. Log-linearizing $R = f'(k)$ yields

$$\check{R} = \frac{k^* f''(k^*)}{f'(k^*)} \check{k},$$

while log-linearizing $w = f(k) - kf'(k)$ yields

$$\check{w} = -\frac{(k^*)^2 f''(k^*)}{f(k^*) - k^* f'(k^*)}\check{k}.$$

Both of these log-linearized equations have $f''(k^*)$ at their core. One way to express the relationship between the two equations is

$$\begin{aligned}\theta\check{R} + (1-\theta)\check{w} &= \frac{k^* f'(k^*)}{f(k^*)}\frac{k^* f''(k^*)}{f'(k^*)}\check{k} + \frac{f(k^*) - k^* f'(k^*)}{f(k^*)}\left(\frac{-(k^*)^2 f''(k^*)}{f(k^*) - k^* f'(k^*)}\right)\check{k} \\ &= 0.\end{aligned} \tag{14.10}$$

Equation (14.10) is the log-linear equation for the factor price possibility frontier. It says that if the real wage goes up, the real rental rate must go down, and vice versa. Why? First, remember that the output price p is the numeraire, so that $p \equiv 1$ and $\check{p} = 0$. With perfect competition, and in the absence of technological progress, movements in the price of output is a weighted average of the movements in the factor prices that determine costs and each factor price should be weighted by that factor's share in costs. Therefore,

$$\theta\check{R} + (1-\theta)\check{w} = \check{p} = 0.$$

Equivalently,

$$\check{w} = -\frac{\theta}{(1-\theta)}\check{R}$$

or

$$\check{R} = -\frac{(1-\theta)}{\theta}\check{w}.$$

In General: Capital/Labor Substitution

By definition, the elasticity of capital/labor substitution σ gives the logarithmic change in the capital/labor ratio $\frac{k}{n}$ with respect to the logarithmic change in the wage/rental ratio $\frac{w}{R}$. Since the quantity of labor is normalized to 1 ($n \equiv 1$),

$$\sigma(\check{w} - \check{R}) = \check{k} - \check{n} = \check{k}.$$

Substituting in $\check{w} = -\frac{\theta}{(1-\theta)}\check{R}$,

$$\sigma\left[-\frac{\theta}{(1-\theta)}-1\right]\check{R} = -\frac{\sigma}{(1-\theta)}\check{R}$$
$$= \check{k}.$$

Solving for $\check{R}$,

$$\check{R} = -\frac{1-\theta}{\sigma}\check{k}. \tag{14.11}$$

Finally, using $\check{w} = -\frac{\theta}{(1-\theta)}\check{R}$,

$$\check{w} = \frac{\theta}{\sigma}\check{k}. \tag{14.12}$$

In words, when the production function is not Cobb-Douglas, equations (14.8) and (14.9) by dividing the coefficient of the capital/labor ratio (here just k) by σ. Factor prices are more sensitive to changes in the capital/labor ratio the lower is the elasticity of substitution between capital and labor. When $\sigma = 1$, (14.11) and (14.12) reproduce the Cobb-Douglas case ((14.8) and (14.9)).

Numerical Example

For a first pass, the Cobb-Douglas case is not unreasonable. With $\sigma = 1$ and $\theta = .3$,

$$\check{R} = -.7\check{k}$$

and

$$\check{w} = .3\check{k}.$$

14.3.7 The Capital Market

The capital market in the Cass model is exceedingly simple. As is implicit in the definitions we have made, arbitrage requires that the real interest rate r equal the net real return on capital $R - \delta$. Since δ is a constant,

$$\tilde{r} = \tilde{R} = R^*\check{R} = (\rho + \delta)\check{R}.$$

The reason this equation is so simple is that with one unit of output being able to make one unit of capital, the price of capital is fixed at 1. Thus, there is no capital gains term in the return to capital. In other models that allow the relative price of capital to fluctuate, the capital gains term must be dealt with. (See Chapter .)

Numerically, with $\rho = .02/\text{year}$ and $\delta = .08/\text{year}$,

$$\tilde{r} = \frac{.10}{\text{year}} \check{R}.$$

In words, this says that a 1% increase in the rental rate raises the real interest rate by .1 %/year or 10 basis points.

14.4 Putting the Model Back Together after Log-Linearization

In order to study the dynamics of the model—in particular, to get the convergence rate—we need log-linear equations for $\dot{\check{k}}$ and $\dot{\check{\lambda}}$ in terms of $\check{k}$ and $\check{\lambda}$. As we put the model back together to get these equations, we will give both algebraic and numerical versions of the equations for the parameter values given above, including two possibilities for the elasticity of intertemporal substitution s.

14.4.1 Accumulation as a Function of k and λ

Since $\check{c} = -s\check{\lambda}$, $\check{y} = \theta\check{k}$ and $\check{g} = 0$,

$$\begin{aligned} \check{\imath} &= \frac{\check{y} - \zeta_c \check{c} - \zeta_g \check{g}}{\zeta_i} \\ &= \frac{\theta \check{k} + s\zeta_c \check{\lambda}}{\zeta_i}. \end{aligned}$$

Numerically, when $s = 1$,

$$\check{\imath} = 1.25\check{k} + 2.92\check{\lambda};$$

when $s = .15$,

$$\check{\imath} = 1.25\check{k} + .44\check{\lambda}.$$

Then, using (14.5) ($\zeta_i = \frac{\delta}{\rho+\delta}\theta$),

$$
\begin{aligned}
\dot{\check{k}} &= \delta[\check{\imath} - \check{k}] \\
&= \frac{\delta}{\zeta_i}[\theta\check{k} + s\zeta_c\check{\lambda}] - \delta\check{k} \\
&= \frac{\rho+\delta}{\theta}[\theta\check{k} + s\zeta_c\check{\lambda}] - \delta\check{k} \\
&= \rho\check{k} + \frac{(\rho+\delta)s\zeta_c}{\theta}\check{\lambda}.
\end{aligned}
$$

(One could also substitute in $1 - \frac{\delta}{\rho+\delta}\theta - \zeta_g$ for ζ_c.) Numerically, when $s = 1$,

$$\dot{\check{k}} = \frac{.02}{\text{year}}\check{k} + \frac{.233}{\text{year}}\check{\lambda};$$

when $s = .15$,

$$\dot{\check{k}} = \frac{.02}{\text{year}}\check{k} + \frac{.035}{\text{year}}\check{\lambda}.$$

14.4.2 The Euler Equation as a Function of k and λ.

Putting the equations above together, starting with the fact that $\check{R} = -\frac{(1-\theta)}{\sigma}\check{k}$,

$$
\begin{aligned}
\dot{\check{\lambda}} &= -\tilde{r} \\
&= -R^*\check{R} \\
&= \frac{R^*(1-\theta)}{\sigma}\check{k} \\
&= \frac{(\rho+\delta)(1-\theta)}{\sigma}\check{k}.
\end{aligned}
$$

Numerically,

$$\dot{\check{\lambda}} = \frac{.07}{\text{year}}\check{k}.$$

14.4.3 The Dynamic Matrix in Log-Linear Form

In matrix, log-linear form, the accumulation and Euler equation can be written together as

$$\begin{bmatrix} \dot{\check{k}} \\ \\ \dot{\check{\lambda}} \end{bmatrix} = \begin{bmatrix} \rho & \frac{(\rho+\delta)s\zeta_c}{\theta} \\ \\ \frac{(\rho+\delta)(1-\theta)}{\sigma} & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \\ \check{\lambda} \end{bmatrix}.$$

Numerically, when $s = 1$, the dynamic matrix *in annual terms* is

$$\begin{bmatrix} \dot{\check{k}} \\ \\ \dot{\check{\lambda}} \end{bmatrix} = \begin{bmatrix} .02 & .233 \\ \\ .07 & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \\ \check{\lambda} \end{bmatrix}.$$

When $s = .15$ the dynamic matrix is

$$\begin{bmatrix} \dot{\check{k}} \\ \\ \dot{\check{\lambda}} \end{bmatrix} = \begin{bmatrix} .02 & .035 \\ \\ .07 & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \\ \check{\lambda} \end{bmatrix}.$$

The only difference between the dynamics in these two cases is that $\dot{k}$ is more sensitive to λ when the elasticity of intertemporal substitution s is large.

14.4.4 The Convergence Rate

The convergence rate is therefore

$$\kappa = \sqrt{\frac{\rho^2}{4} + (\rho+\delta)^2 \frac{s(1-\theta)\zeta_c}{\theta\sigma}} - \frac{\rho}{2}. \tag{14.13}$$

As compared to where we stood at (14.1), we have now justified the way in which $\frac{1-\theta}{\sigma}$ appears and reexpressed the consumption/capital ratio $\frac{c^*}{k^*}$ as

$$\frac{c^*}{k^*} = \frac{c^*}{y^*} \frac{y^*}{R^* k^*} R^* = \frac{\zeta_c(\rho+\delta)}{\theta}.$$

The Dependence of the Convergence Rate on s

Since obtaining the convergence rate was the main goal of this chapter, it seems worthwhile to calculate it for a broader range of values of s than just $s = 1$ and $s = .15$. Straightforward substitution into (14.13) of all of the other parameters yields the equation

$$\kappa = [\sqrt{1 + 163.333s} - 1]\%/\text{year},$$

where % means .01. As can be seen, the convergence rate is roughly proportional to the square-root of the elasticity of intertemporal substitution, at least for values on the order of 1. The following table gives convergence rates in annual terms for various values of s:

s	κ
2	17.1%
1	11.8%
.75	10.1%
.5	8.1%
.25	5.5%
.15	4.0%
.05	2.0%

Remarks

We will see later that the convergence rate measures the duration of the aftereffects of shocks, as those aftereffects are generated endogenously by the model. Thus, a convergence rate on the order of, say 25% per year would imply endogenous dynamics occurring at a business cycle frequency (with a half-life of less than three years), while a convergence rate of .05 has a half-life of $\frac{\ln(2)}{.05} \approx 14$ and suggests dynamics operating only over the course of more than a decade. The lower the elasticity of intertemporal substitution, the less the fluctuations in real models of the economy are about business cycles and the more they are about decadal or slower fluctuations.

Chapter 15

The Quantitative Analytics of Household Behavior

15.1 The Return of the King-Plosser-Rebelo Utility Function

Since the aim of quantitative analytics is to explore the effects of various parameter values, we will typically lean toward generality by extensive use of a nonparametric approach. That is, we will look at Taylor expansions that can represent the behavior of fairly general functional forms. Above all, it is valuable to avoid making an arbitrary assumption of unknown characteristics by the choice of a specific functional form. That said, it is often quite helpful to make an explicit functional form restriction to take care of some issues and focus attention on other, more important issues—especially if the meaning of the functional form restriction is clear. A good practice is to start from a general case and then insist on stating the key functional form restrictions that narrow things down in economic terms, rather than in purely mathematical terms. This makes the economic meaning of the restriction clear and helps one to judge the plausibility and appropriateness of the assumption.

One useful and plausible set of restrictions are those restrictions needed to allow steady-state growth. Over the course of the last century or two, real wages, per capita output and consumption have risen by at least an order of magnitude (that is, by a factor of 10) [some data would be good here]. Yet many key economic quantities and ratios have remained relatively constant in the face of these dramatic changes. In particular, the average workweek has changed little in comparison with the dramatic increases in real wages. Also, the ratio of the real wage to per-capita consumption and per capita output have remained relatively constant in the face of these dramatic changes, as has the capital/output ratio.

In Chapter **??** we mentioned that the restrictions on the utility function necessary to allow steady-state growth restrict the utility function to the King-Plosser-Rebelo form:

$$u(c,n) = \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}$$

with $\beta \geq 0$, $v' > 0$, $v'' > 0$ and u concave in both arguments and

$$u(c,n) = \ln(c) - v(n)$$

as the limit when $\beta \to 1$. In this chapter, we will look at the quantitative analytics of this form of the utility function.

15.2 The Household's Problem

In setting up the household's problem, we will allow for the possibility of a numeraire other than consumption. The results in this chapter will be true for any numeraire, but in particular, it is convenient in later chapters to take investment goods as the numeraire. The multiplicative consumption goods technology shifter will cause the relative price of consumption to investment goods, p_c, to vary. (If one happened to use money as the numeraire, then r would actually be the nominal interest rate and all of the other prices would be nominal prices.)

Consider the partial equilibrium household problem

$$\max_{c,n} \quad \int_0^\infty e^{-\rho t} \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}\, dt$$

subject to

$$\dot{a} = ra + wn - p_c c + \mathcal{T},$$

where a is household assets denominated in terms of investment goods or other numeraire, r is the interest rate in terms of investment goods or other numeraire, and w is the wage in terms of investment goods or other numeraire and $\mathcal{T}$ is net transfers to the household. Using the Greek letter μ for the marginal value of investment (or the marginal value of the numeraire), the current-value Hamiltonian is

$$H = \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)} + \mu[ra + wn - p_c c + \mathcal{T}].$$

The first-order conditions are

$$c^{-\beta} e^{(\beta-1)v(n)} = p_c \mu \tag{15.1}$$

and

$$v'(n) c^{1-\beta} e^{(\beta-1)v(n)} = w\mu. \tag{15.2}$$

Dividing (15.2) by (15.1) reveals that the wage w is given by

$$w = p_c c v'(n). \tag{15.3}$$

The Euler equation is

$$\frac{\dot{\mu}}{\mu} = \rho - r. \tag{15.4}$$

15.3 The Consumption-Constant Elasticity of Labor Supply

Taking logarithms of (15.3),

$$\ln(w) = \ln(p_c) + \ln(c) + \ln(v'(n)).$$

Thus, the log-linearized version of (15.3) is

$$\begin{aligned} \check{w} &= \check{p}_c + \check{c} + \frac{n^* v''(n^*)}{v'(n^*)} \check{n} \\ &= \check{p}_c + \check{c} + \frac{1}{\eta} \check{n}, \end{aligned} \tag{15.5}$$

or

$$\check{n} = \eta[\check{w} - \check{p}_c - \check{c}], \tag{15.6}$$

where η is the consumption-constant labor supply elasticity,

$$\eta = \frac{v'(n^*)}{n^* v''(n^*)}. \tag{15.7}$$

15.4 The Income Expansion Path

The King-Plosser-Rebelo form of the utility function implies equal and opposite income and substitution effects of the real consumption wage $\frac{w}{p_c}$ on labor supply. The parity between income and substitution effects allows labor to be constant in the face of an upward trend in the real consumption wage. The parity between income and substitution effects also implies a close relationship between the size of income effects and the size of the consumption-constant labor supply elasticity η.

15.4.1 The Marginal Expenditure Shares of Consumption and Leisure

Figure **??** shows the income expansion path—or more precisely, the expenditure expansion path—for a particular value of the real consumption wage $\frac{w}{p_c}$. As long as both consumption and leisure are normal goods, consumption and leisure move together as one moves along the income expansion path. That means that consumption and labor move in opposite directions, so that the income expansion path is downward sloping when n and c are on the axes [n is on the horizontal, c is on the vertical axis]. The magnitude of the slope shows how additional expenditure would be split between additional consumption and reduced labor. If dc is the change in consumption and dn is the (negative) change in labor in response to some additional non-labor income, then the total amount of extra expenditure is $p_c\,dc - w\,dn$. Let us use h to denote the share of the additional expenditure devoted to consumption. Then the marginal expenditure share of consumption is

$$h = \frac{p_c\,dc}{p_c\,dc - w\,dn}.$$

The remainder of the additional expenditure is the reduction in labor income from reducing n. Equivalently, the remainder of the additional expenditure is on increased leisure. Therefore, $1 - h$ can be called the marginal expenditure share of leisure. The marginal expenditure share of leisure is

$$1 - h = \frac{-w\,dn}{p_c\,dc - w\,dn}.$$

15.4.2 The Slope of the Income Expansion Path and the Labor Supply Elasticity

Therefore, as noted in Figure **??**, the slope of the income expansion path is

$$\frac{dc}{dn} = -\frac{w}{p_c}\frac{h}{1-h}.$$

Given the real consumption wage $\frac{w}{p_c}$, a steep downward slope indicates a high marginal expenditure share of consumption, while a gradual downward slope indicates a high marginal expenditure share of leisure.

In elasticity form,

$$\begin{aligned} \frac{d\ln(c)}{d\ln(n)} &= \frac{ndc}{cdn} \\ &= -\frac{wn}{p_c c}\frac{h}{1-h}. \end{aligned} \tag{15.8}$$

For comparison, when $\breve{w} - \breve{p}_c = 0$, as it is everywhere along the income expansion path through the steady state, (15.5) implies

$$\breve{c} = -\frac{1}{\eta}\breve{n}.$$

Thus,

$$\frac{1}{\eta} = \frac{w^* n^*}{p_c^* c^*}\frac{h}{1-h}.$$

Or, writing

$$\tau = \frac{w^* n^*}{p_c^* c^*},$$

$$\begin{aligned} \eta &= \frac{p_c^* c^*}{w^* n^*}\frac{(1-h)}{h} \\ &= \frac{1-h}{\tau h}. \end{aligned} \tag{15.9}$$

This equation can also be solved for h or $1-h$, yielding

$$\begin{aligned} h &= \frac{1}{1+\tau\eta} \\ 1-h &= \frac{\tau\eta}{1+\tau\eta}. \end{aligned}$$

15.5 Using the Marginal Expenditure Shares of Consumption and Leisure to Calibrate the Labor Supply Elasticity

There is no general consensus on the magnitude of the labor supply elasticity.[1] As we will see later on, it is difficult to explain macroeconomic fluctuations without a substantial labor supply elasticity. Thus, it is problematic that microeconomic estimates of the labor supply elasticity are often quite low. Much of the Real Business Cycle literature relies on nonconvexities in labor supply to raise the macroeconomic labor supply elasticity. (See for example Rogerson, 1984, Hanson () and Prescott, 1987.) However, many of these models have a glaring defect as an explanation of the discrepancy between microeconomic estimates of the labor supply elasticity and the higher macroeconomic labor supply elasticity that seems apparent in the character of macroeconomic fluctuations: they typically imply not only a higher macroeconomic labor supply elasticity but a higher *microeconomic* labor supply elasticity as well—which should have shown up in the microeconomic estimates.[2]

Another approach to resolving the conflict between the apparent microeconomic and macroeconomic labor supply elasticities begins by noting the weaknesses in the microeconomic estimates of the labor supply elasticity. Many of these estimates rely on regressions involving observed high-frequency variations in observed wages, which may not equal the allocative shadow wage in a long-term labor relationship. Other estimates rely on looking at the effects of life-cycle variations in the real wage—effects that are difficult to disentangle from other age-linked effects.

However, casting doubt on the accuracy of microeconomic estimates of the labor supply elasticity still does not tell us what an appropriate value would be. Rather than leaving this key elasticity as a totally free parameter, it is a

[1] The following discussion draws on Kimball "Labor Market Dynamics," Kimball "Basic Neomonetarist Model," and the currently nonexistent papers Basu and Kimball "Neo-Classical Labor Supply and Demand: Theory and Evidence" and Barsky, Juster, Kimball and Shapiro "Experimental Evidence on Labor Supply."

[2] This is particularly true if the microeconomic estimates include movement in and out of employment. Microeconomic estimates that do not include movement in and out of employment may or may not capture the high labor supply elasticities in a Rogerson-type model, depending on whether or not the horizon over which labor supply is nonconvex for an individual worker is long enough for alternation from one side of the nonconvexity to the other to show up in the official statistics as movement in and out of employment. As Hall (1987) notes, if the nonconvexity stems only from commuting time, it operates only at a daily frequency and so can be convexified over the course of as little as a week.

good idea to have some way to gauge what range of values for the labor supply elasticity are plausible. The relationship

$$\eta = \frac{1-h}{\tau h}.$$

allows us to judge the plausibility of various values of the labor supply elasticity by relating them to the marginal expenditure shares of consumption and leisure.[3]

Before any adjustment for taxes, $\tau = \frac{w^* n^*}{p_c^* c^*}$, taken straight from the national income accounts, is about 1 or 1.1. However, in the household's problem, the relevant wage is the after-tax wage that the household actually sees after paying the marginal tax rate on an extra hour's wage. A reasonable marginal tax rate could easily reduce the after-tax value of τ to as low as .67. If τ is in the range (.67, 1.1), then $\frac{1}{\tau}$ is in the range $(.9, 1.5)$, so that η could be anywhere from slightly below to one-and-a-half times as big as the ratio $\frac{1-h}{h}$ of the marginal expenditure share of leisure to the marginal expenditure share of consumption.

What is the ratio of the marginal expenditure shares of leisure to the marginal expenditure share of consumption? The thought experiment to get these marginal expenditure shares is to think how one would spend the proceeds of lottery winnings. What percentage would you spend on additional consumption and what percentage on reduced time at work? The second wave of the Health and Retirement Study asked just such a question of a broad sample of older households. Barsky, Juster, Kimball and Shapiro (1996) find that the average value of this ratio is (??). A value of .5 for this ratio implies that a marginal expenditure share of .67 for consumption and .33 for leisure. A value of .5 for $\frac{1-h}{h}$ gives a value between .45 and .75 for the consumption-constant elasticity of labor supply η, depending on the value one chooses for τ.

It is helpful to consider what your own marginal expenditure shares are. We suspect that the majority of readers would spend a larger fraction of a

[3] This way of calibrating the labor supply elasticity from the marginal expenditure shares of consumption and leisure is in somewhat the same spirit as Prescott's (1986) calibration of the labor supply elasticity using the *average* expenditure shares of leisure and consumption, with the implicit assumption that the average and marginal expenditure shares are equal. The emphasis on average shares of consumption and leisure gave rise to the question of how to deal with sleep in calculating the average expenditure share of leisure, which made an enormous difference in the implied labor supply elasticity. The answer to the question of how to deal with sleep is to focus on the marginal expenditure shares rather than the average expenditure shares. The *marginal* expenditure shares are well-defined whether or not one treats sleep as leisure.

windfall on additional consumption than on leisure. If the marginal expenditure share of consumption h is larger than the marginal expenditure share of leisure $1-h$, it implies that $h > .5$ and $\frac{1-h}{h} < 1$, making $\eta < 1.5$ even when $\tau = .67$.

One thing that might help getting a somewhat higher labor supply elasticity is that if each household has King-Plosser-Rebelo utility, but households differ in their value of h, then, as shown in the next chapter, the appropriate aggregate labor supply elasticity is a consumption-weighted average of the value of $\frac{1-h}{h}$ for each household divided by (aggregate) τ. Since $\frac{1-h}{h}$ is strongly convex in h, variance in h across households can contribute substantially to the aggregate labor supply elasticity.

Beyond that, in order to get a much higher labor supply elasticity would require one to defend the idea that people would spend a large fraction of a windfalls on reduced work. In the extreme, the only way to get the infinite labor supply elasticity that appears in some Real Business Cycle Models and still have income and substitution effects cancel is to have a marginal expenditure share of leisure equal to 1. That is, those models imply that a household would spend 100% of a windfall on reducing labor supply. This is true even if the model obtains an infinite labor supply elasticity from Rogerson's nonconvexities. These nonconvexities will either operate strongly to raise the size of income effects or they will get in the way of steady-state growth. Even if some kind of externality is involved, the externality must allow income effects commensurate with substitution effects or it will get in the way of steady-state growth.

15.6 Putting the First-Order Condition for Consumption and the Euler Equation Together

Given the level of consumption, the labor supply equation is an intratemporal equation with no intertemporal dimension. Combining the first-order condition for consumption (15.1) with the Euler equation (15.4) gives rise to what is often called the "consumption Euler equation." In particular, (15.1) implies that

$$\ln(p_c) + \ln(\mu) = -\beta \ln(c) + (\beta - 1)v(n). \tag{15.10}$$

Using $w = p_c c v'(n)$ or $v'(n) = \frac{w}{p_c c}$, one can linearize (15.10) around the steady state as

$$\begin{aligned} \check{p}_c + \check{\mu} &= -\beta\check{c} + (\beta - 1)v'(n^*)\tilde{n} \qquad (15.11) \\ &= -\beta\check{c} + (\beta - 1)v'(n^*)n^*\check{n} \\ &= -\beta\check{c} + (\beta - 1)\frac{w^*n^*}{p_c^*c^*}\check{n} \\ &= -\beta\check{c} + (\beta - 1)\tau\check{n}. \end{aligned}$$

It is also useful to solve the equation for $\check{c}$:

$$\begin{aligned} \check{c} &= -\frac{1}{\beta}(\check{p}_c + \check{\mu}) + [1 - \frac{1}{\beta}]\tau\check{n} \qquad (15.12) \\ &= -s(\check{p}_c + \check{\mu}) + (1 - s)\tau\check{n}, \end{aligned}$$

where $s = \frac{1}{\beta}$ is the elasticity of intertemporal subtitution in consumption, holding labor constant.

Linearizing (15.4) around the steady state yields

$$\dot{\check{\mu}} = \rho - r. \qquad (15.13)$$

Once a time derivative is taken of (15.12), it can be combined with (15.13):

$$\begin{aligned} \dot{\check{c}} &= -s(\dot{\check{p}}_c + \dot{\check{\mu}}) + (1 - s)\tau\dot{\check{n}} \qquad (15.14) \\ &= s[r - \dot{\check{p}}_c - \rho] + (1 - s)\tau\dot{\check{n}}. \end{aligned}$$

15.6.1 Empirical Implementation

This is a good chance to show how these expansions around the steady-state relate to a time-derivative in place. Combine (15.4) with (15.10) to get,

$$\begin{aligned} \rho + \frac{\dot{p}_c}{p_c} - r &= \frac{\dot{\mu}}{\mu} \\ &= -\beta\frac{\dot{c}}{c} + (\beta - 1)v'(n)\dot{n}. \end{aligned}$$

Solving for $\frac{\dot{c}}{c}$ and using the wage equation $w = cv'(n)$ to replace $v'(n)$ with $\frac{w}{c}$, one finds that

$$
\begin{aligned}
\frac{\dot{c}}{c} &= \frac{1}{\beta}(r - \frac{\dot{p}_c}{p_c} - \rho) + [1 - \frac{1}{\beta}]nv'(n)\left(\frac{\dot{n}}{n}\right) \qquad (15.15)\\
&= \frac{1}{\beta}(r - \frac{\dot{p}_c}{p_c} - \rho) + [1 - \frac{1}{\beta}]\frac{wn}{p_c c}\left(\frac{\dot{n}}{n}\right)\\
&\approx \frac{1}{\beta}(r - \frac{\dot{p}_c}{p_c} - \rho) + [1 - \frac{1}{\beta}]\frac{w^* n^*}{p_c^* c^*}\left(\frac{\dot{n}}{n}\right)\\
&= \frac{1}{\beta}(r - \frac{\dot{p}_c}{p_c} - \rho) + [1 - \frac{1}{\beta}]\tau\left(\frac{\dot{n}}{n}\right).
\end{aligned}
$$

None of the variables $r - \frac{\dot{p}_c}{p_c}$, $\frac{\dot{c}}{c}$ or $\frac{\dot{n}}{n}$ need to be constant over time. Therefore, this is very close to being an actual regression. Since τ can be drawn from the national income accounts, the two unknown parameters are β and ρ. Define $s = \frac{1}{\beta}$ as above and $\alpha = -\rho/\beta$. After time-aggregation, one can use instrumental variables (from before the two-quarter period covered by the difference of consumption) and make allowances for the MA(1) structure of the error term [we need to decide how much simple rational expectations econometrics we want to include in the book or if we want to assume it] to estimate α and s—the (labor-constant) elasticity of intertemporal substitution in consumption—from the equation

$$\Delta \ln(c) - \tau\Delta \ln(n) = \alpha + s[r - \Delta \ln(p_c) - \tau\Delta \ln(n)].$$

Using this estimating equation, Basu and Kimball (1996) find that s is about .5.

Other work, which takes additive separability as a maintained assumption suggests a very small value of s. See Hall "Intertemporal Substitution in Consumption" and Barsky, Juster, Kimball and Shapiro "An Experimental Approach to Preference Parameters ...." The maintained assumption of additive separability makes it hard to gauge the appropriate value of s, but is not an issue in rejecting $s = 1$, since $s = 1$ implies additive separability and so is nested within both the additively separable case and the King-Plosser-Rebelo case. Therefore, there is substantial evidence that $s < 1$.

15.6.2 Empirical Implications of Nonseparability between Consumption and Labor

Much of the literature on the consumption Euler equation assumes additive separability between consumption and labor. It should be clearer now how restrictive this assumption would be. Additive separability between consumption and labor would require one either (1) to assume that $s = 1$, contrary to the bulk of the evidence, or (2) to give up on the idea of having income and substitution effects on labor supply cancelling (even approximately), forcing one to look for another explanation for why the workweek has had so little trend in the face of dramatic secular increases in the real wage.

In its favor, additive separability has only convenience. With a better understanding of the nonseparable case, that advantage of convenience becomes much less important.

One of the most interesting features of (15.15) is that, despite the nonseparability between consumption and labor, there are still only two parameters to be estimated: the time-discount rate ρ and the elasticity of intertemporal substitution s (or equivalently, $\frac{1}{\beta}$). By (15.15), If s is fairly low, the effect of labor on the consumption Euler equation must be substantial, since we know τ to be somewhere between .67 and 1.1. A King-Plosser-Rebelo utility function is not even approximately additively separable if s is far from 1. As long as there are substantial fluctuations in labor n, assuming additive separability is likely to cause serious misjudgments.

Exercises

1. By introspection, find out what your own personal marginal expenditure shares are. Calculate the implied consumption-constant labor supply elasticity assuming that you have a King-Plosser-Rebelo utility function and that $\frac{w^* n^*}{p_c^* c^*} = .67$.

Chapter 16

Labor Market Equilibrium

In Chapter 1, we used supply and demand in the labor market and the factor-price possibility frontier to analyze the Basic Real Business Cycle Model graphically. In this chapter, we derive the labor supply curve, labor demand curve and the factor price possibility frontier, looking at them algebraically as well as graphically. The first step is to revisit capital-labor substitution. Then we will revisit the factor price possibility frontier, the labor demand curve and the labor supply curve.

16.1 Capital-Labor Substitution when there are Two Sectors and Movements in Technology

Let us return to a dichotomy between production firms and leasing firms and the production functions for investment and consumption goods that we used in Chapter **??** Given the wage w and the rental rate R, a production firm's maximization problem is totally atemporal; we don't need to know anything about the past and the future to solve it. With investment goods as the numeraire, the representative firm in the investment goods sector faces the problem

$$\max_{k_i,n_i} i - Rk - wn = zn_i f(\frac{k_i}{zn_i}) - Rk - wn. \tag{16.1}$$

Similarly, with p_c as the price of consumption goods relative to investment goods, the representative firm in the consumption goods sector faces the problem

$$\max_{k_c, n_c} p_c c - Rk_c - wn_c = p_c \xi z n_c f(\frac{k_c}{zn_c}) - Rk - wn. \tag{16.2}$$

With perfect labor mobility across sectors, the first-order conditions for the optimal quantities of labor are

$$w = z\left[f(\frac{k_i}{zn_i}) - \frac{k_i}{zn_i} f'(\frac{k_i}{zn_i})\right] = p_c \xi z \left[f(\frac{k_c}{zn_c}) - \frac{k_c}{zn_c} f'(\frac{k_c}{zn_c})\right]. \tag{16.3}$$

With perfect capital mobility across sectors, the first-order conditions for the optimal quantities of capital are

$$R = f'(\frac{k_i}{zn_i}) = p_c \xi f'(\frac{k_c}{zn_c}). \tag{16.4}$$

Dividing (16.3) by (16.4) and then by z and writing the effective capital labor ratios $\frac{k_i}{zn_i}$ and $\frac{k_c}{zn_c}$ as x_i and x_c respectively,

$$\frac{w}{zR} = \frac{f(x_i) - x_i f'(x_i)}{f'(x_i)} = \frac{f(x_c) - x_c f'(x_c)}{f'(x_c)}. \tag{16.5}$$

Since, given $f'(x) > 0$ and $f''(x) < 0$, the function $\frac{f(x) - xf'(x)}{f'(x)}$ is monotonically increasing in x (the numerator is increasing in x and the denominator is decreasing in x), (16.5) implies that $x_i = x_c$. Thus, as noted in Chapter 19, the effective capital/labor ratio $\frac{k}{zn}$ must be the same for both sectors and $p_c \xi$ must be equal to 1. Moreover, the effective capital/labor ratio common to both sectors must be a function of the effective wage/rental rate ratio $woverzR$. If we write the inverse of the function $\frac{f(x) - xf'(x)}{f'(x)}$ as χ, then in both the consumption and investment goods sectors,

$$\frac{k}{zn} = \chi(\frac{w}{zR}). \tag{16.6}$$

16.1.1 Log-Linearizing Capital/Labor Substitution

Log-linearizing (16.6) around the steady-state effective wage/rental rate ratio $\omega^* = (\frac{w}{zR})^*$,

$$\begin{aligned} \check{k} - \check{z} - \check{n} &= \frac{\omega^* \chi'(\omega^*)}{\chi(\omega^*)}[\check{w} - \check{z} - \check{R}] \\ &= \sigma[\check{w} - \check{z} - \check{R}]. \end{aligned} \tag{16.7}$$

The local parameter $\sigma = \frac{\omega^* \chi'(\omega^*)}{\chi(\omega^*)}$ is the elasticity of capital/labor substitution in the neighborhood of the steady state. To see this more clearly, consider the fact that when the labor-augmenting technology happens to remain at its original level, so that $\check{z} = 0$, then (16.1.1) reduces to $\check{k} - \check{n} = \sigma[\check{w} - \check{R}]$. More generally, states that for a given level of labor-augmenting technology z, the elasticity of k/n with respect to w/R is equal to σ. Thus, σ must represent the elasticity of capital/labor substitution.

16.1.2 The Relationship Between the Elasticity of Capital-Labor Substitution σ and the Production Function f

Log-linearizing (16.5) directly, around the steady-state effective capital labor ratio $x^* = \left(\frac{k}{zn}\right)^*$, one finds that for both the investment and consumption goods sectors,

$$\begin{aligned}
\check{w} - \check{z} - \check{R} &= x^* \frac{d}{dx} \ln\left(\frac{f(x) - xf'(x)}{f'(x)}\right)\Bigg|_{x=x^*} [\check{k} - \check{z} - \check{n}] \qquad (16.8)\\
&= x^* \frac{d}{dx} [\ln(f(x) - xf'(x)) - \ln(f'(x))]\Big|_{x=x^*} [\check{k} - \check{z} - \check{n}]\\
&= x^* \left[\frac{-x^* f''(x^*)}{f(x) - xf'(x)} - \frac{f''(x^*)}{f'(x^*)}\right] [\check{k} - \check{z} - \check{n}]\\
&= \frac{-f''(x^*)x^*}{f'(x^*)} \left[\frac{x^* f'(x^*)}{f(x^*) - x^* f'(x^*)} + 1\right] [\check{k} - \check{z} - \check{n}]\\
&= \frac{-f''(x^*)x^*}{f'(x^*)} \left[\frac{f(x^*)}{f(x^*) - x^* f'(x^*)}\right] [\check{k} - \check{z} - \check{n}]\\
&= \frac{1}{\sigma} [\check{k} - \check{z} - \check{n}]
\end{aligned}$$

But by (16.3), (16.1) and (16.2), the ratio $\frac{f(x^*) - x^* f'(x^*)}{f(x^*)}$ is equal to labor's share (at least here where returns to scale are constant):

$$\frac{f(x^*) - x^* f'(x^*)}{f(x^*)} = \frac{w^* n_i^*}{i^*} = \frac{w^* n_c^*}{p_c^* c^*} = 1 - \theta. \qquad (16.9)$$

Thus, (16.8) implies that

$$\sigma = \frac{1 - \theta}{\left(\frac{-x^* f''(x^*)}{f'(x^*)}\right)} \qquad (16.10)$$

and

$$\frac{-x^* f''(x^*)}{f'(x^*)} = \frac{1-\theta}{\sigma}. \tag{16.11}$$

16.2 The Factor Price Possibility Frontier

In Chapter 21, we made good use of the Factor Price Possibility Frontier (FPPF) in analyzing the Basic RBC Model graphically. But we have yet to derive the Factor Price Possibility Frontier algebraically. That is the task of this section.

Since the wage and the rental rate are equalized across sectors, there is a single Factor Price Possibility Frontier for the whole economy. With investment goods as the numeraire (and $p_c = \xi$) the possible combinations of w and R are given parametrically in terms of the effective capital labor ratio $x = \frac{k}{zn}$ by

$$R = f'(x) \tag{16.12}$$

$$w = z[f(x) - xf'(x)]. \tag{16.13}$$

One key fact apparent from (16.13) and (16.12) is that an increase in the labor-augmenting technology z causes a proportionate [horizontal?] expansion of the Factor Price Possibility Frontier in the w direction. Figure **??** shows the effect of a doubling of z. [I am beginning to think the Factor Price Possibility Frontier should be drawn with the rental rate on the horizontal axis and the real wage on the vertical axis, which is opposite to how Figure 3 is drawn. There are two reasons. First, to match the convention that k is before n and R before w. Second, because for prices, vertical expansions are what we get in other graphs.]

A second key fact apparent from (16.13) and (16.12) is that (with investment goods as the numeraire), *only* the level of labor augmenting technology shifts the Factor Price Possibility Frontier. Changes in any other variable do not. Of course, movements in k and n (as well as z) induce movements *along* the Factor Price Possibility Frontier.

16.2.1 Log-Linearizing the Factor Price Possibility Frontier

In log-linearizing the Factor Price Possibility Frontier, we need both (16.9) and the complementary relationship for capital's share θ:

$$\frac{x^* f'(x^*)}{f(x^*)} = \frac{R^* k_i^*}{i^*} = \frac{R^* k_c^*}{p_c^* c^*} = \theta. \tag{16.14}$$

Together, (16.9) and (16.14) imply that

$$\frac{f(x^*) - x^* f'(x^*)}{f'(x^*)} = \frac{1-\theta}{\theta}. \tag{16.15}$$

With (16.14), (16.14) and (16.15) in hand, we can log- linearize (16.12) and eqnlabormkt w x as follows:

$$\begin{aligned} \check{R} &= x^* \frac{d}{dx} \ln(f'(x))\Big|_{x=x^*} [\check{k} - \check{z} - \check{n}] \\ &= \frac{x^* f''(x^*)}{f'(x^*)} [\check{k} - \check{z} - \check{n}] \\ &= \frac{1-\theta}{\sigma} [\check{z} + \check{n} - \check{k}] \end{aligned} \tag{16.16}$$

$$\begin{aligned} \check{w} &= \check{z} + x^* \frac{d}{dx} \ln(f(x) - x f'(x))\Big|_{x=x^*} [\check{k} - \check{z} - \check{n}] \\ &= \check{z} - \frac{(x^*)^2 f''(x^*)}{f(x^*) - x^* f'(x^*)} [\check{k} - \check{z} - \check{n}] \\ &= \check{z} - \frac{x^* f''(x^*)}{f'(x^*)} \frac{x^* f'(x^*)}{f(x^*) - x^* f'(x^*)} [\check{k} - \check{z} - \check{n}] \\ &= \check{z} + \frac{1-\theta}{\sigma} \frac{\theta}{1-\theta} [\check{k} - \check{z} - \check{n}] \\ &= \check{z} + \frac{\theta}{\sigma} [\check{k} - \check{z} - \check{n}]. \end{aligned} \tag{16.17}$$

Individually, equations (16.16) and (16.17) determine the real wage w and the rental rate R. Adding capital's share θ times equation (16.16) plus labor's share $1-\theta$ times equation (16.17) yields the equation

$$\theta \check{R} + (1-\theta)\check{w} = (1-\theta)\check{z} \tag{16.18}$$

for the Factor Price Possibility Frontier. Equivalently, solving for $\check{w}$, the Factor Price Possibility Frontier is given by

$$\check{w} = \check{z} - \frac{\theta}{1-\theta} \check{R} \tag{16.19}$$

Figure **??** shows the log-linearized view of the Factor Price Possibility Frontier. We have put $\ln(R)$ and $\ln(w)$ on the axes, but we could just as well have put $\check{R}$ and $\check{w}$ on the axes.

16.2.2 Calibrating the Factor Price Possibility Frontier

With $\theta = .3$ as above, the log-linearized equation for the factor price possibility frontier becomes

$$\check{w} = \check{z} - .43\check{R}.$$

16.3 The Log-Linearized Labor Demand Curve

Without the intermediate steps, (16.17) says that

$$\check{w} = \check{z} + \frac{\theta}{\sigma}[\check{k} - \check{z} - \check{n}].$$

In terms of the log-linearized view of the labor demand curve—with $\ln(n)$ and $\ln(w)$ on the axes shown in Figure **??**, (16.17) means that

1. The labor demand curve slopes down, with a logarithmic slope of $-\frac{\theta}{\sigma}$. This implies a labor demand elasticity of $\frac{\sigma}{\theta}$.

2. An increase in the capital stock shifts the labor demand curve to the right, with an elasticity of 1. (Alternatively, one can say that an increase in the capital stock shifts the labor demand curve up, with an elasticity of $\frac{\theta}{\sigma}$.)

3. An improvement in the labor-augmenting technology z has an ambiguous effect on labor demand. If $\sigma > \theta$ (as with a Cobb-Douglass production function for which $\sigma = 1$), then an increase in z shifts the labor demand curve out. On the other hand, if $\sigma < \theta$ (as with a Leontieff production function for which $\sigma = 0$), then an increase in z shifts the labor-demand curve back to the left! The shift in the labor demand curve can be viewed as a shift *out* with elasticity of $\frac{\sigma}{\theta} - 1$, a shift *up* with an elasticity of $1 - \frac{\theta}{\sigma}$, or in some ways most interestingly, as a shift up *and* back to the left, each with an elasticity of 1. Viewing the shift as a shift directly to the northwest of $(\ln(n), \ln(w))$- space may help clarify the relationship between the slope of the labor demand curve and the direction in which the labor demand curve shifts.

Let us discuss the economics of these results in question and answer format.

Why is the elasticity of labor demand for given k equal to $\frac{\sigma}{\theta}$?

In other words, how is the combination of a 1% fall in the real wage and a $\frac{\sigma}{\theta}$ % increase in labor n consistent with an unchanged capital stock and an unchanged output price? (The price of the investment good must remain unchanged at 1, since it is the numeraire, while the price of consumption must remain unchanged at $\frac{1}{\xi}$.) Here is how. First, as a counterfactual point of reference, if the wage/rental rate ratio were to stay unchanged, the rental rate would also fall by 1% along with the wage and the output price would do likewise. To avoid such a fall in the output price, the rental rate must move to $\frac{1}{\theta}$ % above the reference point of both wage and rental rate falling by 1%. With the real wage still down 1% in all, capital's share times the $\frac{1}{\theta}$ % increase in the rental rate brings the output price back up 1% to where it was. Thus, a 1% fall in the real wage must be associated with a $\frac{1}{\theta}$ % increase in the rental rate/wage ratio. By the definition of the elasticity of capital/labor substitution σ, this corresponds to a $\frac{\sigma}{\theta}$ % increase in n/k, which with k fixed, means a $\frac{\sigma}{\theta}$ % increase in n.

Why is the elasticity of the quantity of labor demanded with respect to the capital stock equal to 1?

With technology held fixed, a given real wage w implies a given rental rate R according to the Factor Price Possibility Frontier. A given real wage thus implies a constant real wage/rental rate ratio w/R, which in turn implies a constant capital/labor ratio k/n as a matter of firm factor demand. Therefore, for a given real wage and given technology, if the capital stock k increases by 1%, the quantity of labor n firms demand must increase by 1% as well.

Why is the effect of a labor-augmenting technology shock on labor demand ambiguous?

There are two opposing effects on labor demand: (a) the productivity effect, and (b) the displacement effect. The fact that each worker is now more productive tends to make a firm want to substitute (effective) labor for capital—the productivity effect. But the fact that the firm can get by with fewer workers in producing any given amount of output (that is, it takes fewer labor hours to get the same amount of effective labor) tends to make the firm want fewer workers—the displacement effect.

Another way to look at the consequences of an improvement in labor-augmenting technology is to focus on the relationship between the real wage

w and quantity of labor n on one hand and the *effective* real wage w/z and *effective* quantity of labor zn on the other. For a given real wage w, a 1% improvement in labor-augmenting technology lowers the *effective* real wage (*i.e.*, the marginal cost of effective labor input) by 1%. This 1% reduction in the *effective* real wage leads to a $\frac{\sigma}{\theta}$ % increase in the quantity of effective labor demanded just as a 1% fall in the real wage raises the quantity of labor demanded by $\frac{\sigma}{\theta}$ % when moving down *along* the labor demand curve. The increase in the quantity of effective labor demanded due to the fall in the effective real wage is the productivity effect. However, the $\frac{\sigma}{\theta}$ % increase in the quantity of effective labor zn is only a $\frac{\sigma}{\theta} - 1$ % increase in the quantity of labor n. The reduction of the effect by 1% is the displacement effect, as a given amount of effective labor gets translated into a smaller amount of actual labor hours n. Graphically, the productivity effect shifts the labor demand curve upward by 1% (as a higher real wage w yields the same *effective* real wage w/z as before); the displacement effect shifts the labor demand curve by 1% back to the left. Whether this amounts to an overall shift outward or an overall shift back depends on the slope of the labor demand curve.

What is the intuition behind the way in which the $\frac{\sigma}{\theta} - 1$ elasticity of the quantity of labor demanded with respect to z depends on the parameters σ and θ?

The productivity effect has the elasticity $\frac{\sigma}{\theta}$ since it is a combination of a substitution effect (substituting effective labor for capital) and a scale effect (increasing the scale of output) which together hold the quantity of capital used constant and lead to an increase in the rental rate which balances out the fall in the price of effective labor, so that the price of the good can remain unchanged. Because of the element of capital/labor substitution in the productivity effect, the size of the productivity effect is positively related to the elasticity of capital/labor substitution σ. (If the technology is Leontieff, making σ equal to zero, there is no productivity effect, since holding the effective capital/labor ratio $\frac{k}{zn}$ *and* the quantity of capital k fixed fixes the effective quantity of labor zn as well.) Because of the scale effect element of the productivity effect, the size of the productivity effect is inversely related to capital's share θ. The smaller capital's share θ, the more the rental rate can be drawn up by an increase in effective labor before the increase in the rental rate balances out the fall in the price of effective labor. By contrast, the displacement effect always has the same size of -1. A 1% improvement in labor-augmenting technology allows a firm to reduce labor hours by 1%

without any overall reduction in output. [Is this paragraph correct?]

16.3.1 Calibrating the Labor Demand Curve

With $\sigma = 1$ (Cobb-Douglas technology) and $\theta = .3$, the log-linearized equation for the constant-capital, constant-price labor demand curve is

$$\check{w} = .3(\check{k} - \check{n}) + .7\check{z}.$$

16.4 The Frisch Labor Supply Curve

In Chapter (**??**) we looked at labor supply for a given level of consumption and the corresponding consumption-constant labor supply elasticity. But in Chapter 1, the labor supply curve we used was labor supply for a given value of ν, the marginal utility of consumption. A marginal-utility-of-consumption-constant labor supply curve is called a Frisch labor supply curve.

In order to find the Frisch labor supply curve, let us go back to the King-Plosser-Rebelo form of the utility function:

$$u(c, n) = \frac{c^{1-\frac{1}{s}}}{1-\frac{1}{s}} e^{[\frac{1}{s}-1]v(n)}.$$

As discussed in Chapter 7, the King-Plosser-Rebelo utility function implies that the real wage in terms of consumption goods must be proportional to consumption:

$$\frac{w}{p_c} = cv'(n). \tag{16.20}$$

Equivalently, the real wage in terms of the investment goods numeraire is proportional to expenditure on consumption:

$$w = p_c c v'(n). \tag{16.21}$$

The marginal utility of consumption ν is

$$\nu = c^{\frac{1}{s}} e^{[\frac{1}{s}-1]v(n)}. \tag{16.22}$$

where s is the (labor-constant) intemporal elasticity of substitution for consumption. Solving for consumption c,

$$c = \nu^{-s} e^{(1-s)v(n)}. \tag{16.23}$$

Substituting this expression for consumption into (16.21) and using the fact that $p_c = \frac{1}{\xi}$,

$$\begin{aligned} w &= p_c \nu^{-s} v'(n) e^{(1-s)v(n)} \\ &= \frac{v'(n) e^{(1-s)v(n)}}{\xi \nu^s}. \end{aligned} \tag{16.24}$$

Functionally, (16.24) says that on the labor supply side the real wage (in terms of the investment goods numeraire) is positively related to the quantity of labor n, and negatively related to the marginal utility of consumption ν and the multiplicative consumption-goods-producing technology ξ. More specifically, for a given quantity of labor n, the supply price of labor is inversely proportional to the product $\xi\nu^s$:

$$w = w(n, \xi\nu^s).$$

Both $v'(n)$ and $v(n)$ are increasing, which makes it obvious that $v'(n)e^{(1-s)v(n)}$ is increasing in n as long as $s \leq 1$. The assumption of concavity of $u(c, n)$ guarantees that $v'(n)e^{(1-s)v(n)}$ is increasing in n in the case of $s > 1$.

Graphically, (16.24) implies that the Frisch labor supply curve (a) is upward-sloping (b) is shifted down and out by either an increase in the marginal utility of consumption ν or by an improvement in the multiplicative consumption-goods-producing technology ξ, and (c) is unmoved by anything other than ν and ξ.

16.4.1 Log-Linearizing the Frisch Labor Supply Curve

Taking logarithms, (16.24) becomes

$$\ln(w) = \ln(v'(n)) + (1-s)v(n) - s\ln(\nu) - \ln(\xi). \tag{16.26}$$

Linearizing around the steady-state in terms of the logarithmic deviations of w, n, ξ and ν (See Appendix A for hints on log-linearization), and using (16.21),

$$\begin{aligned} \check{w} &= \left[\frac{n^* v''(n^*)}{v'(n^*)} + (1-s)n^* v'(n^*)\right] \check{n} - s\check{\nu} - \check{\xi} \\ &= \left[\frac{1}{\eta} + (1-s)\frac{w^* n^*}{p_c^* c^*}\right] \check{n} - s\check{\nu} - \check{\xi} \\ &= \left[\frac{1}{\eta} + (1-s)\tau\right] \check{n} - s\check{\nu} - \check{\xi}, \end{aligned} \tag{16.27}$$

where

$$\eta = \frac{v'(n^*)}{n^* v''(n^*)}$$

is the consumption-constant labor supply elasticity and

$$\tau = \frac{w^* n^*}{p_c^* c^*}$$

is the ratio of labor earnings to consumption spending, as in Chapter 15.

16.4.2 Interpreting the Log-Linearized Frisch Labor Supply Curve

In terms of the log-linearized view of the Frisch labor supply curve—with $\ln(n)$ and $\ln(w)$ on the axes shown in Figure **??**, (16.27) means that

1. The labor demand curve slopes up, with a logarithmic slope of $\frac{1}{\eta} + (1-s)\tau$. This implies a labor supply elasticity of $\frac{\eta}{1+(1-s)\tau\eta}$. When $s < 1$, the Frisch labor supply curve slopes up more steeply than the consumption-constant labor supply curve because higher labor n leads to higher consumption c. (See (15.14).)

2. An increase in the marginal utility of consumption ν shifts the labor supply curve downward with an elasticity of s, which is equivalent to shifting the labor supply curve outward with an elasticity of $\frac{s\eta}{1+(1-s)\tau\eta}$.

3. An improvement in the multiplicative consumption-goods-producing technology ξ shifts the labor supply curve downward with an elasticity of 1, which is equivalent to shifting the labor supply curve outward with an elasticity of $\frac{\eta}{1+(1-s)\tau\eta}$.

As we did with labor demand, let us discuss the economics of these results in question and answer format.

Why is the Frisch labor supply elasticity lower than the consumption-constant labor supply elasticity when $s < 1$?

When the intertemporal elasticity of substitution in consumption is low—in particular, when it is below 1—then the substitution between consumption and leisure is relatively strong compared to intertemporal substitution. If substitution between consumption and leisure dominates, it implies complementarity between consumption and labor. Consumption and labor are a great taste and a bitter taste that taste better together. Thus, a higher real wage is needed to bring forth a greater quantity of labor supplied not only because of the less than infinite labor supply elasticity for a given level of consumption, but also because greater labor tends to lead to higher consumption.

How is the elasticity of labor supply with respect to the marginal utility of consumption ν related to the elasticity of labor supply with respect to the real interest rate?

Since

$$\ln(\nu_t) = \ln(\nu_\infty) + \int_t^\infty (r_{t'} - \frac{\dot{p}_c}{p_c} - \rho)\, dt'$$

the elasticity of anything with respect to ν is equal to the semielasticity of that thing with respect to the cumulative amount by with the real interest rate in terms of consumption goods will be above normal, when one holds fixed what the eventual marginal utility of consumption in the distant future will be.

Why does the elasticity of labor supply with respect to the marginal utility of consumption depend on the elasticity of intertemporal substitution s?

The higher is consumption, the higher the real wage that it will take to bring forth a given amount of labor. If the elasticity of intertemporal substitution s is greater, consumption and therefore the real wage necessary to bring forth a given amount of labor are more sensitive to the marginal utility of consumption ν.

Why does an improvement in consumption-goods-technology lead to a shift outward in labor supply?

Since households care directly only about the consumption wage rather than the wage in terms of investment goods, if the same wage in terms of the investment good numeraire translates into a higher consumption wage because of greater productivity in producing consumption goods, then that same wage in terms of the investment good numeraire will bring forth more labor.

16.4.3 Calibrating the Frisch Labor Supply Curve

It is quite uncertain what the appropriate value for η is, but let us look at what happens when $\eta = 1$. If we choose a consumption share of output ζ_c equal to .7 and capital's share θ equal to .3, then in the absence of distortionary taxation, consistency demands that

$$\tau = \frac{w^* n^*}{p_c^* c^*} = \frac{\frac{w^* n^*}{y^*}}{\frac{p_c^* c^*}{y^*}} = \frac{1-\theta}{\zeta_c}$$

which in this case would be $\frac{.7}{.7} = 1$. Finally, let us take $s = .5$, based on Basu and Kimball (). With $\eta = 1$, $\tau = 1$ and $s = .5$, the log-linearized Frisch labor supply curve becomes

$$\check{w} = 1.5\check{n} - .5\check{\nu} - \check{\xi}.$$

16.5 Labor Market Equilibrium

Labor demand and labor supply together determine labor market equilibrium. To sum up what we did just above, the log-linearized labor demand and labor supply equations (16.17) and (16.27) give

$$\check{w} = -\frac{\theta}{\sigma}\check{n} + \frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z}. \tag{16.28}$$

$$\check{w} = \left[\frac{1}{\eta} + (1-s)\tau\right]\check{n} - s\check{\nu} - \check{\xi}. \tag{16.29}$$

16.5.1 The Equilibrium Quantity of Labor

Equating the logarithmic deviation of the real wage $\check{w}$ as given by labor demand to $\check{w}$ as given by labor supply, then collecting terms, yields the equation

$$\left[\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau\right]\check{n} = \frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z} + s\check{\nu} + \check{\xi}. \tag{16.30}$$

The right-hand-side of (22.12) sets out all of the things that raise the rate of the wage on the labor demand curve to the wage on the labor supply curve, multiplied by the elasticities with which they change this ratio of demand wage to supply wage. The left-hand-side of (22.12) sets out the elasticity with which an increase in the quantity of labor reduces the ratio of the demand wage to the supply wage. This elasticity of the ratio of the demand wage to the supply wage is equal to the sum of the *in*elasticity of labor demand (holding the stock of capital and the price of output fixed) and the *in*elasticity of Frisch labor supply. Solving for $\check{n}$ shows the change in the quantity of labor needed to undo the change in the ratio of demand wage to supply wage induced by k, z, ν, and ξ:

$$\check{n} = \frac{\frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z} + s\check{\nu} + \check{\xi}}{\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau}. \tag{16.31}$$

Thus, k and z (which shift out the labor demand curve) and ν and ξ (which shift out the labor supply curve) all increase the equilibrium quantity of labor n.

For some purposes, it will be convenient to write

$$\alpha = \frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau.$$

Then

$$\check{n} = \frac{\theta}{\sigma\alpha}\check{k} + \frac{s}{\alpha}\check{\nu} + \left(1 - \frac{\theta}{\sigma}\right)\frac{\check{z}}{\alpha} + \frac{\check{\xi}}{\alpha}. \tag{16.32}$$

16.5.2 The Equilibrium Real Wage

Substituting from (16.32) into either the labor demand equation (22.7) or the labor supply equation (16.29) yields an equation for the equilibrium real wage in terms of the investment goods numeraire:

$$\check{w} = \frac{\frac{1}{\eta} + (1-s)\tau}{\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau}\left[\frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z}\right] - \frac{\frac{\theta}{\sigma}}{\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau}\left[s\check{\nu} + \check{\xi}\right] \tag{16.33}$$

Note that if we write ϵ_D for the elasticity of labor demand and ϵ_S for the Frisch elasticity of labor supply, so that

$$\frac{\theta}{\sigma} = \frac{1}{\epsilon_D}$$

$$\frac{1}{\eta} + (1-s)\tau = \frac{1}{\epsilon_S},$$

and

$$\alpha = \frac{1}{\epsilon_D} + \frac{1}{\epsilon_S},$$

then (16.33) can be written as

$$\begin{aligned}
\check{w} &= \frac{\frac{1}{\epsilon_S}}{\frac{1}{\epsilon_D} + \frac{1}{\epsilon_S}} \left[\frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z}\right] - \frac{\frac{1}{\epsilon_S}}{\frac{1}{\epsilon_D} + \frac{1}{\epsilon_S}} \left[s\check{\nu} + \check{\xi}\right] && (16.34) \\
&= \frac{\epsilon_D}{\epsilon_D + \epsilon_S} \left[\frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z}\right] - \frac{\epsilon_S}{\epsilon_D + \epsilon_S} \left[s\check{\nu} + \check{\xi}\right]. && (16.35)
\end{aligned}$$

16.6 Expressing Labor Market Equilibrium in Terms of the Marginal Value of Investment

When differential rates of technological progress in the consumption goods and investment goods industries create a distinction between the marginal utility of consumption ν and the marginal value of investment μ, it is clearest to think of equilibrium between investment demand and saving supply as determining the marginal value of investment μ, particularly since it is the rate of change in μ which gives the real interest rate expressed in terms of the investment goods numeraire:

$$r = \rho - \frac{\dot{\mu}}{\mu}. \qquad (16.36)$$

Since one unit of investment trades off with ξ units of consumption, the marginal value of investment must be ξ times the marginal utility of consumption:

$$\mu = \xi\nu. \qquad (16.37)$$

Thus, given μ, the marginal utility of consumption ν is

$$\nu = \frac{\mu}{\xi}. \qquad (16.38)$$

16.6.1 Expressing the Frisch Labor Supply Curve in Terms of the Marginal Value of Investment

Substituting the expression in for ν into (16.24) yields

$$w = \frac{v'(n)e^{(1-s)v(n)}}{\mu^s \xi^{1-s}}. \tag{16.39}$$

Most interesting is what happens in the important special case when the elasticity of intertemporal substitution s is equal to 1 (logarithmic utility). In that case, given the marginal value of investment μ, the consumption goods technology ξ has no effect on the labor supply curve, because the the direct positive effect of the consumption goods technology ξ on labor supply due to the motivating consumption goods being easier to come by is cancelled out by the fall in the marginal utility of consumption ν for any given value of μ. If it does not affect the labor supply curve, a change in ξ causes no disturbance at all to the labor market; therefore, when $s = 1$, n and w do not depend on ξ.

In the more likely case of $s < 1$, an improvement in the consumption technology ξ still causes the labor supply curve for a given value of μ the shift out and down since the direct positive effect of the consumption goods technology on labor supply overwhelms the effect of the fall in the marginal utility consumption for given μ.

In log-linear form, the labor supply curve in terms of the marginal value of investment $\mu = \xi\nu$ (rather than the marginal utility of consumption ν) is

$$\check{w} = \left[\frac{1}{\eta} + (1-s)\tau\right]\check{n} - s\check{\mu} - (1-s)\check{\xi}. \tag{16.40}$$

16.6.2 Equilibrium n and w

It is easy to calculate the counterparts to (16.32) and (16.33) in terms of the marginal value of investment μ:

$$\check{n} = \frac{\theta}{\sigma\alpha}\check{k} + \frac{s}{\alpha}\check{\mu} + \left(1 - \frac{\theta}{\sigma}\right)\frac{\check{z}}{\alpha} + \frac{(1-s)}{\alpha}\check{\xi} \tag{16.41}$$

$$\check{w} = \frac{\frac{1}{\eta} + (1-s)\tau}{\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau}\left[\frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z}\right] - \frac{\frac{\theta}{\sigma}}{\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau}\left[s\check{\mu} + (1-s)\check{\xi}\right]. \tag{16.42}$$

The most interesting thing to notice here is that when $s = 1$, ξ has no effect on labor market equilibrium. Furthermore, when $s \neq 1$, and

$$\frac{s-1}{s} = 1 - \frac{1}{s} = 1 - \beta$$

times as great logarithmic movement in the marginal value of investment μ will cancel out the effect of ξ on labor market equilibrium.

16.7 Calibrating Equilibrium in the Labor Market

With $\theta = .3$, $\sigma = 1$, $\eta = 1$, $\tau = 1$ and $s = .5$, one can calculate that $\alpha = 1.8$, $\epsilon_D = 3.333$, and $\epsilon = .667$. In terms of the marginal value of investment μ, the equilibrium quantity of labor and the real wage (in terms of the investment goods numeraire) are

$$\check{n} = .167\check{k} + .278\check{\mu} + .389\check{z} + .278\check{\xi}$$

$$\check{w} = .250\check{k} - .083\check{\mu} + .583\check{z} - .083\check{\xi}.$$

Real Business Cycle Theory: A Semiparametric Approach, by Miles S. Kimball

September 7, 2004 at 16:22

Chapter 17

Decentralization and Distortions

17.1 Introduction

The second theorem of welfare economics states that, in the absence of distortions or complications, the solution to a social planner's problem can be decentralized as a competitive equilibrium—or equivalently, that competitive equilibrium will represent the solution to some social planner's problem with appropriate welfare weights. In this chapter, we will illustrate both the undistorted case in which competitive equilibrium represents the solution to a social planner's problem and a case in which taxation distorts economic decisions so that competitive equilibrium is no longer Pareto optimal and no longer represents the solution to a social planner's problem. First, we will start with a social planner's problem, decentralize it, and show the equivalence. Then we will add taxes to the decentralized problem and show how its equilibrium differs from the solution to the social planner's problem.

17.2 A Social Planner's Problem with Endogenous Labor Supply

To allow for a more interesting array of taxes later on, let's start with a model that allows for endogenous labor supply. For now, we can represent felicity (instantaneous utility) as a general concave function $u(c, n)$ of consumption c and labor n. Of course, in the relevant range, $u_c(c, n) > 0$ and $u_n(c, n) < 0$, since people like consumption and either dislike work altogether or (due to the positive wage) work more than what they would be willing to work for free.

The social planner's problem is

$$\max_{c,n} \quad \int_0^\infty e^{-\rho t} u(c,n)\, dt \tag{17.1}$$

$$\text{subject to} \qquad \dot{k} = F(k,n) - \delta k - c - g. \tag{17.2}$$

where $F(k,n)$ is a constant returns to scale production function,

$$F(k,n) = nf(k/n),$$

and g is "government purchases." More precisely, g is the separable part of government purchases. We are implicitly assuming that government purchases can be divided into three components; if there are other types of government spending that interact with the private economy in other ways, a different model is needed. (1) Government purchases that are a perfect substitute for private consumption can be included in c. (2) Government purchases that are a perfect substitute for private investment can be included in investment, which augments the capital stock. (3) Government purchases that enter into the utility function in an additively separable way can be modeled by g above. They do not affect private decisions directly.[1] However, these separable government purchases do have an effect on the private economy through their financing, as we will see. Military spending is probably the most important type of government purchases that can be modeled as separable in this way without too much violence to the facts.

Using the phrase "government purchases" rather than "government spending" is a reminder that transfer payments, which are part of "government spending," should be treated as negative taxes. It is helpful to consider transfer payments part of the system of taxation not only in terms of the flow of funds but also in calculating effective marginal tax rates.

The current-value Hamiltonian for (17.1) is

$$H = u(c,n) + \lambda[F(k,n) - \delta k - c - g].$$

The first order conditions are

[1] See Exercise (1).

$$\begin{aligned} H_c &= u_c(c,n) - \lambda = 0 \\ H_n &= u_n(c,n) + \lambda F_n(k,n) = 0, \end{aligned}$$

and the Euler equation is

$$\dot{\lambda} = \rho\lambda - H_k = [\rho + \delta - F_k(k,n)]\lambda.$$

Thus, the marginal utility of consumption equals the marginal value of capital,

$$u_c(c,n) = \lambda, \tag{17.3}$$

and the marginal disutility of labor is equal to the marginal product of labor times the marginal value of capital:

$$-u_n(c,n) = F_n(k,n)\lambda = [f(k/n) - (k/n)f'(k/n)]\lambda. \tag{17.4}$$

The Euler equation can be simplified to

$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - F_k(k,n) = \rho + \delta - f'(k/n). \tag{17.5}$$

Equations (17.3), (17.4), (17.2) and (17.5) are the key equations describing the social planner's program. These four equations correspond to the four variables c, n, k and λ. We need to pay attention to what happens to these equations (and to (**??**)) when we decentralize the social planner's program and then in Section (17.4.4) add taxes to the mix.

17.3 Decentralizing the Social Planner's Program

In order to decentralize the social planner's program, we can break it up into three pieces—the representative household's problem, the representative firm's problem and the government budget constraint. (Even though the government's actions will be taken as exogenous, we can think about its budget constraint.) Then we need to show that the result is equivalent to the social planner's program.

Part of showing that the decentralized equilibrium is equivalent to the social planner's program is to show that we do not need to make any explicit use of the household, firm and government budget constraints to describe the behavior of the key quantities (and prices)—that to study the behavior of c, n, k, λ and anything determined by these four variables, everything we need to know about budget constraints is contained in the physical accumulation equation (17.2). We will actually do more before the end of the chapter. One of the most important lessons of this chapter is that even when taxes distort the economy, the budget constraints for the separate actors in the economy are not needed to study aggregate behavior. (The government budget constraint is needed in order to determine if the government can get by without lump sum taxes, but that is all.)

17.3.1 The Representative Household

The household's problem is

$$\max_{c,n} \quad \int_0^\infty e^{-\rho t} u(c,n)\, dt \tag{17.6}$$

$$\text{subject to} \qquad \dot{a} = ra + wn - c - \tau, \tag{17.7}$$

where a is the net assets of the household, r is the real interest rate, w is the real wage and τ is the lump-sum tax. Every variable depends on time. There are no pure profits to be distributed to the household because of constant returns to scale, but if there were, they would enter in the same way as the lump-sum tax, except in the positive direction. However, because the representative household owns the representative firm, a includes the value of the representative firm. (The net worth of the represenative firm is zero if it were entirely financed by bonds and k if it were entirely financed by equity.)

The current-value Hamiltonian for the household's problem is

$$H = u(c,n) + \bar{\lambda}[ra + wn - c - \tau].$$

In naming the marginal value of assets $\bar{\lambda}$, we are anticipating the result that the equivalence between the decentralized equilibrium and the social planner's program holds *with the marginal value of assets equal to the marginal value of capital*, so that $\bar{\lambda} = \lambda$.

The first order conditions can be written as

$$u_c = \bar{\lambda} \tag{17.8}$$

$$-u_n = w\bar{\lambda}. \tag{17.9}$$

The Euler equation is

$$\dot{\bar{\lambda}} = \rho\bar{\lambda} - H_a = (\rho - r)\bar{\lambda},$$

or

$$\frac{\dot{\bar{\lambda}}}{\bar{\lambda}} = \rho - r. \tag{17.10}$$

17.3.2 The Representative Firm

The firm's problem is

$$\max_{n,i} \quad \int_0^\infty e^{-\int_0^t r\,dt'}[F(k,n) - wn - i]\,dt \tag{17.11}$$

$$\text{subject to} \qquad \dot{k} = i - \delta k, \tag{17.12}$$

where i is investment and t' is just a dummy variable for time. Here, the objective function is measured in dollars rather than in utils, and the discounting is at the real interest rate rather than at the utility discount rate. The current-value Hamiltonian based on the discounting by the real interest rate is

$$H = F(k,n) - wn - i + q(i - \delta k)$$

where q is Tobin's q: the marginal value of capital in terms of goods. The first order conditions for n and i are

$$\begin{aligned} H_n &= F_n - w = 0 \\ H_i &= -1 + q = 0, \end{aligned}$$

or

$$F_n = w \tag{17.13}$$

$$q = 1. \tag{17.14}$$

The Euler equation is

$$\dot{q} = rq - H_k = (r + \delta)q - F_k.$$

Substituting in $q = 1$ and $\dot{q} = 0$, as implied by (17.14) the firm's Euler equation implies that

$$F_k = r + \delta. \tag{17.15}$$

17.3.3 The Material Balance Condition, the Government Budget Constraint, and the Asset Balance Condition

Equilibrium between aggregate supply and aggregate demand implies the material balance condition

$$F(k, n) = c + i + g. \tag{17.16}$$

As for the financing of government purchases, the government debt b evolves according to

$$\dot{b} = rb + g - \tau. \tag{17.17}$$

In accordance with Walras' law in asset markets, (17.17) is redundant if one imposes the asset balance equation that private assets must equal the sum of government debt and the value of the capital stock:

$$a = b + qk. \tag{17.18}$$

In particular, substituting in $q = 1$ ((17.14)) and differentiating (17.18), then combining it with (17.7), (17.12), (17.16), (17.15) and the fact that (since F has constant returns to scale) $kF_k + nF_n = F$ yields (17.17).

The transversality condition and asymptotic budget constraint for the household (which have not been otherwise used), together with (17.18) insures that the government is unable to finance its spending with a bubble in this economy.

17.3.4 The Equivalence of the Decentralized Economy and the Social Planner's Program

The decentralized economy has ten variables: four corresponding to those that appear in the social planner's program—c, n, k and $\bar{\lambda}$—plus six more—a, r, w, i, q and b. The numbered equations (17.7)–(17.17) form a set of ten independent equations to determine these ten variables. To show that the decentralized economy follows the social planner's program, we need to show that these ten equations determine the same behavior for c, n, and k as the four social planner's equations (17.3), (17.4), (**??**) and (17.5).

If the implied behavior of c, n and k is the same, (17.3) guarantees that the implied behavior of λ will also be the same as in the social planner's program. Since $\bar{\lambda}$ is also equal to the marginal utility of consumption ((17.8)), equivalence between the decentralized economy and the social planner's program necessarily involves $\bar{\lambda} = \lambda$. Showing that the ten equations of the decentralized economy satisfy the other three of the social planner's equations with $\bar{\lambda} = \lambda$ demonstrates that identifying λ and $\bar{\lambda}$ is consistent.

With $\bar{\lambda} = \lambda$, the household's first-order condition for labor (17.9) and the firm's marginal revenue product of labor condition (17.13) together imply the social planner's first-order condition for labor: (17.4). The household's Euler equation (17.10) and (17.15)—an equation derived from the firm's Euler equation and the $q = 1$ condition—together imply the social planner's Euler equation. Finally, the firm's accumulation equation (17.12) and the material balance condition (17.16) together imply the social planner's accumulation equation (17.2). Thus, the decentralized economy follows the social planner's program.

One fact stands out: the household's budget constraint (17.7) and the government's budget constraint (17.17) (or alternatively, (17.18)) are not needed in order to show that the economy follows the social planner's program! More generally, the household and government budget constraints are not needed in order to calculate what the economy would do in the aggregate. They add two additional variables, a and b, that are not needed in order to determine the other eight variables. (This is an example of what Sargent calls "block recursiveness.")

17.4 A Decentralized Economy with Tax Distortions

This section will follow exactly the pattern of Section (17.3.4), but with distortionary taxation injected into the decentralized economy. We will not try to deal here with the almost endless variety of different types of taxes that exist, but only with four simple kinds of taxes, a lump sum tax, and three proportional taxes: a consumption tax τ_c, an employment tax τ_n and an investment tax τ_i. (In this framework, consumption, employment and investment subsidies can be represented by negative tax rates.) The effects of many other, more complex kinds of taxes can be represented by some combination of these. For example, one can often construct an index of the effective rate of taxation of investment based on many elements of the tax structure and use the resulting value for τ_i.

To simplify the accounting, we will have each of these taxes paid by the buyer on top of the pre-tax price or wage. (Here again, the actual variety is great. For example, for accounting purposes, the social security tax in the U.S. is an employment tax is paid half by the worker who sells labor services and half by the employer who buys labor services.)

We will only look here at the qualitative effects of distortionary taxes on the economy, so the three proportional tax rates will be treated as exogenous, with the lump-sum tax adjusting to satisfy the government budget constraint given the desired path of government spending. All of the tax rates can vary over time.

17.4.1 The Taxed Household

The household is a buyer only of consumption goods. Incorporating the tax rate on consumption goods τ_c into the household's problem,

$$\max_{c,n} \quad \int_0^\infty e^{-\rho t} u(c,n)\, dt \tag{17.19}$$

$$\text{subject to} \qquad \dot{a} = ra + wn - (1+\tau_c)c - \tau. \tag{17.20}$$

where, as before, a is the net assets of the household, r is the real interest rate, w is the real wage and τ is the lump-sum tax. The current-value Hamiltonian for the household's problem is

$$H = u(c,n) + \bar{\lambda}[ra + wn - (1+\tau_c)c - \tau].$$

The first order conditions can be written as

$$u_c = (1+\tau_c)\bar{\lambda} \tag{17.21}$$

$$-u_n = w\bar{\lambda}. \tag{17.22}$$

Since there is no direct interest taxation, the household's Euler equation is still

$$\frac{\dot{\bar{\lambda}}}{\bar{\lambda}} = \rho - r. \tag{17.23}$$

17.4.2 The Taxed Firm

The firm must pay both employment and investment taxes. Incorporating these taxes, the firm's problem is

$$\max_{n,i} \int_0^\infty e^{-\int_0^t r\,dt'}[F(k,n) - (1+\tau_n)wn - (1+\tau_i)i]\,dt \tag{17.24}$$

$$\text{subject to} \qquad \dot{k} = i - \delta k. \tag{17.25}$$

The current-value Hamiltonian, discounted by the real interest rate, is

$$H = F(k,n) - (1+\tau_n)wn - (1+\tau_i)i + q(i - \delta k)$$

where q is now a tax-adjusted Tobin's q. The first order conditions for n and i can be written as

$$F_n = (1+\tau_n)w \tag{17.26}$$

$$q = 1+\tau_i. \tag{17.27}$$

The Euler equation is still

$$\dot{q} = (r+\delta)q - F_k.$$

Substituting in $q = 1+\tau_i$ and $\dot{q} = \dot{\tau}_i$, yields the counterpart to equation (17.15):

$$F_k = (r+\delta)(1+\tau_i) - \dot{\tau}_i. \tag{17.28}$$

17.4.3 The Government Budget Constraint

There is no change in the form of the material balance condition $F(k,n) = c + i + g$ or the asset balance condition $a = b + qk$, though the q in the asset balance condition is now more interesting. Including the revenue from the three proportional taxes, the government budget constraint is

$$\dot{b} = rb + g - \tau_c c - \tau_n wn - \tau_i i - \tau. \tag{17.29}$$

Just as before, Walras' Law for assets makes either the asset balance condition or the government budget constraint redundant.

17.4.4 General Equilibrium in the Distorted Economy

As in the absence of taxes, there are ten endogenous variables and ten independent equations. The work of this section is to construct counterparts to the four social planner's equations in order to show how the aggregate behavior of the distorted economy differs from the social planner's program followed by the undistorted economy.

As before, the firm's accumulation equation (17.25) and the material balance condition (**??**) together imply

$$\dot{k} = F(k, n) - \delta k - c - g, \tag{17.30}$$

which is identical to the social planner's accumulation equation (17.2). The fact that this equation remains undistorted by the taxes should not be too surprising, since it is based solely on facts about the technology together with an assumption of technical efficiency and so should be true regardless of how allocations in the economy are determined.

The social planner's other equations all involve the marginal value of capital λ in one way or another. In order to get as close as possible a counterpart to the social planner's marginal value of capital, it is helpful to look at the representative household's marginal value of capital and denote *that* as λ. In view of the asset balance condition $a = b + qk$, the value of one additional unit of capital to the household should be equal to the value of q additional units of assets. Therefore, define

$$\lambda = q\bar{\lambda} = (1 + \tau_i)\bar{\lambda}. \tag{17.31}$$

In combination with (17.21), (17.31) yields the following counterpart to the social planner's marginal utility of consumption condition (17.3):

$$u_c = \frac{1 + \tau_c}{1 + \tau_i}\lambda. \tag{17.32}$$

Similarly, (17.31), the household's first-order condition for labor (17.22) and the firm's marginal revenue product of labor condition (17.26) together imply

$$-u_n = \frac{F_n \lambda}{(1 + \tau_n)(1 + \tau_i)}, \tag{17.33}$$

which reduces to the social planner's first-order condition for labor (17.4) when the proportional tax rates are set to zero.

Finally, to get the counterpart to the social planner's Euler equation, first solve (17.28) for r—

$$r = -\delta + \frac{F_k}{1+\tau_i} + \frac{\dot{\tau}_i}{1+\tau}.$$

—and substitute this expression for r into the household's Euler equation (17.23) to get

$$\frac{\dot{\bar{\lambda}}}{\bar{\lambda}} = \rho + \delta - \frac{F_k}{1+\tau_i} - \frac{\dot{\tau}_i}{1+\tau}.$$

Since from (17.31), $\frac{\dot{\bar{\lambda}}}{\bar{\lambda}} = \frac{\dot{\lambda}}{\lambda} - \frac{\dot{\tau}_i}{1+\tau}$, rewriting this equation in terms of the household's marginal value of capital λ simplifies it to the following distorted version of the social planner's Euler equation (17.5):

$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - \frac{F_k}{1+\tau_i}. \tag{17.34}$$

The household and government budget constraints are again unnecessary in constructing the four equations (17.30), (17.32), (17.33) and (17.34) describing aggregate behavior. As before, these budget constraints add two additional variables, a and b, that are not needed in order to determine the other variables.

Exercises

1. Show that *holding fixed the government's choice of the path of government purchases g*, putting g in the utility function in an additively separable way does not affect competitive equilibrium whether or not there are tax distortions.

Real Business Cycle Theory: A Semiparametric Approach, by Miles S. Kimball

September 7, 2004 at 16:22

Chapter 18

Building on the Golden Rule

18.1 The Golden Rule

The central method of this book is dynamic and stochastic optimization. The simplest kind of dynamic optimization is the choice of an optimal steady state in a dynamic system, as represented by the Golden Rule familiar from elementary growth theory. The Golden Rule itself corresponds to choosing the the optimal steady state when there is no discounting of future utility.

18.1.1 Solving Out for Consumption

One simple way to find the Golden Rule is to solve the steady-state condition

$$\dot{k} = f(k) - \delta k - c = 0$$

for c, yielding

$$c = f(k) - \delta k$$

and then maximizing steady-state consumption c by solving

$$\max_k f(k) - \delta k,$$

yielding the following first-order condition for the optimal capital stock:

$$f'(k^*) - \delta = 0.$$

In words, the net marginal product of capital must be zero. Of course, in this model, the net marginal product of capital equals the real interest rate r, so one can state the Golden Rule condition as the real interest rate being zero.

18.1.2 Using Constrained Optimization

By using the techniques of constrained optimization rather than solving out for consumption, we can take a big step in the direction of the techniques taught in this book. Another step in that direction is to maximize the utility of consumption rather than consumption itself. Because utility is a monotically increasing function of consumption here, it amounts to the same thing, but below we will make utility a function of both consumption and labor. With these two changes in the angle of approach, the Golden Rule in this simple growth model can be found from the maximization problem

$$\max_{c,k} u(c)$$

s.t.,

$$\dot{k} = f(k) - \delta k - c = 0.$$

The Lagrangian for this problem—which we will denote H because it is equal to the Hamiltonian that we will mention later—is

$$H = u(c) + \lambda[f(k) - \delta k - c].$$

The Lagrange multiplier λ is the marginal value of capital accumulation—or in the steady state, one might say that λ is the marginal value of capital *maintenance*. Hereafter, we will shorten the description of λ to "the marginal value of capital."

The standard first-order conditions for c and k are

$$u'(c^*) = \lambda^*.$$

and

$$f'(k^*) - \delta = 0.$$

In words, the marginal utility of consumption must equal the marginal value of capital, and—as before—the net marginal product of capital must equal zero.

18.2 The Modified Golden Rule

We can take another step in the direction of true dynamic optimization by allowing for impatience. The proof relies on later results, but it turns out that there is a very easy way to allow for impatience in looking for the optimal steady state. Think of impatience as measured by the utility discount rate ρ as imposing a required rate of return ρ on capital in the steady state. Then the optimal steady state can be found by maximizing what we will call the "steady-state classic Lagrangian" L:

$$L = H - \rho\lambda k = u(c) + \lambda[f(k) - \delta k - c - \rho k].$$

The first order conditions for maximizing the classical Lagrangian with respect to c and k are

$$u'(c^*) = \lambda^*$$

and

$$f'(k^*) - \delta = \rho.$$

Again, the marginal utility of consumption is equal to the marginal value of capital. But now, the net marginal product of capital must be equal to ρ, or in other words, the real interest rate must equal the utility discount rate instead being set equal to zero. This condition on the real interest rate is known as the *Modified Golden Rule.*

As before, to actually find the Golden Rule steady state, it is important to impose the steady state condition

$$\dot{k} = f(k^*) - \delta k^* - c^* = 0.$$

18.3 Generalizing the Modified Golden Rule

The Modified Golden Rule can be generalized to be of use in finding the optimal steady state in a wide variety of models. Let U be the ultimate objective or utility function, and let A be the capital accumulation function giving $\dot{k}$:

$$\dot{k} = A.$$

Then the general form of the Hamiltonian is

$$H = U + \lambda A,$$

while the general form of the steady-state classic Lagrangian is

$$L = U - \rho\lambda k + \lambda A = U + \lambda[A - \rho k].$$

Properly speaking, the utility discount rate ρ should be seen as modifying the objective function U, not the accumulation function A.

Chapter 19

Steady States and Growth Steady States

19.1 Introduction

Given a model in which the objective function, accumulation equations and constraints of a control problem are all independent of time, except for discounting at a constant rate, a steady state is a set of values for the state variables and costate variables that—in conjunction with the values of the control variables that maximize the Hamiltonian at that point—imply motionlessness for all of these variables. Steady states are important because the optimal program often approaches a steady state as a limit as time goes on.

The concept of a steady state is quite important even in the analysis of models that have no steady state. First, it is often possible to detrend a model with certain kinds of trends in its equations to obtain a detrended model that has a steady state. The steady state of the detrended model corresponds to a steady growth path or "growth steady state" of the original model. Second, even when the equations of a model vary in a more complex way, if the variations from equations that would allow a steady state or growth steady state to exist are not too large one can often look at the behavior of a model very effectively by linearizing a model around the steady state. We must put off linearizing around the steady state, though. Here we will concentrate on steady states and growth steady states themselves.

19.2 The Steady State of the Ramsey-Cass Model

The key equations of the Ramsey-Cass model are

$$u'(c) = \lambda$$
$$\dot{k} = f(k) - \delta k - c$$
$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - f'(k).$$

(See Section **??**.) Using an asterisk to denote the steady-state value of each variables, one can obtain the equations for the steady state by replacing $\dot{k}$ and $\dot{\lambda}$ by 0. That is, a steady state (k^*, λ^*) with its associated value of c^* must satisfy

$$u'(c^*) = \lambda^* \quad (19.1)$$
$$0 = f(k^*) - \delta k^* - c^* \quad (19.2)$$
$$0 = \rho + \delta - f'(k^*). \quad (19.3)$$

In this case, the steady state conditions can be solved one by one, recursively. First, as long as $f(\cdot)$ is strictly concave, $f'(\cdot)$ will be a monotonically decreasing function and $f'(k^*) = \rho + \delta$ (19.3) has a unique solution for the steady-state capital stock:

$$k^* = f'^{-1}(\rho + \delta).$$

Then steady-state consumption c^* can be determined from (19.2):

$$c^* = f(k^*) - \delta k^*.$$

Finally, given c^* the steady-state marginal value of capital λ^* can be determined from (19.1) in the obvious way.

19.3 The Steady State of the Basic Real Business Cycle Model

19.3.1 Laying Out the Basic Real Business Cycle Model

By the "basic real business cycle model" we mean the social planner's problem with endogenous labor supply that we studied in Chapter 4, augmented by including explicit technology parameters and graced by a particular form of the utility function. The particular form of the utility function is

$$u(c,n) = \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}, \tag{19.4}$$

where $\beta > 0$, $v'(\cdot) > 0$, and $v''(\cdot) > 0$ and u is assumed concave in both arguments (which may implicitly place a stricter condition on β than just being positive[1]). This form of the utility function is called a King-Plosser-Rebelo utility function. Its virtues will become more apparent in the next section when we look at growth steady states. After subtracting $\frac{1}{1-\beta}$ (which does not alter the preferences) the limit as $\beta \to 1$ is

$$u(c,n) = \ln(c) - v(n),$$

which is the most familiar special case of the King-Plosser-Rebelo utility function.

The basic real business cycle model has two technology parameters: z, the level of labor-augmenting technology and ξ, the level of additional output-augmenting (multiplicative, Hicks-neutral) technology in the production of (nondurable) consumption goods. Formally, if k_c and n_c are control variables giving the amount of capital and labor used in the production of (nondurable) consumption goods,

$$\begin{aligned} c &= \xi[zn_c f\left(\frac{k_c}{zn_c}\right)] \\ i+g &= z(n-n_c) f\left(\frac{(k-k_c)}{z(n-n_c)}\right). \end{aligned}$$

The answer to Exercise **??** shows that the production possibility frontier for c and $i+g$ for given k and n is equivalent to the single constraint

$$znf\left(\frac{k}{zn}\right) = \frac{c}{\xi} + i + g. \tag{19.5}$$

Therefore, the basic real business cycle model with the two technology parameters z and ξ can be written

$$\max_{c,n,i} \int_0^\infty e^{-\rho t} \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}\, dt \tag{19.6}$$

[1]See Exercise (**??**).

$$\text{subject to} \qquad \dot{k} = i - \delta k \tag{19.7}$$
$$znf\left(\frac{k}{zn}\right) = \frac{c}{\xi} + i + g. \tag{19.8}$$

19.3.2 The Canonical Equations

The (augmented) current-value Hamiltonian for the Basic Real Business Cycle Model is

$$\begin{aligned} H &= \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)} + \lambda[i - \delta k] \\ &\quad + \mu[znf(k/zn) - \frac{c}{\xi} - i - g], \end{aligned}$$

with λ the marginal value of capital and μ the marginal value of investment. (Following a convention that, unless indicated otherwise by parentheses, multiplication takes place before division allows us to write $f(k/zn)$ for $f\left(\frac{k}{zn}\right)$.) The first order conditions can be written as

$$c^{-\beta} e^{(\beta-1)v(n)} = \frac{\mu}{\xi} \tag{19.9}$$
$$v'(n)c^{1-\beta} e^{(\beta-1)v(n)} = z[f(k/zn) - (k/zn)f'(k/zn)]\mu \tag{19.10}$$
$$\mu = \lambda. \tag{19.11}$$

In addition, dividing (19.10) by (19.9) yields the important equation relating the ratio of marginal utilities to the marginal product of labor (in terms of consumption goods):

$$-\frac{u_n(c,n)}{u_c(c,n)} = cv'(n) = \xi z[f(k/zn) - (k/zn)f'(k/zn)] \tag{19.12}$$

The Euler equation is

$$\dot{\lambda} = \rho\lambda - H_k = (\rho + \delta)\lambda - f'(k/zn)\mu,$$

or, in view of (19.11),

$$\frac{\dot{\lambda}}{\lambda} = \rho + \delta - f'(k/zn). \tag{19.13}$$

19.3.3 The Income Expansion Path and the Steady-State Material Balance Condition

In a steady state, the accumulation equation (19.7) and Euler equation (19.13) come down to

$$\begin{aligned} i^* &= \delta k^* &(19.14)\\ f'(k^*/zn^*) &= \rho + \delta. &(19.15) \end{aligned}$$

The other equations of the model remain unaltered, except for added asterisks. The steady-state version of the Euler equation, (19.15), can be solved for the steady-state capital/effective labor ratio $\frac{k^*}{zn^*}$, which we will label κ^*:

$$\kappa^* = \frac{k^*}{zn^*} = f'^{-1}(\rho + \delta). \qquad (19.16)$$

This value for the steady-state capital/effective labor ratio gives us the value of the steady-state real consumption wage on the right-hand side of (19.12). That is, in steady state,

$$c^* v'(n^*) = \xi z[f(\kappa^*) - \kappa^* f'(\kappa^*)]. \qquad (19.17)$$

Also, the steady-state version of the accumulation equation (19.14), in conjunction with the material balance condition (19.8) and the definition of κ^*, yields the equation

$$\begin{aligned} zn^* f(\kappa^*) &= \frac{c^*}{\xi} + i^* + g\\ &= \frac{c^*}{\xi} + \delta k^* + g\\ &= \frac{c^*}{\xi} + \delta\kappa^* zn^* + g, \end{aligned}$$

or

$$\frac{c^*}{\xi} = zn^*[f(\kappa^*) - \delta\kappa^*] - g. \qquad (19.18)$$

Equations (19.17) and (19.18) provide two equations in two unknowns: c^* and n^*. Figure **??** graphs these two equations in the (n, c) plane. Equation (19.17) is the income expansion path for consumption and labor at a given wage ($\xi z[f(\kappa^*) - \kappa^* f'(\kappa^*)]$) as non-wage income varies. It slopes backward because higher (non-wage) income is associated with more consumption and less labor. Equation (19.18) is the steady-state material balance condition. It is upward sloping because greater quantity of labor is associated with more net output and therefore more consumption. Its upward slope is greater than just the marginal product of labor because, in steady state, a greater quantity of labor is associated with a greater amount of capital, as well.

19.3.4 The Steady State

The intersection between the income-expansion path at the steady-state real consumption wage and the steady-state material balance condition gives the steady-state levels of consumption c^* and labor n^*. The steady state values of other variables are found from

$$\begin{aligned} k^* &= \kappa^* z n^* = z n^* f'^{-1}(\rho + \delta) \\ \lambda^* &= \mu^* = \xi (c^*)^{-\beta} e^{(\beta-1)v(n^*)} \\ i^* &= \delta k^* = \delta z n^* f'^{-1}(\rho + \delta). \end{aligned}$$

19.4 Population Growth and Growth Steady States

Of all the types of trend growth, population growth is the easiest to handle. It is natural to begin by representing all quantities in per-capita terms. The accumulation equation for some state variables must be adjusted when expressed in per capita terms. In particular, the "depreciation" rate for *per capita* capital needs to include the dilution of capital that occurs at the population growth rate. Population growth also reduces the rate of accumulation of the per capita assets of the representative household. Thus, for per capita assets, one must replace the interest rate in the accumulation equation by the interest rate minus the population growth rate. Finally, though it is natural to express the utility function in terms of per capita consumption from the start, population growth can easily affect what one considers to be a reasonable discount rate, since people may care more about the average utility level per person in the future if they have many descendants than if they have few descendants. (This effect of population growth on the utility discount rate is in addition to any effects from finite lifetimes as such.)

Other than the need to keep nonreproducible inputs like land and natural resources out of the model if one wishes to study a growth steady state rather than Ricardian immiseration, the shapes of the utility function and the production function are not crucial for the existence of a growth steady state in the face of population growth alone, since as the population increases, the economy can, in effect, replicate itself in every particular.

19.5 Technological Progress and Growth Steady States

In this section we will deal with technological progress alone, assuming that any population growth has already been dealt with in the way described in Section (19.4).

19.5.1 Key Assumptions for the Existence of a Growth Steady State

Several of the assumptions built into the Basic Real Business Cycle above help make a growth steady state possible. First, the King-Plosser-Rebelo utility function (19.4) implies that income and substitution effects on labor supply cancel so that the quantity of labor can be constant in the face of an increasing real wage. Second, technological progress is of one of two restricted types: (1) labor-augmenting technological progress, which acts in certain important respects like population growth and (2) multiplicative (Hicks-neutral) technological progress in the production of consumption goods.

One can begin to get an idea of how these assumptions interact with growth by considering the effects of the corresponding one-time improvements in technology. Equations (19.17) and (19.18) imply that, beginning from a steady state in which these two equations are satisfied, (1) the new steady state after a doubling in the level of consumption technology ξ involves doubling consumption c by the same factor (with the relative price of consumption falling to half its former value and the consumption real wage doubling) and leaving all other quantities unchanged, while (2) the new steady state after a doubling of both the labor-augmenting technology and the level of government spending g involves doubling k, i, c, total output and both the consumption and product real wages, with the relative price of consumption to investment goods and the steady-state capital-labor ratio remaining the same. Steady growth involving the labor-augmenting technology is more complex, since with a continuing improvement in this technology, a higher investment/capital ratio is required, but otherwise, the effects of steady improvement in technology are similar to the effects of these one-time changes.

19.5.2 The Growth Steady State of the Basic Real Business Cycle Model

In response to continuing technological progress, many variables will have time trends. We can use capital letters to denote these trending variables and small letters to denote the corresponding detrended variables. In particular, consider a problem identical to (**??**) above except for the capitalization of the variables that will be trending:

$$\max_{C,n,I} \quad \int_0^\infty e^{-\rho t} \frac{C^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}\, dt \tag{19.19}$$

$$\begin{aligned} \text{subject to} \qquad \dot{K} &= I - \delta K \\ Znf\left(\frac{K}{Zn}\right) &= \frac{C}{\Xi} + I + G. \end{aligned}$$

Using Γ_Ξ to denote the trend growth rate of the consumption technology Ξ and Γ_Z to denote the trend growth rate of the labor augmenting technology and government spending (assumed to trend at the same rate), define detrended values of each of these variables (represented as small letters) by

$$\begin{aligned} \Xi &= \xi e^{\Gamma_\Xi t} \\ Z &= z e^{\Gamma_Z t} \\ G &= g e^{\Gamma_Z t} \\ C &= c e^{(\Gamma_\Xi + \Gamma_Z) t} \\ I &= i e^{\Gamma_Z t} \\ K &= k e^{\Gamma_Z t}. \end{aligned}$$

Since labor does not trend, one could also write $N = n$. Since choices of c and i imply corresponding choices of C and I, substituting these expressions into (19.19) yields

$$\max_{c,n,i} \quad \int_0^\infty e^{-\rho t} e^{(1-\beta)(\Gamma_\Xi + \Gamma_Z)t} \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)}\, dt \tag{19.20}$$

$$\begin{aligned} \text{subject to} \quad \frac{d}{dt}(ke^{\Gamma_Z t}) &= e^{\Gamma_Z t}i - \delta e^{\Gamma_Z t}k \\ e^{\Gamma_Z t}znf\left(\frac{e^{\Gamma_Z t}k}{e^{\Gamma_Z t}zn}\right) &= \frac{e^{(\Gamma_\Xi+\Gamma_Z)t}c}{e^{\Gamma_\Xi}\xi} + e^{\Gamma_Z t}i + e^{\Gamma_Z t}g. \end{aligned}$$

Dividing the constraints through by $e^{\Gamma_Z t}$ and simplifying (including some rearranging of the accumulation equation), this is equivalent to the problem

$$\max_{c,n,i} \quad \int_0^\infty e^{-[\rho+(\beta-1)(\Gamma_\Xi+\Gamma_Z)]t}\frac{c^{1-\beta}}{1-\beta}e^{(\beta-1)v(n)}\,dt \tag{19.21}$$

$$\begin{aligned} \text{subject to} \quad \dot{k} &= i - (\delta + \Gamma_Z)k & (19.22) \\ znf\left(\frac{k}{zn}\right) &= \frac{c}{\xi} + i + g. & (19.23) \end{aligned}$$

Finally, by defining

$$\hat{\rho} = \rho + (\beta - 1)(\Gamma_\Xi + \Gamma_Z) \tag{19.24}$$

and

$$\hat{\delta} = \delta + \Gamma_Z \tag{19.25}$$

we can rewrite this growth model in a way entirely analogous to a stationary model:

$$\max_{c,n,i} \quad \int_0^\infty e^{-\hat{\rho}t}\frac{c^{1-\beta}}{1-\beta}e^{(\beta-1)v(n)}\,dt \tag{19.26}$$

$$\begin{aligned} \text{subject to} \quad \dot{k} &= i - \hat{\delta}k & (19.27) \\ znf\left(\frac{k}{zn}\right) &= \frac{c}{\xi} + i + g. & (19.28) \end{aligned}$$

Thus, the stationary version of the Basic Real Business Cycle model is fully adequate for studying behavior in the face of a trend in both types of technology. If, after detrending, the technology parameters ξ and z and the government spending g are constant, the model will exhibit a growth steady state in which all of the detrended variables are constant and the undetrended variables grow smoothly at a constant rate. If the technologies or government spending fluctuate around their trend, the effects can be analyzed be looking at the response to changes in ξ, z and g in the detrended model.

One other change one must make in interpretation is that, in the stationary model, the real interest rate r is equal to the net marginal product of capital $f'(\cdot) - \delta$. The corresponding variable in the detrended growth model is $f'(\cdot) - \hat{\delta} = f' - \delta - \Gamma_Z$. Thus, in looking at the partial equilibrium pieces of the general equilibrium model, one must also define

$$\hat{r} = r - \Gamma_Z. \tag{19.29}$$

19.5.3 Combining Population Growth and Technological Progress

Combining population growth at rate Γ_L with technological progress means one should modify (19.25) to

$$\hat{\delta} = \delta + \Gamma_Z + \Gamma_L \tag{19.30}$$

and modify (19.29) to

$$\hat{r} = r - \Gamma_Z - \Gamma_L. \tag{19.31}$$

Because there are several ways in which population growth could affect the utility discount rate for per-capita utils, we must express the modification of (19.24) in general terms by making ρ a function of Γ_L:

$$\hat{\rho} = \rho(\Gamma_L) + (\beta - 1)(\Gamma_\Xi + \Gamma_Z), \tag{19.32}$$

with (probably) $-1 \leq \rho'(\Gamma_L) \leq 0$.

It is useful to look at the steady state Euler equation $\hat{r} = \hat{\rho}$, which becomes

$$\hat{r}^* = r^* - \Gamma_Z - \Gamma_L = \rho(\Gamma_L) + (\beta - 1)(\Gamma_\Xi + \Gamma_Z) = \hat{\rho},$$

or

$$r^* = \rho(\Gamma_L) + \Gamma_L + \beta\Gamma_Z + (\beta - 1)\Gamma_\Xi. \tag{19.33}$$

Thus, the regular real interest rate r in steady state should be unambiguously higher in the face of either faster labor-augmenting technical progress or faster population growth ($\rho'(\Gamma_L) + 1 \geq 0$). The effect of faster technological progress in the production of consumption goods on regular steady-state real interest rate r^* is also positive if $\beta > 1$ (the most likely case), but is negative if $\beta < 1$ and zero if $\beta = 1$.

Exercises

1. For a given function $v(\cdot)$, with $v' > 0$, $v'' < 0$, find the condition on β necessary to guarantee that

$$u(c,n) = \frac{c^{1-\beta}}{1-\beta} e^{(\beta-1)v(n)},$$

is jointly concave in c and n for all positive c and n. (Hint: You need to find what is necessary to make the matrix

$$\begin{bmatrix} u_{cc} & u_{cn} \\ u_{cn} & u_{nn} \end{bmatrix}$$

negative definite, that is, to make sure that $u_{cc} < 0$, $u_{nn} < 0$ and $u_{cc}u_{nn} - u_{cn}^2 > 0$.)

2. Using L'Hôpital's rule, show that

$$\lim_{\beta \to 1} \frac{c^{1-\beta} e^{(\beta-1)v(n)} - 1}{1-\beta} = \ln(c) - v(n).$$

3. Analyze the atemporal maximization problem

$$\max_{k_c, n_c} \qquad c = \xi[z n_c f\left(\frac{k_c}{z n_c}\right)]$$

$$\text{subject to} \qquad z(n - n_c) f\left(\frac{(k - k_c)}{z(n - n_c)}\right) = i + g,$$

where ξ, z, k, n, i and g are given, and show that the solution implies

$$\frac{k_c}{n_c} = \frac{k - k_c}{n - n_c} = \frac{k}{n}$$

and

$$z n f\left(\frac{k}{zn}\right) = \frac{c}{\xi} + i + g$$

as the production possibility frontier between c and $i + g$.

Chapter 20

The Basic Neomonetarist Model

The Basic Neomonetarist Model is an attempt to find a counterpart in modern, fully dynamic theory to the old-fashioned static IS-LM model. For readers familiar with the traditional static IS-LM model, the Basic Neomonetarist model offers some surprises. In many ways it is similar to what Tobin and Sargent calling the Dynamic Aggregative Model, except that it has sticky prices insteady of sticky wages.

The distinction between long-run and short-run responses is a familiar one. Mathematically, the language of short-run/long-run distinctions corresponds to approximating different speeds of adjustment for various dimensions of a model by a hierarchy of different time scales. In understanding the Basic Neomonetarist it is helpful to have in mind a hierarchy of different time scales with even more levels. From fastest to slowest (according to our assumptions) these are

1. *almost instantaneous* adjustment of asset markets (putting the economy on the LM curve),

2. *ultra-short-run* adjustment of output (toward short-run equilibrium at the intersection of NRR and LM)

3. *short-run* adjustment of prices (toward medium-run equilibrium at full employment),

4. *medium-run* entry and exit of firms and adjustment of the capital stock (toward long-run equilibrium on the steady state growth path of the

economy), and

5. *long-run* growth.

In this chapter we will formally model only the third level of "short-run" dynamics, discussing the other levels in an informal way.

20.1 The LM Curve

20.1.1 Money Demand

Let P be the aggregate price level, Y the aggregate level of output, M the nominal money supply and let V the *autonomous* component of velocity (the part of velocity not determined by output or the interest rate). For these four variables, use the lower-case letters p, y, m and v to represent the natural logarithms of upper-case letters. Based on the underlying equation $MV/P = Y^h L(r+\pi)$, which allows for any constant elasticity h of money demand with respect to aggregate output and a general dependence of money demand on the nominal interest rate $r + \pi$ generated by the monotonically decreasing function $L(\cdot)$, money demand can be written in logarithmic form as

$$m + v - p = hy + \ln(L(r+\pi)). \tag{20.1}$$

As long as the nominal interest rate is not changing too dramatically, (20.1) can be approximated by

$$m + v - p = hy - \ell \cdot (r+\pi) \tag{20.2}$$

where

$$\ell = \frac{-L'(r^* + \pi^*)}{L(r^* + \pi^*)} > 0$$

is the interest semi-elasticity of money demand at an average or typical level for the nominal interest rate and V has been appropriately normalized to absorb the constant term. If the interest *elasticity* of money demand

$$\frac{-(r+\pi)L'(r+\pi)}{L(r+\pi)}$$

were a constant, as it is, for example, in the Baumol-Tobin case, then the interest *semi-elasticity* of money demand at $r^* + \pi^*$ would be

$$\ell = \frac{\text{constant}}{r^* + \pi^*}.$$

To make a less strong and therefore safer statement, in general, it is likely that the interest semi-elasticity of money demand ℓ is fairly low at high levels of the nominal interest rate—which would be relevant in countries facing high levels of inflation. On the other hand, the interest semi-elasticity of money demand ℓ is likely to be fairly high at low levels of the nominal interest rate—which would be relevant in countries facing low levels of inflation or even deflation. For simplicity, we can log-linearize around $r^* + \pi^*$, but it will be important at a later stage to consider the effects of different average or typical levels of inflation on the interest semi-elasticity of money demand ℓ.

20.1.2 The Dynamics of the Velocity-Adjusted Real Money Supply

Let us give the autonomous-velocity-adjusted log real money supply its own letter, x:

$$x = m + v - p.$$

Also, define μ as the mean growth rate of the autonomous-velocity-adjusted *nominal* money supply:

$$\mu = \frac{d}{dt}[m + v] = \frac{dm}{dt} + \frac{dv}{dt} = \dot{m} + \dot{v},$$

where a dot is used to represent a time derivative. For brevity, we will sometimes drop the adjective "autonomous-velocity-adjusted," and refer to x as "real money balances" or as the "real money supply" and to μ as the "growth rate of the (nominal) money supply."

By definition, the inflation rate π is

$$\pi = \dot{p}.$$

Given these definitions,

$$\dot{x} = \mu - \pi. \tag{20.3}$$

A phase diagram is a useful tool for understanding dynamics because it shows how rates of change in key variables relate to the current levels of those variables at any time. The phase diagram for the Basic Neomonetarist model can be drawn with the autonomous-velocity-adjusted log real money supply $m + v - p$ on the horizontal axis and inflation π on the vertical axis. On the phase diagram, the $\dot{x} = 0$ locus is the horizontal line $\pi = \mu$. Above the line, $\dot{x} < 0$. Below the line, $\dot{x} > 0$. (See the leftward and rightward arrows in Figure 1, indicating these dynamics.)

20.1.3 The LM Curve and the Phase Diagram

The LM curve itself belongs on a graph with output y on the horizontal axis and the real interest rate r on the vertical axis as in Figure 2. Rearranging (20.2),

$$r = \frac{hy - x}{\ell} - \pi.$$

The slope of the LM curve is the coefficient of y, $\frac{h}{\ell}$, which is always positive. Increases in the real autonomous-velocity-adjusted money supply x shift the LM curve outward and down. For example, a .01 increase in x, which represents a 1% increase in the autonomous-velocity-adjusted real money supply, shifts the LM curve outward by $\frac{1}{h}(.01)$, which represents a $\frac{1}{h}$% higher level of output for a given real interest rate. Finally, an increase in inflation π shifts the LM curve down in a one-for-one fashion: an increase in inflation by one percentage point (say from 4% per year to 5% per year) forces the real interest rate at a given level of output to be one percentage point lower (say 1% per year instead of 2% per year).

Notice that the two variables that shift the LM curve—x and π are the two variables on the phase diagram. Thus, a point on the phase diagram indicates a specific position for the LM curve. To be specific, the location of the LM curve in y–r space is determined by the value of $\frac{x}{\ell} + \pi$ in x–π space. The curves $\frac{x}{\ell} + \pi =$ constant on the phase diagram (that is, curves with slope $-\frac{1}{\ell}$) represent a fixed position of the LM curve. (See Figure 3.) Rightward and upward movement in the phase diagram represents an outward shift of the LM curve, while leftward and downward movement in the phase diagram represents an inward shift of the LM curve.

20.1.4 The LM Curve and Other Monetary Rules

The LM curve is the right tool of analysis when the default policy for the central bank (the Federal Reserve in the U.S.) is to keep the autonomous-velocity-adjusted nominal money supply growing at a constant rate, with some tinkering. But if the central bank follows some other policy on a day-to-day basis, it is important to build that rule into the analysis in a more fundamental way.

Real-life central bankers often think of themselves as setting the nominal interest rate. A reasonable, fairly general specification for a monetary rule is for the central bank to set the nominal interest rate $r + \pi$ by

$$r + \pi = a + by + g\pi$$

where b and g are constant coefficients and a evolves gradually in a way that tightens monetary policy if inflation is above its target rate and loosens monetary policy if inflation is below its target rate:

$$\frac{da}{dt} > 0 \qquad \text{when } \pi > \pi^*$$

and

$$\frac{da}{dt} < 0 \qquad \text{when } \pi < \pi^*.$$

An LM curve generated by a nominal money supply growing at a constant rate fits this description of monetary policy with $g = 0$, $a = -\frac{x}{\ell}$, $\pi^* = \mu$ and $b = h/\ell$. (For reasonable values of the income and interest rate elasticities of money demand, $h = 1$ and $\ell = 1$ year, $b = h/\ell$ might be on the order of 1/year.) As an alternative, the Taylor rule fits this description of monetary policy reasonably well with $g = 1.5$ and b (also) on the order of 1/year. The key difference between an LM curve and the Taylor rule becomes clear when one focuses on the determination of the real interest rate. The LM curve implies that

$$r = a + by - \pi$$

while the Taylor rule implies

$$r = a + by + .5\pi.$$

Thus, when inflation falls temporarily as a result of a technological improvement, an LM curve would shift up, while the Taylor rule curve would shift down. In between these two cases is a real interest rate rule given by

$$r = a + by.$$

Although the analysis in the rest of this chapter will continue to use an LM curve as the monetary rule, it is good to keep in mind the alternatives.

20.2 Aggregate Supply and the Net Rental Rate (NRR) Curve

In the absence of investment adjustment costs, the NRR curve is the same thing as the Net Rental Rate curve which gives the rate, net of depreciation, at which capital services can be rented for a given level of aggregate output. When there are investment adjustment costs—a complication that must be deferred until a later chapter—the NRR curve is a fragile path that depends on all of the details of the model and its dynamics, and even on the nature of the shocks hitting the economy.

One of the key features of the Basic Neomonetarist Model is that the location and shape of the Net Rental Rate (NRR) curve—which is the NRR curve in the absence of investment adjustment costs—is intimately related to the details of aggregate supply. Thus, the traditional approach of combining NRR and LM to get aggregate demand and only then bringing in aggregate supply does not work in the Basic Neomonetarist model. Instead, the NRR (=NRR) curve and aggregate supply are joined from the start, almost like two sides of the same coin. It is the combination of NRR and aggregate supply which is then brought together with the LM curve.

20.2.1 Short-Run Aggregate Supply

Both aggregate supply and the net rental curve are based on the technology, preferences and industrial organization of the economy. Medium and long-run aggregate supply are determined as in Real Business Cycle models with some degree of imperfect competition and increasing returns to scale. The actual short-run aggregate supply curve is just the current level of the slowly moving aggregate price—a horizontal line in y–p space. (See Figure 4.) However, the actual (log) aggregate price p need not be the same as the (log) desired price $p^{\#}$ of a typical firm. The instantaneously optimal desired price of a typical firm is defined as the price that a firm would choose if *it* never had any costs or barriers to price adjustment, while all the other firms in the economy remained subject to those limitations in their price-setting. The *notional* short-run aggregate supply curve shown in Figure 4 gives the (log) desired price $p^{\#}$ as a function of (log) aggregate output y:

$$p^{\#} = p + \beta(y - y^f). \tag{20.4}$$

The quantity y^f is log full-employment output, where full-employment output is defined as what output will be after the price adjustment process is complete. Full employment can also be defined as a situation where marginal revenue equals marginal cost for the typical firm—or equivalently, where the desired markup (price/[marginal revenue]) is equal to the actual markup (price/[marginal cost]). The slope β is the elasticity of the desired price with respect to aggregate output.[1]

20.2.2 The Rental Rate of Capital

The net rental rate is simply the gross rental rate of capital R net of the depreciation rate δ. (Note that, contrary to the convention above, r, the real interest rate, is *not* the natural logarithm of R, the gross rental rate. The need to subtract the depreciation rate from R makes it awkward to use natural logarithms indiscriminately at this juncture.) Investment demand depends critically on the gap $R-\delta-r$ between the net rental rate and the real interest rate. Indeed, in the absence of investment adjustment costs the elasticity of investment demand with respect to the gap $R-\delta-r$ is infinite, as illustrated by the horizontal investment demand curve in Figure 5. If $R-\delta > r$, investment demand is infinite. If $R-\delta < r$, demand for *gross* investment is zero. Clearly, infinite investment is impossible in investment market equilibrium (that is, equilibrium in the market for loanable funds). Zero gross investment only occurs in circumstances like those in the Great Depression (discussed later on). In non-depression circumstances, an absence of investment adjustment costs forces the real interest rate to equal the net rental rate in equilibrium:

$$r = R - \delta.$$

Since the depreciation rate δ is a constant, the rental rate R is the key to the determination of the real interest rate r.

If prices were not sticky, the rental rate R could be determined in turn by the marginal revenue product of capital. But with prices sticky, the firm cannot necessarily sell additional output at the temporarily fixed price. Therefore, the rental rate of capital is best seen as determined by the *marginal cost*

[1] In "The Quantitative Analytics of the Basic Neomonetarist Model," Kimball (1995) shows that

$$\beta = \frac{\Omega}{\epsilon\omega}$$

where ϵ is the price elasticity of demand for a typical firm, Ω is the elasticity of (marginal cost)/(marginal revenue) with respect to expansions of aggregate output *and* firm output, ω is the elasticity of (marginal cost)/(marginal revenue) with respect to increases in firm output alone.

product of capital. The *marginal cost product* of capital is the reduction in the cost of other inputs needed to produce a given level of output resulting from an increase in capital services. In the simplified model here, all of the inputs other than capital are represented by the variable input labor, denoted N. Although the assumption is not necessary, nothing important is lost by assuming a generalized Cobb-Douglas production function for each firm i: $Y_i = F\left(K_i^{\theta}(ZN_i)^{1-\theta}\right)$, where K_i is the amount of capital used by firm i, Z is the level of labor-augmenting technology (common to all firms), θ is capital's share in costs and $1-\theta$ is labor's share in costs (both common to all firms). Given the assumption of a generalized Cobb-Douglas production function, cost minimization for firm i to produce output $\bar{Y}_i$ requires the solution of

$$\min_{K_i, N_i} RK_i + WN_i$$

s.t.

$$K_i^{\theta}(Z_i N_i)^{1-\theta} = F^{-1}(\bar{Y}_i)$$

where W is the real wage at which labor is available on a competitive spot market.

A standard result for Cobb-Douglas technology, which can be re-derived here using the Lagrange multiplier technique, is that

$$\frac{RK_i}{WN_i} = \frac{\theta}{1-\theta}.$$

Rearranging and adding up K_i and N_i over all firms to put the equation in terms of the aggregate quantities of capital K and labor N leads to

$$RK = R\sum_i K_i = \frac{\theta}{1-\theta} W \sum_i N_i = \frac{\theta}{1-\theta} WN.$$

Solving for the rental rate R,

$$R = \frac{\theta}{1-\theta}\frac{WN}{K}. \tag{20.5}$$

This means that the real interest rate, equated to the net rental rate, is given by

$$r = R - \delta = \frac{\theta}{1-\theta}\frac{WN}{K} - \delta.$$

20.2.3 The Net Rental Rate (NRR) Curve

Since an increase in output raises the demand for capital services, while the supply of capital remains essentially the same (given how much slower capital accumulation is than typical business cycle movements), one would expect an increase in output to lead to an increase in the rental rate R. (Indeed, a tendency toward a procyclical rental rate of capital is an almost universal feature of serious business cycle models.) Equation (20.6) confirms the idea that the rental rate will increase with output. Given how slowly the quantity of capital moves, in the short run, in the absence of any change in technology, an increase in output requires an increase in the quantity of labor N. In the absence of any change in the representative household's medium-to-long-run situation, getting the standard representative Neoclassical household to supply a larger quantity of labor requires a higher real wage W. (Constancy of the household's medium-to-long-run situation makes this is only a temporary increase in the real wage, and so blocks any substantial wealth or income effect. To be specific, the marginal utility of consumption λ is little affected by the short-run movements of the economy. Therefore, the *Frisch* or marginal-utility-of-consumption-held-constant labor supply curve, which is definitely upward sloping, is the relevant labor supply curve.) With both N and W increasing with output, the rental rate must increase with output.

Changes in technology or in the medium-to-long-run situation of the representative household are *real shocks*, which will be discussed later. For a given level of technology and a given medium-to-long-run situation of the representative household, the level of output determines the rental rate R by way of N and W. Thus, write

$$r = R - \delta = r^f + \phi(y - y^f), \tag{20.6}$$

where r is the real interest rate, r^f is the full-employment net rental rate, $y - y^f$ is the logarithmic gap between output and full-employment output, and the slope ϕ is the semi-elasticity of the real interest rate with respect to the log output gap. Equation (20.6) is depicted as the NRR curve in Figure 6. As noted above, the slope ϕ is positive because an increase in output raises the demand for capital services, a fact confirmed by the increased price and quantity of the alternative input, labor. The upward-slope of the NRR curve makes the NRR curve different from the traditional NRR curve as it is usually drawn. (However, even traditional treatments sometimes admit the possibility that the "investment accelerator" might be so strong that the NRR curve would be upward-sloping.)

The size of the slope ϕ is given by

$$\phi = R\frac{1+\zeta^{-1}}{\Gamma(1-\theta)},$$

where ζ is the Frisch labor supply elasticity and Γ is aggregate degree of returns to scale. In words, the slope of the NRR curve depends on depends on (a) the base level of the gross rental rate R, which indicates how proportional changes in the gross rental rate translates into percentage *point* changes in the real interest rate; (b) the elasticity of labor requirements with respect to output, equal to $\frac{1}{\Gamma(1-\theta)}$ (c) the reciprocal of the Frisch labor supply elasticity, ζ^{-1}, which indicates how the increase in labor N translates into an increase in the real wage W. To explain further, (a) a 1% increase in R might be an increase, from say 15% per year to 15.15% per year, which is not the same as a one percentage *point* increase in R, say from 15% per year to 16 per year; (b) Γ is the elasticity of output with respect to a proportional increase in *all* inputs, but when only labor is increasing, with capital held constant, output does not move as much, so labor has to increase more to get a given increase in output; (c) since ζ^{-1} is the elasticity of W with respect to N, $1+\zeta^{-1}$ is the elasticity of WN with respect to N.

The NRR curve also differs from the traditional NRR curve in being unaffected by short-run movements in any variable other than technology (since short-run movements in any variable cannot change the medium-to-long-run situation of the representative household). This is discussed more below in connection with the effects of real shocks.

20.3 Dynamic Aggregate Supply

Price-setting is an ongoing process, with some firms changing their prices at the same time other firms are leaving their prices fixed. This pattern of staggered price setting and overlapping fixed prices is something worth capturing in a model of sticky prices. However, because introducing sticky prices into a model creates many complications, it is a good idea to begin by introducing them in a simple way. Calvo (1982) pioneered a particularly convenient way of dealing with overlapping fixed prices set on a staggered schedule. Calvo's simplified model of price setting gets at the idea of time-dependent staggered price setting in a way that leaves the frequency of price adjustment by firms—which we will denote α—as a flexible parameter of the model.

The simplifying trick in Calvo's model is to think of firms as getting the chance to consider and adjust their prices after random intervals of time, determined by a Poisson process. This insures that the small subset of firms changing their price at any point in time are a random sample of all firms–with an average *old* price equal to the aggregate price. Assume also that all of the firms are essentially identical other than each firm's old price carried over from the past. Then, in setting a new price, each firm is in essentially the same situation, and there is a single *optimal reset price* B for all firms that get the opportunity to change their prices at a given time.

20.3.1 The Evolution of the Aggregate Price Level

In order to keep the model of dynamic aggregate supply simple, we will also make some fairly innocuous approximations. To begin with, regardless of the elasticities of substitution between the prices of the varieties of goods, the logarithmic (percentage) change in the aggregate price level is equal, to a first order approximation, to a weighted average of the logarithmic (percentage) changes in the individual prices—weighted by the share of each variety in total spending. The share of each variety in total spending will remain close enough to the average share that changes in this weighting create only second-order adjustments in the behavior of the aggregate price level. Thus, given underlying symmetry between all the firms, to a first-order approximation, the logarithmic change in the aggregate price level is equal to a simple average of the logarithmic changes in individual prices. Since, on average, firms are changing their log prices from the log aggregate price p to the log reset price b at the Poisson rate of price adjustment α, the rate of inflation π is

$$\pi = \dot{p} \approx \alpha(b - p). \tag{20.7}$$

20.3.2 The Determination of the Optimal Reset Price

The next key issue is the determination of the optimal reset price. Intuitively, the optimal reset price should be a weighted average of the (instantaneously optimal) desired prices over the period of time for which the newly reset price will be fixed. As long as the microeconomic rate of price adjustment α is large relative to the difference between the real interest rate and the growth rate of the economy, such a weighted average of future desired prices is very close to correct. In the present context, the small gap having to do with discounting of future considerations at the real interest rate can be appropriately neglected.

A first-order approximation also makes it possible to state the relationship in similar form in terms of log prices:

$$b_t = \int_t^\infty [\alpha e^{-\alpha(\tau - t)}] p_\tau^\# \, d\tau. \tag{20.8}$$

Since $\int_t^\infty [\alpha e^{-\alpha(\tau-t)}] \, d\tau = 1$, (20.8) states that the log reset price b is a weighted average of future log desired prices $p^\#$. Desired prices in the more distant future are discounted at rate α simply because it is unlikely that the firm will still be stuck with the price it is currently resetting for that long. A desired price beyond the time of the next opportunity to adjust prices is irrelevant to the decision of how to set the price now.

20.3.3 Digression on Leibniz' Rule

It is sometimes desirable to convert integral equations, such as (20.8), into differential equations. For this purpose, Leibniz' rule is often useful. Leibniz' rule states that if

$$g(t) = \int_{A(t)}^{Z(t)} f(\tau, t) \, d\tau,$$

then

$$\begin{aligned} \frac{dg(t)}{dt} &= \int_{A(t)}^{Z(t)} \frac{\partial f(\tau, t)}{\partial t} \, d\tau \\ &\quad + \frac{dZ(t)}{dt} f(Z(t), t) \\ &\quad - \frac{dA(t)}{dt} f(A(t), t). \end{aligned} \tag{20.9}$$

Note that t acts as a parameter in the integration while τ is the dummy variable over which integration is performed. The function $g(t)$ is the area under the curve of $f(\tau, t)$ as τ goes from $A(t)$ to $Z(t)$. To get some intuition for Leibniz' rule, consider what the picture looks like if $A(t)$, $Z(t)$ and $f(\tau, t)$ all increase with the parameter t. Multiplying the right-hand side of (20.9) by dt yields the additional area when t is increased to $t + dt$. On the right-hand side of (20.9), the first line indicates the additional ribbon of area added on top. (See Figure 7.) The second line indicates the additional sliver of area added at the right as $Z(t)$ increases. The third line indicates the sliver of area sliced off at the left as $A(t)$ increases. The area of each sliver is equal to base times height.

The base is proportional to $\frac{dZ(t)}{dt}$ on the right and proportional to $\frac{dA(t)}{dt}$ on the left. The height is equal to $f(Z(t),t)$ on the right and to $f(A(t),t)$ on the left. The small squarish shapes in the corners can be ignored because they are small, being proportional to dt^2. Similar pictures can be drawn when one or more of $A(t)$, $Z(t)$ and $f(\tau,t)$ decreases with the parameter t.

20.3.4 Differential Equations for the Optimal Reset Price and for Inflation

Applying Leibniz' rule to (20.8),

$$\begin{aligned} \dot{b}_t &= \frac{d}{dt}\int_t^\infty [\alpha e^{-\alpha(\tau-t)}]p^\#_\tau \, d\tau \\ &= \int_t^\infty [\alpha^2 e^{-\alpha(\tau-t)}]p^\#_\tau \, d\tau - \alpha p^\#_t \\ &= \alpha(b_t - p^\#_t). \end{aligned} \tag{20.10}$$

With the dummy variable τ eliminated from the equation, the time subscript can once again be safely omitted: $\dot{b} = \alpha(b - p^\#)$.

Equation (20.10) leads to a differential equation for inflation. Since $\pi = \alpha[b-p]$

$$\begin{aligned} \dot{\pi} &= \alpha[\dot{b} - \dot{p}] \\ &= \alpha[\alpha(b-p^\#) - \alpha(b-p)] \\ &= \alpha^2(p - p^\#). \end{aligned}$$

Using the notional short run aggregate supply curve, (20.4), to relate the "inflationary gap" $p^\# - p$ to the "output gap" $y - y^f$,

$$\dot{\pi} = -\alpha^2\beta(y - y^f). \tag{20.11}$$

Equation (20.11) is the *dynamic aggregate supply* equation. Note that $\dot{\pi} = 0$ when $y = y^f$; the dynamic aggregate supply equation is consistent with any constant level of inflation when the economy is at full employment. In Phillips curve space with log output y on the horizontal axis and inflation on the vertical axis, the dynamic aggregate supply curve is vertical at y^f. (See Figure 8.) In the absence of new information, when output is above full employment, inflation must be falling; below full employment, inflation must be rising. This may sound strange, but it actually makes sense in a model with rational, forward-looking agents. As shown below, in dynamic general equilibrium, output above full employment tends to be associated with *high and falling* inflation, while output below full employment tends to be associated with *low and rising* inflation. The mantra *"high and falling* or *low and rising"* is invaluable in making sense of the negative sign in the dynamic aggregate supply equation (20.11).

20.4 Short-Run Equilibrium

Figure 9 illustrates short-run equilibrium at the intersection of the LM and NRR curves. To solve for this equilibrium algebraically, substitute from the equation for the NRR curve (20.6) into the equation for the LM curve (20.2):

$$\begin{aligned} m + p - v = x &= hy - \ell[r^f + \phi(y - y^f)] - \ell\pi \\ &= (h - \ell\phi)(y - y^f) - \ell\pi + [hy^f - \ell r^f]. \end{aligned} \tag{20.12}$$

Solving for the output gap $y - y^f$,

$$y - y^f = \frac{x + \ell\pi \ - \ [hy^f - \ell r^f]}{h - \ell\phi}. \tag{20.13}$$

Thus, assuming the denominator $h - \ell\phi$ is positive, output depends positively on the autonomous-velocity-adjusted real money supply x and on the rate of inflation π, which tends to lead to economizing of transaction money balances, making a given amount of money go further. (Other than noting that they are not independent, we will defer discussion of changes in full-employment output and the full-employment real interest rate until later, when we discuss real shocks.)

The denominator $h - \ell\phi$ in (20.13) is closely related to the slopes of the NRR and NRR curves. With log output y on the horizontal and the real interest rate r on the vertical axis, the slope of the NRR curve is ϕ, while the slope of the LM curve is h/ℓ. Thus

$$h - \ell\phi = \ell[\frac{h}{\ell} - \phi] = \ell[(\text{slope of LM}) - (\text{slope of NRR})].$$

With the NRR curve upward-sloping, stability of the short-run equilibrium requires the LM curve to have a steeper upward slope than the NRR curve. Stability is easier to guarantee the lower is ℓ. If ℓ becomes too great, as it may if inflation is very low, it calls the stability of the short-run equilibrium into question as the LM curve becomes relatively flat. In effect, the upward-sloping NRR curve makes a kind of liquidity trap possible even for finite values of ℓ that might occur for strictly positive values of the nominal interest rate. This is very relevant in thinking about the Great Depression. But for now, assume that $h - \ell\phi > 0$, ensuring stability.

20.5 The Phase Diagram for the Basic Neomonetarist Model

20.5.1 The Dynamics of Inflation

Equation (20.13) makes it possible to express $\dot{\pi}$, the rate of change in inflation, as a function of $x = m + v - p$ and π:

$$\begin{aligned}\dot{\pi} &= -\alpha^2\beta(y - y^f) \\ &= -\frac{\alpha^2\beta}{h - \ell\phi}[x + \ell\pi - (hy^f - \ell r^f)]. \qquad (20.14)\end{aligned}$$

On the phase diagram, the $\dot{\pi} = 0$ locus is downward-sloping, with slope $-\frac{1}{\ell}$. It intercepts the x-axis at $x = hy^f - \ell r^f$. To the right of the $\dot{\pi} = 0$ locus, π is falling. To the left of the $\dot{\pi} = 0$ locus, π is rising. (See Figure 10.)

Note that since it has the slope $-\frac{1}{\ell}$, $x + \ell\pi$ is constant along the $\dot{\pi} = 0$ locus and the LM curve is in the same position everywhere along the $\dot{\pi} = 0$ locus—the position of the LM curve that puts the short-run equilibrium at full employment output y^f.

20.5.2 The Convergence Rate

Figure 11 combines the dynamics for x and π. The intersection of the $\dot{x} = 0$ and $\dot{\pi} = 0$ loci is the phase diagram's steady state. We will call this point "full employment," the "medium-run steady-state," or "medium-run equilibrium" in contrast to the "long-run steady state," which is where the economy ends up after both prices *and* the capital stock have adjusted.

The short-run dynamics shown in Figure 11 imply an upward-sloping saddle path. (The rule guaranteeing a downward-sloping saddle path applies only when the two axes are the state and costate variables of an optimizing model.)

In matrix form, the dynamic equations are

$$\begin{bmatrix} \dot{x} \\ \dot{\pi} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ \frac{-\alpha^2\beta}{h-\ell\phi} & \frac{-\ell\alpha^2\beta}{h-\ell\phi} \end{bmatrix} \begin{bmatrix} x \\ \pi \end{bmatrix} + \begin{bmatrix} \mu \\ \frac{\alpha^2\beta(hy^f - \ell r^f)}{h-\ell\phi} \end{bmatrix}.$$

The characteristic equation for an eigenvalue Λ is

$$\Lambda^2 - \text{trace}\ \Lambda + \text{determinant} = \Lambda^2 + \frac{\ell\alpha^2\beta}{h-\ell\phi} \cdot \Lambda - \frac{\alpha^2\beta}{h-\ell\phi} = 0.$$

By the quadratic formula,

$$\Lambda = -\frac{\ell\alpha^2\beta}{2(h-\ell\phi)} \pm \sqrt{\left(\frac{\ell\alpha^2\beta}{2(h-\ell\phi)}\right)^2 + \frac{\alpha^2\beta}{h-\ell\phi}}.$$

Subtracting the square root yields the negative root. The convergence rate κ is the absolute value of this negative root. Simplifying,

$$\kappa = \frac{\ell\alpha^2\beta}{2(h-\ell\phi)} \left[1 + \sqrt{1 + \frac{4(h-\ell\phi)}{\ell^2\alpha^2\beta}}\right]. \tag{20.15}$$

20.5.3 The Slope of the Saddle Path

The eigenvector associated with the convergence rate shows the slope of the saddle path. To find the convergent eigenvector, it is necessary to solve the equation

$$\begin{bmatrix} \dot{x} \\ \dot{\pi} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ \frac{-\alpha^2\beta}{h-\ell\phi} & \frac{-\ell\alpha^2\beta}{h-\ell\phi} \end{bmatrix} \begin{bmatrix} 1 \\ \sqcup \end{bmatrix} = -\kappa \begin{bmatrix} 1 \\ \sqcup \end{bmatrix}.$$

for $\sqcup$. Focusing on the top scalar equation indicates that

$$\begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ \sqcup \end{bmatrix} = -\sqcup = -\kappa.$$

Thus, $\sqcup = \kappa$ and the convergent eigenvector is

$$\begin{bmatrix} 1 \\ \kappa \end{bmatrix}.$$

It is straightforward but tedious to show that this eigenvector solves the bottom equation as well. Therefore, the slope of the saddle path is κ.

In words, the argument proving that the slope of the saddle path is κ is equivalent to the following argument. If autonomous velocity-adjusted real money balances x are 1% above their full employment level, they must decline by κ percentage points per year. To accomplish this, inflation must be κ percentage points per year above normal.

20.5.4 Determinants of the Convergence Rate, the Slope of the Saddle Path and the Contract Multiplier

Knowing the equality of the convergence rate and the slope of the saddle path, we can kill at least two birds with one stone in analyzing the determination of κ in (20.15). The slope of the saddle path is important because it shows the size of the response of inflation to a one-shot increase in the autonomous-velocity-adjusted money supply $m + v$—or an increase in $m + v$ coming out of an i.i.d. time series of changes. The convergence rate is important because it is the *macroeconomic rate of price adjustment.* It is also of interest to look at the contract multiplier $\frac{\alpha}{\kappa}$—the extent to which the macroeconomic rate of price adjustment is slower than the microeconomic rate of price adjustment.

In dynamic general equilibrium the sensitivity of *output* to a one-shot (or i.i.d.) increase in the autonomous-velocity-adjusted real money supply is given by

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} + \kappa \frac{\partial y}{\partial \pi} = \frac{1 + \kappa \ell}{h - \ell \phi} = \frac{\kappa^2}{\alpha^2 \beta}.$$

The last line is an implication of the characteristic equation satisfied by $-\kappa$. It implies

$$\kappa = \alpha \sqrt{\beta \frac{dy}{dx}} = \alpha \sqrt{\beta \left(\frac{1 + \kappa}{h - \ell \phi} \right)}. \tag{20.16}$$

The contract multiplier, $\frac{\alpha}{\kappa}$ is then

$$\frac{\alpha}{\kappa} = \beta^{-.5} \left(\frac{dy}{dx} \right)^{-.5}$$

In the quantity theory case with $\ell = 0$ and $h = 1$, $\frac{dy}{dx} = 1$ and the contract multiplier is $\frac{\alpha}{\kappa} = \beta^{-.5}$, making the question of how to get a larger contract multiplier the same as the question of how to get a low value of β. Ball and Romer (1990) give the name "real rigidity" to having a low value of β. Therefore, real rigidity is a key to getting a large contract multiplier. More generally, the contract multiplier is increased by real rigidity and by a small dynamic general equilibrium sensitivity of output to money $\frac{dy}{dx}$.

20.6 Changes in the Growth Rate of Money

Since i.i.d. monetary shocks move the economy around on the saddle path, they can create the illusion of a stable Phillips curve, with inflation and the level of output covarying. But attempting to exploit this "Phillips curve," with the idea of accepting higher inflation in exchange for higher output, would lead to an increase in the expected rate of autonomous-velocity-adjusted money growth μ. As can be seen in Figure 12, this shifts the medium-run steady state to a higher level of inflation π and a lower level of real money balances x. The new saddle path, leading to the new steady state, shifts up to a higher level of inflation for any value of real money balances x. The shift in the saddle path also implies a higher level of inflation for any value of output y—in effect an upward shift of the Phillips curve, as shown more directly in Figure 13. The advantage of Figure 12 as compared to Figure 13 is the knowledge that the initial jump must be vertical in Figure 12. In Figure 13, the initial jump is on a slant.

The dynamics of the short-run adjustment process go as follows. Since there has been no sudden change in $m + v$, but only a change in the growth rate, and the price level p cannot jump, $x = m+v-p$ does not jump. Therefore, without any jump in x, inflation π must jump up to the new saddle path. After that, the economy follows the new saddle path down to the medium-run steady state.

During this whole process, the NRR curve does not move. Every point on the phase diagram corresponds to a position for the LM curve. Lines parallel to the $\dot{\pi} = 0$ locus ($y = y^f$ locus) are sets of points with the same position for the LM curve. The upward jump in π causes the LM curve to shift to the right, pushing the economy above full employment. (See Figure 14.) The phase diagram treats the adjustment to the short-run equilibrium at the intersection of NRR and the new LM as already completed. As the economy follows the saddle path down to the left, the LM curve gradually shifts back to the left until the economy returns to full employment.

In the *very* short-run, output has not yet adjusted to the NRR curve. Before output has had any chance to adjust, the economy jumps down to the new LM curve at the old level of output. Since the LM curve shifts down by exactly the change in inflation $\Delta\pi$, this initial very short-run equilibrium has a real interest rate r that is lower by $\Delta\pi$ but an unchanged nominal interest rate $r + \pi$. Then the economy follows the new LM curve up to the right until it meets the NRR curve.

A one-shot or i.i.d. increase in the money supply, which moves the economy up along the saddle path, induces a similar pattern of short-run adjustment in the LM-NRR graph. The LM curve initially shifts to the right, then gradually moves back to the left as prices adjust. In this case, the LM curve moves down by more than $\Delta\pi$, so that both the real and the nominal interest rates fall in the very short run, then rise again through the very short-run dynamics as the economy moves toward the NRR curve. Because of the upward slope of the NRR curve, the real interest rate (and *a fortiori*, the nominal interest rate) is higher in short-run equilibrium at the intersection of NRR and LM than it was before the shock.

20.7 Real Shocks

So far we have discussed *monetary* shocks which have little or no effect on the real variables in the economy. By contrast a "real shock" is a shock that causes a significant change in the level of full-employment output y^f or the full-employment real interest rate r^f—typically both in tandem. By convention, a *positive* real shock is one that increases full employment output. (In the unlikely case in which only r^f changes, a positive real shock would be one that increased r^f.) Where monetary shocks shift the $\dot{x} = 0$ locus, real shocks shift the $\dot{\pi} = 0$ locus.

A set of four types of real shocks gives a good sense of the range of possible reactions of the economy to real shocks: labor supply shocks, technology shocks, government purchase shocks and industrial organization shocks.

20.7.1 Labor Supply Shocks

The phrase "labor supply shocks" refers to a wide variety of different shocks that shift the labor supply curve. Regardless of the cause of the shift of the labor supply curve, it results in similar short-run dynamics as the economy adjusts to a new level of full-employment output and of the full-employment real interest rate. Labor supply as seen by firms may shift out for any of the following reasons:

1. an increase in patience—that is, an increase in households' willingness to defer gratification and work for the future;

2. fears of hard times (or the certainty of hard times) in the future that households want to prepare for by saving more;

3. an increase in the long-term real interest rate, making saving look more attractive;

4. an influx of new potential workers into the economy, either by immigration or by the maturation of a large cohort of young people;

5. a reduction in the effective marginal tax rate on labor;

6. a reduction in the market power of unions;

7. an increase in the desire for consumption goods relative to leisure, making households willing to work harder (to be able to purchase those goods);

8. a reduction in the need to spend time at home, whether because of household labor-saving devices or because of having fewer children at home.

The first three sources of labor supply shocks all operate through the marginal utility of consumption λ for a fixed felicity (instantaneous utility) function. These movements in λ follow the logic of the Basic Real Business Cycle Model with essentially no change to that model, as long as we assume that the desired markup of price over costs at full employment is small. Studying the other sources of labor supply shocks formally would require additional machinery.

The Effect on y^f

The Basic Real Business Cycle Model, with or without peripheral add-ons, implies that in medium-run (full-employment) equilibrium, a shift outward in labor supply leads to an increase in full-employment output y^f. To see why, remember the characterization of full employment as a situation in which the typical firm has marginal revenue equal to marginal cost. The shift outward and therefore downward in the labor supply curve means that at the old level of full employment output, the real wage is lower than before the shock. The lower real wage results in lower marginal cost at the old level of full-employment output, while marginal revenue at that level of output remains unchanged. Since marginal revenue is above marginal cost at the old level of full-employment output, the new level of full employment output must be higher.

The Effect on r^f

Because it means more output for each unit of capital, this increase in full-employment output leads to an increase in the full-employment rental rate unless the full-employment profit rate rises very fast with the level of full employment output.[2] Because the rental rate is likely to increase, it is likely that a positive labor supply shock will increase the full-employment real interest rate r^f.

The Shift of the NRR Curve

A shift outward and downward in the labor supply curve means that for any given quantity of labor, the real wage is lower. With no change in technology or in the capital stock, there is no change in the amount of labor needed to produce a given amount of output. (See Figure 15.) Therefore, an increase in labor supply implies a lower real wage for any given amount of output, *and* a lower value of

$$R = \frac{\theta}{1-\theta}\frac{WN}{K}$$

for a given amount of output. As a consequence, the NRR curve, which gives the mapping from output y to the net rental rate $R-\delta$, shifts downward. (See Figure 16.)

The Shift of the $\dot{\pi} = 0$ locus

On the LM-NRR diagram, the y^f–r^f point indicates full employment. Define the *full-employment LM curve* as the LM curve that exists for some values of x and π which passes through the y^f–r^f point. The $\dot{\pi} = 0$ locus shows the set of points on the $x-\pi$ plane that yield a full-employment LM curve. Therefore (given the positive relationship between x and the LM curve), a rightward shift of the full-employment LM curve corresponds to a rightward shift of the $\dot{\pi} = 0$ locus. Conversely, a leftward shift of the full-employment LM curve corresponds to a leftward shift of the $\dot{\pi} = 0$ locus.

In response to a positive labor supply curve, *both the full-employment LM curve and the $\dot{\pi} = 0$ locus shift to the right.* Here is why. The new y^f–r^f point must be on the new NRR curve. The increase in full-employment output y^f, combined with the downward shift in the NRR curve puts the new y^f–r^f point unambiguously to the right of the old full-employment LM. Therefore, the new full-employment LM curve after a positive labor supply shock must be to the right of the old full-employment LM curve. (See Figure 17.)

[2]Given a Cobb-Douglas production function, $R = \theta(1-\Pi)(Y/K)$, where Π is the actual shadow profit rate at any point in time. Since $1-\Pi =$ [degree of returns to scale]/[desired markup], for the full-employment profit rate to rise that fast, either the degree of returns to scale must fall very fast or the desired markup must rise very fast, or both.

The Short Run Dynamics on the x–π Phase Diagram

On the x–π phase diagram, the medium-run (full-employment) equilibrium is at the intersection of the $\dot{\pi} = 0$ locus with the $\dot{x} = 0$ locus. The $\dot{x} = 0$ locus is unchanged at $\pi = \mu$. Thus, the rightward shift of the $\dot{\pi} = 0$ locus in reaction to an increase in labor supply implies a higher value of x^f, the autonomous-velocity-adjusted real money supply at full employment. The full-employment level of inflation, π^f is unchanged.[3] From the standpoint of the money demand equation, the higher level of full-employment output y^f, with no change in $r^f + \pi^f$ requires a higher level of real money balances x^f.

Figure 18 shows the short-run dynamic general equilibrium effects of a permanent rightward shift in the $\dot{\pi} = 0$ locus. On impact, inflation jumps down to the new saddle path. Then the economy follows the new saddle path up to the new medium-run steady state. The moment the point on the phase diagram crosses the old $\dot{\pi} = 0$ locus is distinguished as the moment the shifting LM curve passes over the original LM curve. Beyond that point, the LM curve continues shifting toward the y^f–r^f point on the new NRR curve. (See Figure 19.)

The Contractionary Effects of an Increase in Labor Supply in the Short Run

The position of the LM curve is fully determined by the real money supply x and the inflation rate π; the mapping from the phase diagram to the position of the LM curve is unaffected by real shocks. In dynamic general equilibrium, in the absence of any change in autonomous-velocity-adjusted money $m + v$, $x = m + v - p$ will not jump on impact. In this case, the initial shift of the LM curve in reaction to any real shock will be determined by the jump in the inflation rate, which is forward-looking.

Figure 18 shows the initial fall in the inflation rate in response to an increase in labor supply. This disinflationary response is typical of positive real shocks. A fall in inflation causes an upward shift in the LM curve, as a given nominal interest rate gets mapped into a higher real interest rate.

Figure 19 shows this initial upward jump in the LM curve. The upward shift in LM, combined with the downward shift in NRR, leads to an unambiguous short-run reduction in output y. Thus, in response to an increase in

[3] For π^f to be unchanged, the medium-run dynamics of the capital stock must be *much* slower than the macroeconomic rate of price adjustment. When this is not a good approximation, one must allow for possible changes in π^f. Analyzing changes in π^f requires a good grasp of the medium-run effects of a given real shock.

labor supply, which *increases full employment output* y^f, output falls in the short run! This perverse short-run effect on output is typical of real shocks.[4]

Other Monetary Policy Rules

What if the central bank follows a policy that automatically changes the money supply in reaction to a shock? First, if the central bank fully understood the nature of the shock, it could increase the money supply enough to immediately reach the new equilibrium level of the real money supply $(x^f)'$ instead of getting to $(x^f)'$ through an extended disinflationary period. However, this would require a lot of knowledge on the part of the central bank. Suppose the central bank does not have that much knowledge, but follows a Taylor rule. Then the Monetary Policy (MP) curve, which otherwise looks like an LM curve, shifts in the opposite direction from an LM curve in response to a drop in π. (See Figure 20.) The $x - \pi$ phase diagram is now replaced by an a–π phase diagram. The $\dot{\pi} = 0$ locus still shifts to the right, along with the saddle path. But now on impact, the economy moves to a point on the new saddle path to the right of the old $\dot{\pi} = 0$ locus, since the MP curve jumps down and to the right in response to a fall in π. (See Figure 21. Note that π cannot jump up, since then the economy would have to jump to a point on the saddle point to the left of the old $\dot{\pi} = 0$ locus, which is impossible if π jumps up.) From that point, the economy moves along the saddle path toward medium-run equilibrium in a familiar way. The initial jump downward in the monetary policy curve combined with the downward shift in the NRR curve makes it ambiguous what happens to output in the short run. Finally, suppose the central bank follows a real interest rate rule, so that the MP curve does not jump in response to inflation. This case is left as an exercise for the reader.

20.7.2 Technology Shocks

Because it increases their permanent income, an improvement in technology can shift the labor supply curve back, but because it raises the long-run real interest rate, an improvement in technology can shift the labor supply curve out. A sufficient statistic for both of these effects is the movement in the marginal utility of consumption λ induced by the improvement in technology.

[4]The perverse short-run effect of this and other positive real shocks on output depends crucially on a non-zero interest elasticity of money demand; if the LM curve were vertical, real shocks would initially have no effect on output.

For simplicity, we will assume in this subsection that the wealth effect of the increase in permanent income is cancelled out by the long-run real interest rate effect. That is, we will assume that the improvement in technology has no impact effect on λ and therefore no impact effect on labor supply. To the extent that a technology shock does have an impact effect on labor supply, the responses discussed in the previous subsection about labor supply shocks can be added to the direct effects of the technology shock discussed in this subsection.

The effects of technology shocks are quite similar to the effects of labor supply shocks.

One of the most unambiguous results from the Basic Real Business Cycle model is that a permanent improvement in technology results in an increase in output on impact. Since the Basic Real Business Cycle model has perfectly flexible prices, this increase in output corresponds to an increase in medium-run equilibrium (full-employment) output y^f.

For any given capital stock K an increase in the level of labor-augmenting technology Z reduces the amount of labor N needed to produce any given amount of output. With an unchanging labor supply curve, this reduction in the labor requirements for given Y reduces the real wage W associated with that value of Y. (See Figure 22.) Since both N and W for given output fall, the rental rate

$$R = \frac{\theta}{1-\theta}\frac{WN}{K}$$

for given Y must fall. As a consequence, the NRR curve, which gives the mapping from (log) output y to the net rental rate $R - \delta$, shifts downward.

The downward shift in the NRR curve, combined with the increase in full-employment output y^f, guarantees that the full-employment LM curve must shift to the right. As a consequence, the new $\dot{\pi} = 0$ locus must be to right of the old one. The dynamics on the short-run x–π phase diagram are very similar to those induced by an increase in labor supply. Inflation π jumps down on impact, resulting in an upward jump in the LM curve. In concert with the downward shift of the NRR curve, the upward jump in the LM curve leads to a short-run reduction in output y, even though the improvement in technology increases full-employment output y^f.

20.7.3 Government Purchase Shocks

Government spending is divided into *transfers* (such as social security payments) which can be though of as negative taxes, and government purchases of goods and services. Turning to government purchases of goods and services, some types of government purchases are close substitutes for private purchases, and can be lumped in with private consumption or investment. Other types of government purchases, such as most military spending, have little interaction with the private economy *except for the use of resources.* This is the type of government purchases we focus on here.

Long-lasting increases in government purchases, whether financed by current taxes or by borrowing backed by future taxes, makes households significantly poorer. This impoverishment of households makes them eager to earn more money by working more. Thus, the impoverishment effect of government purchases tends to increase labor supply.

If the additional government purchases are financed by current lump-sum taxes and by borrowing backed by future taxes, the impoverishment effect is the key effect on labor supply. However, if the additional government purchases are financed in part by increases in the current labor income tax rate, this increase in the marginal labor income tax rate has an adverse effect on labor supply that acts in the opposite direction.

Given households obeying the permanent income hypothesis, relatively brief increases in government purchases—lasting, say, four to five years time—might have only a trivial impoverishment effect on labor supply.

In this subsection, we will analyze the effects of an increase in government purchases is either brief enough to have little effect on labor supply, or is financed by an increase in labor income taxes that cancels out its effect on labor supply. Depending on how long it takes prices to adjust, it may be possible to have an increase in government purchases that lasts long enough for prices to adjust to a new medium-run (full-employment) equilibrium, but not long enough to have a significant effect on labor supply. To the extent an increase in government purchases affects labor supply, the effects of the previous subsection on labor supply shocks can be added to the direct effects of an increase in government purchases discussed here.

In the absence of investment adjustment costs, the direct effect of an increase in government purchases is simply to crowd out investment one-for-one, with no effect on medium-run equilibrium output y^f. Moreover, in the absence of any change in labor supply, the increase in government purchases has no effect on the level of N or W for a given level of output and therefore no effect on the NRR curve. The full-employment LM curve does not move, and so the $\dot{\pi} = 0$ locus remains where it started. Other than the increase in government purchases G and the corresponding reduction in investment I, nothing happens, either in the medium-run equilibrium or in the short-run! Of course, the qualifications we have made, point to some of the places one might look to restore the idea that government purchases should have an effect on output. But restoring the traditional Keynesian story is much more difficult than one might think.

20.7.4 Industrial Organization Shocks

Suppose the desired markup falls because of an increase in consumers' readiness to switch between brands.[5] This results in lower desired prices in an otherwise similar economic situation. Ignoring any effect on labor supply, this results in an outward shift in the $\dot{\pi} = 0$, with consequent dynamics as shown on the phase diagram. The outward shift in the $\dot{\pi} = 0$ locus guarantees an increase in x^f and an outward shift in the full-employment LM curve. With no shift in labor supply or in technology, the NRR curve stays put. Thus, the outward shift in the full-employment LM curve implies a shift in the y^f–r^f point up along the NRR curve. Other than the absence of a shift in the NRR curve, the analysis is similar to that for an increase in labor supply or an improvement in technology.

20.7.5 Open Economy Shocks

Events abroad can have an important effect on an economy. There is no way to deal will all of the important open economy issues here, but it is useful to relate open economy shocks to the shocks we have discussed so far while tacitly assuming a closed economy.

1. *The Balance of Trade.* An increase in the trade surplus acts much like an increase in government purchases, since, like government purchases, it widens the gap between output and $C + I$. Since exports may not generate as much money demand as other forms of output, a change in the balance of trade can also cause a change in what we have been calling autonomous velocity v.

2. *The Terms of Trade.* An improvement in the terms of trade—the quantity of imports than can be obtained from a given amount of exports—

[5] Ultimately, the increased competitive pressure will cause a shakeout in which many firms will exit, leaving fewer, larger firms that can take advantage of greater scale economies to avoid losses in the more competitive environment. But by the hierarchy of time-scales we have assumed, this occurs after the economy has already reached the new medium-run equilibrium.

acts much like an improvement in technology—which increases the quantity of output that can be obtained from a given amount of input. Changes in the terms of trade often also cause changes in the balance of trade.

3. *Trade Barriers.* A reduction in trade barriers is like an improvement in the terms of trade. Like other improvements in the terms of trade, a reduction in trade barriers can also result in an increase in the effective level of competition—yielding a positive industrial organization shock.

4. *The Exchange Rate.* An increase in the exchange rate causes an improvement in the terms of trade, and often causes a reduction in the trade surplus. If the exchange rate will go back to its previous value over time, the fall in the exchange rate from its high value tends to push down the long-run real interest rate. (A reduction in the long-run real interest rate tends to reduce labor supply.)

5. *International Oil and Raw Materials Prices.* Increases in international oil and raw materials prices represent a worsening of the terms of trade—often with the associated effects on the balance of trade and on autonomous velocity. To complicate matters further, increases in oil and raw materials prices are sometimes caused by a worldwide increase in long-term real interest rates.

The main message to be drawn from this itemization is that international shocks—including oil shocks—can have quite complex effects. Open economy macroeconomics is inherently more difficult than closed economy macroeconomics and is correspondingly less well understood.

In its effects (though not of course in its origins), a terms of trade shock is like a technology shock, since trade is like a production function that converts exports (the inputs) into imports (the outputs). To treat a change in the terms of trade as a technology shock, one must add trade to the list of sectors whose production function one considers. In its effects, an increase in the trade surplus is similar to an increase in autonomous government purchases, since both affect the material balance condition in the same way.

20.8 The Composition of Output

So far, we have used the Basic Neomonetarist Model to determine output and other variables without regard to the composition of output as between consumption and investment. The key to the analysis of the composition of output is to realize that with no investment adjustment costs, investment is passively determined as a residual *after* consumption is determined. (Government purchases are exogenous, and net exports are determined in a complex way in an open economy model.)

The short-run movements in consumption are entirely governed by interaction between consumption and labor in the utility function. With additively separable utility between consumption and labor, consumption is always very close to its medium-run equilibrium value C^f, since the marginal utility of consumption λ is determined by medium-run considerations, not by short-run considerations.

In response to a monetary expansion, consumption only increases to the extent that consumption is complementary to labor in the utility function. With additively separable utility, consumption would move very little.

The effect on full-employment consumption of an increase in labor supply depends on the source of the increase in labor supply:

1. an increase in patience increases labor supply but reduces C^f;

2. fears of hard times (or the certainty of hard times) in the future increase labor supply but reduce C^f;

3. an increase in the long-term real interest rate increases labor supply, but reduces C^f;

4. an influx of new potential workers into the economy, either by immigration or by the maturation of a large cohort of young people raises both labor supply and C^f;

5. a reduction in the effective marginal tax rate on labor raises both labor supply and C^f;

6. a reduction in the market power of unions or in other labor market distortions raises both labor supply and C^f;

7. an increase in the desire for consumption goods relative to leisure, raises both labor supply and C^f;

8. a reduction in the need to spend time at home raises labor supply and increases C^f if it is because of household labor-saving devices, but reduces C^f if it is because of having fewer children at home.

Other than any effects operating through changes in labor supply from sources 2 and 3, an improvement in technology causes C^f to change only insofar as consumption is complementary with labor. But even then, the change depends on the sign of the change in N^f, which is ambiguous. If utility is additively separable between consumption and labor, consumption is determined by λ alone.

Other than its effects on λ, an increase in government purchases has no effect on consumption.

Finally, a reduction in the desired markup tends to increase C^f.

Even when medium-run equilibrium consumption C^f increases, consumption can fall in the short-run if consumption is complementary with labor.

Further analysis of consumption and investment in medium-run equilibrium requires a deeper investigation of medium-run dynamics and long-run equilibrium.

Exercises

1. If $Y = F(K, ZN)$, show that the marginal product of capital $F_K(K, ZN)$ can be expressed as a function of Y and K that is invariant to changes in Z.

2. Analyze the effects of a temporary decrease in the growth rate of autonomous-velocity-adjusted money μ.

3. Analyze the effects of an anticipated decrease in the growth rate of money μ.

Chapter 21

Graphical Analysis of the Basic Real Business Cycle Model

In order to give a small taste of what dynamic optimization methods can do, we would like to present, here and now, a graphical treatment of the Basic Real Business Cycle Model. Figuratively, we would like to give an aerial view of the near side of the forest, before hiking into the woodlands to study the trees.

In addition to illustrating some of the power of these methods, an aerial view makes it easier to give a sense of what the first half of the book is all about. Everything in the first half of the book is aimed at a better understanding of the foundations and workings of the Basic Real Business Cycle Model and its extensions.

The key extensions of the Basic Real Business Cycle Model are (1) the addition of the investment adjustment costs that allow one to formalize the Q theory of investment, and (2) the addition of the imperfect price flexibility that allows one to think more deeply about the short-run effects of monetary policy.

A better understanding of the workings of any of these models involves examining (a) the partial equilibrium determinants of household behavior, firm behavior and the resulting supply and demand in each market, (b) the interaction of markets with each other and (c) the interaction of the past and the anticipated future with the present. It involves *quantitative* as well as *qualitative* analysis, from the calibration of the underlying parameters of a model to the calculation of how all of the forces in a model work together in general equilibrium.

In the broad-brush treatment of the Basic Real Business Cycle Model that follows, we make many claims that will not be fully backed up until later chapters. But every statement *is* fully documented later on. In effect, we are stating a result, with later chapters providing the proof. In this initial view of the model, it is better to concentrate on the overall picture, without getting bogged down in every detail, even details presented immediately below.

21.1 The Outlines of the Basic Real Business Cycle Model

Because it has no distortions in it the Basic Real Business Cycle Model is the solution to a social planner's problem:

$$\max_{c,n,y,i} \int_0^\infty e^{-\rho t} u(c,n) dt$$

s.t.,

$$\begin{aligned} \dot{k} &= i - \delta k, \\ y &= c + i + g, \end{aligned}$$

and

$$y = F(k, zn),$$

The quantities c, n, y, i are aggregate consumption, labor, output and investment. ρ is the utility discount rate (measuring "impatience"). u is a concave utility function, increasing in c and decreasing in n. k is the capital stock and δ is the depreciation rate. g is government purchases. Finally, F is a constant returns to scale production function and z is the level of labor-augmenting technology. Government purchases g and the level of labor-augmenting technology z are the exogenous driving variables of the model. (In order to avoid introducing distortions into the economy, the government purchases must be financed by lump-sum taxes.)

Though they are only implicit in the social planner's problem, a set of prices arise in describing the Real Business Cycle Model as a competitive equilibrium. The key prices are the real wage w, the rental rate or marginal revenue product of capital R, the real interest rate r and the marginal value of capital λ. Of these prices, the marginal value of capital λ is probably the least familiar. As an alternative to thinking of λ as the marginal value of capital, λ can be thought of as the marginal value of output in terms of utils, whether consumption or investment. Since each unit of investment adds one unit to the capital stock annually, and output can be used equally well for either investment or consumption, the marginal value of capital λ is equal to the marginal value of investment and *the marginal utility of consumption.*

In the Basic Real Business Cycle Model, the capital stock k indicates everything that needs to be known about the past. The marginal value of capital λ indicates everything that needs to be known about expectations of the future in order to determine what happens now.

If new information alters expectations of the future, the marginal value of capital λ can jump in response to that information; but in the absence of new information, λ can only evolve according to the real interest rate. A high real interest rate r makes goods in the future cheaper relative to goods now. Thus, λ (the price of goods in utils) must fall if the real interest rate is high. A low real interest rate r makes goods in the future more expensive relative to goods now. Thus, λ (the price of goods in utils) must rise if the real interest rate is low.

21.2 The Phase Diagram and its Economic Geography

For given values of the exogenous variables g and z, the dynamics of the capital stock and the marginal value of capital are given by the phase diagram shown in Figure 1 with k on the horizontal axis and λ on the vertical axis. Demonstrating as much (that is, showing that we have drawn the phase diagram correctly) requires examining the behavior of the economy at various points on the phase diagram—what one might call the *economic geography* of the phase diagram.

Two pieces of equipment are essential before setting out on our explorations. First, we need to pack the powerful tool of supply and demand in the labor market. As shown in Figure 2, except for the real wage w, which is indicated on the vertical axis, there is a clean separation between the determinants of labor supply and labor demand. The position of the labor supply curve on the graph only depends on λ, the marginal value of capital, while the position of the labor demand curve only depends on k and z, the capital stock and the level of technology. A higher marginal value of capital λ increases labor supply because it makes households want to work harder in order to save and invest. A higher capital stock always increases labor demand. Better technology also typically increases labor demand, something we will assume here.[1]

Second, we need the factor price possibility frontier shown in Figure 3. For any given level of technology z, only certain combinations of w and R are possible; a high capital/labor ratio k/n implies a high real wage w and a low marginal revenue product of capital R, while a low capital/labor ratio k/n implies a low real wage w and a high marginal revenue product of capital R. In this model, the only way both w and R can increase is if technology improves. The effect of the capital/labor ratio k/n on the real wage w is already implicit in the labor demand curve, but the factor price possibility frontier clarifies the linked effects of k/n on w *and* R.

21.2.1 The Great White North: The Effects of Higher λ

Journeying first to the north, let's examine what happens when the marginal value of capital λ is higher. (See Figure 4.)

First, consumption falls, since the marginal utility of consumption, which equals λ, has increased.

Second, the labor supply curve shifts out because the increase in the marginal value of capital leads to an increased eagerness to save and invest. The increase in labor supply leads to an increase in the quantity of labor n and a fall in the real wage w. (See Figure 5.)

On the quantity side, the increase in labor input n leads to more output y, which with the reduction in consumption c implies more total saving and an increase in investment i. The increase in investment in turn raises the rate of increase $\dot{k}$ in the capital stock k.

On the price side, the fall in the real wage w, which arises from the fall in the capital/labor ratio k/n as n increases, implies an increase in the marginal revenue product of capital R. The increase in R can also be seen as a reflection of the downward slope of the factor price possibility frontier. (See Figure 6.) By increasing the rate of return to capital, the increase in the marginal revenue product of capital R causes an increase in the real interest rate r. This increase in the real interest rate r tends to make future goods cheaper relative to goods now, lowering the rate of increase $\dot{\lambda}$ of the marginal value of capital.

To the south, a lower marginal value of capital leads to opposite effects.

[1] As shown in Chapter **??**, in the face of a low enough elasticity of substitution between capital and labor, labor-augmenting technological progress can reduce labor demand until the capital stock has a chance to accumulate in response to that technological progress.

21.2.2 To the East: The Effects of Higher k

Now, heading due east, let's examine what happens when the capital stock k is higher, with λ held fixed. (See Figure 7.)

First, nothing happens to consumption if it is additively separable from labor. What happens to consumption c is ambiguous if consumption is not additively separable from labor.[2] However, since consumption is first and foremost a function of the unchanging marginal utility of consumption λ, consumption will not change enough to overturn the effects of changes in the quantity of labor on household saving.

Second, labor demand increases because of the higher capital stock. More labor is needed to work the additional factories and machines. The increase in labor demand leads to an increase in the quantity of labor n and a rise in the real wage w. (See Figure 8.)

On the quantity side, the increase in the inputs of both capital and labor leads to more output y. The increase in output overwhelms any change in consumption c to increase the total resources available for investment, leading to an increase in i. Although the higher capital stock leads to more depreciation, since dynamic optimization by a representative agent guarantees dynamic efficiency, the extra depreciation is less than the extra output produced by the higher capital stock. Adding in the extra output produced by the additional labor, minus any change in consumption (always smaller), the rate of increase $\dot{k}$ in the capital stock is guaranteed to increase.

On the price side, the increase in the real wage w implies a fall in the marginal revenue product of capital R as can be seen from the downward slope of the factor price possibility frontier.[3] (See Figure 9.) By reducing the rate of return to capital, the reduction in R causes a reduction in the real interest rate r. This reduction in the real interest rate r tends to make future goods more expensive relative to goods now, raising the rate of increase $\dot{\lambda}$ of the marginal value of capital.

To the west, a lower capital stock leads to opposite effects.

---

[2]We argue in chapter **??** that consumption is likely to be positively related to the real wage, which would make it increase in response to either higher k or an improvement in technology.

[3]The capital/labor ratio k/n increases because the change in n is less than the change in k that induces it.

21.2.3 The Implied Dynamics

The preceding arguments about the economic geography of the phase diagram imply that $\dot{k} > 0$ above and to the right (north and east) of the line on which $\dot{k} = 0$; while $\dot{k} < 0$ below and to the left (south and west) of the $\dot{k} = 0$ locus. Therefore, the $\dot{k} = 0$ locus must be downward-sloping, as shown in Figure 1.

Similarly, the preceding arguments imply that imply that $\dot{\lambda} > 0$ *below* and to the right (south and east) of the line on which $\dot{\lambda} = 0$; while $\dot{\lambda} < 0$ *above* and to the left (north and west) of the $\dot{\lambda} = 0$ locus. Therefore, the $\dot{\lambda} = 0$ locus must be upward-sloping, as shown in Figure 1.

Noting the dynamics for each of the four quarters marked out by the $\dot{k} = 0$ locus and $\dot{\lambda} = 0$ locus, it is clear from the phase diagram that the saddle path must be downward-sloping. In fact, it must be more steeply downward-sloping than the $\dot{k} = 0$ locus. If the capital stock is low to begin with, the marginal value of capital starts high because of that shortage of capital, then falls gradually as the capital stock approaches its higher steady-state value at the intersection of the $\dot{k} = 0$ and $\dot{\lambda} = 0$ locus. If the capital stock is high to begin with, the marginal value of capital starts low because of that surplus of capital, then rises gradually as the capital stock approaches its lower steady-state value.

21.3 The Effects of an Increase in Government Purchases

If government purchases g increase, the $\dot{k} = 0$ locus shifts up, while the $\dot{\lambda} = 0$ locus stays put. (See Figure 10.) Why? The key is that except for the crowding out of investment i, all of the effects of an increase in government purchases work *through* the capital stock k and the marginal value of capital λ. In the short-run, before the effects on the capital stock take hold, most of the effects result from agents expecting to be poorer because of the higher government purchases and higher lump-sum taxes in the future. But in determining economic behavior now, λ is a sufficient statistic for all expectations of the future. Any expected impoverishment due to higher government spending in the future will be reflected in a change in the marginal value of capital (and marginal utility of consumption) λ.

For a *given* capital stock k, level of technology z and marginal value of capital (and marginal utility of consumption) λ, an increase in government purchases leaves current consumption, current labor supply and current labor demand unchanged. (Remember that the position of the labor supply curve is fully determined by λ, while the position of the labor demand curve is fully determined by k and z.) With labor supply and demand unchanged, n and w remain unchanged (Figure 2 with no change).

On the price side, the marginal revenue product of capital R remains unchanged (think of the factor price possibility frontier, which has not shifted), leaving the real interest rate r and the rate of increase $\dot{\lambda}$ in the marginal value of capital. The fact that for given k and λ an increase in government purchases has no effect on $\dot{\lambda}$ implies that the $\dot{\lambda} = 0$ locus does not shift.

On the quantity side, with both inputs, k and n, unchanged, output is unchanged. However, investment i is given by $i = y - c - g$ and must fall, dragging down $\dot{k}$ with it. The fact that for given k and λ an increase in government purchases causes $\dot{k}$ to fall implies that the $\dot{k} = 0$ locus must shift up to leave points where $\dot{k}$ was zero previously underneath the *new* $\dot{k} = 0$ locus, in the region where $\dot{k} < 0$.

21.4 The Effects of Improved Technology

If the labor-augmenting technology z improves, the $\dot{k} = 0$ locus shifts down, while the $\dot{\lambda} = 0$ locus shifts down and to the right. Since, if technology improves permanently, the new steady state must involve a higher capital stock, the shift in the $\dot{\lambda} = 0$ locus must dominate in its effect on the value of k at the intersection of the two curves. (See Figure 11.)

Here is why the $\dot{k} = 0$ and $\dot{\lambda} = 0$ curves shift as claimed. For *given* k and λ, the main effect of an improvement in technology is (at least in the usual case) to increase labor demand. The increase in labor demand raises both n and w. (See Figure 12.)

On the quantity side, output y increases both because of the direct effect of the improved technology and because of the increase labor input n. The movement in consumption c is ambiguous, but it is not enough to overturn the positive effect of the increased labor input on household saving, let alone the direct effect of the improved technology on output and therefore the resources available for investment. Therefore, investment i and the rate of increase $\dot{k}$ in the capital stock go up. The increase in $\dot{k}$ for given λ and k means that the $\dot{k} = 0$ locus must shift down, to leave points where $\dot{k}$ was zero before *above* the new $\dot{k} = 0$ locus where $\dot{k} > 0$.

On the price side, an outward shift in the factor price possibility frontier allows the marginal revenue product of capital R to go up as well as the real wage. (See Figure 13.) Indeed, with both z and n increasing, effective labor input zn definitely increases, raising the effective amount of labor per unit of capital. This ensures that R increases. The increase in the rate of return to capital raises the real interest rate r, which in turn makes goods in the future cheaper relative to goods now, leading to a fall in the rate of increase $\dot{\lambda}$ in the marginal value of capital. The fall in $\dot{\lambda}$ for given k and λ means that the $\dot{\lambda}$ locus must shift to the right, to leave points where $\dot{\lambda}$ was zero before to the *left* of the new $\dot{\lambda} = 0$ locus where $\dot{\lambda} < 0$.

Exercises

1. With the economy initially in steady state, graph the dynamic response on the phase diagram to a permanent increase in government purchases.

2. Graph the dynamic responses on the phase diagram to temporary increases in government purchases. Show that the dynamic response is qualitatively different for a long-lived increase in government purchases than it is for a short-lived increase in government purchases.

3. Graph the dynamic response on the phase diagram to an anticipated increase in government purchases. Is there a qualitative difference in the response to an anticipated increase in the distant future as opposed to an anticipated increase in the near future?

4. Graph the dynamic response on the phase diagram to a permanent improvement in technology. There are at least two different possible cases that need to be shown.

5. Graph the dynamic response on the phase diagram to a temporary improvement in technology, showing all qualitatively different possibilities.

Chapter 22

The Contemporaneous Equilibrium of the Basic Real Business Cycle Model

Knowing the location of an economy on the phase diagram, together with the current values of the exogenous variables is enough to determine the current behavior of the economy. When the capital stock k is the only state variable, k tells everything one needs to know about the past, while the marginal value of capital λ tells everything one needs to know about the future, and the exogenous variables tell everything else one needs to know about the present in order to determine the behavior of all of the control variables. Thus, determining the contemporaneous equilibrium of an economy is a matter of expressing the behavior of all of the variables of interest (including the growth rates of k and λ) in terms of the contemporaneous values of k, λ and the exogenous variables.

22.1 The Approach of Using Focal Variables

The models we are looking at are quite intricate. It is easy to get bogged down in complexity. One technique that helps greatly in finding a path through the thicket of complexity is the use of focal variables. By this we mean expressing variables in terms of one or two key contemporaneous variables in addition to k, λ and the exogenous variables, while carrying along the equations that determine those key contemporaneous variables. Consider how, in analyzing the Basic Real Business Cycle model graphically, how much we deduced from what happened in the labor market. Using the quantity of labor n as a focal variable accomplishes something similar—algebraically rather than graphically.

Choosing a focal variable is a matter of judgment. The behavioral implications must be equivalent regardless of which variable one chooses as a focal variable, but calculations will be simplified most by the choice of a variable that is part of several of the most fundamental relationships of a model. The variables that are central to a graphical analysis provides a good clue to where to look for focal variables. For the Basic Real Business Cycle model, n is an excellent choice for a focal variable. In the Q-RBC model, it works well to choose n and the marginal value of investment $\mu = \xi\nu$ as the two focal variables, as one might expect from the centrality of labor market equilibrium (with n on the horizontal axis) and equilibrium of investment demand and saving supply (with μ on the vertical axis) in the Q-RBC model.

Since many relationships involving the marginal value of investment $\mu = \xi\nu$ carry over from the Basic RBC model to the Q-RBC model, it is helpful to begin by using μ as a focal variable even in analyzing the Basic RBC model, even though in the Basic RBC model we will soon substitute in λ for μ.

22.2 Focal Variable Equations for c, y, i, $\dot{k}$, w, R, r and $\dot{\lambda}$

22.2.1 Quantities: c, y, i and $\dot{k}$

Consumption c

Since $p_c = \frac{1}{\xi}$ means that $\check{p}_c = -\check{\xi}$, (15.12) implies that

$$\check{c} = -s\check{\mu} + (1-s)\tau\check{n} + s\check{\xi}. \tag{22.1}$$

Output with Rescaled Consumption y

As noted in Chapter 19 [?], if we use y to represent output with rescaled consumption,

$$y = \frac{c}{\xi} + i + g,$$

the production relations for both sectors add up to

$$y = znf\left(\frac{k}{zn}\right),$$

where n and k are the aggregate quantities of labor and capital and z is the overall level of labor-augmenting technology. This equation is formally the same as the one we log-linearized to

$$\check{y} = \theta\check{k} + (1-\theta)\check{n} + (1-\theta)\check{z} \tag{22.2}$$

as long as one is assuming constant returns to scale.

Investment i

Log-linearizing the aggregate expenditure equation $y = (c/\xi) + i + g$ yields

$$\check{y} = \zeta_c[\check{c} - \check{\xi}] + \zeta_i\check{\imath} + \zeta_g\check{g}.$$

Solving for $\check{\imath}$,

$$\check{\imath} = \frac{\check{y} - \zeta_c[\check{c} - \check{\xi}] - \zeta_g\check{g}}{\zeta_i}. \tag{22.3}$$

Using the fact that the steady-state investment share in output is

$$\zeta_i = \frac{\delta\theta}{\rho+\delta},$$

and the fact that

$$\zeta_c = \frac{p_c^* c^*}{y^*} = \frac{p_c^* c^*}{w^* n^*}\frac{w^* n^*}{y^*} = \frac{\frac{w^* n^*}{y^*}}{\frac{w^* n^*}{p_c^* c^*}} = \frac{1-\theta}{\tau},$$

but still holding the fact that $\zeta_c + \zeta_i + \zeta_g = 1$ in reserve,

$$\begin{aligned}
\check{\imath} &= \frac{\rho+\delta}{\delta\theta}\left\{\check{y} - \frac{(1-\theta)}{\tau}[\check{c} - \check{\xi}] - \zeta_g\check{g}\right\} \qquad (22.4)\\
&= \frac{\rho+\delta}{\delta\theta}\left\{\theta\check{k} + (1-\theta)[\check{n} + \check{z}] - \frac{(1-\theta)}{\tau}[-s\check{\mu} + (1-s)\tau\check{n} + s\check{\xi} - \check{\xi}] - \zeta_g\check{g}\right\}\\
&= \frac{\rho+\delta}{\delta\theta}\left\{\theta\check{k} + (1-\theta)[\check{n} - (1-s)\check{n} + \frac{s}{\tau}\check{\mu} + \check{z} + \frac{(1-s)}{\tau}\check{\xi}] - \zeta_g\check{g}\right\}\\
&= \frac{\rho+\delta}{\delta}\check{k} + \frac{\rho+\delta}{\delta}\frac{1-\theta}{\theta}[s\check{n} + \frac{s}{\tau}\check{\mu} + \check{z} + \frac{(1-s)}{\tau}\check{\xi} - \frac{\zeta_g}{1-\theta}\check{g}].
\end{aligned}$$

The coefficient on $\check{n}$ is proportional to s because s is the fraction of the income from additional labor that is saved rather than spent on consumption. Indeed, a useful concept is what we can call "primary household saving":

$$\text{primary household saving} = wn - p_c c.$$

The divisia index of the change in the *quantity* of primary household saving (holding the prices w and p_c fixed) is

$$\begin{aligned} w\tilde{n} - p_c\tilde{c} &= wn\left[\frac{\tilde{n}}{n} - \frac{c}{wn}\frac{\tilde{c}}{c}\right] \qquad (22.5)\\ &\approx w^* n^*[\check{n} - \frac{\check{c}}{\tau}] \\ &= w^* n^* \left\{\check{n} - [-s(\check{\mu} - \check{\xi}) + (1-s)\tau\check{n}\right\} \\ &= w^* n^* \left\{s\check{n} + \frac{s}{\tau}(\check{\mu} - \check{\xi})\right\} \\ &= w^* n^* \frac{s}{\tau}[\tau\check{n} + \check{\mu} - \check{\xi}]. \end{aligned}$$

This index of "primary saving" is proportional to the actual reduction in felicity caused by changes in c and n. The coefficient of ξ in (22.4) is different from what one would get from (22.5) because in addition to the household's pains, there is an expansion of production possibilities when the consumption technology improves.

Capital Accumulation $\dot{k}$

Substituting from (22.4) into the simple log-linearized accumulation equation

$$\dot{\check{k}} = \delta(\check{\imath} - \check{k})$$

(which is valid even when there are investment adjustment costs) yields

$$\begin{aligned} \dot{\check{k}} &= (\rho+\delta)\check{k} + \frac{(\rho+\delta)(1-\theta)}{\theta}[s\check{n} + \frac{s}{\tau}\check{\mu} + \check{z} + \frac{(1-s)}{\tau}\check{\xi} - \frac{\zeta_g}{1-\theta}\check{g}] - \delta\check{k} \qquad (22.6)\\ &= \rho\check{k} + \frac{(\rho+\delta)(1-\theta)}{\theta}[s\check{n} + \frac{s}{\tau}\check{\mu} + \check{z} + \frac{(1-s)}{\tau}\check{\xi} - \frac{\zeta_g}{1-\theta}\check{g}]. \end{aligned}$$

22.2.2 Factor Prices: w and R

The Real Wage (in Terms of the Investment Goods Numeraire) w

Both the labor demand equation and the labor supply equation from the previous chapter are focal variable equations for the real wage. This illustrates the fact that focal variable equations are somewhat arbitrary. Whether we choose the labor demand equation or the labor supply equation as the focal variable equation for the real wage, we will get the same results in the end. However, the simpler the equation we begin with, the simpler will be the path to the end result. Of these two equations, the labor demand equation is simpler, so we will work with the labor demand equation:

$$\check{w} = \frac{\theta}{\sigma}(\check{k} - \check{n}) + \left(1 - \frac{\theta}{\sigma}\right)\check{z}. \tag{22.7}$$

(Note, however, that in Chapter **??**, sticky prices will invalidate the labor demand equation in the short run, so we will use the still-valid labor supply equation.)

The Real Rental Rate

From (16.16),

$$\check{R} = \frac{1-\theta}{\sigma}[\check{z} + \check{n} - \check{k}]. \tag{22.8}$$

22.2.3 The Real Interest Rate and the Euler Equation

All of the focal variable equations we have looked at so far are valid in both the Basic RBC model and the Q-RBC model. The only difference so far is that in analyzing the Basic RBC model we will want to substitute $\check{\lambda}$ everywhere that $\check{\mu}$ appears, since in the absence of investment adjustment costs $\mu = \lambda$.

In addition to $\mu = \lambda$, there are two equations of importance that are valid in the Basic RBC model but not in the Q-RBC model: the Basic RBC equation for the real interest rate and the Basic RBC Euler equation for the rate of change in the marginal value of capital λ.

The Real Interest Rate (in Terms of the Investment Goods Numeraire) r

The real interest rate is the one variable we have dealt with for which it is more convenient thinking about the *absolute deviation* from the steady-state value rather than the logarithmic deviation. Linearizing the equation

$$r = R - \delta$$

around the steady state yields the equation

$$\tilde{r} = \tilde{R} = R^* \check{R} = (\rho + \delta)\check{R}$$

Substituting in $\check{R}$ from (**??**),

$$\tilde{r} = (\rho + \delta)\frac{1-\theta}{\sigma}[\check{z} + \check{n} - \check{k}]. \tag{22.9}$$

The Euler Equation: The Growth Rate of the Marginal Value of Capital λ

Linearizing the equation

$$\frac{\dot{\lambda}}{\lambda} = \rho - r \tag{22.10}$$

around the steady state readily yields

$$\dot{\check{\lambda}} = -\tilde{r}.$$

Substituting in from (22.9),

$$\dot{\check{\lambda}} = \frac{(\rho + \delta)(1-\theta)}{\sigma}[\check{k} - \check{z} - \check{n}]. \tag{22.11}$$

22.2.4 Rounding Out the Circle: From n to n

Finally, it is necessary to include the equation that determines n itself. Equation (22.12) is what we need. Rearranging to put 0 on the left, and writing $\check{\mu} - \check{\xi}$ for $\check{\nu}$,

$$0 = \frac{\theta}{\sigma}\check{k} + \left(1 - \frac{\theta}{\sigma}\right)\check{z} + s\check{\mu} + (1-s)\check{\xi} - \left[\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau\right]\check{n}. \tag{22.12}$$

22.2.5 Focal Variable Equations in Matrix Form

Assembled in matrix form, the forgoing focal variable equations that are valid in both the Basic RBC model and the Q-RBC model are

$$\begin{bmatrix} \dot{\check{k}} \\ 0 \\ \check{c} \\ \check{y} \\ \check{i} \\ \check{w} \\ \check{R} \end{bmatrix} = \begin{bmatrix} \rho & \frac{(\rho+\delta)(1-\theta)s}{\theta\tau} & \frac{(\rho+\delta)(1-\theta)s}{\theta} \\ \frac{\theta}{\sigma} & s & -\frac{\theta}{\sigma} - \frac{1}{\eta} - (1-s)\tau \\ 0 & -s & (1-s)\tau \\ \theta & 0 & (1-\theta) \\ \frac{\rho+\delta}{\delta} & \frac{(\rho+\delta)(1-\theta)s}{\delta\theta\tau} & \frac{(\rho+\delta)(1-\theta)s}{\delta\theta} \\ \frac{\theta}{\sigma} & 0 & -\frac{\theta}{\sigma} \\ -\frac{(1-\theta)}{\sigma} & 0 & \frac{(1-\theta)}{\sigma} \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{\mu} \\ \check{n} \end{bmatrix} + \begin{bmatrix} \frac{(\rho+\delta)(1-\theta)}{\theta} & \frac{(\rho+\delta)(1-\theta)(1-s)}{\theta\tau} & -\frac{(\rho+\delta)\zeta_g}{\theta} \\ 1-\frac{\theta}{\sigma} & 1-s & 0 \\ 0 & s & 0 \\ 1-\theta & 0 & 0 \\ \frac{(\rho+\delta)(1-\theta)}{\delta\theta} & \frac{(\rho+\delta)(1-\theta)(1-s)}{\delta\theta\tau} & -\frac{(\rho+\delta)\zeta_g}{\delta\theta} \\ 1-\frac{\theta}{\sigma} & 0 & 0 \\ \frac{(1-\theta)}{\sigma} & 0 & 0 \end{bmatrix} \begin{bmatrix} \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix} \tag{22.13}$$

Now, assuming away adjustment costs means that $\check{\mu} = \check{\lambda}$, and gives us the additional structure we need in order to add equations for $\tilde{r}$ and $\dot{\check{\lambda}}$. It is convenient to arrange the equations as two matrix equations:

$$\begin{bmatrix} \dot{\check{k}} \\ \dot{\check{\lambda}} \\ 0 \end{bmatrix} = \begin{bmatrix} \rho & \frac{(\rho+\delta)(1-\theta)s}{\theta\tau} & \frac{(\rho+\delta)(1-\theta)s}{\theta} \\ \frac{(\rho+\delta)(1-\theta)}{\sigma} & 0 & -\frac{(\rho+\delta)(1-\theta)}{\sigma} \\ \frac{\theta}{\sigma} & s & -\frac{\theta}{\sigma} - \frac{1}{\eta} - (1-s)\tau \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{\lambda} \\ \check{n} \end{bmatrix} + \begin{bmatrix} \frac{(\rho+\delta)(1-\theta)}{\theta} & \frac{(\rho+\delta)(1-\theta)(1-s)}{\theta\tau} & -\frac{(\rho+\delta)\zeta_g}{\theta} \\ -\frac{(\rho+\delta)(1-\theta)}{\sigma} & 0 & 0 \\ 1-\frac{\theta}{\sigma} & 1-s & 0 \end{bmatrix} \begin{bmatrix} \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix} \tag{22.14}$$

$$\begin{bmatrix} \check{c} \\ \check{y} \\ \check{i} \\ \check{w} \\ \check{R} \\ \tilde{r} \end{bmatrix} = \begin{bmatrix} 0 & -s & (1-s)\tau \\ \theta & 0 & (1-\theta) \\ \frac{\rho+\delta}{\delta} & \frac{(\rho+\delta)(1-\theta)s}{\delta\theta\tau} & \frac{(\rho+\delta)(1-\theta)s}{\delta\theta} \\ \frac{\theta}{\sigma} & 0 & -\frac{\theta}{\sigma} \\ -\frac{(1-\theta)}{\sigma} & 0 & \frac{(1-\theta)}{\sigma} \\ -\frac{(\rho+\delta)(1-\theta)}{\sigma} & 0 & \frac{(\rho+\delta)(1-\theta)}{\sigma} \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{\lambda} \\ \check{n} \end{bmatrix} + \begin{bmatrix} 0 & s & 0 \\ 1-\theta & 0 & 0 \\ \frac{(\rho+\delta)(1-\theta)}{\delta\theta} & \frac{(\rho+\delta)(1-\theta)(1-s)}{\delta\theta\tau} & -\frac{(\rho+\delta)\zeta_g}{\delta\theta} \\ 1-\frac{\theta}{\sigma} & 0 & 0 \\ \frac{(1-\theta)}{\sigma} & 0 & 0 \\ \frac{(\rho+\delta)(1-\theta)}{\sigma} & 0 & 0 \end{bmatrix} \begin{bmatrix} \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix} \tag{22.15}$$

22.3 Eliminating the Focal Variable

As will become apparent, for many purposes, the focal variable form of the contemporaneous equations is especially convenient. However, it is also a straightforward matter to eliminate the focal variable and express every variable in terms of $\check{k}$, $\dot{\check{\lambda}}$ and the exogenous variables. In this case, using the focal variable form as a way station is still helpful for accuracy and for interpretive purposes.

Appendix **??** shows how to eliminate focal variables in general. The key is the fact that from the third row of (22.14),

$$\check{n} = \frac{1}{\frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau} \left[\frac{\theta}{\sigma} \quad s \quad 1-\frac{\theta}{\sigma} \quad 1-s \quad 0 \right] \begin{bmatrix} \check{k} \\ \check{\lambda} \\ \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix}. \tag{22.16}$$

Thus, to the direct effects of $\check{k}$, $\check{\lambda}$ and the exogenous variables on the contemporaneous variables variables, the endogenous adjustment of n adds the column vector of elasticities with respect to n times this expression for $\check{n}$:

$$\begin{bmatrix} \dot{\check{k}} \\ \dot{\check{\lambda}} \\ \check{n} \\ \check{c} \\ \check{y} \\ \check{i} \\ \check{w} \\ \check{R} \\ \tilde{r} \end{bmatrix} = \text{direct effects} + \begin{bmatrix} \frac{(\rho+\delta)(1-\theta)s}{\theta} \\ -\frac{(\rho+\delta)(1-\theta)}{\sigma} \\ 1 \\ (1-s)\tau \\ 1-\theta \\ \frac{(\rho+\delta)(1-\theta)s}{\delta\theta} \\ -\frac{\theta}{\sigma} \\ \frac{1-\theta}{\sigma} \\ \frac{(\rho+\delta)(1-\theta)}{\sigma} \end{bmatrix} [\frac{\theta}{\sigma}+\frac{1}{\eta}+(1-s)\tau]^{-1} \begin{bmatrix} \frac{\theta}{\sigma} & s & 1-\frac{\theta}{\sigma} & 1-s & 0 \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{\lambda} \\ \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix}.$$

Abbreviating the sum of labor inelasticities by α,

$$\alpha = \frac{\theta}{\sigma} + \frac{1}{\eta} + (1-s)\tau, \tag{22.17}$$

and adding together the direct effects and the indirect effects mediated by the focal variable n yields

$$\begin{aligned} \begin{bmatrix} \dot{\check{k}} \\ \dot{\check{\lambda}} \end{bmatrix} &= \begin{bmatrix} \rho + \frac{(\rho+\delta)(1-\theta)s}{\alpha\sigma} & \frac{(\rho+\delta)(1-\theta)s}{\theta}\left(\frac{1}{\tau}+\frac{s}{\alpha}\right) \\ \frac{(\rho+\delta)(1-\theta)}{\sigma}\frac{\left(\frac{1}{\eta}+(1-s)\tau\right)}{\alpha} & -\frac{(\rho+\delta)(1-\theta)s}{\alpha\sigma} \end{bmatrix} \begin{bmatrix} \check{k} \\ \check{\lambda} \end{bmatrix} \\ &+ \begin{bmatrix} \frac{(\rho+\delta)(1-\theta)}{\theta}\left(1+\frac{s}{\alpha}(1-\frac{\theta}{\sigma})\right) & \frac{(\rho+\delta)(1-\theta)(1-s)}{\theta}\left(\frac{1}{\tau}+\frac{s}{\alpha}\right) & -\frac{(\rho+\delta)\zeta_g}{\theta} \\ -\frac{(\rho+\delta)(1-\theta)}{\sigma}\left(\frac{1+\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & -\frac{(\rho+\delta)(1-\theta)(1-s)}{\alpha\sigma} & 0 \end{bmatrix} \begin{bmatrix} \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix} \end{aligned} \tag{22.18}$$

and

$$
\begin{bmatrix} \check{n} \\ \check{c} \\ \check{y} \\ \check{i} \\ \check{w} \\ \check{R} \\ \tilde{r} \end{bmatrix}
=
\begin{bmatrix}
\frac{\theta}{\alpha\sigma} & \frac{s}{\alpha} \\
\frac{(1-s)\tau\theta}{\alpha\sigma} & -\frac{s}{\alpha}\left(\frac{\theta}{\sigma}+\frac{1}{\eta}\right) \\
\theta\left(1+\frac{1-\theta}{\alpha\sigma}\right) & \frac{(1-\theta)s}{\alpha} \\
\frac{\rho+\delta}{\delta}\left(1+\frac{(1-\theta)s}{\alpha\sigma}\right) & \frac{(\rho+\delta)(1-\theta)s}{\delta\theta}\left(\frac{1}{\tau}+\frac{s}{\alpha}\right) \\
\frac{\theta}{\sigma}\left(\frac{\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & -\frac{s\theta}{\alpha\sigma} \\
-\frac{(1-\theta)}{\sigma}\left(\frac{\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & \frac{s(1-\theta)}{\alpha\sigma} \\
-\frac{(\rho+\delta)(1-\theta)}{\sigma}\left(\frac{\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & \frac{s(\rho+\delta)(1-\theta)}{\alpha\sigma}
\end{bmatrix}
\begin{bmatrix} \check{k} \\ \check{\lambda} \end{bmatrix}
$$
$$
+\begin{bmatrix}
\frac{1-\frac{\theta}{\sigma}}{\alpha} & \frac{(1-s)}{\alpha} & 0 \\
\left(1-\frac{\theta}{\sigma}\right)\frac{(1-s)\tau}{\alpha} & s+\frac{(1-s)^2\tau}{\alpha} & 0 \\
(1-\theta)\left(\frac{1+\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & \frac{(1-s)(1-\theta)}{\alpha} & 0 \\
\frac{(\rho+\delta)(1-\theta)}{\delta\theta}\left(1+\frac{s}{\alpha}(1-\frac{\theta}{\sigma})\right) & \frac{(\rho+\delta)(1-\theta)(1-s)}{\delta\theta}\left(\frac{1}{\tau}+\frac{s}{\alpha}\right) & -\frac{(\rho+\delta)\zeta_g}{\delta\theta} \\
\left(1-\frac{\theta}{\sigma}\right)\left(\frac{\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & -\frac{(1-s)\theta}{\alpha\sigma} & 0 \\
\left(\frac{1-\theta}{\sigma}\right)\left(\frac{1+\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & \frac{(1-s)(1-\theta)}{\alpha\sigma} & 0 \\
\frac{(\rho+\delta)(1-\theta)}{\sigma}\left(\frac{1+\frac{1}{\eta}+(1-s)\tau}{\alpha}\right) & \frac{(1-s)(\rho+\delta)(1-\theta)}{\alpha\sigma} & 0
\end{bmatrix}
\begin{bmatrix} \check{z} \\ \check{\xi} \\ \check{g} \end{bmatrix}
\qquad (22.19)
$$

22.4 Calibrating the Contemporaneous Equilibrium of the Basic Real Business Cycle Model

With the parameter values $\rho = .02$/year, $\delta = .08$/year, $\theta = .3$, $\sigma = 1$, $s = 1$, $\eta = 1$ and $\tau = 1$ (implying $\zeta_c = .7$, $\zeta_i = .24$ and $\zeta_g = .06$), the annual elasticities of $\dot{\check{k}}$, $\dot{\check{\lambda}}$, the simple elasticities of n, c, y, i, w, R and the semielasticity of r with respect to k, λ, z, ξ and g in the contemporaneous general equilibrium of the Basic Real Business Cycle Model are given in the accompanying table.

Chapter 23

Graphical Analysis of the Basic Q-RBC Model

23.1 Four Levels of Integration

A dynamic general equilibrium model has at least four levels of integration or levels of comprehensiveness that can be distinguished:

1. Household and Firm Optimization
2. Contemporaneous Market Equilibrium
3. Contemporaneous General Equilibrium
4. Dynamic General Equilibrium

In the Basic Real Business Cycle Model outlined in the previous chapter, (1) household optimization goes into the construction of the labor supply curve and some other basic relationships, while firm optimization goes into the construction of the labor demand curve and the factor-price possibility frontier. (2) Supply and demand in the labor market, put together, illustrate contemporaneous market equilibrium. (3) Contemporaneous general equilibrium is completed by the determination of output, consumption, investment and the real rental rate based on the real wage and quantity of labor that come out of labor market equilibrium. (4) Dynamic general equilibrium is illustrated by the path taken on the phase diagram, but includes the time-paths of variables other than the dynamic variables k and λ shown directly on the phase diagram.

There is an intricate interplay between the different levels of integration for a model. One of the best ways to understand how to approach the different levels of integration is to think about the distinction in the Marshallian Cross of supply and demand between a movement along a curve and the shift of a curve. For example, when tastes shift in a good's favor, then the quantity demanded *at any given price* increases, shifting the entire demand curve to the right. This rightward shift in the demand curve leads to a new equilibrium price and quantity. This new market equilibrium can be found because in constructing both the new demand curve and the preexisting supply curve, the economist had already contemplated the responses to a whole range of different possible prices—even prices that seemed irrelevant before the change in tastes occurred.

The Marshallian Cross of supply and demand is an elegant way of dealing with the interplay between the first two levels of integration—individual household and firm optimization and market equilibrium. In analyzing a dynamic general equilibrium model, there are higher levels of integration to deal with as well. But the same principles apply. Ignoring the distinction between the various different levels of integration has the same kinds of consequences as ignoring the distinction between a movement along a curve and a shift in a curve in supply and demand analysis.

To find out what happens in dynamic general equilibrium, one must first determine the characteristics of the contemporaneous general equilibrium at a substantial variety of different (k, λ) points on the phase diagram. To find out what happens in contemporaneous general equilibrium, one must first determine a variety of market equilibria which in turn depend on knowing the supply and demand relations that describe the way in which households and firms would respond to a wide variety of different circumstances.

23.2 The Debut of the Q-Theory Real Business Cycle (Q-RBC) Model

In the Basic Real Business Cycle Model, contemporaneous general equilibrium is determined primarily by labor market equilibrium. In this chapter, we will illustrate more of the potential of the techniques we will use in this book by analyzing graphically the Q-Theory Real Business Cycle (Q-RBC) Model, which requires two (sub-)levels of market equilibrium stacked on top of each other in order to determine contemporaneous general equilibrium.

We will not try to do as complete an analysis of the Q-RBC model as we did of the Basic RBC Model. We will focus this chapter on demonstrating one surprisingly strong result at the highest level of comparative dynamic general equilibrium: a permanent increase in either the general labor augmenting technology z or in government purchases g (either of which raises the steady state capital stock) unambiguously raises the (*ex ante*) real interest rate above normal during the transition to the new steady state.

The details of dealing with investment adjustment costs are tackled in Chapter **??**. Here, the beginning of wisdom about the Q-Theory Real Business Cycle Model is to realize that investment adjustment costs create a distinction between the marginal utility of consumption (MUC)—which we will call ν—and the marginal value of capital (MVK) λ. To discuss the distinction between the marginal utility of consumption and the marginal value of capital, we need one more concept: the marginal value of investment (MVI). In the Q-RBC Model here, in the absence of differential technological progress between the production of consumption goods and the production of investment goods, the relative price of consumption and investment will be fixed, so that by choosing appropriate units of measurement, we can say that one unit of consumption trades off with one unit of investment. Thus, the marginal utility of consumption will still be equal to the marginal value of investment (MUC = MVI). But, due to investment adjustment costs, on unit of investment is no longer guaranteed to produce one unit of capital. If a firm invests at an especially fast pace, there will be some wastage and mistakes, so that—on the margin—one unit of investment produces less than the normal one unit of capital; conversely, if a firm invests at an especially slow pace, the less-than-normal wastage and mistakes make it so that—on the margin—one unit of investment produces more than the normal one unit of capital. Thus, in general, the marginal value of investment is not equal to the marginal value of capital (MVI $\neq$ MVK).

Since at the normal (here read the adjective "normal" as "steady-state") rate of investment, one unit of investment produces one unit of capital, the price of investment goods is an appropriate notion of the "replacement cost" of capital. Thus, ratio $\frac{\text{MVK}}{\text{MVI}}$, which is equal to the relative price between capital and investment, is equal to (marginal) Tobin's q. Thus, as long as MUC=MVI,

$$q = \frac{\text{MVK}}{\text{MVI}} = \frac{\text{MVK}}{\text{MUC}} = \frac{\lambda}{\nu}. \tag{23.1}$$

23.3 The Investment Demand (II) Curve

Turning (23.1) around, $\nu = \lambda/q$. Therefore, the familiar positive relationship between (marginal) Tobin's q and the rate of investment relative to the capital stock, i/k, means that for a given value of λ, there is a negative relationship between i/k and ν. Figure 14 illustrates this relationship. The downward-sloping "II" curve is the "investment demand curve." The marginal efficiency with which investment spending is converted into capital falls as the rate of investment increases. As long as gross investment is positive, q is inversely proportional to this "marginal efficiency of investment"; if firms are willing to invest at such a high rate that investment is turned into capital quite inefficiently, it must be because the relative price of capital is high. At the normal, steady-state rate of investment, which is $\frac{i}{k} = \delta$, $q = 1$ and $\nu = \lambda$, as shown in Figure 14.

Having used the words "marginal efficiency of investment" to explain why the investment demand curve slopes down, it is important to distinguish our investment demand curve from the one Keynes drew.

In both Keynes' marginal efficiency of investment schedule and our investment demand curve, expectations of the future are crucial. Keynes described a large chunk of those expectations as a matter of "animal spirits." As late twentieth-century macroeconomists, we assume that the expectations of the future encapsulated in λ are rational. An increase in future prospects, as indicated by an increase in λ, shifts the entire investment demand curve to the right.

It is only a matter of convenience that we put the investment rate i/k on the horizontal axis rather then investment i itself as Keynes does.

The most important difference between Keynes' marginal efficiency of investment schedule and our investment demand curve is that Keynes' marginal efficiency of investment schedule has the real interest rate on the vertical axis, while we have the marginal utility of consumption. There is an important distinction, but also a family relationship between the real interest rate r and the marginal utility of consumption ν. Loosely speaking, ν is a lot like a long-term real interest rate, except that it also includes a component akin to a wealth effect. To be precise, as we will show, household optimization implies that (in expectation, using a certainty equivalence approximation)

$$r = \rho - \frac{\dot{\nu}}{\nu}, \tag{23.2}$$

or equivalently,

$$\ln(\nu_t) - \ln(\nu_T) = \int_t^T [r_{t'} - \rho]\, dt' \tag{23.3}$$

and

$$\ln(\nu_t) = \ln(\nu_\infty) + \int_t^\infty [r_{t'} - \rho]\, dt', \tag{23.4}$$

where r is the real interest rate stated in terms of consumption goods, ρ is the utility discount rate (pure time preference rate) and T can be any time in the future. Thus, the real interest rate is just ρ plus the proportional rate at which ν declines. The total decline in $\ln(\nu)$ over any period of time is proportional to the excess of a long-term real interest rate beyond the normal long-term real interest rate given by ρ. Finally, the behavior of $\ln(\nu_t)$ is the sum of a wealth effect as indicated by what happens to $\ln(\nu_\infty)$ and a kind of long-term real interest rate.

Equation (23.2) means that to prove that the real interest rate is above normal in response to a shock, all we need to do is to show that ν will be falling after the first impact effect of the shock. The impact effect itself is a different matter, since that is an unexpected response to new information that would not be reflected in the *ex ante* real interest rate (though it might be reflected in *ex post* stock returns or the like).

23.4 The Saving Supply (SS) Curve

Figures 15 and 16 show the saving supply curve, labelled SS. In the Basic RBC model, we have $q = 1$ and therefore $\nu = \lambda$ all the time. In the Q-RBC model, $\nu \neq \lambda$. This gives rise to a serious question. In the previous chapter, when we look at the effects of an increase in λ, is it really the effects of the marginal value of capital λ we are looking at, or is it really the effects of the marginal utility of consumption—which happens to equal the marginal value of capital in that model. Of course, the only coherent way to answer such a question is to look at what happens in a model like the Q-RBC model in which there is a distinction between the marginal value of capital and the marginal utility of consumption.

The simple answer is that labor supply, being a household decision, depends directly on the marginal utility of consumption ν, not the marginal value of capital λ. Thus, the effects of "λ" in the previous chapter are better thought of as effects of the marginal utility of consumption ν. The marginal utility of consumption ν determines labor supply, as shown in Figure 17. Labor demand, arising from firm optimization for a given value of the capital stock, is unaffected by the presence of investment adjustment costs. Labor supply and demand together determine equilibrium w and n. From these labor market equilibrium values it is also easy to determine consumption c, output y and thence investment i, as well as the real rental rate of capital R. Thus, with the single exception of the real interest rate r, all of the key variables that we were concerned about in the previous chapter can be determined from a knowledge of ν, k and exogenous variables alone, without reference to λ. Graphically, this means that, by and large, what happens at a given point on the saving supply (SS) curve can be determined without regard to how one got there.

In the Basic RBC model, an increase in λ lead to an increase in n, c, y, i and R, and a decrease in w. In the Q-RBC model, the corresponding result is that moving up the saving supply (SS) curve to a higher value of ν, the variables n, c, y, i and R increase, while w decreases. The fact that i increases as one moves up along the saving supply curve ensures that i/k will also increase in that direction. This guarantees that the saving supply (SS) curve will be upward-sloping.

23.5 Comparative Statics in the (II-SS) Market for Loanable Funds

In the Basic RBC model, for a given value of λ, an increase in k raises the accumulation rate $\dot{k}$ and the closely related investment rate i/k. In the Q-RBC model, the corresponding result is that, for a given value of ν, an increase in k raises $\dot{k}$ and the closely related investment rate i/k; therefore, an increase in k shifts the SS curve to the right. (See Figures 18 and 19.) Similarly, the fact that in the Basic RBC model, labor augmenting technology z raises $\dot{k}$ and i/k for a given value of λ means that in the Q-RBC model, an increase in z raises $\dot{k}$ and i/k for a given value of ν—shifting the SS curve to the right. (See Figures 20 and 21.) Finally, the fact that in the Basic RBC model, government purchases g lowers $\dot{k}$ and i/k for a given value of λ means that in the Q-RBC model, an increase in z lowers $\dot{k}$ and i/k for a given value of ν—shifting the SS curve to the left. (See Figures 22 and 23.)

In all of these aspects of the Q-RBC model, the marginal utility of consumption ν takes the place that λ had in the Basic RBC model. Where then does the marginal value of capital λ come in? As mentioned above, the one object that the marginal value of capital λ affects is the investment demand curve. An increase in λ raises the investment demand curve, shifting it outward. Indeed, since the intersection of the investment demand curve with the vertical $(i/k) = \delta$ line is at height λ, the vertical shift in the investment demand (II) curve on that line is equal to the change in λ. (See Figure 24.)

23.6 The Adjustment to a Higher Capital Stock and the Real Interest Rate r

We do not need to know everything about the phase diagram in order to analyze the effect of a permanent increase in the general labor-augmenting technology z or government purchases. Besides the fact that the new steady state is at a higher level of capital, the only fact we need is that after any initial jump in the marginal product of capital, the marginal value of capital gradually falls as the capital stock increases toward a higher value at the new steady state. In other words, what we need to know is that the saddle path slopes downward. It does. (See Figure 25.) (As it happens, the phase diagram in the Q-RBC model often looks quite similar to the phase diagram in the Basic RBC model. But we do not need this fact for the result we are trying to get.)

Intuitively the saddle path must slope downward because in dynamic general equilibrium, having a higher initial capital stock should induce a lower marginal value of capital. If the exogenous variables are constant after their initial jump, then accumulating more capital puts the economy in the same position (though at a later date) as if the initial capital stock had been higher.

After the initial impact effects, the accumulation of capital causes a gradual rightward shift of the saving supply (SS) curve. On the other side of the scissors, the fall in the marginal value of capital as the economy follows the saddle-path down leads to a gradual downward shift of the investment demand (II) curve. Both of these shifts tend to cause the equilibrium marginal utility of consumption ν to fall. By (23.2), with ν falling, the real interest rate r must be above its normal level of ρ during the adjustment process. (See Figure 26. Since k is increasing along the saddle path, $\dot{k}$ and the closely related i/k must also be above their normal values of 0 and δ during the adjustment process.)

All of this is just the overview of the treatment of the Q-RBC model in later chapters.