INTERTEMPORAL OPTIMIZATION IN GENERAL EQUILIBRIUM: A PRACTICAL INTRODUCTION

Peter J. Wilcoxen
Department of Economics
The University of Texas at Austin

This is a reprint of Impact Preliminary Working Paper IP-45 (1989), which was written while the author was at the Impact Research Centre at the University of Melbourne. It contains minor revisions.

August 1995

Intertemporal Optimization in General Equilibrium:
A Practical Introduction

Table of Contents

Exercises

Appendicies

Preface

This paper is intended to be a practical guide to intertemporal modeling, with particular emphasis on how intertemporal optimization can be incorporated into computable general equilibrium models. It focuses on the practical details of building intertemporal models: how to set up and solve intertemporal optimization problems, how to analyze such models in partial equilibrium, and how to link them to computable general equilibrium models. For the most part, nuances of theory have been relegated to footnotes, but plenty of references have been provided to enable an interested reader to pursue the subject in more depth. A reading guide for further study appears at the end.

To make the paper as useful as possible, a number of exercises, complete with answers, have been included. These present supplementary material or go into particular topics in more depth. They can be used in the ordinary way to practice techniques discussed in the text, or they can be used as rather lengthy footnotes on certain topics. In any event, they form an important part of the paper and should not be neglected. Overall, with this structure and emphasis, we hope the paper will be a useful introduction to the rapidly expanding field of intertemporal general equilibrium modeling.

1 Introduction

Most applied economic models are designed to represent economies at particular points in time. For long run models the point of interest is far in the future, when all immobilities have vanished and all transient behavior has died out. On the other hand, the only period of interest in short run models is the immediate present. In both cases, however, only one period is captured by the model, so they are both essentially static. No information is included on how the economy changes over time, so it is impossible to solve for the sequence of equilibria between the short and long run solutions.

In contrast, intertemporal models specifically include equations describing how the economy evolves. These allow the models to be used to find the economy's trajectory through time. Unfortunately, this versatility comes at a price: intertemporal models are somewhat harder to build − and much harder to solve − than static models. However, there are two circumstances in which the extra effort is worthwhile. The first arises when the trajectory itself is of interest apart from the short and long run equilibria. Policy makers, for example, are often keenly interested in how fast the economy moves toward the long run, and whether or not the transition is smooth. This is especially true when the short and long run effects of the policy are very different. Furthermore, some models, especially those in which lags play a prominent role, show cycles in certain variables between the short and long run. If so, it is often important to know the timing and amplitude of the oscillations. A final occasion in which the trajectory might be of interest occurs when the model is to be used to evaluate the effect of different policies on the rate of growth. Thus, in a number of circumstances it is necessary to be able to compute the path of the economy over an extended period of time, and in these cases it would be worthwhile to build an intertemporal model.

The other reason for building an intertemporal model is to incorporate intertemporal optimization by agents. If some of the agents in the model choose current variables to optimize intertemporal objective functions, even short run results will require some form of intertemporal modeling. For example, households might be modeled as lifecycle savers whose consumption is based in part on their human wealth. Since human wealth is the discounted sum of expected future labor earnings, any shock that changes expected wage rates or hours worked in the future will change human wealth and hence change current consumption. Since the lifecycle model is used more and more often as the basis of savings behavior, it has stimulated the development of intertemporal modeling.

Perhaps the strongest motivation for developing intertemporal models, however, has been the desire to integrate recent theories of investment behavior into applied general equilibrium work. In such investment models, each firm chooses its level of investment to maximize the stock market value of its equity. Market value, in turn, depends on the earnings a firm is expected to generate in the future. Thus, changes in expectations about a firm's prospects can change its market value and hence its level of investment. To a large extent, this is nothing more than a formalization of the common sense principle that firms whose prospects look good will invest more than others. Investment is *prima facie* an intertemporal decision, so treating it as such in general equilibrium models is very appealing.

In the remainder of this paper we present a small general equilibrium model incorporating investment behavior derived from value maximization.[1] We begin by setting up and solving a simple investment problem. Next, we analyze the investment model thoroughly in partial equilibrium to gain insight into how it works, and to demonstrate some powerful analytical techniques that can be used with intertemporal models of all types. After that, we link the investment problem to a small, static general equilibrium model to produce an intertemporal general equilibrium model. Finally, we will use the resulting model to analyze a number of different policies.

2 A Simple Intertemporal Model of Investment

We begin building our intertemporal model of investment by making two assumptions: first, that each firm chooses its level of investment to maximize its stock market value, and second that an arbitrage equation governs the relationship between returns on debt and returns on equities. The first assumption establishes the basis for the firm's investment behavior. There are many other ways in which a firm might choose its level of investment, but this is the only one which is likely to be optimal for its shareholders. The second assumption is needed to define how the firm's market value is determined by asset holders. Together, the two assumptions will allow us to construct the firm's objective function.

The next step in setting up the model is to choose a particular arbitrage equation and use it to find an expression for the market value of a firm. In this paper we will use the arbitrage equation below:

$$r(t)V(t) = D(t) + V'(t) \,, \tag{2.1}$$

where $V(t)$ is the value of the firm at time t, $r(t)$ is the rate of interest on bonds at t, $D(t)$ is the dividend paid by the firm, and $V'(t)$ is the derivative of the firm's value with respect to time.[2] The left side of the equation gives the return that could be earned by holding V dollars of bonds. The right side is the return received by holding all of the firm's equity and is equal to dividends plus capital gains. Arbitrage will occur as long as the returns on the two assets differ, so in equilibrium, equation (2.1) must hold.

Many extensions and modifications to equation (2.1) are possible. If equity is thought to be riskier than debt, a risk premium could easily be added. If dividends, capital gains and interest income are taxed differently, those taxes could also be included in the equation. We have used expression (2.1) to keep our exposition as clear as possible, but it could easily be modified without changing the substance of our analysis.

Expression (2.1) is a differential equation in the value of the firm. It can be solved by collecting terms in V on the left, finding an appropriate integrating factor, and integrating both sides.

1. The investment model incorporates adjustment costs in the spirit of Eisner and Strotz (1963), Lucas (1967), Gould (1968) and Treadway (1969). In this formulation, investment will depend on the marginal value of Tobin's q (see Tobin (1969) for a discussion of q).

2. For typographical convenience we will denote time derivatives using the prime symbol (′) rather than the usual dot.

Collecting terms produces the following equation:

$$V'(t) - r(t)V(t) = -\, D(t)\ . \tag{2.2}$$

When r is constant, an appropriate integrating factor for (2.2) is:[3]

$$e^{-rt}\ . \tag{2.3}$$

Multiplying both sides of (2.2) by (2.3) produces the following:

$$(\ V'(t) - rV(t)\ )e^{-rt} = -\, D(t)e^{-rt}\ . \tag{2.4}$$

From inspection, it is clear that the left side of (2.4) is the differential of the product of V and the integrating factor, so (2.4) can be rewritten as:

$$\frac{d(\ V(t)e^{-rt}\ )}{dt} = -\, D(t)e^{-rt}\ . \tag{2.5}$$

Equation (2.5) shows how the value of the firm must change over time if the arbitrage equation is to hold. At this point, the expectations of investors become important. If investors have information allowing them to form expectations about the future path of dividends, and they believe the arbitrage condition will always hold, then (2.5) can be integrated to give the value of the firm.[4] To see this, suppose investors at time τ have an information set Ω_τ that leads them to expect the path of dividends at all future times $t \geq \tau$ to be given by a function $D(t;\Omega_\tau)$. That is, $D(t;\Omega_\tau)$ is the dividend expected for time t given information set Ω_τ. Similarly, let $V(t;\Omega_\tau)$ be the expected value of the firm at t given information Ω_τ. Then, integrating both sides of (2.5) from τ to an arbitrary future time T (and rearranging slightly) gives the expression below:

$$V(\tau;\Omega_\tau) = V(T;\Omega_\tau)e^{-r(T-\tau)} + \int_\tau^T D(t;\Omega_\tau)e^{-r(t-\tau)}\ dt\ . \tag{2.6}$$

Equation (2.6) has a clear and intuitive interpretation. If we let τ be the present, "today", then the left term, $V(\tau;\Omega_\tau)$, is the value of the firm today, given today's information. Moving to the right, $V(T;\Omega_\tau)$ is the expected value of the firm at time T given information available today, so the first term on the right side is the present value today of owning the firm at T. Finally, the rightmost term is the present value of the dividends expected to be paid between τ and T. Thus, equation (2.6) shows that the value of the firm today is equal to the present value of owning it at

---

3. When r is not constant the integrating factor becomes a bit more complicated; for this model it would be $exp(\int r(v)dv)$. However, none of the subsequent results would change significantly.

4. These expectations need not be correct, but they must exist; investors must have some belief about the dividends the firm will pay in the future.

T plus the present value of the dividends it will pay between τ and T.

If we knew what investors today thought about the value of the firm at some particular time T in the future, and what dividends they expected the firm to pay between now and then, we could use (2.6) to compute the value of the firm today. Unfortunately, for most points in the future, we do not know *a priori* what investors expect the firm's value to be, so we do not have an appropriate value of $V(T;\Omega_\tau)$. One solution, however, is to let T go to infinity. Then, by making a plausible assumption about the rate of growth of V far in the future, the middle term in (2.6) can be evaluated.

To see how this works, observe that when T goes to infinity the middle term in (2.6) becomes the following:

$$\lim_{T\to\infty} V(T;\Omega_\tau)e^{-r(T-\tau)} \ . \tag{2.7}$$

As long as $V(T;\Omega_\tau)$ remains bounded as $T \to \infty$, expression (2.7) will be zero. Thus, if V remains finite for all time, the middle term in (2.6) can be dropped from the equation. In fact, (2.7) will be zero under the more general condition that as T goes to infinity, the rate of growth of V is strictly less than the interest rate. That is, (2.7) will be zero as long as the following holds:[5]

$$\lim_{T\to\infty} \frac{V'(T;\Omega_\tau)}{V(T;\Omega_\tau)} < r \ . \tag{2.8}$$

Thus, if we are willing to assume that (2.8) holds, we can solve (2.6) for the value of the firm at time t:

$$V(\tau;\Omega_\tau) = \int_\tau^\infty D(t;\Omega_\tau)e^{-r(t-\tau)}\ dt \ . \tag{2.9}$$

This says that the value of the firm today is the present value of the dividend stream it is expected to pay in the future, given today's information.

Expression (2.8) is formally known as a "transversality condition" because it is a requirement imposed on the limit of V as time goes to infinity. It is an assumption made in order to obtain equation (2.9), and is not an implication of the model because there is nothing in the problem we have specified so far that requires (2.8) to hold. However, it has a sensible economic interpretation, and is not an unreasonable assumption. In essence, expression (2.8) rules out Ponzi schemes that go on forever.[6] To see why, notice that if (2.8) is violated, the arbitrage condition can only hold if the firm pays negative dividends. Otherwise, (if dividends were zero or positive)

5. In fact, the following analysis can also be applied, with slight modifications, when (2.8) holds with equality.

6. Ponzi schemes take their name from Charles Ponzi who perpetrated a famous chain letter swindle in the 1920's. Today the term is used for any pyramid scheme that operates by continually drawing in new people at the bottom.

the return on equity would be higher than the return on debt, so no one would be willing to hold bonds. This would force the interest rate up until (2.8) held. The only time (2.8) can be violated and still have the arbitrage equation hold is if investors are willing to pay money into a firm forever without receiving any dividends. Thus, assuming that (2.8) holds is nothing more than ruling out infinitely-lived Ponzi schemes.

In the remainder of the paper the values of all variables in future periods will always be expectations based on an information set Ω_τ. To keep our notation as simple as possible, we will often suppress Ω_τ from variable names and write, for example, $V(t;\Omega_\tau)$ as $V(t)$. It is important to remember, however, that all variables at future times are expectations implicitly derived from a particular information set.

At this point we have derived the objective function in the firm's investment problem: equation (2.9) gives the firm's value in terms of its expected future dividends. Furthermore, (2.9) has a clear, intuitive interpretation: the value of the firm at time t is the discounted present value of its dividend stream. This interpretation is so compelling that it is often tempting to begin building investment models by assuming that (2.9) holds and dispensing with its derivation from arbitrage. Starting with the arbitrage equation, however, provides a rigorous basis for (2.9). More importantly, the approach can also be used for much more complicated models (such as those with taxes or risk premia) for which the form of the value function is not obvious from inspection.

The next step in setting up the investment problem is to specify how dividends depend on the firm's choice variables. To keep our model simple, we assume the firm pays out everything it earns except what it uses for investment. In addition, we also assume that all investment is internally financed: the firm does not issue new debt or equity to pay for investment. These assumptions are fairly innocuous: introducing other means of finance such as corporate bonds or new share issues alters the problem relatively little. In fact, the financial decision makes no difference at all if capital markets are perfect.[7] Nonetheless, it is straightforward to incorporate finance into the model if necessary.

Under the assumptions above, dividends are equal to the difference between the firm's revenue and the total of its variable costs, its investment costs, and any taxes it pays. To put this symbolically, if K is a vector of capital stocks, L is a vector of variable inputs, I is a vector of investments in the capital stocks, P is a vector of prices and wages, and Z is a vector of taxes, dividends are given by some function $D(K, L, I, P, Z)$.

To get much further, we need to specify the actual form of D. For the purposes of this section, we will assume there is a single tax which falls on dividends, T^d, and that D is additively separable into a short run profit function and an investment cost function.[8] The short run profit function gives the profit on a particular vector of capital stocks after variable inputs are chosen optimally. Since this corresponds closely to the accounting idea of earnings, we will represent it

7. This is a consequence of the Modigliani-Miller theorem, first described in Modigliani and Miller (1958). Blanchard and Fischer (1989) provides a clear discussion of this point in chapter 6.

8. Formally, we have assumed that the short run variable input decision is separable from the long run decision on investment, and also that the capital stock does not enter the investment cost function. Both of these are controversial in the literature, and neither is really necessary to the model. However, they do keep the exposition much clearer.

by function E, and will often refer to it as earnings. Using C to represent the investment cost function, dividends can be written as shown below:

$$D(K, L, I, P, Z) = (\ E(K, P) - C(I, P)\ )(1 - T^d)\ . \tag{2.10}$$

To keep our exposition as clear as possible, we will assume there is no sign constraint on dividends (or that if there is one, it is never binding). In particular, we will not prohibit the firm from spending more on investment than it earns at a particular date.

At this point we can set up and solve the firm's investment problem under fairly general conditions. Inserting (2.10) into (2.9) gives us the firm's objective function. In addition, the firm is subject to an accumulation constraint which specifies how the capital stock evolves as a consequence of the firm's investment. Here we will make the usual assumption that the time derivative of the capital stock is given by the difference between gross investment and depreciation. Thus, the firm's investment problem at time τ is to choose a path of investment, $I(t)$ for $t \geq \tau$, that solves:[9]

$$\max \int_{\tau}^{\infty} (\ E(K, P) - C(I, P)\ )(1 - T^d) e^{-r(t-\tau)}\ dt\ ,$$

$$\text{subject to } K' = I - \delta K\ . \tag{2.11}$$

Problem (2.11) requires dynamic optimization and can be solved using the method of optimal control. Although a complete treatment of optimal control is well beyond the scope of this book,[10] a simple problem, such as the one above, can be solved in the following way. Suppose the problem has the form:

$$\max \int f(s, c, t)\ dt\ ,$$

$$\text{subject to } s' = g(s, c, t)\ ,$$

where s and c are variables and t indicates time. It is customary to refer to s as the problem's "state variable" and c as its "control variable". In problem (2.11), for example, the state variable is the capital stock (K) and the control variable is the rate of investment (I). To find necessary conditions for an optimum, we form the problem's Hamiltonian function, H, as shown:

$$H = f(s, c, t) + \Lambda(t) g(s, c, t)\ ,$$

9. In writing the problem this way we have implicitly assumed that any constraint on the sign of investment would not be binding. Moreover, we have also ignored certain boundary conditions (such as the initial capital stock) which constrain the firm. We will discuss the boundary conditions in detail later in this section.

10. A very lucid treatment of applied optimal control and other methods of dynamic optimization is Kamien and Schwartz (1981).

where $\Lambda(t)$ is a multiplier much like the Lagrange multiplier of static problems.[11] It is possible to show that the solution must satisfy the following first-order conditions:[12]

$$\frac{\partial H}{\partial c} = 0, \qquad \frac{\partial H}{\partial s} = -\Lambda', \qquad \frac{\partial H}{\partial \Lambda} = s'. \, .$$

Thus, it is possible to characterize the optimal path of the state and control variables by constructing the Hamiltonian and taking first-order conditions.

To apply this approach to our current problem, we start by constructing the Hamiltonian for (2.11):

$$H = (E - C)(1 - T^d)e^{-r(t-\tau)} + \Lambda(I - \delta K) \; . \tag{2.12}$$

Thus, the necessary conditions for optimality of a particular path of investment are the following:

$$\frac{\partial H}{\partial I} = 0 \; , \tag{2.13}$$

$$\frac{\partial H}{\partial K} = -\Lambda' \; , \tag{2.14}$$

$$\frac{\partial H}{\partial \Lambda} = K' \; . \tag{2.15}$$

Differentiating (2.12) as required by (2.13) through (2.15) produces the first order conditions for the problem:

$$-\frac{\partial C}{\partial I}(1 - T^d)e^{-r(t-\tau)} + \Lambda = 0 \; , \tag{2.16}$$

$$-\frac{\partial E}{\partial K}(1 - T^d)e^{-r(t-\tau)} + \delta\Lambda = \Lambda' \; , \tag{2.17}$$

$$I - \delta K = K' \; . \tag{2.18}$$

The multiplier, $\Lambda(t)$, can be interpreted as the change in the value of the firm today (at time τ) due to a marginal increase in the capital stock at time t in the future.[13] For convenience, we can introduce a new function, $\lambda(t)$, which is defined as shown:

11. Unlike the static case, however, a dynamic optimization problem will have a whole sequence of multipliers, one for each point in time. Hence, Λ is a function of t.

12. Refer to Kamien and Schwartz or another textbook on dynamic optimization for proofs of these.

13. That is, $\Lambda(t)$ shows how the stock market value of the firm would change if, at the optimum investment plan, investors suddenly discovered that the firm was going to be given an extra unit of capital at time t

$$\Lambda(t) = \lambda(t)e^{-r(t-\tau)} \ . \tag{2.19}$$

The interpretation of λ is straightforward: it is the value at time t of having an additional unit of capital at that time, based on information held at time τ. To distinguish between it and Λ, λ is often called the "current value" multiplier. Substituting (2.19) into (2.16) and (2.17) and rearranging produces the expressions below:

$$\lambda = \frac{\partial C}{\partial I}(1 - T^d) \ , \tag{2.20}$$

$$\lambda' = (r+\delta)\lambda - \frac{\partial E}{\partial K}(1 - T^d) \ . \tag{2.21}$$

Equations (2.20), (2.21) and (2.18) are the first order conditions for any problem in which the dividend function has the separability properties discussed above. To solve a particular problem, we will need to insert specific functions for E and C. However, a number of conclusions can be drawn from the general structure above.

Starting with (2.20), if we assume that investment costs are a continuous, strictly convex function of I (so that C_I and C_{II} are positive), then we know from the implicit function theorem that there must be an inverse function F such that:

$$I = F(\lambda, T^d, P) \ . \tag{2.22}$$

Thus, (2.20) determines the level of investment as an implicit function of λ: if λ were known, I could be calculated from (2.22).[14] In contrast, equation (2.21) is a first order differential equation in λ which does not depend on I. This allows it to be solved using the method of integrating factors described above for the arbitrage equation.[15] In this case, the resulting expression is:

$$\lambda(t) = \int_t^\infty \frac{\partial E}{\partial K}(1 - T^d)e^{-(r+\delta)(s-t)}ds \ . \tag{2.23}$$

Equation (2.23) shows how the market value of the firm changes in response to marginal changes in its capital stock. Moreover, it has an interesting and useful interpretation. The right side of (2.23) is the present value at time t of the additional future post-tax earnings that would be

14. This property has spawned dozens of empirical papers. Under conditions derived in Hayashi (1981), λ can be linked to stock market data, and hence can be observed. (This will be discussed in detail in exercise E6). Thus, if λ can be taken to be exogenous, choosing a particular functional form for C allows the adjustment cost model to be tested econometrically using only the first of the necessary conditions, in this case the rewritten expression (2.22); one such example is Summers (1981). What is often overlooked in these papers, however, is that since λ and I are simultaneously determined, it is inappropriate to assume that λ is exogenous. See McLaren (1989) for a more complete discussion.

15. No loss of generality is implied by this. Were it not independent of investment, equation (2.22) could be used to eliminate I by substitution.

generated by an extra unit of capital received at time t. Thus, (2.23) determines the stock market value of an extra unit of capital, while (2.22) selects the optimal level of investment given that market valuation.

The next step in building a practical investment model is to derive the earnings and investment cost functions for a firm with a particular technology. Suppose the firm's output is produced according to a constant returns to scale Cobb-Douglas function of labor and capital:

$$q = L^{\varepsilon} K^{1-\varepsilon} \ . \tag{2.24}$$

In the short run, the capital stock is fixed and the firm chooses labor to maximize the difference between its revenue and its variable costs. That is, it solves the following problem:

$$\max \ pq - wL \ ,$$

$$\text{subject to } q = L^{\varepsilon} K^{1-\varepsilon} \ . \tag{2.25}$$

The earnings function E can be found by inserting the optimal labor input found from (2.25) into the maximand. The result is the expression below:

$$E(K,P) = (\frac{1-\varepsilon}{\varepsilon}) \left(\frac{\varepsilon p}{w} \right)^{1/(1-\varepsilon)} wK \ . \tag{2.26}$$

For convenience, we can define a function β which captures the price and wage effects, so (2.26) can be rewritten as:

$$E(K,P) = \beta(P) K \ , \tag{2.27}$$

were P is a vector of wages and prices. The function $\beta(P)$ gives the short run return on a unit of capital, so it can be thought of as the rental price of the capital stock. We will often refer to earnings functions that are linear in K as having constant returns to scale.

Turning now to the investment cost function, we adopt the adjustment cost formulation and require that C and its first two derivatives be positive. This means that investment is costly, that it becomes more expensive as more of it is done, and that its marginal cost increases as I rises. To derive a particular investment cost function, we can start with an assumption about how new capital goods are produced.

Suppose that new capital goods have to be installed (in some sense) before they can be used in production. Machines, for example, have to be installed in factories before they can be used. Each firm might produce its own installed capital good by buying raw capital goods and hiring special workers to install it. If raw capital goods and installation services have to be combined in fixed proportions, the production function for installed capital goods would have the following form:

$$I = \min\{X_k, S\} , \tag{2.28}$$

where X_k is the quantity of raw capital goods, and S is a measure of installation services chosen to have the same units as X_k. Next, suppose that S is produced from labor as follows:

$$S = \left(\frac{L^i}{\theta} \right)^{1/2} , \tag{2.29}$$

where L^i is the labor hired for installation. Thus, the total cost of investment is the expenditure on raw capital goods plus the cost of labor for installation. If the firm chooses X_k and L^i to minimize the cost of attaining any particular level of I, the investment cost function shown below is obtained:

$$C(I, P) = P_k I + \theta w I^2 . \tag{2.30}$$

By inspection, (2.30) has the properties we required of an investment cost function: it is positive, increasing and convex. As long as θ is greater than zero, there will be costs of adjustment in investment since the marginal cost increases as I rises.

Next, the earnings function in (2.27) and the adjustment cost function in (2.30) can be inserted into the general first order conditions derived above − (2.20), (2.21) and (2.18)− to produce the necessary conditions for this particular problem:

$$\lambda = (P_k + 2w\theta I)(1 - T^d) , \tag{2.31}$$

$$\lambda' = (r + \delta)\lambda - \beta(P)(1 - T^d) , \tag{2.32}$$

$$K' = I - \delta K . \tag{2.33}$$

As suggested above, (2.31) can be solved for the optimal level of investment given a particular value of λ:

$$I = \frac{1}{2w\theta} \left(\frac{\lambda}{1 - T^d} - P_k\right) . \tag{2.34}$$

Equation (2.34) can be used to eliminate I from (2.33), producing a final pair of differential equations in λ and K:

$$\lambda' = (r + \delta)\lambda - \beta(P)(1 - T^d) \ , \tag{2.35}$$

$$K' = \frac{\lambda}{2w\theta(1 - T^d)} - \delta K - \frac{P_k}{2w\theta} \ . \tag{2.36}$$

Equations (2.35) and (2.36) fully characterize the solution to the firm's investment problem. Solving them simultaneously would produce paths of λ and K that were consistent with the first order conditions for the problem laid out in (2.31) through (2.33). Unfortunately, it is impossible to solve them analytically for arbitrary time paths of wages, prices and taxes. Instead, they must usually be solved numerically. Later we will describe several numerical methods suitable for problems like this. First, however, we will demonstrate a number of techniques that can be used to provide a lot of intuition about the model's behavior without actually solving for the optimal path of investment.

3 Graphical Analysis

Ideally we would like to solve differential equations (2.35) and (2.36) for the time paths of λ and K, insert the resulting λ into (2.34), and solve for the path of investment over time. However, some of the terms in the equations, such as tax rates, can be arbitrary functions of time. That means that (2.35) and (2.36) must be solved explicitly for each policy to be modeled. For most policies, it will be difficult or impossible to solve the equations analytically, so numerical methods must be used to obtain explicit results for investment. On the other hand, (2.34) through (2.36) do contain all relevant economic information about the solution, albeit implicitly. This makes it possible to explore many properties of the model without actually solving for the explicit path of investment. In the remainder of this section, we demonstrate how such an analysis might proceed. The methods we use are routine analytical tools in the study of differential equations and can be found in textbooks such as Birkhoff and Rota (1978).

3.1 The Steady State

The difficult integration required to solve (2.35) and (2.36) becomes easy when the model reaches the steady state. For example, suppose that the exogenous variables eventually settle down to stable values P^{ss}, $T^{d^{ss}}$, w^{ss} and P_k^{ss} at some point in the future. We can find the steady state corresponding to these by setting λ' and K' to zero in (2.35) and (2.36) and solving the equations simultaneously. This produces the expressions below:

$$\lambda^{ss} = \beta(P^{ss})\left(\frac{1 - T^{d^{ss}}}{r + \delta} \right), \tag{3.1}$$

$$K^{ss} = \frac{1}{2\theta\delta w^{ss}} \left(\frac{\beta(P^{ss})}{r+\delta} - P_k^{ss} \right). \tag{3.2}$$

From these, steady state investment can be found by inserting (3.1) into (2.34). As might be expected, the result confirms that $I^{ss} = \delta K^{ss}$.

Equations (3.1) and (3.2) can be used to examine the effects of different shocks on the model's steady state. This can be accomplished most easily by total differentiation of both expressions with respect to prices and taxes:

$$d\lambda^{ss} = (\frac{1-T^{d^{ss}}}{r+\delta})\ \nabla_p \beta \cdot dP^{ss} - \frac{\beta(P^{ss})}{r+\delta}\ dT^{d^{ss}}\ , \tag{3.3}$$

$$dK^{ss} = \frac{1}{2\theta\delta(r+\delta)w^{ss}}\ \nabla_p \beta \cdot dP^{ss} - \frac{1}{2w^{ss}\theta\delta}\ dP_k^{ss}\ , \tag{3.4}$$

where $\nabla_p \beta$ is a vector of partial derivatives of β with respect to the elements of P, and dP^{ss} is a vector of changes in elements of P. From (3.3) and (3.4) it is easy to see what happens when one of the model's exogenous variables changes. For example, if the dividend tax, $T^{d^{ss}}$, were to rise, steady state λ would fall, while the capital stock would be unchanged. Capital is unaffected because T^d is a pure profits tax in the long run, falling only on profits and not affecting any decisions at the margin. In contrast, if the price of capital, P_k^{ss}, were to rise, the steady state capital stock would fall while λ^{ss} would be unchanged. λ^{ss} is present value of future earnings on an additional unit of capital, so it is not affected by a change in the cost of new capital goods, P_k. When the price of capital increases with no accompanying rise in λ, the capital stock must fall. Finally, changes in the prices and wages embodied in vector P affect λ and K in the same way. For example, from (2.26) an increase in the firm's output price would increase β, so λ and K would both rise. On the other hand, a rise in the wage rate would lower β, so λ and K would both fall.

This sort of analysis can be applied to a wide variety of models, not just those that have an explicit steady state. In fact, it can be used with any model that can be transformed to have a steady state. For example, if the original model did not have a steady state because of exogenous population growth, it could be transformed onto a per capita basis. The result would then have a steady state. Models which do not themselves have steady states but which can be transformed to have them are often said to attain balanced growth in the long run. Thus, steady state analysis can be applied both to models with steady states, and to models with long-run balanced growth. Moreover, this is likely to encompass all models with interesting long run behavior, since any model which does not asymptotically attain balanced growth will eventually exhibit very peculiar features (such as negative budget shares in consumption).

3.2 Constructing a Phase Diagram

Knowing how the steady state responds to changes in the exogenous variables is helpful, but it does not provide any information about the model's dynamic behavior. For that, another tool can be used: the Poincare phase plane.[16] A phase plane is a two dimensional graph of a model's dynamic behavior that is constructed in the following way. Two of the model's variables are chosen to be axes. Usually these will be the two variables of most dynamic interest; in the investment model above, they would be λ and K. Then, for each point in the resulting plane, the time derivatives of the two variables are evaluated. Again using the investment model as an example, these would be λ' and K'. Together, these derivatives define a vector which indicates the direction the system would move if it ever happened to reach that point. Thus, by using the finished phase plane, it would be possible to trace out the complete trajectory of the system given initial values of the dynamic variables. It would be a tedious but straightforward process: start at the given point, evaluate the derivatives, take a small step in the indicated direction, and repeat. As a practical matter, computing the derivatives of the dynamic variables at all points in the plane is unnecessary. Indeed, most of the details of the model's dynamics can be found by computing a few carefully chosen loci, as we will show.

The first step in constructing a phase diagram is to solve for and plot the model's steady state.[17] At the steady state, the derivatives of both variables are zero, so if the system ever gets there, it will stay there. Next, find the locus of points where the derivative of the first dynamic variable is zero. This will include at least the steady state. Then, find the locus of points where the derivative of the other variable is zero. Again, this should include the steady state. These loci divide the plane up into several regions. In each region the derivatives of the dynamic variables will have a particular sign, so if the system ever enters that region, it will unambiguously evolve in a particular direction. From this it is possible to conclude a great deal about qualitative aspects of the model's dynamics. To illustrate how a phase plane can be used, we will now construct one for the investment model.

The variables of most dynamic interest in the investment model are λ and K, so we will use those as the axes. From equations (3.1) and (3.2) we know the model has a unique steady state, (λ^{ss}, K^{ss}), which is plotted in figure 3.1 and marked A.[18] Next, we solve for the locus of all points for which the time derivative of λ is zero. This can be accomplished by totally differentiating (2.35) with respect to λ and K, and setting $d\lambda'$ to zero.[19] This produces an equation showing what change must be made in λ in order to keep λ' zero when K changes:

$$0 = (r + \delta)d\lambda \; . \tag{3.5}$$

16. Phase planes, also called phase diagrams, are described in more detail in textbooks on differential equations such as Birkhoff and Rota. They can be constructed for any system that is stationary, or "autonomous", in which the variables do not explicitly depend on time.

17. A model without a steady state can often be transformed fairly easily to have one; for example, by converting to a *per capita* basis. In the remainder of this section we will assume a steady state exists.

18. Multiple steady states can be accommodated, if necessary, although linear models will have only one.

19. For this particular model, another way to find the $\lambda' = 0$ locus is to set $d\lambda$ to zero in (2.35). However, this relies on an unusual property of the model−the absence of K from (2.35)−which will not be true in general.

Thus, for λ' to be zero, $d\lambda$ must be zero, so λ must always remain at its steady state value. This means that the $\lambda' = 0$ locus is a horizontal line through the steady state in figure 3.1.[20] Next, we apply the same procedure to equation (2.36) to obtain a locus of points where K' is zero. This produces the following:

$$0 = \frac{1}{2w\theta(1 - T^d)}\ d\lambda - \delta dK\ . \tag{3.6}$$

Forming the ratio $d\lambda/dK$ shows that:

$$\frac{d\lambda}{dK} = 2w\theta(1 - T^d)\delta\ . \tag{3.7}$$

This indicates that if K increases slightly, λ must also rise in order for K' to remain zero. Put another way, the $d\lambda/dK$ locus must be upward sloping. Since it includes the steady state, it must look like the $K' = 0$ locus in figure 3.1.

These loci allow the model's dynamic behavior to be inferred without solving for the system's direction of motion at every point in the plane. To see why, we can consider the two loci in turn. By construction, the $\lambda' = 0$ locus contains all points in the plane where λ' is zero. Points not on the locus, therefore, have nonzero derivatives. Since both equations in the system are continuous in λ and K, regions of positive and negative derivatives must be separated by the $\lambda' = 0$ locus. Thus, the derivative of λ must have the same sign in regions I and II, and must be nonzero. Similarly, λ' in regions III and IV must be nonzero, and of the opposite sign to that of regions I and II.[21] Inserting an arbitrary value of λ greater than λ^{ss} into equation (2.35) reveals that λ' is positive when λ is above its steady state value. In the same way it can be shown that λ' is negative for values of λ below the steady state. These facts can be summarized on the phase diagram by small arrows pointing up in regions I and II, and down in regions III and IV.

The same technique can be applied to the $K' = 0$ locus. Points to the right of $K' = 0$, in regions I and IV, must have the same sign for K'. Inserting an arbitrary value of K above its steady state into (2.36) shows that K' must be negative in those regions. Similarly, it can be shown that the derivative of K must be positive in regions II and III. This information can be included in the phase diagram by small arrows pointing to the right in regions II and III and to the left in regions I and IV. The results are shown in figure 3.2.

The phase diagram is now almost complete and can be used to reveal a great deal about the dynamic behavior of the model. For example, suppose the economy is initially somewhere in region II. We know from the analysis above, as summarized by the arrows drawn in region II, that the system will move upward and to the right as long as it is in that region. In economic terms, this means that λ and K will both grow indefinitely, so the system will move farther and

20. As noted in the previous footnote, the absence of the capital stock from equation (3.5) is a particular property of this model and will not be true in general. Exercise E5, for example, discusses a model which does not have this property.

21. We assume the $\lambda' = 0$ locus does not lie along a local extremum.

farther away from the steady state as long as it remains in region II. The only event that could possibly change the trajectory would be for the system to move into one of the other regions, but we can show that never happens. The system cannot move from region II into either region III or region IV because in region II the derivative of λ is strictly positive. Thus, the system can only move upward. On the other hand, the derivative of K is positive in region II, so the model could move to the right toward quadrant I. As it did so, however, the derivative of K would become closer and closer to zero, so the rightward motion would slow down. Finally, at the $K' = 0$ locus, the system would not be moving to the right at all, and the upward motion of increasing λ would push it back into region II. Thus, the system cannot move from region II into any of the other regions.

From this argument we know that if the system ever entered region II it would move upward and to the right forever. Similar reasoning shows that if the model entered region IV, it would move down and to the left forever. Regions I and III are a little more difficult because it turns out that the system does leave those sectors eventually, but in both cases it is possible to show that the system will move farther and farther away from the steady state as time goes on.

The final step in constructing the phase diagram is to identify any dynamic paths that lead to the steady state. So far, the situation does not look too promising: the path cannot run through any of the regions I through IV because we have established that once the model enters those regions it will never return to the steady state. However, two possible paths remain: the $\lambda' = 0$ locus and the $K' = 0$ locus. Of these, the $K' = 0$ locus can be ruled out because λ tends to move even farther away from its steady state value at each point along the locus. On the other hand, the economy could move along the $\lambda' = 0$ locus and would eventually converge to the steady state. Thus, only a single path leads to the steady state and it lies along the $\lambda' = 0$ locus. This is illustrated on figure 3.3 by a heavy line with several arrows.

A trajectory leading to the steady state is usually called a "stable path" since by proceeding along it, the model eventually attains the steady state. The stable path plays a crucial role in the dynamic behavior of the economy. As will be discussed in section (3.3), in most models it will be unique. That is, it associates a single value of λ with each value of the capital stock. Moreover, if the economy starts at some point on the stable path, as time passes it will remain on the path and move closer to the steady state. At the same time, if the economy starts somewhere off the stable path, it never attains the steady state. Together, these properties mean that if the system is to attain the steady state from an arbitrary initial capital stock, there will be a unique value of λ associated with that stock. That is, the marginal value of an additional unit of capital is unique at any particular capital stock. The essence of dynamic modeling is to determine the stable path, and hence λ, correctly.

At this point, the phase diagram is complete. It shows the dynamic behavior of the system at any point (λ,K), given the model's parameters and the expected values of the exogenous variables. The next step will be to use the phase diagram to determine what happens when new information arrives that causes investors to change their expectations of future variables. Such information might be a government announcement about future tax rates, a new discovery by the firm that will lower its costs, a change in the regulatory structure, or any number of other events. Before analyzing any shocks, however, we will digress briefly to discuss a very important property of models: uniqueness of the stable path.

3.3 Saddle Path Stability and Uniqueness

At least near the steady state, most interesting economic models possess a characteristic known as "saddle path stability" which guarantees uniqueness of the stable path. This uniqueness turns out to be essential in many models because it allows the transversality condition discussed in section (2) to be used to tie down the value of the firm at some point in the future. If there were many paths leading to the steady state, the transversality condition alone would not be enough to determine the value of the firm at an earlier point in time.

To understand the role of the transversality condition more deeply, and to see why it is important that the stable path be unique, it is useful to think about exactly what the equations of motion tell us. As a group, they describe how the model would evolve from any particular point in the (λ, K) plane, given an underlying information set Ω_τ. Thus, if we knew where the system was at a particular moment, and no news had occurred to change Ω_τ, the equations of motion would tell us where the system was going next. However, at the instant that new information arrives, investors may change their expectations about the firm's prospects. If they do, λ will change discontinuously from its value under the old information set, say $\lambda(\tau; \Omega_1)$ to a figure appropriate given the new information, say $\lambda(\tau; \Omega_2)$. As a practical matter, this behavior is very familiar since it occurs in the stock market every day. When a company unexpectedly announces an innovative new product, for example, its stock market value (and normally its marginal value of new capital) jumps upward. The path of λ up to the instant that news arrives provides essentially no guidance about what λ will be just after a shock. Moreover, the size of the jump cannot be determined from the equations of motion alone; an additional piece of information is needed. Formally, one of the problem's boundary conditions is missing.

Boundary conditions are special requirements imposed on the solution in order to make it satisfy certain facts known about the model. For example, we might require the solution path to begin at the existing capital stock. The role of boundary conditions is to determine the constants of integration that arise in solving (integrating) the model's equations of motion. If the model consists of two differential equations, for example, two constants of integration will appear and two boundary conditions will be needed. Sometimes these conditions can be derived from fairly obvious facts. In particular, since state variables, such as the capital stock, do not change discontinuously, their values should not change at the instant of the shock. Since the initial post-shock values of state variables should be exactly equal to their values just before the shock, the pre-shock values of state variables provide some of the needed boundary conditions.

Unfortunately, state variables alone do not provide enough information. In the investment model, for instance, two boundary conditions are required but there is only one state variable. Since costate variables such as λ can change discontinuously when news arrives, their values before the shock provide no information about the condition of the model just after the shock. In technical terms, the initial condition for λ is unknown, so we must look for another fact to use for the second boundary condition.

One possibility is to impose something on the long run behavior of the model. We may not know the initial value of λ, but we might be willing to assume that in the long run λ will eventually approach its steady state value. Technically, this is another transversality condition like the one used at the beginning of section (2). (In fact, it is actually the same condition slightly disguised.) If the stable path is unique, transversality conditions can be used to provide the missing boundary conditions. On the other hand, transversality conditions are inadequate when there are several stable paths leading to the steady state − steady state properties alone cannot determine which stable path the model will follow. Thus, uniqueness of the stable path is a very important property of a model because it allows information about the steady state to be used to determine some of the model's integration constants. If the stable path is unique, there will be a single post-shock value for each costate variable, so it is possible to determine exactly which dynamic path the system will follow after a shock.

All models discussed in here, and virtually all rational expectations models appearing in the literature, have unique stable paths. Showing that a particular model has this property is a fairly technical exercise in linear algebra and differential equations. In the remainder of this section we sketch how it can be done, but the material is not needed to understand the rest of the paper. If you are not comfortable with differential equations, we suggest you skip to section (3.4).

Saddle path stability holds when the linear form of the model has eigenvalues that are distinct, nonzero and of mixed sign.[22] Hence, to check whether a model has a unique stable path we must solve for its eigenvalues. Suppose the model is a system of first order differential equations that can be written in the form:

$$x' = Ax + b \ , \tag{3.8}$$

where x is a vector of variables whose time derivatives are given by x', A is a matrix and b is a vector (both A and b may be functions of time). The eigenvalues we require are those of matrix A. Thus, to see if the model of section (2) has a unique stable path, we would write equations (2.35) and (2.36) as shown:

$$\begin{bmatrix} \lambda' \\ K' \end{bmatrix} = \begin{bmatrix} r+\delta & 0 \\ \dfrac{1}{2w\theta(1-T^{d})} & -\delta \end{bmatrix} \begin{bmatrix} \lambda \\ K \end{bmatrix} + \begin{bmatrix} -\beta(P)(1-T^{d}) \\ -P_k/2w\theta \end{bmatrix} . \tag{3.9}$$

The eigenvalues of this expression are $r+\delta$ and $-\delta$, which are distinct, nonzero and have mixed signs. The model in section (2), therefore, has the saddle path property.

To see why the eigenvalues are so important, consider solving (3.8) near the steady state (so that A and b are essentially constant). The first step is to look for a solution to the homogeneous form of the problem in which b is zero:

22. Strictly speaking, these are sufficient conditions for saddle path stability. For a review of the properties of eigenvalues, or of linear algebra in general, see Strang (1980). In this context, we are using "linear" in the differential equations sense only.

$$x' = Ax \; . \tag{3.10}$$

If all of the eigenvalues of A are distinct and nonzero, then any vector x can be written as a linear combination of the eigenvectors of A. Thus, if Γ is a matrix whose columns are the eigenvectors, the following must be true:

$$x = \Gamma c \; , \tag{3.11}$$

where c is a vector of coefficients. Since Γ is constant near the steady state (because A is constant), differentiating (3.11) with respect to time gives the following:

$$x' = \Gamma c' \; . \tag{3.12}$$

Substituting (3.11) and (3.12) into (3.10) gives:

$$\Gamma c' = A \Gamma c \; . \tag{3.13}$$

However, since Γ is composed of the eigenvectors of A, it must be the case that:

$$A \Gamma = \Gamma \psi \; , \tag{3.14}$$

where ψ is a diagonal matrix of eigenvalues. Inserting (3.14) into (3.13) and multiplying through by the inverse of Γ produces the following:

$$c' = \psi c \; . \tag{3.15}$$

Thus, the original differential equation can be transformed into (3.15), which is much easier to solve: ψ is diagonal, so (3.15) is nothing more than a collection of unrelated differential equations, each of which can be solved by the method of integrating factors. The solution to (3.15) is:

$$c = e^{\psi t} \cdot \gamma \; , \tag{3.16}$$

where γ is a vector of integration constants.[23] From (3.11), this means that x has the solution:

$$x = \Gamma e^{\psi t} \cdot \gamma \; . \tag{3.17}$$

The complete solution to the original equation (3.8) is the sum of (3.17), which solves the homogeneous equation, and a particular solution to (3.8), such as the steady state. Thus, a final

23. In case this notation is unfamiliar, a scalar raised to a matrix power is a matrix whose elements are the scalar raised to the power of each element of the original matrix.

expression for x is the following:

$$x = \Gamma e^{\psi t} \cdot \gamma + x^{ss} \ . \tag{3.18}$$

Equation (3.18) allows us to infer a great deal about the behavior of the solution from the eigenvalues alone. For example, if all the eigenvalues had positive real parts, then the model would only converge to x^{ss} if every element of γ were zero. A model whose eigenvalues were all positive, therefore, could only reach the steady state if it started out there initially. For this reason, such models are said to be "unstable". On the other hand, if the eigenvalues all had negative real parts, the solution would converge to the steady state from any starting vector. This comes about because the first term on the left in (3.18) will always approach zero as time tends to infinity when the eigenvalues are negative. Models with this property are said to be "globally stable".

Many economic models, however, have a mixture of positive and negative eigenvalues. Since they are neither unstable nor globally stable, such models are often said to be "saddle-path stable". At the most fundamental level, saddle-path stability means that a model will converge to its steady state from some initial vectors but not from others. In a sense, this is intermediate between the unstable and globally stable cases: an unstable model will only reach the steady state from a single point − the steady state itself − while a globally stable model will get there from any starting point. Saddle path models are in between: they converge to the steady state from some, but not all, of the possible initial vectors.

To put this more formally, suppose x is a vector of length n in $\mathbf{R}^n$. If all the eigenvalues are distinct, there will be n of them. In general, if m of the eigenvalues have negative real parts, the model will converge to the steady state from an m dimensional subspace of $\mathbf{R}^n$. If, for example, all n eigenvalues are positive ($m = 0$), the model will converge to the steady state from a subspace of dimension 0 − a single point. This is the unstable case discussed above. If all n eigenvalues are negative ($m = n$), the steady state can be reached from any point in $\mathbf{R}^n$, so the model will be globally stable. However, if there are m negative eigenvalues, where $0 < m < n$, the system can reach the steady state from an m dimensional subspace of $\mathbf{R}^n$. In practical terms, this means that if we want the system to converge to the steady state, we can only choose m elements of x independently − the other $n - m$ terms will be implied by the model.

This property of saddle path models is very useful because we usually do not know the initial values of all elements of vector x. In the model of section (2), for example, we do not know the value of λ immediately after a shock. Since the model is saddle-path stable, however, we know that it can only attain the steady state from certain (λ, K) pairs. Furthermore, because one of the model's two eigenvalues is negative, we know that if we choose a value for either λ or K, the value of the other will be implied. Thus, knowing the initial value of K and requiring that the model eventually attain the steady state is enough, in principle, to let us calculate the value of λ. As we will discuss in section (4), however, computing λ usually requires numerical integration.

In summary, checking that a particular model has a unique stable path requires computing the model's eigenvalues and verifying that they are distinct, nonzero and of mixed sign. If the stable path is unique, however, it allows features of the steady state to be imposed on the solution as boundary conditions. For models with foresight, this is often essential.

3.4 Analyzing an Experiment

Once the phase diagram is complete it can be used to analyze the effects of changes in the exogenous variables. As we will show, the phase diagram is a very powerful tool: it provides a complete qualitative description of the effects of a shock. The only thing it does not provide, of course, is a set of numerical results for the variables. Sometimes numerical results are necessary, so we will discuss numerical analysis in section (4). Keep in mind, however, that all of the qualitative features of the solution will be revealed by the phase diagram.

Before going into the details of analyzing experiments it is useful to consider what sorts of shocks can be studied with an intertemporal model. In a static model a shock consists of a single change in an exogenous variable. A tax rate, for example, might rise. In intertemporal modeling, however, the entire path of the tax rate over time matters. This makes it convenient to group experiments into four categories depending on: (1) whether the shock is permanent or temporary, and (2) whether the change is announced in advance or implemented immediately. One of the most interesting features of intertemporal modeling is that a given shock can have substantially different effects depending on how it is enacted over time. A temporary tax increase, for example, can produce effects that are completely different from the permanent version of the same policy.

These distinctions between policies mean that each shock has (at least implicitly) three dates associated with it: its announcement, its implementation, and its repeal. The announcement date is the time at which the public first becomes aware of the policy. It is often quite a bit earlier than the date of implementation, which is the time at which the policy (and the relevant exogenous variable) actually changes. Temporary policies also have a repeal date, at which the shock ceases and the exogenous variables return to their original values.

All four categories of experiment can be analyzed using phase diagrams. A natural place to start is with immediate, permanent changes in policy. The first step in using the phase plane is to determine how the shock affects the zero-derivative loci. Most shocks will shift one or both of the loci, resulting in a new steady state. The second step is to find the new stable path. Usually this will be straightforward once the new loci have been found. At this point, the post-shock phase diagram is complete. It governs the motion of the system when all of the exogenous variables have their post-shock values. Thus, the overall diagram consists of two superimposed phase planes: one which applies when the exogenous variables have their initial values, and one which applies after the shock.

The remaining step is to trace out the motion of the system over time. At the instant the policy is implemented, which in this case is immediately, the state variables (K in the example above) are fixed and cannot change. This means that if the economy is ever to get to the new steady state, the costate variables must immediately jump to the new stable path. Once on the stable path, the economy evolves over time toward the steady state. In the phase diagram this will appear as a vertical jump in the costate variable, followed by gradual movement along the new stable path.

To make this discussion more concrete we will now show how an immediate, permanent increase in the dividend tax in the model of section (3) could be analyzed. First, consider what happens to the λ' locus. For convenience, equation (2.35) is repeated below:

$$\lambda' = (r + \delta)\lambda - \beta(P)(1 - T^d) \; . \tag{2.35}$$

As T^d rises, the rightmost term in equation (2.35) becomes closer to zero. For λ' to remain zero, therefore, λ must fall. Thus, the $\lambda' = 0$ locus must shift downward. The new location of the locus can be found by solving for the new steady state value of λ. To find the effect of the tax increase on the $K' = 0$ locus, we begin with equation (2.36), repeated below:

$$K' = \frac{\lambda}{2w\theta(1 - T^d)} - \delta K - \frac{P_k}{2w\theta} \; . \tag{2.36}$$

The increase in T^d causes the leftmost term on the right hand side to rise. For K' to remain zero at constant λ, K must rise. Thus, the $K' = 0$ locus shifts to the right. The result of shifting both loci is shown in figure 3.4.

The location of the new steady state can be found by setting λ' and K' equal to zero and solving for λ and K in equations (2.35) and (2.36), using the new values of the exogenous variables. In this model, the dividend tax has a particularly interesting effect: it does not change the steady state value of the capital stock. This can be verified by solving for λ^{ss}, inserting it into equation (2.36), and solving for K^{ss}. The dividend tax cancels out, so it has no effect on the steady state capital stock. After a moment's thought the intuition behind this result is clear: T^d is a pure profits tax which does not affect any margins. An increase in the tax lowers the post-tax dividends that firms can pay, so λ falls, but it does not affect the optimal level of investment, so the capital stock is unchanged.

For this policy it is easy to trace out the motion of the system over time. When the tax rises, λ jumps down to the new stable path. However, the capital stock is already at its steady state level, so the jump in λ brings the model instantly to the new steady state. Thus, the only change in the system is an immediate drop in λ; the capital stock and investment are completely unaffected. This path is shown in figure 3.5 by a grey line.

Interestingly, the results are quite different when the shock is anticipated. Suppose that instead of implementing the tax increase immediately, the government announces that it will occur after three years. The initial and final steady states are exactly the same as in the previous case, so the basic phase diagram is the same as figure 3.4. The path of the model over time, however, is more complicated because the policy change does not occur immediately.

When the policy is announced, λ falls part of the way toward its new steady state value, but not as far as it would if the policy took effect immediately. It stays higher initially because the dividend tax stays at its old value for three years, so dividends paid during that time will not be taxed any more heavily than they were before. However, λ does drop below its original value because the firm will eventually pay lower dividends.

After the initial drop in λ, the system evolves according to the equations of motion associated with the original steady state. These equations apply because they depend only on current tax rates and none of the taxes have actually changed yet. Thus, the system moves down and to the left. It continues to move in that direction until the tax change occurs in year three. At that time, the model becomes governed by the new equations of motion. These have the same form as the original equations, but differ in that they are evaluated at the new value of T^d. Since the model is required to attain the steady state eventually, it must be on the new stable path when the tax is implemented in year three. After year three, the system evolves along the stable path toward the new steady state. The path of the model is shown in figure 3.6.

Notice that figure 3.6 shows no jump in λ at the instant when the tax is implemented. Instead, λ evolves smoothly and reaches the new stable path precisely at the moment of implementation without jumping. This reflects an important feature of intertemporal models with perfect foresight: there are no windfall gains or losses from the *implementation* of anticipated policies. There can be windfalls associated with the *announcement* of a policy--in this example, λ falls at the announcement--but there are no windfalls from events that have been anticipated.

This point can be understood intuitively by thinking about what happens to λ near implementation. Recall from equation (2.23) that λ is the present value of the after tax earnings of a marginal unit of capital. Once the tax has actually increased, all subsequent earnings are evaluated at the new rate. Shortly before implementation, however, the value of an extra unit of capital is what it will earn after implementation plus a small amount more obtained before the tax change. As time becomes closer and closer to implementation, the extra amount of earnings becomes smaller and smaller, so λ approaches its post implementation value smoothly.

It is straightforward to demonstrate this mathematically. Let the time of implementation be τ and the value of λ at that point be $\lambda(\tau)$. Now consider the value of λ at an instant Δ before implementation. Equation (2.23) can be written:

$$\lambda(\tau - \Delta) = \int_{\tau-\Delta}^{\tau} \frac{\partial E}{\partial K}(1 - T^d)e^{-(r+\delta)(s-\tau-\Delta)}\, ds + \lambda(\tau)e^{-(r+\delta)\Delta} \,. \tag{3.19}$$

As Δ approaches zero, $\lambda(\tau - \Delta)$ approaches $\lambda(\tau)$. Thus, λ must be continuous at implementation and cannot jump. Notice that this argument does not depend on features of the model such as the form of the earnings or investment cost function. It is a very general property of perfect foresight models.

Figure 3.6 shows something very striking: the capital stock falls during the period between announcement and implementation of the policy. This occurs because firms respond to the policy by paying higher dividends before implementation while the dividend tax is low. Higher dividends limit investment and drive down the capital stock. Once the tax is in place, however, investment returns to normal and dividends drop. Since the capital stock was driven down, restoring the normal level of investment causes the amount of capital to rise. This, in turn, causes the value of the firm to increase and allows investors to receive part of their return as capital gains.[24] Since the capital gains tax is unchanged (and zero), investors benefit from shifting part of their return from heavily taxed dividends to lightly taxed capital gains.

The difference between implementing the dividend tax increase immediately and announcing the change in advance highlights one of the most interesting aspects of intertemporal modeling. There is no way that the adverse effect of the announced policy on the capital stock could have been discovered using a static model. This kind of unexpected result arises frequently in the analysis of announced policies, and also in the study of temporary policies. Since it is rare for shocks to the economy to come as a complete surprise, explicit modeling and analysis of announcement effects is essential to understanding the impact of government policy and the consequences of other kinds of shocks.

We hope this discussion has demonstrated the value of phase plane analysis in the study of intertemporal models. All qualitative features of the model's response to any shock can be obtained using the phase plane. The only details it does not provide are the coordinates of particular points in the plane. We might want to know, for example, exactly how far λ initially drops in figure 3.6, or exactly how much the capital stock has fallen by implementation of the tax increase. To obtain these values it is necessary to find an explicit solution to the model using numerical methods.

4 Numerical Methods

To solve the model numerically, we must obtain explicit numerical paths for the dynamic variables (λ and K) throughout time. Once these are known, the paths of other variables can be found easily by applying equations from the model. For example, in section (2) investment can be calculated from λ using equation (2.34). This means that finding a numerical solution to the model boils down to solving the model's equations of motion. These are a set of simultaneous differential equations, so solving them requires some form of numerical integration.

If the initial values of λ and K were known, it would be easy to integrate (2.35) and (2.36). Differential equations for which all boundary conditions are known at the initial point in time are called initial value problems, and there are many methods available to solve them.[25] A simple, intuitive approach is Euler's method, which works in the following way.[26] Let the first instant

24. Refer to the arbitrage condition shown in equation (2.1).

25. There is an enormous range of methods available for solving initial value problems. For more information consult Press, *et al.* (1986).

26. Although it is intuitively appealing, Euler's method is not usually satisfactory in practice. For a complete discussion, refer to Press, *et al.* (1986).

after the shock be called time 0. If λ and K were known to take values λ_0 and K_0 at $t = 0$, those values could be used in (2.35) and (2.36) to calculate $\lambda'(0)$ and $K'(0)$. Multiplying the derivatives by a tiny increment of time, say Δt, would show approximately how much λ and K changed over that interval. Adding these changes to λ_0 and K_0 would give approximate values for λ and K at time Δt. These could then be used in (2.35) and (2.36) to obtain $\lambda'(\Delta t)$ and $K'(\Delta t)$. By applying this process repeatedly, the entire future path of λ and K could be calculated. Moreover, the solution could be made arbitrarily accurate by making the step size Δt sufficiently small.

Unfortunately, the initial post-shock values of any costate variables in the model (such as λ above) will usually be unknown.[27] This leaves the model without enough boundary conditions to determine the solution uniquely. To understand this intuitively, recall the phase diagram in figure 3.6. If $\lambda(0)$ is not known, there is no way to determine where the system will be immediately after the shock, except that it will be somewhere in the vertical line of points above $K(0)$. Since different points above $K(0)$ lead to drastically different paths of the economy over time, the solution is not completely determined.

Fortunately, this indeterminacy can be eliminated for models having the saddle path property discussed in section (3.3). Any absent initial conditions can be replaced by conditions on the long run behavior of the costate variables. Typically this is accomplished by imposing transversality conditions which require the costate variables to approach their steady state values as time tends to infinity. These replace the missing initial conditions and allow the solution to be uniquely determined. This produces a system, however, in which some of the boundary conditions hold at the initial time, and some at the steady state. When the boundary conditions are scattered among several points in time, the system is formally described as a "two-point boundary value problem", and it cannot be solved using techniques for initial value problems.

In fact, two-point boundary value problems are much harder to solve than their initial value counterparts, and require special numerical methods. The next few sections describe some of these methods and how they perform for economic models. To keep the discussion concise and fairly concrete, we will focus on models that have a single costate variable, such as the model in section (2). However, all of the methods can be extended without difficulty to handle multiple costate variables.

4.1 Shooting

One intuitive way to solve a two-point boundary value problem would be to guess the missing initial condition, integrate the system forward as though it were an initial value problem, and check whether the transversality condition was satisfied. If it was not, the guessed condition could be revised and the entire process repeated. Eventually an initial condition would be found which led to the steady state when the system was integrated forward. This approach has been used extensively in engineering and the physical sciences, and is known as "shooting".[28]

---

27. By this we mean that the costate variables may jump initially, taking on new values which cannot be determined without solving the entire model.

28. Shooting is described in Press, *et al.* (1986), Roberts and Shipman (1972), and in most textbooks on numerical analysis.

In practice, shooting is usually implemented in the following way. Let the missing initial condition be denoted λ_0. For each guess of λ_0, the model is integrated forward using Euler's method to a large but finite time T. This generates a $\lambda(T)$ which we will refer to as an "achieved" value and denote λ_T^a. Next, the transversality condition is tested by comparing λ_T^a with its steady state value λ^{ss}.[29] For convenience, we can define a function M which measures how close the solution is to the steady state:

$$M(\lambda_0) = \lambda_T^a(\lambda_0) - \lambda^{ss} \ . \tag{4.1}$$

Since the achieved value of λ at T depends on the guess of λ_0, both λ_T^a and M are written as functions of λ_0. When a guess of λ_0 has been found for which M less than a specified tolerance, a solution has been obtained.

Defining M as in (4.1) makes it clear that the object of shooting is to choose a value of λ_0 that sets M to zero. This suggests using Newton's method to update the guess at each iteration. At iteration $k+1$, a first-order Taylor series expansion of M about the previous guess λ_0^k for a trial solution λ_0^{k+1} gives:

$$M(\lambda_0^{k+1}) = M(\lambda_0^k) + \frac{dM(\lambda_0^k)}{d\lambda}\,(\lambda_0^{k+1} - \lambda_0^k) \ . \tag{4.2}$$

Assuming that λ_0^{k+1} is indeed a solution, $M(\lambda_0^{k+1})$ can be set to zero. This allows the equation to be rearranged as shown:

$$\lambda_0^{k+1} = \lambda_0^k - \frac{M(\lambda_0^k)}{dM(\lambda_0^k)/d\lambda} \ . \tag{4.3}$$

Thus, by evaluating both M and its first derivative at λ_0^k, a revised guess of λ can be constructed.

Unfortunately, in economic applications shooting suffers from severe numerical instability and can rarely be used. Small errors made in the initial guess of the missing costate variable lead to intertemporal paths that move far from the steady state after only a few years. The model of section (2) provides a typical example of why this problem occurs. Recall equation (2.35), one of the model's equations of motion:

$$\lambda' = (r+\delta)\lambda - \beta(P)(1 - T^d) \ . \tag{2.35}$$

If the true post-shock value of λ_0 is inserted on the right side of the equation, the true value of $\lambda'(0)$ can be calculated and the system integrated forward toward the steady state. On the other hand, consider what happens if the guess of λ_0 is too high. Evaluating (2.35) would give a value

29. The value of λ^{ss} is calculated in advance using the approach discussed in section (3.1).

of λ' that was also above its true value. Thus, if λ starts out too high, it will grow too fast as well. As it grows, λ' increases, so λ moves farther and farther away from the stable path.

In fact, for the model of section (2) the error in λ grows exponentially over time at rate $r + \delta$. Thus, if the initial guess of λ is too high by Δ, after T years the error will have grown to $\Delta e^{(r+\delta)T}$ which could be a gigantic number.[30] For example, if the interest rate were 5 percent and depreciation rate 10 percent, after 100 years the error in λ would exceed $\Delta \times 10^{10}$. A similar problem arises if the initial guess of λ is too small. This means that a small error in the initial guess of λ will set the model on a dynamic path that leads far from the steady state. In the terminology introduced above, M will be huge for even very small errors in λ_0. This sensitivity makes shooting very vulnerable to the rounding errors introduced by computer programs and prevents it from being useful for most economic models.

4.2 Multiple Shooting

Multiple shooting is a refinement of simple shooting that helps control models with explosive tendencies.[31] The full period over which the model is to be solved is divided into a number of subintervals and the model is then shot over each. Shooting over shorter periods keeps the model from drifting too far from the stable path in any one interval. This limits the numerical damage done by rounding errors, so multiple shooting can be used with economic models. However, using more intervals means that rather than searching for a single missing initial condition, the algorithm must find a vector of such conditions spread out across time.

As a example of how multiple shooting is used, consider solving the investment model over two adjoining intervals: $[0, \tau]$ and $[\tau, T]$. The first step is to guess what values λ will take at 0 and τ. Like simple shooting, multiple shooting is an iterative procedure, so let the guesses at iteration k be denoted by λ_0^k and λ_τ^k. The next step is to integrate the model forward from 0 to τ starting at the known initial capital stock, K_0, and the guess λ_0^k. This produces a pair of achieved values of $K(\tau)$ and $\lambda(\tau)$ which we will denote by K_τ^a and λ_τ^a. Using K_τ^a and λ_τ^k as initial conditions, the model is then integrated forward from τ to T. From this, an achieved value of λ at T will be obtained.

The key feature of multiple shooting is that the integration over $[\tau, T]$ starts from the guess λ_τ^k, and not from λ_τ^a, the achieved value from the first integration. Starting from the achieved value of λ would be exactly the same as integrating the model over $[0, T]$, which is ordinary shooting. When the guess of λ_0^k is incorrect, however, λ will have drifted very far from its true value by time τ. This makes λ_τ^a a terrible estimate of what λ should actually be at τ, so replacing it with a guessed value − even a bad guess − vastly reduces the error in the second integration.

To see this intuitively, recall the example of error propagation described in section (4.1): after T years, an initial error Δ had compounded to a miss distance of $\Delta e^{(r+\delta)T}$. Dividing the

30. The error's growth rate is given by the model's positive eigenvalue, $r + \delta$; see section (3.3).

31. Multiple shooting was introduced to economics by Lipton, *et al.* (1982), but has a long history of use in other disciplines. It is

period $[0, T]$ into two subintervals of equal length reduces the problem enormously. An error Δ in the guess of λ_0 only grows to $\Delta e^{(r+\delta)T/2}$ by the end of the first interval, which is the square root of its previous value. Of course this reduction comes at a cost: a guess for λ is also required at the beginning of the second interval. Assuming that another error of the order of Δ is made in the second guess, λ will miss its steady state value by $\Delta e^{(r+\delta)T/2}$. Thus, dividing the interval into two equal parts reduces the total miss distance M to roughly $2M^{1/2}$. When M is large, this will be an enormous improvement. Moreover, the number of subintervals is not limited to two. Using more subintervals reduces the error propagation problem even further, so any explosive tendencies of the model can be completely controlled.

Since the costate variable has to be guessed at the beginning of each interval, the revision rule used to generate new guesses at each iteration is slightly more complicated than the one used for simple shooting. In the case of two subintervals, the guess of λ for the beginning of the second interval, λ_τ, would be revised until M_2 in following equation became zero:

$$M_2(\lambda_0, \lambda_\tau) = \lambda_T^a(\lambda_0, \lambda_\tau) - \lambda^{ss} \ . \tag{4.4}$$

As with ordinary shooting, M_2 is a miss distance. The subscript 2 has been added to indicate that it is the miss distance for the second interval. Just as in shooting, M_2 and λ_T^a both depend on λ_τ, the guess of λ at the beginning of the interval. However, they now also depend on an earlier λ, λ_0. This occurs because λ_0 affects K_τ, the starting capital stock for the second interval.

A second rule is needed to guide revision of λ_0. Since the model will not necessarily have reached the steady state by τ, trying to reduce M_2 in (4.4) to zero would be inappropriate. Instead, λ_0 is revised until M_1 in the expression below becomes zero:

$$M_1(\lambda_0, \lambda_\tau) = \lambda_\tau^a - \lambda_\tau \ . \tag{4.5}$$

That is, λ_0 is varied until a value is found that can be integrated forward to attain the starting guess of λ for the next interval. As a result, when the correct value of λ_τ has been found (so that M_2 in (4.4) is zero), a value of λ_0 that makes M_1 in (4.5) zero must be the true initial value of λ; if it were inserted into the model's equations of motion, the system could be integrated forward to time T and the transversality condition would be satisfied.

Thus, dividing the original period into two subintervals means that two variables (λ_0 and λ_τ) must now be chosen to satisfy two equations ($M_1 = 0$ and $M_2 = 0$). This suggests using the multivariate version of Newton's method to compute an updated vector of guesses at each step of the algorithm. If λ^k is the vector of guesses at iteration k, then a new guess could be constructed as follows:

$$\lambda^{k+1} = \lambda^k - J^{-1} M(\lambda^k) \ , \tag{4.6}$$

described clearly and in detail in Roberts and Shipman (1972).

where J is the Jacobian matrix of partial derivatives of M evaluated at λ^k and λ^k itself is the guess of λ at iteration k. Equation (4.6) can be used with any number of shooting intervals, and with multiple costate variables, so multiple shooting is a fairly robust and versatile algorithm.

Unfortunately, it also consumes a great deal of computer time. Moreover, like all algorithms based on Newton's method, multiple shooting is not guaranteed to converge. Worse yet, it is particularly unsuitable for intertemporal general equilibrium models because it requires solving the intraperiod part of the model thousands of times in the course of finding a full intertemporal solution.[32] A single intraperiod equilibrium solution requires solving a static short run general equilibrium model, which for even moderate sized models will require a noticeable amount of computer time. Having to compute hundreds or thousands of these solutions makes multiple shooting of very limited use for general equilibrium work.

4.3 The Fair-Taylor Method

A third method for solving two-point boundary value problems, and one which is often used for intertemporal general equilibrium models, is known as the Fair-Taylor algorithm after its originators.[33] It is much easier to use than multiple shooting, controls explosive tendencies in the solution equally well, and requires somewhat less computer time.

The algorithm itself is very simple. First, a guess is made of the entire path of the unknown costate variable. That is, instead of guessing a single λ as in shooting, or a handful of λ's as in multiple shooting, values of λ are guessed for each point in the set $\{0, 1, \ldots, T\}$. For convenience, let the guess at iteration k be denoted by the vector λ^k. If T is chosen to be year 100, λ^k will usually have 101 elements (the extra one is for year zero).[34] The final element is always chosen to satisfy the transversality condition. Using λ^k and the equation of motion of the capital stock, the model is integrated forward from the initial point to the terminal time. During this process, the costate variable's equation of motion is temporarily ignored. The result is a vector of the capital stocks, K^k for iteration k, that is consistent with λ^k. It is the path the economy would follow if λ actually had the sequence of values in λ^k. However, it is not necessarily a solution to the model because λ_k does not necessarily satisfy λ's equation of motion. This means the algorithm must iterate over vectors λ until one is found that satisfies both equations of motion.

Revising the guess vector between iterations is accomplished by using the equation of motion for λ in a special way. The technique can best be explained by example, so consider a

32. Thousands of intratemporal solutions are needed because each iteration of the algorithm requires solving every period in the time interval several times. If the model is to be solved over [0,100], for example, 101 intraperiod solutions are required just to integrate the path forward from 0 to T once. Much worse, however, is that the Jacobian matrix will usually have to be computed numerically. That requires perturbing each of the elements of λ and computing an entire solution path from 0 to T. If, for example, there are five shooting intervals, the entire path of 101 intratemporal solutions would have to be computed six times in order to evaluate M and J. Since over 600 intraperiod solutions would have to be found for a single iteration of the intertemporal algorithm, the method is not useful for more than very small models.

33. The Fair-Taylor algorithm was originally proposed by Fair (1979), and later extended by Fair and Taylor (1983). This section describes Fair and Taylor's "type II" iteration. They also proposed a "type III" procedure which can be used when the terminal condition cannot be computed easily.

34. Simulating periods one year apart is not necessary for the Fair-Taylor algorithm to work. However, it is the most common approach.

slightly more general version of the model introduced in section (2). There will be one state variable (K), one costate variable (λ), a vector of exogenous variables (Z), and two equations of motion, one in K' and one in λ'. The equation of motion for λ will have the following form:

$$\lambda'_t = f(\lambda_t, K_t, Z_t) , \tag{4.7}$$

where f is a function that depends on the structure of the model. Equation (2.35), for example, is a special case of (4.7) in which a particular functional form has been imposed for f. The derivative of λ at time t can be approximated by the difference between two consecutive values of λ:[35]

$$\lambda'_t \approx \lambda_{t+1} - \lambda_t . \tag{4.8}$$

Inserting this into (4.7) produces the equation below:

$$\lambda_{t+1} - \lambda_t \approx f(\lambda_t, K_t, Z_t) . \tag{4.9}$$

This expression holds at all points along the path of λ. Rearranging it slightly and dropping the implied error terms produces the following:

$$\lambda_t = \lambda_{t+1} - f(\lambda_t, K_t, Z_t) . \tag{4.10}$$

Equation (4.10) must hold at all points along the solution path so it suggests a revision rule for guesses of λ. By inserting values of λ_{t+1}, λ_t and K from iteration k into the right side of (4.10), an implied value of λ_t could be calculated. To use the terminology introduced in section (4), an achieved value λ_t^a could be computed as shown:

$$\lambda_t^a = \lambda_{t+1}^k - f(\lambda_t^k, K_t^k, Z_t) . \tag{4.11}$$

At the solution, λ_t^a will be exactly equal to λ_t^k because the solution vector must satisfy λ's equation of motion by definition. Away from the solution, however, λ_t^a will not be the same as λ_t^k. The Fair-Taylor algorithm uses λ_t^a to update the guess of λ_t in the following way:

$$\lambda_t^{k+1} = \alpha\lambda_t^k + (1-\alpha)\lambda_t^a , \tag{4.12}$$

where α is a parameter used to ensure the algorithm converges smoothly. Its takes on values in the interval $[0, 1]$ and is typically around one-half.[36]

---

35. There are a number of ways to approximate the derivative of a function at a point, but this formulation is particularly convenient for this algorithm. For further details, refer to exercise E7.

36. Choosing α carefully is fairly important. It is not a good idea to make α too close to one because that puts undue emphasis on λ_t^a, which is not necessarily closer to the true solution than λ_t^k. That is, the true value of λ_t may lie between λ_t^k and λ_t^a, but be much closer to λ_t^k than λ_t^a. In that case, $\alpha \approx 1$ would tend to make the algorithm diverge. On the other hand, if α is too close

Thus, for a simple investment model the algorithm would be applied in the following way. Given a guess of the path of λ, the first step would be to compute the corresponding path of K by integrating forward from the initial point. Next, using the guess of λ and the resulting path of K, construct a sequence of achieved values λ^a. Finally, create a new guess by taking a convex combination of the old guess and the achieved values. The solution has been obtained when the guess and achieved values differ by less than a given tolerance.

On the surface, this technique seems to differ quite a bit from multiple shooting. At a deeper level, however, the algorithms are very similar. Multiple shooting proceeds by guessing λ at a handful of points, integrating forward, and employing a fairly sophisticated updating rule. Fair-Taylor works by guessing λ at a vast number of points, integrating forward, and using a very simple updating rule. In a sense, the Fair-Taylor algorithm is a version of multiple shooting in which there are T shooting intervals−one for each year of the solution. It can, however, be somewhat faster than multiple shooting because it does not require computing the Jacobian matrix of the miss distances at each iteration. In practice, this means that Fair-Taylor will require more iterations to converge than multiple shooting, but each iteration will be much faster to compute.

4.4 Finite Differences

A fourth approach to solving two-point boundary value problems is the finite difference method.[37] It differs from the three previous algorithms because it does not operate by guessing the missing costate variable and integrating forward to see if the transversality condition is satisfied. Instead, a system of overlapping difference equations is constructed which approximates the model's equations of motion. This system is then solved simultaneously to give the paths of the model's dynamic variables. The initial and terminal conditions will both be satisfied exactly, and the algorithm is completely immune to the numerical instability that plagues shooting methods. In addition, since finite differences is not iterative, it will always find a solution if one exists: it will never fail to converge.[38]

The first step in using finite differences is to replace all of the derivatives in the model by finite difference formulae. These formulae are local approximations to derivatives, and are constructed in a straightforward way from Taylor series expansions. For example, one approximation for a first derivative might be constructed as follows. First, expand the function of interest about a particular time t for an adjacent time $t+h$:

$$f(t+h) = f(t) + f'(t)h + O(h^2) \ , \tag{4.13}$$

where $O(h^2)$ represents the Taylor series terms of order h^2 and above. Rearranging (4.13) and dropping the higher-order terms shows that:

---

to zero, the algorithm will converge very slowly.

37. Finite differences is a standard method in engineering and physics. It was first applied to economic models by Wilcoxen (1985a).

38. Strictly speaking, finite differences is not iterative in the sense that multiple shooting iterates over guesses of the costate variable. It may, however, require iteration to solve the difference equations if they are nonlinear.

$$f'(t) \approx \frac{f(t+h) - f(t)}{h} \ . \tag{4.14}$$

The term on the right is a finite difference approximation to the derivative of f evaluated at time t. Since it was constructed using current and future values of f, it is technically known as a forward difference. Dividing through by h reduced the terms that were dropped to $O(h)$, so the approximation itself will be accurate to that order.[39] This means that if (4.14) were used to replace a derivative, the resulting equation would be accurate only when h was fairly small. Thus, it is inappropriate to use a single difference equation to approximate the original model over a long period of time.

Long periods of time can, however, be modeled using a series of expressions like (4.14), each holding over successive intervals of time. If the total period were 100 years, for example, it could be broken up into two intervals of fifty years each. Then, one equation like (4.14) could link years 0 to 50, while another connected years 50 to 100. Of course 50 years is still a very large value for h, so it would usually be necessary to break the original interval up into even smaller segments using many more equations. If necessary, the solution can be made arbitrarily accurate by using a sufficiently small step size h. No matter how many intervals are actually used, in the end the original differential equation will have been replaced by a system of difference equations which link values of f at different points in time. Solving this system simultaneously would yield the entire path of f. For many models the equations will be linear or easily linearized, so often the solution can be found using a variant of Gaussian elimination. This makes finite differences very fast and a natural choice for use with general equilibrium models that employ Johansen's method.

To provide a more concrete example of how the method is actually implemented, consider solving a fairly general investment model with equations of motion as shown below:

$$\lambda'(t) = a(t)\lambda(t) + b(t)K(t) - c(t) \ , \tag{4.15}$$

$$K'(t) = d(t)\lambda(t) + e(t)K(t) - f(t) \ . \tag{4.16}$$

The model of section (2) is a special case of this in which $b(t) = 0$. Converting (4.15) and (4.16) into finite difference form using forward differences produces the following:

$$\frac{\lambda(t+h) - \lambda(t)}{h} = a(t)\lambda(t) + b(t)K(t) - c(t) \ , \tag{4.17}$$

$$\frac{K(t+h) - K(t)}{h} = d(t)\lambda(t) + e(t)K(t) - f(t) \ . \tag{4.18}$$

---

39. A number of other difference formulae will be discussed in exercise (E4), some of which are accurate to higher orders.

For convenience, equations (4.17) and (4.18) can be written in matrix notation as shown:

$$\begin{bmatrix} a(t)+\frac{1}{h} & b(t) & \frac{-1}{h} & 0 \\ d(t) & e(t)+\frac{1}{h} & 0 & \frac{-1}{h} \end{bmatrix} \begin{bmatrix} \lambda(t) \\ K(t) \\ \lambda(t+h) \\ K(t+h) \end{bmatrix} = \begin{bmatrix} c(t) \\ f(t) \end{bmatrix}. \tag{4.19}$$

This system approximates the true equations of motion, (4.15) and (4.16), in the neighborhood of t. The complete solution requires a set of such equations, one for each interval of time h. If there are N intervals, collecting the approximations together produces a set of equations with the structure below:

$$\begin{bmatrix} a(t^0)+\frac{1}{h} & b(t^0) & \frac{-1}{h} & 0 & 0 & 0 \\ d(t^0) & e(t^0)+\frac{1}{h} & 0 & \frac{-1}{h} & 0 & 0 \\ 0 & 0 & a(t^1)+\frac{1}{h} & b(t^1) & \frac{-1}{h} & 0 & \cdots \\ 0 & 0 & d(t^1) & e(t^1)+\frac{1}{h} & 0 & \frac{-1}{h} \\ & & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \lambda(t^0) \\ K(t^0) \\ \lambda(t^1) \\ K(t^1) \\ \cdots \\ \lambda(t^N) \\ K(t^N) \end{bmatrix} = \begin{bmatrix} c(t^0) \\ f(t^0) \\ c(t^1) \\ f(t^1) \\ \cdots \\ c(t^{N-1}) \\ f(t^{N-1}) \end{bmatrix}. \tag{4.20}$$

This is a system of $2N$ equations and $2(N+1)$ variables. However, two of the variables are known from the model's boundary conditions: $K(t^0)$ and $\lambda(t^N)$ (using λ^{ss} as the approximate value of $\lambda(t^N)$). Moving the corresponding columns over to the right side of the equation and simplifying produces:

$$\begin{bmatrix} a(t^0)+\frac{1}{h} & \frac{-1}{h} & 0 & 0 & 0 \\ d(t^0) & 0 & \frac{-1}{h} & 0 & 0 \\ 0 & a(t^1)+\frac{1}{h} & b(t^1) & \frac{-1}{h} & 0 & \cdots \\ 0 & d(t^1) & e(t^1)+\frac{1}{h} & 0 & \frac{-1}{h} \\ & \cdots \end{bmatrix} \begin{bmatrix} \lambda(t^0) \\ \lambda(t^1) \\ K(t^1) \\ \cdots \\ \lambda(t^{N-1}) \\ K(t^{N-1}) \\ K(t^N) \end{bmatrix} = \begin{bmatrix} c(t^0)-b(t^0)K(t^0) \\ f(t^0)-(e(t^0)+\frac{1}{h})K(t^0) \\ c(t^1) \\ f(t^1) \\ \cdots \\ c(t^{N-1})+\frac{\lambda(t^N)}{h} \\ f(t^{N-1}) \end{bmatrix} \tag{4.21}$$

This direct use of the boundary conditions eliminates the need for iteration and removes the numerical instability problem associated with shooting methods. Equation (4.21) can be written compactly as:

$$\Theta F = B\,, \tag{4.22}$$

where Θ is a matrix of coefficients, F is a vector of unknown values of λ and K, and B is a vector resulting from applying the boundary conditions. Solving the model requires finding the unknown values of F, which can be accomplished by computing:[40]

$$F = \Theta^{-1} B \ . \tag{4.23}$$

Since approximations were used for derivatives in the model, the results obtained using (4.23) will contain a certain amount of error. This is known as truncation error because it arises from dropping high-order terms in the Taylor series expansions used to form the difference approximations. For the difference formulae used above, the truncation error will be $O(h)$. It is possible to construct formulae that are accurate to higher orders (see exercise (E7), for example), but all will introduce some truncation error. The severity of the problem depends on the step size h between adjacent points of the solution. The dates at which these points are placed are often called a "net" or a "grid", and grid spacing is crucial to the numerical accuracy of finite difference solutions.

As the distance between grid points approaches zero, a finite difference approximation converges to the true solution,[41] so with enough grid points, the results can be made arbitrarily accurate. However, the size of Θ goes up with the square of the number of points, and the steps required to solve (4.23) rise even more rapidly. As a practical matter there will usually be a upper limit on the number of grid points that can be used. For this reason, the grid must be chosen carefully to attain maximum accuracy at minimum cost.

Two features of the grid play key roles in determining the accuracy of the solution: the total number of grid points used, and the location of those points in the interval $[0, T]$. To see why the sheer number of grid points is important, consider solving a model over a uniform grid of N intervals using difference formulae accurate to $O(h)$. This would require $N + 1$ grid points (the extra one is for time 0) at a distance h apart, where h is given by:

$$h = \frac{T}{N} \ . \tag{4.24}$$

The resulting grid would have points at times $\{0, h, 2h, \cdots, (N-1)h, T\}$, and at each point the solution would be subject to truncation error of order $O(h)$. In this situation, doubling N would cut h in half and reduce the error at each point by roughly a factor of two. Thus, doubling the density of the grid is a powerful tactic for reducing truncation error. Moreover, comparing solutions on grids of N and $2N$ intervals gives a good indication of the extent to which truncation error has affected the results. Also, it is possible to use Richardson's extrapolation[42] to exploit this fact to obtain even more accurate results.

---

40. In practice, the solution would never be computed using (4.23). Instead, Gaussian elimination would be applied to the system in (4.22). Elimination is much faster than matrix inversion.
41. For a proof of this, see Isaacson and Keller (1966).
42. See Birkhoff and Rota (1969).

Another way to reduce truncation error is to abandon using a uniform grid. It is often possible to make a solution much more accurate by rearranging the locations of the grid points in time. This has the advantage of keeping the size of the problem fairly small. Intuitively, the way it works is as follows. In certain periods of time, often late in the solution as the model nears the steady state, the model's dynamic variables will be changing very slowly. Moving grid points from these regions to periods where the variables are changing rapidly improves the finite difference approximation's ability to capture the model's true dynamic behavior.

In practice, shifting grid points around is an extremely powerful tool for improving the accuracy of finite difference solutions. The reason behind this stems from the Taylor series expansions used to construct the difference formulae. To see why, consider the Taylor series expansion below:

$$f(t+\varepsilon) = f(t) + f'(t)\varepsilon + \frac{f''(t)\varepsilon^2}{2!} + \cdots . \tag{4.25}$$

In forming the forward difference formula used above, the terms above first order were discarded. This introduced the following error:

$$\frac{f''(t)\varepsilon^2}{2!} + \cdots . \tag{4.26}$$

Ignoring higher-order terms, this means that on a uniform grid, truncation error would be highest where f'' was largest. Similarly, the solution would be very accurate in regions where f'' was small. Thus, shifting points from regions of low curvature to regions of high curvature would improve the solution by reducing overall truncation error.

Thus, there are two related techniques for reducing truncation error in finite difference solutions. One approach is to increase the number of grid points used, thereby reducing the distance between adjacent points and shrinking truncation error. When this is costly or inconvenient, a second tactic is to place a given number of grid points at strategic times in the solution period. One way to do this would be to transfer points from regions of low curvature to areas of high curvature. Another approach would be to move points from uninteresting parts of the solution to periods of more interest. Most of the time, of course, these two reallocations will about the same.

Overall, the finite difference method is versatile, robust to numerical instability, fairly easy to implement and very fast to solve. Furthermore, it is particularly suitable for Johansen-style general equilibrium models because it results in a system of equations which can be integrated directly into the Johansen solution procedure.[43] For these reasons, it was the method we chose to solve the intertemporal general equilibrium model presented in the next few sections.

43. We will return to this point in section (9). This approach to intertemporal general equilibrium modeling is due to Wilcoxen (1987).

5 An Intertemporal General Equilibrium Model

We now turn to the central topic of this paper: how intertemporal behavior can be included in a general equilibrium model and what benefits that produces. As in the rest of the paper, we will build the discussion around a particular model − in this case, a five sector general equilibrium model with intertemporal investment. Subsequent sections describe the structure of the model, explain how it was solved, and present a number of simulations showing how the inter- and intratemporal parts of the model interact. The simulation results demonstrate two things: that general equilibrium effects have a strong impact on investment behavior, and that changes in investment brought about by intertemporal optimization have a significant effect on general equilibrium variables. Thus, integrating intertemporal behavior into general equilibrium improves both types of model. The remainder of this section presents an overview of the model.

The finished model consists of a sequence of short run general equilibrium models linked together by an adjustment-cost investment model. All of the general equilibrium models have the same structure, but each represents the economy at a different point in time. There are five sectors of production denoted A, B, 1, 2, and 3. Sectors A, 1 and 2 produce consumption goods, sector B produces capital services, and sector 3 produces raw capital goods. There are two types of capital, K_a and K_b, and one type of labor, L. K_a is created by industry A's investment and is used solely in the production of good A. K_b is created by sector B's investment and is rented out to sectors 1, 2 and 3. Sectors 1, 2 and 3 are traditional general equilibrium industries which use a malleable capital stock (K_b) and do not do any investment. The sectors are shown schematically in figure 5.1, and their attributes are summarized in table 5.1.

Table 5.1: Characteristics of the Sectors

Sector	Invests In	Capital Used	Output Produced
A	K_a	K_a	Consumption good A
B	K_b	−	K_b capital services
1	−	K_b	Consumption good 1
2	−	K_b	Consumption good 2
3	−	K_b	Raw capital goods

Sector A is fully integrated into both the general equilibrium and investment models. It uses a capital stock that is specific to it and must solve both a short-run production and a long-run investment problem. Sector B, however, operates more like a bank. It invests to build up a stock of capital which it then rents out to other industries for use in production. Thus, sector B solves a long-run investment problem but its short-run problem is trivial − it rents out whatever it has.[44] Between them, sectors A and B account for all of the investment in the model, so all investment is the outcome of intertemporal optimization. Finally, sectors 1 and 2 are traditional zero-profit industries which rent their capital from sector B and differ only in capital intensity. Both sectors

44. By requiring that sector B always rent out all of its capital we are ruling out monopolistic behavior on its part.

produce consumption goods. Sector 3 is also a zero-profit industry which rents capital from sector B, but it makes raw capital goods. Separating it from the others facilitates experiments involving the price of raw capital goods.

The sectors were given these characteristics to emphasize that intertemporal behavior could be added gradually to an existing general equilibrium model. Suppose a particular short-run model had four sectors, all of which used a single capital good. The first step in adding intertemporal investment would be to include a fifth industry like sector B. That would allow the overall capital stock to be determined by intertemporal optimization but without requiring the structure of the existing four sectors to be changed in any way. Later, it might be useful to relax the assumption that capital is freely mobile between sectors. This could be done by asserting that one or more of the industries use industry-specific capital stocks. Each such sector would then have to solve its own investment problem, and so would end up having the characteristics of sector A. Thus, intertemporal behavior can be added to an existing model in stages; it is certainly not necessary to rebuild the model completely.

Choosing the sectors to have the characteristics shown in table 5.1 resulted in the model including two explicit intertemporal investment problems: one each for sectors A and B. These were linked to a sequence of short-run general equilibrium models by a simple but flexible model of expectations formation. This allowed simulations to be conducted under a variety of assumptions about the accuracy of expectations. In all, eleven general equilibrium models were used. The first corresponded to the present while the eleventh was 100 years in the future. The exact locations of the other equilibria in the interval $[0, 100]$ will be discussed at length in section (9). The next section presents the investment submodel in more detail.

6 The Investment Submodel

As in section (2), each firm chooses its investment path to maximize the stock market value of its equity. Assuming once more that the firm's short and long run optimizations are separable, the outcome of its short run decision can be summarized by an earnings function $E(K, P)$, where K is its capital stock and P is a vector of short run variables such as the price of the firm's output. All investment is internally financed, so dividends are short run profits less investment expenditure. Writing $C(I, P)$ for the investment cost function, where I is investment, and following the method described in (2), the firm's investment problem at time τ can be shown to be:

$$\max \int_{\tau}^{\infty} (\ E(K, P) - C(I, P)\ )(1 - T^d)e^{-rt}\ dt\ ,$$

$$\text{subject to } K' = I - \delta K\ , \qquad (6.1)$$

where r is the interest rate, δ is the rate of depreciation, and T^d is the dividend tax rate. This is precisely the problem described in section (2). Given specific earnings and investment cost functions, the principles of optimal control could be applied to generate first-order conditions for value maximization. The completed model will contain two of these investment submodels, one each for industries A and B.

6.1 Investment by Sector A in Firm-Specific Capital

To formulate firm A's investment problem, the first step is to derive its short run profit function from its production function. Production of good A is taken to be a Cobb-Douglas function of labor and capital, and the firm takes prices as given, so from the production function:

$$X_a = (L_a^P)^{\varepsilon_a}(K_a)^{1-\varepsilon_a} \ , \qquad (6.2)$$

where L_a^P is the labor used in production by industry A, the short run profit function can be shown to be:

$$E_a(K_a, P) = \frac{1-\varepsilon_a}{\varepsilon_a}\left(\frac{\varepsilon_a P_a}{W}\right)^{1/(1-\varepsilon_a)} W K_a \ , \qquad (6.3)$$

where P_a is the price of the firm's output and W is the wage rate. Notice that the firm must form expectations about both the future course of wages and the price of its product in order to be able to compute the earnings of its future capital stock.

In addition to the earnings function, the firm's investment cost function is needed in order to solve the optimization problem. An intuitive way to obtain the function is to derive it from a particular choice of the firm's investment-good production function. In this model we assume the firm produces its own investment good by purchasing raw capital (X_3^a) and hiring labor (L_a^I) to install it. The amount of labor required is proportional to the square of the amount of raw capital. This description can be summarized by the Leontief production function below:

$$I_a = \min\{\ X_3^a\ ,\ \left(L_a^I/\theta_a\right)^{\frac{1}{2}}\ \}\ , \qquad (6.4)$$

where X_3^a is raw capital, L_a^I is labor used by industry A in the construction of its investment good, and θ_a is a parameter. The corresponding cost function is:

$$C_a(X_3^a, P) = (P_3 X_3^a + W L_a^I)(1 - T^s)\ , \qquad (6.5)$$

where T^s is an investment subsidy. Minimizing investment costs given the production function above requires that the following hold:

$$I_a = X_3^a = \left(L_a^I / \theta_a \right)^{\frac{1}{2}} . \tag{6.6}$$

Solving for X_3^a and L_a^I in terms of I_a gives:

$$X_3^a = I_a \tag{6.7}$$

and

$$L_a^I = \theta_a I_a^2 . \tag{6.8}$$

Finally, this means the investment cost function can be written as shown:

$$C_a(I_a, P) = (P_3 I_a + W\theta_a I_a^2)(1 - T^s) . \tag{6.9}$$

Three important remarks must be made about this equation. First, because θ_a is not zero, the firm faces internal costs of adjustment: the cost of new capital is convex in investment. Second, adjustment costs depend on gross rather than net investment. In the steady state, gross investment will be equal to depreciation, so steady state adjustment costs depend on the size of the capital stock. This feature will be relevant for a simulation presented in section (10), but it could easily be removed by rewriting the problem in terms of net investment. Third, adjustment costs depend on investment but not on the capital stock. In the investment literature, adjustment costs are often assumed to be a function of the ratio I/K, not of I alone. Our formulation comes about as a consequence of the installation function we used in (6.4). To make adjustment costs a function of I/K, (6.4) would have to be modified.

Returning to development of the investment model, finding the path of the capital stock requires solving an optimal control problem using the short run profit and investment cost functions above. The result is a system of differential equations--the problem's first-order conditions--which must be solved to produce an explicit expression for the capital stock over time. For sector A, these first-order conditions are:

$$\lambda_a = (P_3 + 2W\theta_a I_a)(1 - T^d)(1 - T^s) , \tag{6.10}$$

$$\lambda_a{}' = (r + \delta)\lambda_a - \beta(1 - T^d) , \tag{6.11}$$

$$K'_a = I_a - \delta K_a . \tag{6.12}$$

where

$$\beta = (\frac{1-\varepsilon_a}{\varepsilon_a})\left(\frac{\varepsilon_a P_a}{W}\right)^{1/(1-\varepsilon_a)} W \ . \tag{6.13}$$

Solving for investment as a function of λ_a and the exogenous variables (see equation (2.22)) produces the following:

$$I_a = \frac{1}{2W\theta_a} \left(\frac{\lambda_a}{(1-T^d)(1-T^s)} - P_3 \right) . \tag{6.14}$$

Inserting this into the capital accumulation condition gives the equations of motion for sector A:

$$\lambda'_a = (r+\delta)\lambda_a - \beta(1-T^d) \ , \tag{6.15}$$

$$K'_a = \frac{1}{2W\theta_a} \left(\frac{\lambda_a}{(1-T^d)(1-T^s)} - P_3 \right) - \delta K_a \ . \tag{6.16}$$

Finally, we will eventually need an expression for the steady state value of λ_a. Setting λ'_a to zero in (6.15) and solving for λ_a produces the required formula:

$$\lambda_a^{ss} = \frac{\beta(1-T^{de})}{r+\delta} \ . \tag{6.17}$$

Of course, (6.17) will only hold at the steady state.

6.2 Investment by Sector B in General Purpose Capital

The other investment sector, industry B, produces capital services which it rents to other sectors. It takes prices as given, so its earnings depend only on its capital stock and the corresponding rental price:

$$E_b(K_b, P) = \rho K_b \ , \tag{6.18}$$

where ρ is the rental price of a unit of general purpose capital. The sector's investment cost function is identical in form to that of industry A, except that parameter θ_a has been replaced by θ_b:

$$C_b(I_b, P) = (P_3 I_b + W\theta_b I_b^2)(1-T^s) \ . \tag{6.19}$$

The first-order conditions for this problem are given below:

$$\lambda_b = (P_3 + 2W\theta_b I_b)(1 - T^d)(1 - T^s) \ , \tag{6.20}$$

$$\lambda_b{}' = (r + \delta)\lambda_b - \rho(1 - T^d) \ , \tag{6.21}$$

$$K'_b = I_b - \delta K_b \ . \tag{6.22}$$

For this sector, investment is given by

$$I_b = \frac{1}{2W\theta_b} \left(\frac{\lambda_b}{(1 - T^d)(1 - T^s)} - P_3 \right) . \tag{6.23}$$

Thus, sector B's equations of motion are the following:

$$\lambda'_b = (r + \delta)\lambda_b - \rho(1 - T^d) \ , \tag{6.24}$$

$$K'_b = \frac{1}{2W\theta_b} \left(\frac{\lambda_b}{(1 - T^d)(1 - T^s)} - P_3 \right) - \delta K_b \ . \tag{6.25}$$

Finally, the steady state value of λ_b can be shown to be the following:

$$\lambda_b^{ss} = \frac{\rho^e(1 - T^{de})}{r + \delta} \ . \tag{6.26}$$

7 The Short Run General Equilibrium Model

The general equilibrium model includes the two investment sectors (A,B), three "traditional" industries (1,2,3), one consumer and the government. The traditional industries rent capital from sector B and earn no short run profits. Consumption goods are produced by industry A and by traditional sectors 1 and 2. The third traditional sector, 3, produces raw capital goods used in investment. All prices in the model are those received by producers, except for that of raw capital goods which is the purchaser's price. The following subsections present the model's equations, which are also summarized in appendix (2).

7.1 Investment Sectors

As discussed above, production in sector A is a Cobb-Douglas function of labor used in production and the industry's capital stock:

$$X_a = (L_a^P)^{\varepsilon_a}(K_a)^{1-\varepsilon_a} \ . \tag{7.1}$$

Maximizing profits on existing capital implies the labor demand equation shown below:

$$L_a^P = \left(\frac{\varepsilon_a P_a}{W}\right)^{1/(1-\varepsilon_a)} K_a \ . \tag{7.2}$$

Cost minimization in production of investment goods generates the demands shown below for raw capital and investment labor:

$$X_3^a = I_a \ , \tag{7.3}$$

$$L_a^I = \theta_a I_a^2 \ . \tag{7.4}$$

Finally, revenue less wage costs in production less investment costs gives pre-tax dividends:

$$D_a = P_a X_a - W L_a^P - \left(P_3 X_3^a + W L_a^I\right)(1 - T^s) \ . \tag{7.5}$$

Industry B produces only capital services, so its behavior is entirely determined by the optimal path of its capital stock. Deriving the demands for raw capital and investment labor produces the equations shown below:

$$X_3^b = I_b \ , \tag{7.6}$$

$$L_b^I = \theta_b I_b^2 \ . \tag{7.7}$$

Gross dividends are simply revenue less investment costs:

$$D_b = \rho K_b - (P_3 X_3^b + W L_b^I)(1 - T^s) \ . \tag{7.8}$$

The equations above fully describe the short run behavior of the special investment sectors in the model.

7.2 Other Production

Three other sectors are included in the model: industries 1, 2 and 3. These sectors rent their capital from industry B at price ρ and do not invest. Production in each sector is Cobb-Douglas, as shown below, where $i \in \{1, 2, 3\}$:

$$X_i = \gamma_i (L_i)^{\varepsilon_i} (K_b^i)^{1-\varepsilon_i} \ . \tag{7.9}$$

Straightforward optimization generates the factor demand equations shown below:

$$L_i = \frac{1}{\gamma_i} X_i \left(\frac{\rho \varepsilon_i}{W(1-\varepsilon_i)} \right)^{1-\varepsilon_i} , \tag{7.10}$$

$$K_b^i = \frac{1}{\gamma_i} X_i \left(\frac{W(1-\varepsilon_i)}{\rho \varepsilon_i} \right)^{\varepsilon_i} . \tag{7.11}$$

Finally, each sector is constrained to earn zero pure profits. For industries 1 and 2 this condition is:

$$X_i P_i = W L_i + \rho K_b^i \ . \tag{7.12}$$

Because the price of raw capital goods is the purchaser's cost, the zero pure profit condition for industry 3 is slightly different, as shown below:

$$X_3 P_3 = (1 + T_s^3)(W L_3 + \rho K_b^3) \ , \tag{7.13}$$

where T_s^3 is the sales tax on capital goods.

7.3 The Consumer

The single consumer in the model supplies labor and owns both investment firms, so income includes wages, dividends and lump sum payments from the government. There is no explicit saving, so all income in each period is spent on consumption. (Implicitly, however, the consumer saves whatever earnings the firms retain for new investment.) Thus, the consumer's budget constraint is the following:

$$C = WL(1 - T^w) + (D_a + D_b)(1 - T^d) + LS \ , \tag{7.14}$$

where C is consumption expenditure, T^w is the tax on wages, and LS is a lump sum payment from the government. Utility is a Cobb-Douglas function of the consumption of goods A, 1 and 2, and labor is supplied inelastically, so the demand system below can be derived from utility maximization subject to the budget constraint in equation (7.14):

$$X_a^C P_a (1 + T_s^a) = \alpha_C^a C \ , \tag{7.15}$$

$$X_1^C P_1(1 + T_s^1) = \alpha_C^1 C \ , \tag{7.16}$$

$$X_2^C P_2(1 + T_s^2) = \alpha_C^2 C \ . \tag{7.17}$$

where the α's are Cobb-Douglas exponents, and T_s^a, T_s^1, and T_s^2 are sales taxes on goods A, 1 and 2 respectively.

An important consequence of formulating the consumer problem this way is that the supply of savings is perfectly elastic: consumers save whatever firms want to borrow at the prevailing interest rate. Many applied general equilibrium models use the opposite assumption, that capital accumulation is driven entirely by savings behavior. The truth is likely to be somewhere in between. One way to give the model more realistic savings behavior would be to formulate the consumer problem as an explicit intertemporal optimization; for example, consumers could be modeled as life-cycle savers. All methods presented in this paper apply as easily to consumption as they did to investment. In particular, the finite difference method works in exactly the same way. Thus, a more sophisticated model could include intertemporal optimization by both consumers and firms with little additional difficulty. For the purposes of this paper, however, we will confine our model to intertemporal investment.

7.4 The Government

The government is constrained to balance its budget, so spending is equal to tax revenue less lump sum payments and subsidies. Revenue is raised by dividend taxes, wage taxes and sales taxes, while lump sum payments are made to the consumer and subsidies are paid on investment expenditure. Thus, the government's budget is given by the following equation:

$$\begin{aligned} G = T^d(D_a + D_b) - T^s\Big(P_3(X_3^a + X_3^b) + W(\theta_a I_a^2 + \theta_b I_b^2)\Big) \\ + T_s^a P_a X_a + T_s^1 P_1 X_1 + T_s^2 X_2 P_2 + T_s^3 P_3 X_3 \\ + T^w WL - LS \ . \end{aligned} \tag{7.18}$$

The government demand system is derived from a Cobb-Douglas utility function and consists of the following equations:

$$X_a^G P_a(1 + T_s^a) = \alpha_G^a G \ , \tag{7.19}$$

$$X_1^G P_1(1 + T_s^1) = \alpha_G^1 G \ , \tag{7.20}$$

$$X_2^G P_2(1 + T_s^2) = \alpha_G^2 G \ . \tag{7.21}$$

7.5 Market Clearing

The final group of equations necessary to define the model is the set of market clearing conditions. For goods A, 1 and 2, total demand is the sum of private and government demand. Demand for good 3 is the sum of raw capital demand by the two investment sectors. The four equations are:

$$X_a = X_a^C + X_a^G \ , \tag{7.22}$$

$$X_1 = X_1^C + X_1^G \ , \tag{7.23}$$

$$X_2 = X_2^C + X_2^G \ , \tag{7.24}$$

$$X_3 = X_3^a + X_3^b \ . \tag{7.25}$$

In addition, factor market clearing for labor and capital B requires the following:

$$L = L_a^P + L_a^I + L_b + L_1 + L_2 + L_3 \ , \tag{7.26}$$

$$K_b = K_b^1 + K_b^2 + K_b^3 \ . \tag{7.27}$$

7.6 Other Equations

In addition to all of the equations above, a price deflator was also incorporated into the model. The index, ζ, was defined as the cost of the current bundle of consumption and government goods at current prices divided by its cost at period zero's initial prices:

$$\zeta = \frac{X_a P_a (1 + T_s^a) + X_1 P_1 (1 + T_s^1) + X_2 P_2 (1 + T_s^2)}{X_a [P_a (1 + T_s^a)]_b + X_1 [P_1 (1 + T_s^1)]_b + X_2 [P_2 (1 + T_s^2)]_b} \ , \tag{7.28}$$

where the variables in parentheses subscripted by "b" are base case values.

8 Expectations

The two investment models described in section (6) depend on a number of future variables that the firms take as given. Strictly speaking, what appears in the optimizations are firms' *expectations* of those variables. This means that we must make an assumption about how expectations are formed in order to be able to link the investment and general equilibrium models. One possibility is to assume that the expectations are "rational", by which we mean that in the absence of any unforeseen shocks, firms can predict the course of the economy perfectly. To implement rational expectations the variables needed in the investment model could be taken directly from their counterparts in the short-run general equilibrium models. The price of capital appearing in an investment model, for example, would be exactly equal to the price generated by the general equilibrium model for the appropriate date. Solving the complete model simultaneously would yield a path of wages and prices consistent with firms' planned capital stocks, and also a capital accumulation plan consistent with wages and prices. Thus, one possible assumption about expectations is that they are rational, which is straightforward to implement.

On intuitive grounds, rational expectations might seem implausible − it appears to require excessively sophisticated behavior on the part of agents. However, it has one compelling characteristic: it is the only expectations mechanism that is not dominated from an agent's point of view; forming expectations any other way means an agent would systematically be wrong about the future. As long as there are no costs to forming rational expectations, it will always be in an agent's interest to do so. For this reason, we will adopt the rational expectations assumption for many of the simulations described in this paper. There is nothing about the model, however, that makes this necessary; any expectations mechanism could have been used.

In fact, the actual mechanism in the model contains provisions for introducing particular departures from complete rationality. The investment problems outlined in section (6) depend on expectations about ρ, W, P_a, P_3, T^d and T^s. For each of these, an expectation was formed by combining its true general equilibrium value with an exogenous component. For example, the expected wage in the investment submodel, W^e, was formed out of the true general equilibrium wage, W, and a fixed expectation W^x, as shown below:

$$W^e = (W)^{\lambda_n} (W^x)^{1-\lambda_n} \ , \tag{8.1}$$

where λ_n was a parameter ranging from zero to one. When $\lambda_n = 1$, firms have perfect foresight; when $\lambda_n = 0$, the expected wage is set to the exogenous value W^x. This procedure was also carried out for expectations of the exogenous variables T^d and T^s, but a separate parameter, λ_x, was used.

Parameters λ_n and λ_x allow simulations to be run under different assumptions about the extent to which firms can predict future variables. When both λ_n and λ_x are set to 1, firms have perfect foresight. On the other hand, if $\lambda_n = 0$ and $\lambda_x = 1$, firms understand what tax changes are planned for the future, but they are unable to correctly predict the general equilibrium consequences. Setting λ_n and λ_x both to zero converts the model to a set of linked static equilibria in which investment is not affected by anything. The remaining case, $\lambda_n = 1$ and $\lambda_x = 0$ is of little interest.

9 Implementing the Model

We have now completed the economic specification of our small intertemporal general equilibrium model. Before it can be used to analyze experiments, however, it must be implemented on a computer. In this section, we describe one way that could be done. The method we present is straightforward and produces a versatile model that solves quickly. It is by no means the only way the model could have been implemented, however. Thus, it is important to distinguish between the model, described in sections (5) to (8), and the solution method, which is described below.

There are four tasks to accomplish in implementing a model: selecting the solution algorithm, constructing the data set, partitioning the variables into endogenous and exogenous sets, and testing the final program. The next four sections describe each of these steps in detail.

9.1 The Solution Algorithm

Solving the two investment problems requires solving a system of differential equations, while solving the short-run general equilibrium model requires solving a large system of nonlinear equations. Both components of the model must be solved simultaneously, so the entire process is not trivial. This section will set out the basic approach used and discuss how close the numerical solution will be to the true solution.

In order to solve the investment problems of the firms in sectors A and B, the equations of motion for the two problems were converted to their finite difference equivalents. For sector B this produced the following expressions:

$$\frac{\lambda_b(t+h) - \lambda_b(t)}{h} = (r+\delta)\lambda_b(t) - \rho^e(t)[1-(T^d(t))^e] \,, \tag{9.1}$$

$$\frac{K_b(t+h) - K_b(t)}{h} = \frac{1}{2W^e(t)\theta_b} \left(\lambda_b^*(t) - P_3^e(t) \right) - \delta K_a(t) \,, \tag{9.2}$$

where $\lambda_b^*(t)$ has been introduced solely for notational convenience, and is given by the expression:

$$\lambda_b^*(t) = \frac{\lambda_b(t)}{(1-(T^d(t))^e)(1-(T^s(t))^e)} \,. \tag{9.3}$$

Equations (9.1) and (9.2) were obtained by inserting the appropriate terms from (6.24) and (6.25) into the general expressions (4.17) and (4.18). The results for sector A are very similar. Notice that we have been careful to write all the variables as functions of time, and to mark all of the expected variables with a superscript "e". This will be essential when we link the investment models to the short-run general equilibrium system.

Equations (9.1) and (9.2) are linear in K and λ_b, so if the time paths of the other variables were known, the complete system of finite difference equations could be solved easily using Gaussian elimination. Thus, if we only wanted a partial-equilibrium investment model we could stop here. In a general equilibrium analysis, however, many of those variables are endogenous and are not known prior to solving the investment problem. To impose rational expectations, for example, the expected endogenous variables at all future times must be consistent with the short-run general equilibrium model.

In practice, expected values are needed for each grid point in the finite difference approximation. If there are N grid points, there will be N systems like (9.1) and (9.2), each holding at a different point in time. Every system will require expected values for the endogenous variables, so the short-run model will have to be solved at each point on the grid. To accomplish this, we converted the short-run model to its percentage change form[45] and then required that the resulting system hold at each grid point. This produced a set of N identical linear short-run models, each holding at a different point in time. In addition, we required that linearized versions of the steady state formulae for λ_a and λ_b given in equations (6.17) and (6.26) hold in the final period. Linearizing the short-run model made it necessary to linearize the investment model as well, but that had an extremely useful consequence: the set of short-run models and the investment model had both become systems of linear equations. This made it possible to obtain an intertemporal solution simply by applying Gaussian elimination to a very large system of equations.

The complete model thus entails Taylor series expansions in both time (finite differences) and variables (linearization). This means the solution is only approximate, so steps must be taken to ensure that truncation error is kept adequately small. However, the results may be made arbitrarily accurate by decreasing the step size used in each expansion, so at least in principle this is not an insurmountable problem. We will return to this topic in section (9.4).

For the simulations described below, we implemented the model with the following features. The terminal time − the point at which the steady state value of λ was imposed − was chosen to be 100 years in the future. This made the overall solution period of the model $[0, 100]$. Eleven grid points were used in the finite difference approximation: one at year 0, one at year 100, and nine scattered in between. To obtain adequate numerical accuracy, a nonuniform grid was used. Thus, the grid points were not necessarily located at multiples of ten years. The subject of grid spacing will be discussed further in section (9.4).

9.2 The Data Set

The next step in implementing the model was to construct a data set to be used as the base case. For a static model a base case can usually be assembled from a single input-output table and a handful of parameters. Intertemporal models, however, require much more data. In principle, an intertemporal base case consists of an entire string of equilibria stretching far into the future. This presents a formidable problem because the future equilibria cannot be observed.[46] Since the base case cannot be built from observable data, it must be constructed by postulating future paths for the exogenous variables and finding a corresponding solution to the model. The difficulty of this depends on what behavior is required of the base case exogenous variables.

45. This is an application of Johansen's method. For further details on Johansen linearization, refer to chapter 3 of Dixon, Parmenter, Powell and Wilcoxen.

46. For retrospective (counterfactual) simulations, it may be possible to observe a number of the initial equilibria, but there will always be some that are unobservable.

The easiest base case to construct, and the one that has dominated intertemporal modeling to date, is a steady state. In this approach, the base case values of future exogenous variables are set to particular constants. Then, the data set for the first year of the base case (which is usually obtained from historical data) is adjusted so that the model will replicate itself from year to year as long as the exogenous variables remain at their original values. The result is a base case which consists of an arbitrary string of future periods which are identical to the initial year. This kind of scenario is fairly easy to construct because it only requires obtaining a steady state solution to the model.

For many experiments, a steady state base case is perfectly acceptable. Often the most important question about a particular shock is how far it pushes the economy away from the base case path, not how the base case itself is evolving. In this situation, starting from a steady state is a minor liability which is more than offset by the ease with which the base case can be constructed. For this reason, we have chosen to use a steady state base case for the simulations presented below. Methods for building other kinds of base case are beyond the scope of this book, but are discussed in Wilcoxen (1988) and Codsi, Pearson and Wilcoxen (1990).

The actual data set we used is presented in appendix A1. It has a number of interesting features, but does not represent any particular economy. One of its most important characteristics is that the patterns of private consumption and government spending are identical, so no compositional effects arise when changes in taxes induce transfers of income between the private and government sectors. We built this feature into the data set deliberately so it would be easier to interpret the model's results.

9.3 Partitioning

Once the solution algorithm and the data set have been prepared, the next step is to decide on a partitioning of the model's variables into endogenous and exogenous sets. One of the advantages of using the Johansen approach is that it is easy to change the partition for individual experiments, so the partition established at this stage does not constrain future simulations at all. For other solution algorithms, however, the partitioning done at this stage will be permanent; it will be impossible to switch the endogenous and exogenous variables later. Table 9.1 shows the basic list of exogenous variables used in the experiments discussed below.

Since the price deflator is exogenous, it will be the model's numeraire. The role of the numeraire in an intertemporal model differs quite a bit from its role in a static setting. In particular, the numeraire may have to establish the rate of pure inflation in addition to setting the overall price level. In the model above, for example, there is no equation describing how the price level evolves from one year to the next. This means that the rate of pure inflation will be determined by the path of the numeraire over time. Thus, if the numeraire were constant (as it is in the base case data set), there would be no pure inflation − the price of the aggregate consumption bundle would be

Table 9.1: Exogenous Variables

Symbol	Description
K_a	Sector A capital (period 0 only)
K_b	Sector B capital (period 0 only)
L	Total labor supply
G	Government spending
T_w	Tax on wages
T_s^a	Sales tax on good A
T_s^1	Sales tax on good 1
T_s^2	Sales tax on good 2
T_s^3	Sales tax on good 3
T^d	Dividend tax
T^s	Investment subsidy
γ_1	Technical change parameter, industry 1
γ_2	Technical change parameter, industry 2
γ_3	Technical change parameter, industry 3
ρ^x	Exogenous expectations, rental price
W^x	Exogenous expectations, wage rate
P_3^x	Exogenous expectations, raw capital price
P_a^x	Exogenous expectations, price of good A
T^{d^x}	Exogenous expectations, dividend tax rate
T^{s^x}	Exogenous expectations, investment subsidy
r	Interest rate
ζ	Price deflator

constant from one year to the next. On the other hand, if the base case numeraire rose by 5 percent a year, the model would embody a 5 percent rate of pure inflation. Thus, it is possible to think of the numeraire as having two distinct roles: determining the level of prices in the first period, and selecting the rate of growth of the price level over time. A more detailed model might include a money demand equation and an exogenous supply of money, in which case the price level and the rate of pure inflation would be determined by the (exogenous) money supply. In the absence of an explicit model of the money market, the numeraire performs the same function.

A second consequence of this choice of partition is that since government spending is exogenous and the lump sum payment is endogenous, any revenue accruing from changes in tax rates will be passed back to households. This is convenient for the simulations we discuss below, but there are other alternatives. In some applications, for example, it may be more useful to make the lump sum tax exogenous and government spending endogenous. As we mentioned earlier, however, one of the advantages of the Johansen method is that the partition can be changed at any time.

A final important feature of the partition is that the interest rate is exogenous. This results from the structure of the consumer problem described in section (7.3). Since the consumer always saves exactly what firms want to invest, the supply of savings is perfectly elastic with respect to the interest rate. Thus, we have little choice but to make the interest rate exogenous. If we wanted the financial market to be more realistic, we could introduce an upward-sloping savings supply curve by changing the consumer model to include intertemporal optimization. In this paper, however, we limit intertemporal modeling to investment in order to keep the exposition as clear as possible.

9.4 Testing the Complete Model

Once the two finite difference investment modules were combined with the eleven general equilibrium models, and the entire system was linearized, a number of special experiments were run to check that the model was programmed correctly. This is an essential step if the numerical results of the model are to be believed. Three simulations were used: a homogeneity test, a surprise dividend tax at time zero, and an increase in the tax on wages.

The homogeneity test consists of a simultaneous increase in the price deflator and the nominal lump sum payment. The effect of this should be to raise all nominal variables by the amount of the increase, and to leave all real variables unchanged. The model produced this result correctly which indicated that it was free of gross programming errors. The second test was slightly more interesting. As noted in section (3), increasing the dividend tax at time zero is an unavoidable pure profits tax. As such, it should have no effect on the capital stock or output of any industry, although dividends and firm values should fall by the amount of the tax. There should also be a large shift of income from consumers to the government, but since both sectors have the same patterns of demand, there will be no compositional effects. These results were also correctly generated by the model. The final test was an increase in the wage tax paid by consumers. Since labor supply is completely inelastic, the effect of a wage tax should be a simple transfer of income to the government. Again the model produced the correct results. Together these experiments provide strong (albeit indirect) evidence that the model's implementation was free of programming errors. Having verified this, it was necessary to check that the linearized model would correctly converge on less trivial experiments.

In a purely mathematical sense, the model is a system of partial differential equations in time and variables which are solved by integrating over time using finite differences and over variables using Euler's method.[47] The accuracy of the solution depends on the step size used in these integrations: the solution will approach the true solution as the step size in both time and variables is made infinitesimal.[48] To verify that the model was formulated correctly, it was necessary to show that the numerical solution could be made arbitrarily close to the true analytical solution for a particular experiment. Incidently, this shows how accurate results can be obtained with a fairly small number of steps in each dimension.

As discussed in section (4), for most experiments it is impossible to obtain an analytical solution to the investment problem of either firm. One exception, however, is an announced increase in the dividend tax. Under partial equilibrium (no feedback from the firm's decision to the variables it takes as given), the analytical solution to a dividend tax experiment can be obtained in a straightforward manner.[49] By comparing the results of numerical simulations to the analytical solution, the accuracy of the former can be assessed.

The actual experiment we used was an increase in the dividend tax rate from 10 percent to 20 percent, to take effect ten years in the future. The expectation parameters discussed in section (8) were set to $\lambda_n = 0$ and $\lambda_x = 1$, so firms ignored any feedback effects from their actions to the price variables in their investment decisions. The results of the experiment were discussed in detail in section (3), where the analysis went roughly as follows. The announcement of the tax causes firms to pay large dividends immediately before the tax takes effect. This drives down investment, so when the tax is implemented the capital stock will be lower than it would have been. Once implemented, however, the tax falls on pure profits, so firms return to their pre-announcement behavior and the economy gradually returns to the original steady state capital stock, although owners of capital have suffered a windfall loss.

To assess the accuracy of different numerical solutions, we examined how well they captured the true value of the capital stock in year ten. This is a good measure of the overall accuracy of a solution because the true path has a cusp at that point. Recall from section (4) that truncation error will be large in regions where the high-order derivatives dropped from the difference formulae are large. At a cusp, the first derivative changes discontinuously and the second derivative goes through infinity. In a numerical simulation, this will manifest itself as rounding of the solution near where the cusp should be.

The results of several experiments are shown in table 9.2.[50] The rows indicate how many

47. See chapter 3 of Dixon, Powell, Parmenter and Wilcoxen for an explanation of how Johansen linearization is related to Euler's method.

48. As shown in chapter 3 of Dixon, Parmenter, Powell and Wilcoxen, truncation errors in the Johansen linearization can be made arbitrarily small by applying the shock in a series of small steps.

49. This is discussed in detail in one of the exercises.

50. The values in this and subsequent tables were computed using slightly different difference formulae from those shown in section

steps were used to impose the exogenous shock, while the columns show how many grid intervals were used in the finite difference approximation. Each entry gives the value of $K(10)$ obtained with a particular combination of iterations and grid points. Step sizes decrease downward and to the right, so the solution should become more accurate in those directions. The bottom row was obtained by solving the investment model without linearizing. Since that is equivalent to using an infinitesimal step size, the row is labeled "infinite" iterations.

Table 9.2: The Effect of Grid Density and Iterations
on the Value of the Capital Stock at Period 10*

	Number of Grid Intervals			
Iteration	10	20	40	80
1	.9623	.9376	.9247	.9180
2	.9605	.9361	.9236	.9174
4	.9596	.9353	.9233	.9171
8	.9591	.9350	.9229	.9169
∞	.9586	.9346	.9227	.9168

* The true value of $K(10)$ is .9113

Several things are readily apparent. First, increasing either grid density or iterations improves the solution. Second, because first-order Taylor expansions form the basis of the linearizations in each dimension, the difference between values obtained from successive halvings of the step size decreases by roughly a factor of two. Third, from the initial 1-iteration, 10-grid-interval solution, accuracy increases most rapidly by increasing the number of finite difference grid points. This indicates that the error introduced by the finite difference approximation overwhelms that of the Johansen linearization. In fact, doubling the grid density improves the solution by more than increasing the number of Johansen iterations to infinity. This shows that Johansen linearization error is trivial, but that introduced by finite differences may not be.

As noted in section (4), for large models it may not be feasible to eliminate finite difference truncation error by using a vast number of grid points. For such models, however, it is possible to improve the approximation by shifting grid points from regions of low curvature to regions of high curvature. In practice this involves moving points from late years, say year 80 or 90, to times nearer implementation of the policy. Table 9.3 shows five possible allocations of nine points to times between 0 and 100 years.

For an announced tax change which is to be implemented in year 10, the system will almost be back to the steady state at late years like 90. High-order derivatives of the model's variables will be close to zero there, so it might be desirable to move such points to an earlier time where

(9). Implementing the model as it is described in this paper would produce slightly different, but qualitatively similar, results.

the derivatives are large. One possibility would be to shift the point from year 90 to year 5; this is shown in grid B above. Continuing the rearrangement produces the set of grids shown in the table. The dividend tax was then simulated over each grid, producing the results shown in table 9.4. Again, results are shown for several iterations over the shock.

Table 9.3: A Selection of Grid Spacings

	Grid				
Point	A	B	C	D	G
0	0	0	0	0	0
1	10	5	5	5	5
2	20	10	7	7	7
3	30	20	10	9	9
4	40	30	20	10	10
5	50	40	30	20	15
6	60	50	40	30	20
7	70	60	50	40	35
8	80	70	60	50	50
9	90	80	70	60	75
10	100	100	100	100	100

Table 9.4: The Effect of Grid Choice and Iterations on the Value of the Capital Stock at Period 10*

	Grid				
Iterations	A	B	C	D	G
1	.9623	.9407	.9323	.9233	.9199
2	.9605	.9392	.9310	.9226	.9192
4	.9596	.9384	.9303	.9222	.9187
8	.9591	.9381	.9300	.9219	.9185
∞	.9586	.9377	.9297	.9217	.9183

* The true value of $K(10)$ is .9113

Table 9.4 demonstrates two important properties. First, using a non-uniform grid does not harm convergence when the number of iterations is increased. This can be seen by reading down the columns. Second, rearranging a limited number of grid points can produce a solution almost as accurate as increasing the density of a uniform grid by a factor of eight. This means that using a limited number of grid points does not necessarily produce an unreasonable amount of finite difference truncation error.

At this point, we have verified that our computer implementation of the model behaves correctly for a variety of experiments. It has produced correct results for the homogeneity test, the surprise dividend tax test and the wage tax test. In addition, we have subjected it to a more difficult test by computing numerical solutions to an experiment whose analytical results are known. Since it has passed all of these tests, it can now be used to run simulations.

10 Some Illustrative Simulations

We had two motives for presenting the model above. The first was to demonstrate how an intertemporal general equilibrium model can be built; that task has now been accomplished. The second goal was to show that building such models is worthwhile; it is to that topic that we will now turn. The evidence we present will be a number of simulations whose results could not have been obtained from either a short-run general equilibrium model or a partial-equilibrium investment model. Of course, there are many experiments for which one or the other of those techniques is perfectly adequate. We will not argue that an intertemporal general equilibrium model is always necessary; just that it is a very useful approach for certain problems.

10.1 The Importance of Foresight

Having gone to all the trouble of integrating the two kinds of model, one question we might ask is whether we have improved the original investment model. Does it matter that general equilibrium linkages have been added, or was the partial-equilibrium model just as good? The answer depends on what assumption we want to make about agents' expectations. If we are content to give them beliefs about the future that are exogenous and completely fixed, then nothing has been gained by moving to an integrated model. On the other hand, if we want agents to have rational expectations, adding the short-run general equilibrium models makes that possible. This suggests an important question: does it matter whether expectations are rational?

Using our integrated model, it is possible to test precisely that point. The expectations parameters, λ_n and λ_x, allow us to run simulations under either perfect foresight or fixed expectations. Setting both parameters to one produces perfect foresight; setting either of them to zero produces some form of fixed expectations. One case of fixed expectations is somewhat appealing: perhaps agents have perfect foresight with respect to taxes but have fixed beliefs about prices. To be concrete, a firm might know the future path of the dividend tax accurately from government proclamations, but be completely unaware that the tax will end up changing the wages and prices it faces. This form of expectations can be simulated by setting λ_x to one and λ_n to zero. To see whether this differs significantly from perfect foresight, we can run a typical experiment under the two assumptions about λ_n and λ_x.

The experiment we chose was an announced change in the dividend tax from 10 percent to 20 percent to be implemented in period ten. In this experiment, and all of the others described below, ten finite difference intervals were used with grid points placed according to column G in table 9.3. The dividend tax experiment was chosen because it has no permanent effect on prices and wages. Thus, even the agents without perfect foresight will not be wrong forever.

The results of these two simulations are shown in figure 10.1. Each graph gives the percentage change in a particular variable from its base case value in the corresponding year. Different panels show capital stocks A and B, investment A and B, the wage, the rental price of capital B, the price of good A, dividends paid by sectors A and B and consumption. Consumption is unchanged in the long run because the government returns any extra revenue through lump-sum payments. The paths marked "P" are for perfect foresight; those marked "F" are for fixed expectations.

The striking feature of figure 10.1 is that perfect foresight with respect to prices and wages attenuates the response of the model by about 50 percent. This affect even occurs in period zero: investment falls less than half as much under perfect foresight. The reason the simulations are so different is that as the capital stock declines, the price of each firm’s output rises in response. This keeps returns higher in the period before implementation, so the firms with foresight will not let their capital stock deteriorate as fast as those expecting the price to be unchanged.

One other important feature of the results is that many of the variables change substantially in the period before the tax is implemented. This demonstrates the second part of our assertion about the usefulness of building intertemporal models: intertemporal optimization by investors can have a substantial effect on the economy even before an anticipated event occurs. In a static model, expectations of future policy changes can never affect current variables.

From this simulation we conclude that adding general equilibrium effects has changed the investment model substantially. To the extent that agents have rational expectations, a partial equilibrium investment model will overstate the response of investment and the capital stock to any given shock. Moreover, the inaccuracy can be as much as 50 percent. Finally, both sets of results also show that intertemporal effects can have significant consequences for ordinary general equilibrium variables.

10.2 Indirect Dynamic Effects

Another advantage of the integrated model is that it can be used to study the dynamic effects of policies which only influence the investment problem indirectly. One example of this would be a change in sales taxes. Sales taxes do not appear in either firm’s investment problem because the firms are only interested in the price they receive, not in what purchasers actually pay. In general equilibrium, of course, changing sales taxes would usually change the prices faced by a producer. Thus, even though sales taxes do not enter the investment model explicitly, they can still affect it by changing the prices that do appear in the problem. Using the integrated model allows these effects to be captured. This section presents two sales tax simulations, each of which highlights a different characteristic of the model.

Figure 10.2 shows the effect of an announced increase in the sales tax on good A to be implemented at year ten. The general equilibrium effect of the tax is to raise purchaser prices and lower producer prices. This makes investment in sector A less attractive, so its capital stock begins to decline. The path of sector A's investment and capital stock is very similar to that displayed in exercise (E1), and comes about for the same reason: there is a decline in the producer price of good A. The shock has a modest effect on industry A, reducing its long run capital stock by about 6 percent.

The decline in sector A has effects throughout the economy. It frees up labor that used to be employed in sector A, causing a drop in the wage rate. Lower wages benefit sector B by lowering its investment costs, so I_b increases and K_b rises over time. This drives down the rental price of capital B, so producer prices of goods 1, 2 and 3 must fall since the wage also fell. The producer price of good A also falls, although the purchaser's price has risen because of the tax.

The behavior of sector A's dividend stream is striking: before the tax is implemented, D_a actually increases. As discussed in exercise (E1), this comes about because investment drops in anticipation of the tax, leaving more earnings to be distributed as dividends. Once the tax is in place, however, fewer dividends can be paid so D_a drops below its base case value. The present value of the change in the dividend stream is negative, so the tax results in a windfall loss to owners of the firm.

To summarize this experiment, an anticipated change in a sales tax can produce interesting intertemporal effects that could not be captured by either a partial-equilibrium investment model or a static general equilibrium model. This point is further emphasized by a second sales tax experiment which introduces a shock that is even further removed from the investment models: an increase in the tax on good 2. Sector 2 does not do any investment itself, so the effect on investment of increasing the sales tax on good 2 will only occur through indirect general equilibrium linkages.

The results of simulating an increase in T_s^2 are shown in figure 10.3. The main difference between this and the previous experiment is that the capital stock in both sectors A and B rises over time. The source of this curious result is that sector 2 is very large and very labor intensive. A small contraction in its output leads to a substantial drop in the wage. This lowers investment costs for both sectors A and B, so investment rises and both capital stocks grow. Finally, growth of the capital stocks causes the rental price of capital B and the price of good A to fall. In the end, sector A has gained, producing slightly higher dividends, while sector B's dividends have fallen considerably.

Overall, the three simulations presented in this section indicate the wide variety of experiments that can be analyzed using an intertemporal general equilibrium model which could not be studied in a static or partial equilibrium context. For many policy questions, this will easily justify the extra work required to build an intertemporal model.

11 Goals, Reading Guide and References

In this paper we have tried to illustrate the basic techniques used in intertemporal modeling and to show how they can be applied in building an intertemporal general equilibrium model. We believe that by reading it and working through the exercises, you will develop the skills needed to understand the intertemporal models you come across in the literature and to be able to build your own. In particular, we hope you will

(1) be able to discuss what circumstances call for the use of an intertemporal model;

(2) understand how to build theoretical models of intertemporal decisions from basic principles such as arbitrage, and to understand q-theoretic investment models in particular;

(3) be able to use graphical techniques such as phase planes to describe in qualitative terms how an intertemporal model will respond to a shock;

(4) be familiar with numerical methods that can be used when it is necessary to obtain quantitative results from an intertemporal model;

(5) know how to integrate intertemporal decisions into general equilibrium models, and be able to discuss what costs and benefits that entails.

Intertemporal modeling uses a number of mathematical methods that you may not have encountered before. The reading guide for this paper is intended to help you fill in gaps in your knowledge of optimal control, differential equations, numerical methods and linear algebra. It also includes a number of references to particularly important or useful parts of the economic literature on intertemporal analysis.

Reading Guide, part 1

Reading Guide, part 2

References

Birkhoff, Garrett, and Gian-Carlo Rota (1978), *Ordinary Differential Equations*, New York: John Wiley and Sons.

Blanchard, Olivier Jean, and Stanley Fischer (1989), *Lectures on Macroeconomics*, Cambridge: MIT Press.

Bryson, Arthur E. Jr, and Yu-Chi Ho (1975), *Applied Optimal Control: Optimization, Estimation and Control*, Washington: Hemisphere.

Codsi, George, Ken Pearson and Peter J. Wilcoxen (1990), "Dynamic Models in GEMPACK," Impact Draft Preliminary Working Paper, University of Melbourne.

Dixon, Peter B., B.R. Parmenter, Alan A. Powell and Peter J. Wilcoxen (1992), *Notes and Problems in Applied General Equilibrium Economics*, Amsterdam: North-Holland.

Dixon, Peter B., B.R. Parmenter, John Sutton and D.P. Vincent (1982), *ORANI: A Multisectoral Model of the Australian Economy*, Amsterdam: North-Holland.

Eisner, R., and R.H. Strotz (1963), "Determinants of Business Investment," in *Impacts of Monetary Policy*, Englewood Cliffs, New Jersey: Commission on Money and Credit, pp. 59-233.

Fair, Ray C. (1979), "An Analysis of a Macro-Econometric Model with Rational Expectations in the Bond and Stock Markets," *American Economic Review*, 69(4): 539-552.

Fair, Ray C., and John B. Taylor (1983), "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models," *Econometrica*, 51(4): 1169-1185.

Fox, L. (1962), *Numerical Solution of Ordinary and Partial Differential Equations*, London: Addison-Wesley.

Gould, J.P. (1968), "Adjustment Costs in the Theory of Investment of the Firm," *Review of Economic Studies*, 35(1): 47-55.

Hayashi, F. (1982), "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50(1): 213-224.

Isaacson, E., and H.B. Keller (1966), *Analysis of Numerical Methods*, New York: John Wiley and Sons.

Kamien, Morton I., and Nancy L. Schwartz (1981), *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, Amsterdam: North-Holland.

Leitmann, G. (1981), *The Calculus of Variations and Optimal Control: An Introduction*, New York: Plenum Press.

Lipton, D., J. Poterba, J. Sachs and L. Summers (1982), "Multiple Shooting in Rational Expectations Models," *Econometrica*, 50(2): 1329-1333.

Lucas, R.E. (1967): "Optimal Investment Policy and the Flexible Accelerator," *International Economic Review*, 8(1): 78-85.

McLaren, Keith (1989), "A Reappraisal of the Neoclassical Approach to Modelling Business Investment," mimeo.

Modigliani, Franco and Merton Miller (1958), "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, 76(3): 297-313.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, (1986), *Numerical Recipes*, New York: Cambridge University Press.

Roberts, S.M. and J.S. Shipman (1972), *Two-Point Boundary Value Problems: Shooting Methods*, New York: American Elsevier.

Strang, Gilbert (1980), *Linear Algebra and Its Applications*, New York: Academic Press.

Summers, L. H. (1981), "Taxation and Corporate Investment: A q-Theory Approach," *Brookings Papers on Economic Activity*, no. 1, pp. 67-127.

Tobin, J. (1969), "A General Equilibrium Approach to Monetary Theory," *Journal of Money, Credit and Banking*, 1(1): 15-29.

Treadway, A. (1969), "On Rational Entrepreneurial Behavior and the Demand for Investment," *Review of Economic Studies*, 3(2): 227-39.

Wilcoxen, P.J. (1985a), "Numerical Methods for Investment Models with Foresight," IMPACT Preliminary Working Paper No. IP-23, University of Melbourne.

Wilcoxen, P.J. (1985b), "Computable Models of Investment with Foresight," University of Melbourne, Department of Economics, Working Paper No. 138.

Wilcoxen, P.J. (1987), "Investment With Foresight in General Equilibrium," IMPACT Preliminary Working Paper No. IP-35, University of Melbourne.

Wilcoxen, P.J. (1988), "The Effects of Environmental Regulation and Energy Prices on U.S. Economic Performance," Ph.D. Thesis, Harvard University.

Wilcoxen, P.J. (1989), "Intertemporal Optimization in General Equilibrium: A Practical Introduction," Impact Preliminary Working Paper No. IP-45, University of Melbourne.

E1. The Effects of a Price Shock

Section (3) explored the consequences of a dividend tax change for the model of section (2). There are, however, several other interesting experiments that can be examined with the model. One of these is an increase in the price of the firm's output, which will be the subject of this exercise. For convenience, the model's investment equation and equations of motion are shown below:

$$I = \frac{1}{2w\theta} \left(\frac{\lambda}{1 - T^d} - P_k \right) , \tag{2.34}$$

$$\lambda' = (r + \delta)\lambda - \beta(P)(1 - T^d) , \tag{2.35}$$

$$K' = \frac{\lambda}{2w\theta(1 - T^d)} - \delta K - \frac{P_k}{2w\theta} . \tag{2.36}$$

(a) Using a phase diagram, show the effect of an unexpected immediate, permanent increase in the price of the firm's output. Also, sketch the paths of the multiplier, investment and the capital stock over time. Briefly explain the results.

(b) Using another phase diagram, show the effect of a permanent increase in the firm's output price expected to occur in two years. Again, sketch the time paths of the model's variables.

(c) Finally, use a third phase diagram to analyze the effect of an anticipated temporary increase in the price of the firm's output lasting for three years. That is, investors believe that the price will rise immediately, stay high for three years, and then fall back to its original value. Sketch the paths of the variables.

******** Answer ********

(a) The phase diagram and time paths of the variables for a surprise permanent increase in the price of the firm's output are shown in figure E1.1. The increased output price immediately raises the marginal value of additional capital λ, so the system jumps from point A to point B. Since the price of raw capital goods has not changed, the firm expands investment until higher adjustment costs raise the price of installed capital to the new value of λ. With higher investment, the capital stock rises asymptotically toward its new steady state level at C.

(b) The phase diagram and graphs of the dynamic variables are shown in figure E1.2. These results differ quite a bit from those of the previous experiment. When news of the price increase first arrives, λ jumps part way toward its new steady state value. It does not jump all the way, however, because the higher price will not be obtained for several years. In the period between announcement and implementation of the new price, λ rises exponentially toward the new stable path, arriving there just as the price increase occurs. Investment follows the path of λ, jumping upward at the announcement and then rising steadily toward its new steady state value. The capital stock is drawn upward by the higher level of investment, reaching a peak rate of growth at the instant of implementation. After that it continues to grow, but at a slower rate. Over time, it asymptotically approaches the steady state.

(c) This experiment is the most interesting of the three. The phase diagram and variable paths are shown below in figure E1.3. In this case, a temporary increase in the output price leads to an upward jump in λ from point A to B at the instant the price rises. This occurs because the increased output price raises the returns to capital, at least for a while.

After the initial jump, the system moves downward and to the right under the control of the equations of motion holding at the higher price. The intuitive reason for this is quite interesting. When λ rises, so does investment. Higher investment leads to growth in the capital stock, so the system moves toward the right. At the same time, the remaining period of higher prices becomes steadily shorter. This causes λ to fall back toward its initial value, pushing the system downward. As λ falls, however, so does investment. Eventually, a point is reached where investment just covers depreciation of the higher capital stock. On the phase diagram, that point occurs where the dynamic path crosses the $K' = 0$ locus, as indicated by point C in the figure.

At point C, the system is moving straight down. Investment is just enough to maintain the captial stock, but λ is still declining. Past C, the model begins moving downward and to the left. The value of λ has fallen so much that investment is no longer enough to offset depreciation and the capital stock begins to erode. This continues until the price finally returns to its original value. At that time the system will have reached point D. At D, λ and investment are back to their original values, but the capital stock is somewhat higher than its steady state value. Thus, K continues to erode, gradually returning to its initial value.

An interesting feature of this experiment is that the dynamic path crosses the $K' = 0$ locus. This demonstrates that the loci do not necessarily confine the model to a particular quadrant. However, if the system does cross one of the loci, the derivative of the corresponding variable must be zero (by definition of the loci). In this experiment, for example, the derivative of the capital stock at point B must be zero; the dynamic path can only cross the $K' = 0$ locus while moving vertically. This feature also allows an interesting fact to be deduced about the solution: point D cannot lie to the left of the initial steady state. If it did, the system would have to cross the $K' = 0$ locus a second time. Since that can only occur while the path is moving vertically, λ would have to be rising. However, λ falls continuously after its initial jump, so the path cannot rise and cannot recross the $K' = 0$ locus. It must, therefore, intersect the original stable path at or to the right of the original steady state.

E2. The Effects of a Rise in the Price of Capital Goods

A second interesting experiment that can be explored with the model of section (2) is a change in the price of capital goods. Again, the model's key equations are repeated below for convenience:

$$I = \frac{1}{2w\theta}\left(\frac{\lambda}{1 - T^d} - P_k\right), \tag{2.34}$$

$$\lambda' = (r + \delta)\lambda - \beta(P)(1 - T^d), \tag{2.35}$$

$$K' = \frac{\lambda}{2w\theta(1 - T^d)} - \delta K - \frac{P_k}{2w\theta}. \tag{2.36}$$

(a) Using a phase diagram, explain the consequences of an unexpected immediate permanent fall in the price of capital goods. Sketch the paths of λ, investment and the capital stock over time.

(b) Now use a second phase diagram to illustrate what happens if the fall in P_k is anticipated several years in advance. Are the results what you expected? Discuss.

******** Answer ********

(a) A fall in the price of capital goods shifts the $K' = 0$ locus to the right, as shown in figure E2.1. The $\lambda' = 0$ locus is completely unaffected, however, so λ is already at its new steady state value and does not jump when the price change occurs. From equation (2.34), the lower price of capital leads to a higher level of investment, even with no change in λ. This, in turn, leads to capital accumulation, so the system moves gradually from the initial steady state at A toward the new steady state at B.

(b) Surprisingly, knowing about the price decline in advance does not change the behavior of firms at all. As shown in figure E2.2 below, the model remains at the initial steady state until the change actually occurs. After that, it proceeds in precisely the manner described in part (a).

At first this result appears peculiar. Intuitively, it seems as though the firm should be able to gain by postponing some investment just before the price decline and doing more investment later. The key, however, is adjustment costs. Recall that λ is unaffected by the shock. Since λ is the marginal benefit of additional capital goods, investment after the price change proceeds until adjustment costs rise just enough to make the cost of installed capital equal to λ. At that point, it would not be optimal for the firm to do more investing because the marginal cost of installed capital would be greater than its benefit. Thus, the firm would not want to move marginal units of investment from the instant before the price drop to the instant after it because those units will be just as expensive when adjustment costs are taken into account.

This point is illustrated in figure E2.3 which shows the marginal cost and marginal benefit of investment before and after the change in the price of capital goods. The marginal cost curve does, indeed, shift down. However, switching investment from before the change to after it entails saving an amount shown by box A, while spending the amount shown by box B. Clearly, this is a net loss to the firm, so investment will not be shifted across time. Thus, although the purchase price of new capital goods has fallen, the marginal cost of investment faced by the firm *at its new optimum* is unchanged. This means that it is not profitable for the firm to postpone investment.

E3. Adding More Taxes to the Investment Model

Now consider an economy in which there are two financial assets: government bonds and equities issued by corporations. Bonds pay a fixed rate of return and there is no inflation. The government levies three taxes: a dividend tax, a tax on interest payments, and a tax on capital gains. The earnings (short run profits) of firms are a function of wages, prices and the capital stock, but not a function of investment. On the other hand, investment costs depend on wages, prices and investment, but not on the capital stock. Firms take wages, prices, tax rates and the interest rate to be exogenous. Thus, this economy is similar to that of section (2), except that there are two additional taxes.

(a) Write down the arbitrage equation for this economy and explain what it means. Using the arbitrage equation, solve for an explicit expression for the value of the firm in terms of the earnings and investment cost functions and the model's exogenous variables. What transversality equation did you use? How should it be interpreted?

(b) Assuming the firm chooses investment to maximize its stock market value, write down the investment problem, form the Hamiltonian, and find the first order conditions.

(c) Suppose the earnings and investment cost functions have the form shown below, where β is function of wages and prices, K is the capital stock, P_k is the price of new capital goods, I is the level of investment, and θ is a parameter:

$$E = \beta(P)K \ , \tag{E3.1}$$

$$C = \frac{P_k I^2}{2\theta} \ . \tag{E3.2}$$

Using this information and the results obtained in part (b), find the first order conditions for this particular problem. Then find an expression for investment in terms of other variables. Finally, using the investment equation to eliminate investment from the other first order equations, show that the model's equations of motion are the following:

$$\lambda' = \left(r \, \frac{1 - T^i}{1 - T^c} + \delta \right)\lambda - B(P)\, \frac{1 - T^d}{1 - T^c} \ , \tag{E3.3}$$

$$K' = \frac{\lambda\theta}{P_k}\left(\frac{1 - T^c}{1 - T^d}\right) - \delta K \ . \qquad \text{(E3.4)}$$

******** Answer ********

(a) Arbitrage will equate after-tax returns on bonds and equity, so in equilibrium the following expression must hold:

$$(1 - T^i)rV = D(1 - T^d) + V'(1 - T^c) \ , \qquad \text{(E3.5)}$$

where r is the interest rate on bonds, V is the value of the firm, D is the dividend paid by the firm, and the T's are the three tax rates. The left side of (E3.5) is the after-tax return on V dollars of bonds, while the right side is the after-tax return on V dollars of equity.

The arbitrage condition in (E3.5) is a differential equation describing the evolution of the value of the firm. To find an explicit expression for the firm's value, start by rearranging (E3.5) to obtain:

$$V' - \frac{1 - T^i}{1 - T^c}\ rV = -\ \frac{1 - T^d}{1 - T^c}\ D \ . \qquad \text{(E3.6)}$$

This form suggests the equation can be solved using the integrating factor shown below:

$$e^{-R(0,t)} \ , \qquad \text{(E3.7)}$$

where $R(a, b)$ is defined by:

$$R(a, b) = \int_a^b r(v)\ \frac{1 - T^i(v)}{1 - T^c(v)}\ dv \ . \qquad \text{(E3.8)}$$

Expression (E3.8) takes the form of an integral because the two tax rates (and the interest rate) are not necessarily constant over time. If they were, (E3.8) would simplify to:

$$R(a, b) = r\ \frac{1 - T^i}{1 - T^c}\ (b - a) \ , \qquad \text{(E3.9)}$$

which is similar in form to the integrating factor used in section (2).

Multiplying (E3.6) by (E3.7) converts the left side of (E3.6) into an exact differential, so it is easy to show that the following is true:

$$(V' - \frac{1-T^i}{1-T^c}\, rV)e^{-R(0,t)} = \frac{d(Ve^{-R(0,t)})}{dt}\ . \tag{E3.10}$$

Thus, after multiplying by the integrating factor, (E3.6) can be integrated over [t,∞) to give:

$$\lim_{s\to\infty} V(s)e^{-R(0,s)} - V(t)e^{-R(0,t)} = -\int_t^\infty \frac{1-T^d}{1-T^c}\, De^{-R(0,s)}ds\ . \tag{E3.11}$$

At this point, we assume that the limit in the left-most term of (E3.11) is zero. This is known as a transversality condition, and it will be true as long as the value of the firm grows more slowly than the tax-adjusted interest rate as time tends toward infinity. The best way to interpret it is to look at what behavior it rules out. If the value of the firm were to grow more rapidly than the interest rate, for the arbitrage condition (E3.5) to hold, the firm would have to pay *negative* dividends; otherwise, no one would be willing to hold bonds. Thus, the transversality condition rules out firms whose value grows more rapidly than the tax-adjusted interest rate forever even though they pay negative dividends. In recognition of the Ponzi swindle, the transversality condition is often said to prohibit infinitely-lived Ponzi schemes.

After applying the transversality condition and rearranging slightly, we obtain an explicit equation for the value of the firm at any time t:

$$V(t) = \int_t^\infty \frac{1-T^d}{1-T^c}\, De^{-R(t,s)}ds\ . \tag{E3.12}$$

Note that use has been made of the following property:

$$e^{-R(0,s)} \cdot e^{R(0,t)} = e^{-R(t,s)}\ , \tag{E3.13}$$

which can be shown to be true from the definition of $R(a,b)$.

(b) From (E3.12) and the capital accumulation constraint, the firm's problem at time t can be stated as follows:

$$\max \int_t^\infty \frac{1-T^d}{1-T^c}\, De^{-R(t,s)}ds\ , \tag{E3.14}$$

$$\text{subject to } K' = I - \delta K\ . \tag{E3.15}$$

The appropriate Hamiltonian for this problem is:

$$H = \frac{1 - T^d}{1 - T^c}\ De^{-R(t,s)} + \Lambda(I - \delta K)\ . \tag{E3.16}$$

The first-order conditions are obtained by differentiation:

$$\frac{\partial H}{\partial I} = \frac{\partial D}{\partial I}\ (\frac{1 - T^d}{1 - T^c})\ e^{-R(t,s)} + \Lambda = 0\ , \tag{E3.17}$$

$$\frac{\partial H}{\partial K} = \frac{\partial D}{\partial K}\ (\frac{1 - T^d}{1 - T^c})\ e^{-R(t,s)} - \delta\Lambda = -\Lambda'\ , \tag{E3.18}$$

$$\frac{\partial H}{\partial \Lambda} = I - \delta K = K'\ . \tag{E3.19}$$

It is convenient to introduce the following transformation of Λ to eliminate the discount factors:

$$\Lambda(s) = \lambda(s) e^{-R(t,s)}\ . \tag{E3.20}$$

Differentiating this with respect to future time s gives:

$$\Lambda'(s) = \left(\ \lambda'(s) - r(s)\,\frac{1 - T^i(s)}{1 - T^c(s)}\ \lambda(s)\ \right) e^{-R(t,s)}\ . \tag{E3.21}$$

Inserting (E3.20) and (E3.21) into (E3.17) and (E3.18) gives:

$$\frac{\partial D}{\partial I}\ (\frac{1 - T^d}{1 - T^c}) + \lambda = 0\ , \tag{E3.22}$$

$$\frac{\partial D}{\partial K}\ (\frac{1 - T^d}{1 - T^c}) - \delta\lambda = -\lambda' + r\ \frac{1 - T^i}{1 - T^c}\ \lambda\ . \tag{E3.23}$$

Equations (E3.19), (E3.22) and (E3.23) are the model's first-order conditions.

(c) Equations (E3.1) and (E3.2) can be used to obtain the dividends function for the firm. Since dividends are the difference between earnings and investment, the following must be true:

$$D = E - C\ . \tag{E3.24}$$

Inserting (E3.1) and (E3.2) gives:

$$D = \beta(P)K - \frac{P_k I^2}{2\theta} . \tag{E3.25}$$

Equation (E3.25) can be differentiated to provide the differentials needed in (E3.22) and (E3.23):

$$\frac{\partial D}{\partial I} = - \frac{P_k I}{\theta} , \tag{E3.26}$$

$$\frac{\partial D}{\partial K} = \beta(P) . \tag{E3.27}$$

Inserting these into (E3.22), (E3.23) and (E3.19), and rearranging a bit, produces the particular first-order conditions for this problem:

$$\lambda = \frac{P_k I}{\theta} \left(\frac{1 - T^d}{1 - T^c}\right) , \tag{E3.28}$$

$$\lambda' = \left(r \frac{1 - T^i}{1 - T^c} + \delta \right)\lambda - \beta(P) \frac{1 - T^d}{1 - T^c} , \tag{E3.29}$$

$$K' = I - \delta K . \tag{E3.30}$$

Equation (E3.28) can be solved for investment as a function of K and λ:

$$I = \frac{\lambda\theta}{P_k} \left(\frac{1 - T^c}{1 - T^d}\right) . \tag{E3.31}$$

Using this to eliminate investment from (E3.30) produces the system’s equations of motion:

$$\lambda' = \left(r \frac{1 - T^i}{1 - T^c} + \delta \right)\lambda - \beta(P) \frac{1 - T^d}{1 - T^c} , \tag{E3.32}$$

$$K' = \frac{\lambda\theta}{P_k} \left(\frac{1 - T^c}{1 - T^d}\right) - \delta K , \tag{E3.33}$$

where equation (E3.32) is just (E3.29) repeated for clarity.

E4. The Effects of a Capital Gains Tax

This exercise explores the qualitative effects of a change in the capital gains tax introduced in exercise (E3). Use the equations of motion from part (c) of that exercise to answer the following questions.

(a) Construct a phase diagram for the model and label the important features of it clearly. Be sure that each locus has the correct slope.

(b) Draw another phase diagram and use it to analyze the effects of an unexpected permanent decrease in the capital gains tax. Show the initial and final steady states and the transition path. Sketch the paths of the multiplier (the costate variable), investment and the capital stock over time. Identify any important characteristics of the path and interpret it briefly.

(c) Draw a third phase diagram and use it to illustrate the effects of a permanent decrease in the capital gains tax announced several years in advance. Sketch the paths of important variables and interpret the solution. Does anticipation of the shock lead to any interesting or perverse effects?

(d) Finally, suppose the government surprises investors with a temporary drop in the capital gains tax. The tax is lowered immediately, kept low for several years, and then returned to its original level. Investors understand the new policy, and realize that the tax change is temporary. Analyze this shock using an appropriate phase diagram.

******** Answer ********

(a) The phase diagram for the model is shown in figure E4.1. It is very similar to the phase diagram derived in section (2), except that the loci are located in somewhat different positions due to the additional taxes.

(b) The phase diagram and intertemporal paths of capital, the multiplier (λ) and investment are shown in figure E4.2. The most interesting consequence of the shock is that λ falls while the capital stock rises. The latter comes about because even though λ falls, the implicit cost of investment drops more because of the reduction in the capital gains tax. In terms of equation (E3.31), the drop in T^c pushs investment up more than the reduction in λ lowers it.

(c) The diagrams for this section are shown in figure E4.3. An interesting and important feature of the result is the decline in the capital stock which occurs after the policy is announced but before it is implemented. The intuition behind this result is exactly the same as in the case of an announced dividend tax: after the policy has been implemented, the tax rate on capital gains relative to dividends is lower, so firms find it optimal to shift shareholder returns toward capital gains and away from dividends. This is accomplished by paying higher dividends before the policy is implemented, which drives down the capital stock, allowing it to grow rapidly after implementation. This could be described as a perverse effect because the policy causes a short term deterioration in K even though it increases K in the long run. Thus, anticipation causes the short and long term effects to have opposite sign.

(d) The diagrams are shown in figure E4.4. When the tax change is only temporary, there is no permanent effect on the capital stock. However, in the short run there will be a burst of growth in K until the tax is returned to its initial level. At that point, the capital stock begins to decline back to its original value. The intuition behind this is another variation on the theme discussed in part (c): the temporary drop in the capital gains tax makes capital gains (rather than dividends) a more effective way of transferring earnings to the stockholders. Thus, when the tax is low, firms do a lot of investment, raising the capital stock and producing capital gains for the stockholders. Once the tax is removed, investment returns to its original level and the capital stock begins to deteriorate back to the initial steady state.

E5. Diminishing Returns

Now consider an economy similar to that of exercise (3) but with diminishing returns to capital in the earnings function. In particular, suppose everything is the same except that earnings are given by:

$$E = \beta(P)\ln(K)\ , \tag{E5.1}$$

where β, P and K have the same interpretation as before. Since K has been replaced by its natural logarithm, the second derivative of the earnings function will be negative. This means that marginal earnings decrease as the capital stock increases, so from the firm's point of view, there are diminishing returns to capital.

(a) Starting from the general results obtained in part (b) of exercise (E3), derive the firm's first order conditions. From those, solve for investment as a function of other variables. Finally, solve for the model's equations of motion.

(b) Construct a phase diagram for the model and discuss how it compares to the one obtained in part (a) of exercise (E4).

(c) Using a phase diagram, analyze the effects of an unexpected permanent increase in the tax on interest payments. Show the initial and final steady states, and also the transition path. Sketch the paths of the multiplier, investment and the capital stock over time.

(d) Now draw a third phase diagram and use it to illustrate the effects of a permanent increase in the interest tax announced several years in advance. Sketch the paths of important variables and interpret the solution.

******** Answer ********

(a) In part (b) of exercise (E3), the general first order conditions for optimization in investment models of this type were shown to be the following:

$$I - \delta K = K' \ , \tag{E3.19}$$

$$\frac{\partial D}{\partial I} \left(\frac{1 - T^d}{1 - T^c}\right) + \lambda = 0 \ , \tag{E3.22}$$

$$\frac{\partial D}{\partial K} \left(\frac{1 - T^d}{1 - T^c}\right) - \delta \lambda = -\lambda' + r \ \frac{1 - T^i}{1 - T^c} \ \lambda \ . \tag{E3.23}$$

To obtain the specific first order equations for this model, the next step is to construct the dividend function. As in part (c) of exercise (E3), dividends are earnings less investment costs:

$$D = E - C \ . \tag{E5.2}$$

Applying equations (E5.1) and (E3.2) gives the following:

$$D = \beta(P) \ln(K) - \frac{P_k I^2}{2\theta} \ . \tag{E5.3}$$

Differentiating (E5.3) provides the terms needed in equations (E3.22) and (E3.23):

$$\frac{\partial D}{\partial I} = - \ \frac{P_k I}{\theta} \ , \tag{E5.4}$$

$$\frac{\partial D}{\partial K} = \beta \ \frac{(P)}{K} \ . \tag{E5.5}$$

Of these, only the second equation has changed from exercise (E3). Since $\partial D / \partial I$ has not changed, it is straightforward to show that the investment function and capital accumulation equation will be identical to those found in part (c) of (E3):

$$I = \frac{\lambda\theta}{P_k} \left(\frac{1 - T^c}{1 - T^d} \right), \tag{E3.31}$$

$$K' = \frac{\lambda\theta}{P_k} \left(\frac{1 - T^c}{1 - T^d} \right) - \delta K \ . \tag{E3.33}$$

However, the difference in $\partial D/\partial K$ changes the remaining first order condition considerably. Inserting (E5.5) into (E3.23) produces the following:

$$\lambda' = \left(r \frac{1-T^i}{1-T^c} + \delta \right)\lambda - \beta \frac{(P)}{K} \left(\frac{1-T^d}{1-T^c} \right). \tag{E5.6}$$

Thus expression (E3.31) gives the investment function for this model, while (E3.33) and (E5.6) give its equations of motion.

(b) The phase diagram for this model is shown in figure E5.1. It differs from figure E4.1 in two respects: both the $\lambda' = 0$ locus and the stable path are now downward sloping. These changes are due to the existence of diminishing returns in the earnings function. The $\lambda' = 0$ locus becomes hyperbolic because of the $1/K$ term in equation (E5.6). This reflects the fact that marginal earnings decrease as the capital stock becomes larger.

When the $\lambda' = 0$ locus changes, so does the stable path. To see why, consider where the model would go from an arbitrary point located horizontally to the left of the steady state. Such a point is no longer on the $\lambda' = 0$ locus and, in fact, is in a region where λ' is negative. This means that λ will begin decreasing as the system moves to the right. Moreover, an inspection of the phase diagram shows that the model will continue moving downward forever. Thus, if the system were to start at a point directly to the left of the steady state, λ would quickly fall below its steady state value and remain below it forever. A similar analysis applies for points to the right of the steady state, except that those points lead to perpetually increasing values of λ. The stable path, therefore, is no longer horizontal. In fact, it must be downward sloping, and will lie between the $\lambda' = 0$ locus and a horizontal line through the steady state. Only from points along such a path could the model eventually reach the steady state. The path will be unique under the conditions discussed in section (3.3)

(c) An increase in the tax on interest payments shifts the $\lambda' = 0$ locus to the right but leaves the $K' = 0$ locus unchanged. The steady state, therefore, moves upward and to the right, as shown by point C in figure E5.2. Since the tax change is permanent and occurs immediately, the model jumps instantly from the original steady state, point A, to point B on the new stable path. Then, as time passes, the system moves downward and to the right along the stable path from B to C. This produces the time paths of λ, I and K that are shown in figure E5.2.

At first, this result may seem surprising. Why should a tax increase lead to a rise in the capital stock? The reason can best be understood from equation (E3.5), the arbitrage condition for the model. When the tax on interest payments rises, the after-tax return on bonds falls. The tax does not apply to dividends or capital gains, however, so the return on equity is unchanged. Thus, the initial effect of the policy is to make the after-tax return on equity higher than the return on bonds. This produces a windfall gain to the holders of equity, which shows up in figure E5.2 as a jump from A to B. The increase in λ produces a subsequent rise in investment, so the capital stock begins to grow and the model moves toward the new steady state at C.

(d) When the tax increase is announced in advance, the system follows the path shown in figure E5.3. At the moment of the announcement, the model jumps from A to B because of the windfall benefit to holders of equity. It does not, however, move all the way to the new stable path because the actual tax change will not occur for some time. From point B the model evolves according to the original equations of motion (since the tax has not yet changed). It reaches point C at the exact instant the tax change is implemented. After that, the model moves along the new stable path toward the steady state at D. The paths of the model's variables over time are shown at the bottom of the figure.

E6. The Stock Market and the Costate Variable

Intertemporal investment models, such as the ones discussed in this paper, often generate equations giving the optimal value of investment as a function of the capital stock, a costate variable (or multiplier), and variables that the firm takes as given. In the model of section (2), for example, equation (2.34) gives investment as a function of K, λ and a number of prices:

$$I = \frac{1}{2w\theta} \left(\frac{\lambda}{1 - T^d} - P_k \right) . \tag{2.34}$$

If λ were observable, equation (2.34) and others like it could be estimated econometrically. This would allow values to be obtained for some of the parameters in the model, such as θ in the expression above. More importantly, however, it would also allow the statistical performance of the model to be assessed.

In what has become a very influential paper, Hayashi (1982) presented conditions under which the marginal value of additional capital (λ, in the notation of this paper) would be exactly equal to the average value of the capital stock. This finding allowed observable stock market data to be used to construct the unknown variable λ, which in turn allowed investment equations such as (2.34) to be estimated. A number of studies along those lines were conducted, a good example of which is Summers (1981). Because of its empirical importance, the remainder of this exercise will be devoted to deriving the Hayashi result.

So far in this paper, we have always assumed that dividends were additively separable into an earnings function, which was independent of investment, and an investment cost function, which was independent of the capital stock. Now we will relax that assumption and solve the investment problem under very general conditons. In particular, assume that dividends are a function of capital, investment, and a vector of short-run variables, P, that the firm takes as given:

$$D = D(K, I, P) . \tag{E6.1}$$

A single restriction will be imposed on D: it must be homogeneous of degree one in capital and investment. For convenience, assume the interest rate is constant and there are no taxes.

(a) Write down the firm's investment problem and derive the first-order conditions that must hold along the optimal investment path.

(b) Differentiate the following function with respect to future time s, where λ, as usual, is the current value multiplier associated with the capital accumulation constraint:

$$F(s) = \lambda(s) K e^{-r(s-t)} \ . \tag{E6.2}$$

Use the conditions found in part (a) to eliminate the terms in λ, λ' and K'. Then, apply Euler's theorem to simplify the result.

(c) Integrate the equation from part (b) over the interval $[t, \infty)$ and discuss the result.

******** Answer ********

(a) Since there are no taxes, the arbitrage condition used in section (2) applies. This means that the firm's value function can be obtained by inserting (E6.1) into (2.9) to give:

$$V(t) = \int_t^\infty D(K, I, P) e^{-r(s-t)} ds \ . \tag{E6.3}$$

Thus, the firm's investment problem is to choose I to maximize (E6.3) subject to the accumulation equation below:

$$K' = I - \delta K \ . \tag{E6.4}$$

The Hamiltonian for this problem is particularly simple:

$$H = D(K, I, P) e^{-r(s-t)} + \Lambda(I - \delta K) \ . \tag{E6.5}$$

Taking first-order conditions and converting the multiplier to its current value equivalent produces the following:

$$\frac{\partial D}{\partial I} + \lambda = 0 \ , \tag{E6.6}$$

$$\frac{\partial D}{\partial K} - \delta \lambda = -\lambda' + r \lambda \ , \tag{E6.7}$$

$$I - \delta K = K' . \tag{E6.8}$$

Equations (E6.6) through (E6.8) must hold along the optimal path of investment.

(b) Differentiating F with respect to s is straightforward and produces the following expression:

$$\frac{dF}{ds} = (\ \lambda' K + \lambda K' - r\lambda K\ ) e^{-r(s-t)} . \tag{E6.9}$$

Using (E6.7) to eliminate λ', (E6.6) to eliminate λ, and (E6.8) to eliminate K' gives, after collecting terms:

$$\frac{dF}{ds} = -\left(\frac{\partial D}{\partial K} K + \frac{\partial D}{\partial I} I \right) e^{-r(s-t)} . \tag{E6.10}$$

Finally, since D is homogeneous of degree one, Euler's theorem states that the following holds:

$$\frac{\partial D}{\partial K} K + \frac{\partial D}{\partial I} I = D(K, I, P) . \tag{E6.11}$$

Thus, (E6.10) can be simplified to:

$$\frac{dF}{ds} = -\ D(K, I, P) e^{-r(s-t)} . \tag{E6.12}$$

(c) Since F is known, integrating (E6.12) over an interval $[a, b]$ is straightforward and produces:

$$\lambda(b)K(b)e^{-r(b-t)} - \lambda(a)K(a)e^{-r(a-t)} = -\int_a^b D(K, I, P)e^{-r(s-t)} ds . \tag{E6.13}$$

Choosing the limits of integration to be t and ∞, and making use of the usual transversality condition on the behavior of λ as time tends toward infinity gives:

$$\lambda(t)K(t) = -\int_t^\infty D(K, I, P)e^{-r(s-t)} ds . \tag{E6.14}$$

The right side of this equation is exactly the same as the right side of equation (E6.3), so the following must be true:

$$\lambda(t)K(t) = V(t) \ . \tag{E6.15}$$

Thus, $\lambda(t)$ can be calculated using the formula below:

$$\lambda(t) = \frac{V(t)}{K(t)} \ . \tag{E6.16}$$

Equation (E6.16) shows that under the assumptions made above, the marginal value of an additional unit of capital, λ, is exactly equal to the average value of a unit of capital, V/K. Thus, λ can be calculated by simply dividing the firm's stock market value by its capital stock. Assuming that the capital stock can be observed or computed, (E6.16) provides a way of obtaining the multiplier, λ. However, this approach depends heavily on the assumption that the dividend function is homogeneous of degree one in capital and investment. If it is not, the average and marginal values of the capital stock will differ.

E7. Constructing Finite Difference Formulae

Difference formulae accurate to high orders can be constructed by using combinations of Taylor series expansions at several adjacent points. For example, a first-order difference accurate to $O(h^2)$ can be constructed by subtracting the expansion for $f(t-h)$ from that for $f(t+h)$. This is known as a "central" difference. Using more expansions, it is possible to construct formulae accurate to even higher orders.

It is also possible to construct difference approximations to higher-order derivatives. This is done by applying the method used for first-order differences recursively. For example, an approximation for a second-order derivative could be constructed as follows. First, difference formula for f'' in terms of f' would be built. Then, inserting an appropriate difference formulae for f' would produce the desired difference formula.

Finally, difference formulae can also be constructed for unevenly spaced grids in which adjacent points are separated by varying distances. This is done in exactly the same manner as for uniform grids, but with the appropriate distances inserted wherever h appears. As discussed in the text, uneven grid spacing can be a very powerful tool for reducing truncation error. However, it does introduce an additional source of truncation error at points where there are sharp changes in grid spacing. This point will be discussed in detail in part (c).

(a) Construct a first-order central difference formula. How does its accuracy compare with the forward difference presented in the text?

(b) Using the results of part (a), construct a second-order central difference formula. What is its order of accuracy?

(c) Construct the analog of a first-order centered difference for an unevenly spaced grid. Discuss its order of accuracy.

******** Answer ********

(a) As suggested at the beginning of the exercise, a first-order central difference is constructed from Taylor series expansions around time t for times $t+h$ and $t-h$. To fourth order, these expansions are the following

$$f(t+h) = f(t) + hf'(t) + \frac{h^2 f''(t)}{2!} + \frac{h^3 f'''(t)}{3!} + O(h^4) \, , \tag{E7.1}$$

$$f(t-h) = f(t) - hf'(t) + \frac{h^2 f''(t)}{2!} - \frac{h^3 f'''(t)}{3!} + O(h^4) \, . \tag{E7.2}$$

Subtracting produces:

$$\begin{aligned} &f(t+h) - f(t-h) = \\ &\left(f(t) + hf'(t) + \frac{h^2 f''(t)}{2!} + \frac{h^3 f'''(t)}{3!} + O(h^4) \right) - \\ &\left(f(t) - hf'(t) + \frac{h^2 f''(t)}{2!} - \frac{h^3 f'''(t)}{3!} + O(h^4) \right) . \end{aligned} \tag{E7.3}$$

Notice that the even-order terms all cancel out. Rearranging (E7.3) and dividing through by $2h$ produces the central difference formula below:

$$f'(t) = \frac{f(t+h) - f(t-h)}{2h} + O(h^2) \, . \tag{E7.4}$$

This expression is accurate to $O(h^2)$, an order more accurate than the simple forward and backward differences presented in the text.

(b) To construct a second-order central difference, start by expanding $f'(t+a)$ and $f'(t-a)$ around $f(t)$:

$$f'(t+a) = f'(t) + af''(t) + \frac{a^2 f'''(t)}{2!} + \frac{a^3 f''''(t)}{3!} + O(a^4) \, , \tag{E7.5}$$

$$f'(t-a) = f'(t) - af''(t) + \frac{a^2 f'''(t)}{2!} - \frac{a^3 f''''(t)}{3!} + O(a^4)\ . \qquad \text{(E7.6)}$$

Subtracting these, dividing through by $2a$, and rearranging produces an expression analogous to (E7.4):

$$f''(t) = \frac{f'(t+a) - f'(t-a)}{2a} + O(a^2)\ . \qquad \text{(E7.7)}$$

Substituting (E7.4) for the derivatives produces:

$$f''(t) = \frac{1}{2a}\left(\frac{f(t+2a) - f(t)}{2a} - \frac{f(t) - f(t-2a)}{2a} \right) + O(a^2)\ . \qquad \text{(E7.8)}$$

Since all of the terms at $t+a$ and $t-a$ have cancelled out, we can define a new step size, h, with the property that $h = 2a$. This allows (E7.8) to be rewritten as:

$$f''(t) = \frac{f(t+h) - 2f(t) + f(t-h)}{h^2} + O(h^2/4)\ . \qquad \text{(E7.9)}$$

For the purposes of error analysis, the factor of 1/4 in the error term is ignored, so (E7.9) is accurate to $O(h^2)$.

(c) A first-order central difference for a non-uniform grid is constructed almost exactly as shown in part (a), except that care must be taken about the spacing of the points. The first step is to construct Taylor series expansions around t for $t+a$ and $t-b$. To fourth order, these expansions are the following:

$$f(t+a) = f(t) + af'(t) + \frac{a^2 f''(t)}{2!} + \frac{a^3 f'''(t)}{3!} + O(a^4)\ , \qquad \text{(E7.10)}$$

$$f(t-b) = f(t) - bf'(t) + \frac{b^2 f''(t)}{2!} - \frac{b^3 f'''(t)}{3!} + O(b^4)\ . \qquad \text{(E7.11)}$$

Subtracting these produces:

$$f(t+a) - f(t-b) =$$

$$\left(f(t) + af'(t) + \frac{a^2 f''(t)}{2!} + \frac{a^3 f'''(t)}{3!} + O(a^4) \right) -$$

$$\left(f(t) - bf'(t) + \frac{b^2 f''(t)}{2!} - \frac{b^3 f'''(t)}{3!} + O(b^4) \right) . \tag{E7.12}$$

In this case, the even-order terms do not cancel out. Rearranging (E7.12) and dividing through by $a+b$ produces the following:

$$\frac{f(t+a) - f(t-b)}{a+b} = f'(t) + \frac{1}{2!}(a-b) f''(t) + \frac{1}{3!}\left(\frac{a^3+b^3}{a+b}\right) f'''(t) + \cdots . \tag{E7.13}$$

This suggests using the difference formula below:

$$f'(t) \approx \frac{f(t+a) - f(t-b)}{a+b} , \tag{E7.14}$$

which will have an error term given by the following:

$$\frac{1}{2!}(a-b) f''(t) + \frac{1}{3!}\left(\frac{a^3+b^3}{a+b}\right) f'''(t) + \cdots . \tag{E7.15}$$

When a and b are close in magnitude, the even-order terms in (E7.15) will be negligible. In that case, the error will be essentially $O(a^2)$ (or $O(b^2)$ for that matter, since $a \approx b$), which is the same as for a uniform grid. On an uneven grid, in regions where the spacing between points changes suddenly, a and b may differ substantially. When that occurs, the first term in (E7.15) will be significant, so the error will be more like $O(a)$. This effect can be minimized by avoiding sharp jumps in grid spacing, and by locating any such changes in the regions where $f''(t)$ is small. By constructing more elaborate difference formulae, the term in $(a-b)$ can be eliminated entirely. For more discussion of difference formulae refer to Fox (1962).

E8. Solving the Dividend Tax Analytically

For most experiments with most models, it is difficult or impossible to obtain an analytic solution to the model's equations of motion. The eigenvector transformation described in section (3.3) can be used when the model's coefficients are constant, such as near the steady state, but no general method exists for solving problems with time-varying coefficients. However, it is sometimes possible to obtain analytic solutions to particular models for particular experiments. The investment models in this paper, for example, can be solved analytically for the case of an announced increase in the dividend tax.

Consider sector A's investment problem from section (6) of the text. Its equations of motion were the following:

$$\lambda'_a = (r + \delta)\lambda_a - \beta(1 - T^d) , \tag{6.11}$$

$$K'_a = I_a - \delta K_a , \tag{6.12}$$

and investment was given by:

$$I_a = \frac{1}{2W\theta_a} \left(\frac{\lambda_a}{(1 - T^d)(1 - T^s)} - P_3 \right) . \tag{6.14}$$

Ordinarily, a model's equations of motion are mutually interdependent and must be solved simultaneously. In the problem above, however, the capital stock does not appear in equation (6.11). As a result, (6.11) can be integrated in isolation to obtain a function giving $\lambda_a(t)$ in terms of exogenous variables. This function can then be inserted into (6.14) to obtain investment. Investment, in turn, can be used in (6.12) to find the capital stock. Thus, for this model subjected to a dividend tax shock, it is possible to find an analytic solution.

(a) Assuming that β and W are constant, use a suitable integrating factor to solve (6.11) for a closed-form expression for $\lambda_a(t)$ in terms of the other variables.

(b) Now suppose the government announces that the dividend tax will rise from T_1^d to T_2^d at date τ in the future. Using this information, evaluate the integral obtained in part (a). It will help to recall the following property of integrals:

$$\int_a^b f(x)dx = \int_a^\tau f(x)dx + \int_\tau^b f(x)dx . \tag{E8.1}$$

(c) Insert the results of part (b) into (6.14) to obtain an expression for investment in terms of the exogenous variables. Also, derive a function $\Delta I_a(t)$ giving the change in investment at each point in time from its value before the policy change was announced. Assume that the model was initially at the steady state.

(d) Using a second integrating factor, solve (6.12) for the capital stock as a function of I_a and the exogenous variables.

(e) Combine the results of parts (c) and (d) to obtain an expression for $K_a(t)$ when $t \leq \tau$.

******** Answer ********

(a) By collecting terms in λ_a on the left side, equation (6.11) can be rearranged as shown below:

$$\lambda'_a - (r+\delta)\lambda_a = -\beta(1-T^d) \ . \tag{E8.2}$$

This suggests the integrating factor below:

$$e^{-(r+\delta)s} \ . \tag{E8.3}$$

Multiplying both sides of (E8.2) by (E8.3) gives:

$$(\ \lambda'_a - (r+\delta)\lambda_a\ )e^{-(r+\delta)s} = -\beta(1-T^d)e^{-(r+\delta)s} \ . \tag{E8.4}$$

If both sides of (E8.4) are multiplied by dt, the left side becomes an exact differential, so the entire equation can be rewritten as shown:

$$\frac{d}{dt}\left(\lambda_a e^{-(r+\delta)s}\right) = -\beta(1-T^d)e^{-(r+\delta)s} \ . \tag{E8.5}$$

Integrating both sides from t to ∞ and assuming that λ_a grows more slowly than $r+\delta$ as time tends to infinity, produces the following:

$$\lambda_a(t) = \int_t^\infty \beta(1-T^d)e^{-(r+\delta)(s-t)}ds \ . \tag{E8.6}$$

Equation (E8.6) is the desired closed-form solution for $\lambda_a(t)$.

(b) To evaluate (E8.6) for an announced dividend tax change occuring at time τ, start by splitting the integral into two parts:

$$\lambda_a(t) = \int_t^\tau \beta(1-T^d)e^{-(r+\delta)(s-t)}ds + \int_\tau^\infty \beta(1-T^d)e^{-(r+\delta)(s-t)}ds \ . \tag{E8.7}$$

This expression is valid when $t \le \tau$ (the case of $t > \tau$ will be covered below). In each integral, β and T^d are constant, although T^d differs between the two integrals, so (E8.7) can be rewritten as shown below:

$$\lambda_a(t) = \beta(1 - T_1^d) \int_t^{\tau} e^{-(r+\delta)(s-t)} ds + \beta(1 - T_2^d) \int_{\tau}^{\infty} e^{-(r+\delta)(s-t)} ds \ , \tag{E8.8}$$

where the two values of T^d have now been inserted. Evaluating (E8.8) is straightforward and produces the following expression:

$$\lambda_a(t) = \frac{\beta(1 - T_1^d)(1 - e^{-(r+\delta)(\tau-t)})}{r+\delta} + \frac{\beta(1 - T_2^d)e^{-(r+\delta)(\tau-t)}}{r+\delta} \ . \tag{E8.9}$$

Finally, (E8.9) can be simplified to give:

$$\lambda_a(t) = \frac{\beta(1 - T_1^d)}{r+\delta} \left(1 + \frac{T_1^d - T_2^d}{1 - T_1^d} \ e^{-(r+\delta)(\tau-t)} \right). \tag{E8.10}$$

This gives $\lambda_a(t)$ when $t \leq \tau$.

Obtaining the value of $\lambda_a(t)$ when $t > \tau$ is somewhat easier because (E8.6) can be integrated directly. Thus, for $t > \tau$ the following must hold:

$$\lambda_a(t) = \frac{\beta(1 - T_2^d)}{r+\delta} \ . \tag{E8.11}$$

Together, equations (E8.10) and (E8.11) give the value of λ_a for any point in time.

(c) Finding an expression for investment is a straightforward matter of inserting (E8.10) and (E8.11) into (6.14). When $t \leq \tau$, this produces:

$$I_a = \frac{1}{2W\theta_a} \left[\frac{\beta}{(r+\delta)(1 - T^s)} \left(1 + \frac{T_1^d - T_2^d}{1 - T_1^d} \ e^{-(r+\delta)(\tau-t)} \right) - P_3 \right]. \tag{E8.12}$$

In contrast, when $t > \tau$ investment is given by:

$$I_a = \frac{1}{2W\theta_a} \left(\frac{\beta}{(r+\delta)(1 - T^s)} - P_3 \right) . \tag{E8.13}$$

Notice that (E8.13) does not depend on the dividend tax, and is precisely the same as the expression giving investment before the tax change was announced. This is true because the dividend tax falls purely on profits, so once it has been implemented firms return to their original investment behavior. Thus, the difference between investment after implementation of the tax and investment before the tax change was announced is zero.

In the period between the announcement of the change and its implementation, investment does differ from its pre-announcement value. A function giving this difference is:

$$\Delta I_a(t) = I_a(t) - I_a^{ss} \ , \tag{E8.14}$$

where the pre-announcement (and post-implementation) value of investment has been written as I_a^{ss}. As noted above, I_a^{ss} is given by (E8.13), so comparing (E8.12) and (E8.13) shows that the following must be true:

$$\Delta I_a(t) = \frac{1}{2W\theta_a}\left(\frac{\beta}{(r+\delta)(1-T^s)}\right)\left(\frac{T_1^d - T_2^d}{1-T_1^d}\ e^{-(r+\delta)(\tau-t)}\right). \tag{E8.15}$$

(d) Solving (6.12) requires the same sequence of steps used to solve (6.11) in part (a). First, rearrange the equation as shown:

$$K'_a + \delta K_a = I_a \ . \tag{E8.16}$$

Now, introduce the integrating factor below:

$$e^{\delta s} \ , \tag{E8.17}$$

and multiply both sides of (E8.16) by (E8.17). As before, this converts the left side into an exact differential, so the equation can be written:

$$\frac{d}{dt}\left(K_a e^{\delta s}\right) = I_a \ . \tag{E8.18}$$

Integrating (E8.18) from 0 to t (since $K(0)$ is known) and rearranging gives the following:

$$K_a(t) = K(0)e^{-\delta t} + \int_0^t I_a e^{\delta(s-t)} ds \ . \tag{E8.19}$$

(e) The remaining step is to eliminate I_a from (E8.19) using the results of part (c). Start by noting that from the definition of ΔI_a in (E8.14), the following is true:

$$I_a(t) = I^{ss} + \Delta I_a(t) \ . \tag{E8.20}$$

Inserting (E8.20) into (E8.19) produces:

$$K_a(t) = K(0)e^{-\delta t} + \int_0^t I_a^{ss} e^{\delta(s-t)} ds + \int_0^t \Delta I_a e^{\delta(s-t)} ds \ . \qquad \text{(E8.21)}$$

If the model was initially at the steady state, the sum of the first two terms on the right gives the original steady state capital stock. To see this, notice that the second term on the right is the net amount of capital constructed by investing at rate I_a^{ss} over the period $[0, t]$. The first term on the right is the amount of capital left after depreciation over $[0, t]$, so the sum of the first two terms is the amount of capital that would exist at time t given an initial stock $K(0)$ and a rate of investment I_a^{ss}. Since the model was initially at the steady state, I_a^{ss} is exactly the investment needed to perpetuate the original capital stock, so the first two terms must add to $K(0)$. Thus, (E8.21) can be rewritten as shown:

$$K_a(t) = K(0) + \int_0^t \Delta I_a e^{\delta(s-t)} ds \ . \qquad \text{(E8.22)}$$

Inserting ΔI_a from expression (E8.15) gives:

$$K_a(t) = K(0) + \int_0^t \frac{\beta(T_1^d - T_2^d)e^{-(r+\delta)(\tau-s)} \ e^{\delta(s-t)}}{2W\theta_a(r+\delta)(1-T^s)(1-T_1^d)} \ ds \ . \qquad \text{(E8.23)}$$

Finally, evaluating the integral in (E8.23) gives the following expression for $K_a(t)$:

$$K_a(t) = K(0) + \frac{\beta(T_1^d - T_2^d)(e^{-(r+\delta)(\tau-t)} - e^{-(r+\delta)\tau-\delta t})}{2W\theta_a(r+\delta)(1-T^s)(1-T_1^d)(r+2\delta)} \ . \qquad \text{(E8.24)}$$

Equation (E8.24) can be used to find the capital stock in year 10 for the dividend tax experiment discussed in section (9). Inserting the parameters given in appendix A1 shows that an announced increase in T^d from 10 percent to 20 percent to occur in year 10 causes $K(10)$ to fall from 1 to 0.9113, the value mentioned in the text.

A1. The Trial Data Set

The data set used in section (9) is shown in the tables below. Since the base case was a steady state, these values were used for each grid point.

Table A1: Variables in the Trial Data Set

Symbol	Definition	Value
K_a	Capital stock A, specific to industry A	1.0
β	Short run profit on a unit of K_a	0.25
λ_a	Marginal value of K_a	1.5
K_b	Capital stock B, nonspecific	10.0
ρ	Rental price of a unit of K_b	0.25
λ_b	Marginal value of K_b	1.5
K_b^1	Type B capital used by industry 1	3.177778
K_b^2	Type B capital used by industry 2	4.622222
K_b^3	Type B capital used by industry 3	2.2
W	Wage rate	1.0
L	Total labor supply	5.0
L_a^P	Labor used in production by industry A	0.25
L_a^I	Labor used in investment by industry A	0.042593
L_b^I	Labor used in investment by industry B	0.425926
L_1	Labor used by industry 1	0.264815
L_2	Labor used by industry 2	3.466667
L_3	Labor used by industry 3	0.55
P_a	Price of good A	1.0
P_1	Price of good 1	1.0
P_2	Price of good 2	1.0
P_3	Price of raw capital goods	1.0
X_a	Production of good A	0.5
X_1	Production of good 1	1.059259
X_2	Production of good 2	4.622222
X_3	Production of raw capital goods	1.1
I_a	Investment by industry A	0.1
I_b	Investment by industry B	1.0
D_a	Dividends paid by industry A	0.121667
D_b	Dividends paid by industry B	1.216667
C	Private consumption	5.404500
G	Government spending	0.776981
T_w	Tax on wages	0.2
T_s^a	Sales tax on good A	0.0
T_s^1	Sales tax on good 1	0.0

continued ...

Table A1: Variables in the Trial Data Set, continued

Symbol	Definition	Value
T_s^2	Sales tax on good 2	0.0
T_s^3	Sales tax on good 3	0.0
LS	Lump sum payment	0.2
T^d	Dividend tax	0.10
T^s	Investment subsidy	0.10
γ_1	Technical change parameter, industry 1	0.620403
γ_2	Technical change parameter, industry 2	1.240806
γ_3	Technical change parameter, industry 3	1.0
ρ^x	Exogenous expectation, rental price	0.25
W^x	Exogenous expectation, wage rate	1.0
P_3^x	Exogenous expectation, raw capital price	1.0
P_a^x	Exogenous expectation, price of good A	1.0
T^{dx}	Exogenous expectation, dividend tax rate	0.10
T^{sx}	Exogenous expectation, investment subsidy	0.10
ρ^e	Actual expectation, rental price	0.25
W^e	Actual expectation, wage rate	1.0
P_3^e	Actual expectation, raw capital price	1.0
P_a^e	Actual expectation, price of good A	1.0
T^{de}	Actual expectation, dividend tax rate	0.10
T^{se}	Actual expectation, investment subsidy	0.10
r	Interest rate	0.05
ζ	Price deflator	1.0

Table A2: Parameters in the Trial Data Set

Symbol	Definition	Value
δ	Depreciation rate	0.10
θ_a	Investment parameter, industry A	4.259259
θ_b	Investment parameter, industry B	0.425926
ε_a	Labor exponent, industry A	0.5
ε_1	Labor exponent, industry 1	0.25
ε_2	Labor exponent, industry 2	0.75
ε_3	Labor exponent, industry 3	0.5
α_C^a	Share of private consumption, good A	0.080887
α_C^1	Share of private consumption, good 1	0.171360
α_C^2	Share of private consumption, good 2	0.747753
α_G^a	Share of government spending, good A	0.080887
α_G^1	Share of government spending, good 1	0.171360
α_G^2	Share of government spending, good 2	0.747753

A2. The Model's Equations

The appendix lists the equations of the intertemporal general equilibrium model discussed in section (9) in the form used to implement the model on a computer. For the most part, the equations appear exactly as they did in the text. However, three changes have been made: derivatives have been replaced by finite difference approximations, expected variables have been used in the investment submodels, and several variables have been eliminated by algebraic substitution. Where possible, the equations have been numbered as they were in the text.

A2.1. Dynamic Equations

The model contains four true dynamic equations which link different points in time; two are sector A's equations of motion, and two are sector B's. Together these equations govern the evolution of four variables: λ_a, λ_b, K_a and K_b. Since each equation of motion holds over each grid interval, if there are N intervals there will be $4N$ dynamic equations in all.

Equations of motion for industry A:

$$\frac{\lambda_a(t+h) - \lambda_a(t)}{h} = (r+\delta)\lambda_a - \beta(1 - T^{de}) \ , \tag{6.15}$$

$$\frac{K_a(t+h) - K_a(t)}{h} = I_a - \delta K_a \ . \tag{6.16}$$

Equations of motion for industry B:

$$\frac{\lambda_b(t+h) - \lambda_b(t)}{h} = (r+\delta)\lambda_b - \rho^e(1 - T^{de}) \ , \tag{6.24}$$

$$\frac{K_b(t+h) - K_b(t)}{h} = I_b - \delta K_b \ . \tag{6.25}$$

A2.2. Boundary Conditions

The model has four boundary conditions: the two initial capital stocks, $K_a(0)$ and $K_b(0)$, and the two steady state multipliers, λ_a^{ss} and λ_b^{ss}. The capital stocks can be obtained from observable data, but the multipliers must be calculated using the equations below:

$$\lambda_a^{ss} = \frac{\beta(1 - T^{de})}{r+\delta} \ , \tag{6.17}$$

$$\lambda_b^{ss} = \frac{\rho^e(1 - T^{de})}{r + \delta} \ . \tag{6.26}$$

A2.3. Intraperiod Equations

Each period's submodel consists of the 27 equations listed below. From Walras Law, however, we know that one of these is redundant, so we will drop equation (7.25) to leave 26 independent equations. Since these hold at each grid point, a grid of N intervals will have $26(N+1)$ intraperiod equations.

Short run profit on a unit of K_a:

$$\beta = (\frac{1 - \varepsilon_a}{\varepsilon_a})\left(\frac{\varepsilon_a P_a^e}{W^e}\right)^{1/(1-\varepsilon_a)} W^e \ . \tag{6.13}$$

Investment by sector A:

$$I_a = \frac{1}{2W^e\theta_a}\left(\frac{\lambda_a}{(1 - T^{de})(1 - T^{se})} - P_3^e \right). \tag{6.14}$$

Output of sector A:

$$X_a = (L_a^P)^{\varepsilon_a}(K_a)^{1-\varepsilon_a} \ . \tag{7.1}$$

Labor demanded for production in sector A:

$$L_a^P = \left(\frac{\varepsilon_a P_a}{W}\right)^{1/(1-\varepsilon_a)} K_a \ . \tag{7.2}$$

Labor demanded for investment by sector A:

$$L_a^I = \theta_a I_a^2 \ . \tag{7.4}$$

Pretax dividends of sector A:

$$D_a = P_a X_a - WL_a^P - \left(P_3 I_a + WL_a^I\right)(1 - T^s) \ . \tag{7.5}$$

Investment by sector B:

$$I_b = \frac{1}{2W^e\theta_b}\left(\frac{\lambda_b}{(1-T^{de})(1-T^{se})} - P_3^e\right). \tag{6.23}$$

Labor demanded by sector B:

$$L_b^I = \theta_b I_b^2 \ . \tag{7.7}$$

Pretax dividends of sector B:

$$D_b = \rho K_b - (P_3 I_b + W L_b^I)(1-T^s) \ . \tag{7.8}$$

Labor demanded by sector i, $i \in \{1, 2, 3\}$:

$$L_i = \frac{1}{\gamma_i} X_i \left(\frac{\rho\varepsilon_i}{W(1-\varepsilon_i)}\right)^{1-\varepsilon_i} \ . \tag{7.10}$$

Capital B demanded by sector i, $i \in \{1, 2, 3\}$:

$$K_b^i = \frac{1}{\gamma_i} X_i \left(\frac{W(1-\varepsilon_i)}{\rho\varepsilon_i}\right)^{\varepsilon_i} \ . \tag{7.11}$$

Zero-profit condition for sector i, $i \in \{1, 2\}$:

$$X_i P_i = W L_i + \rho K_b^i \ . \tag{7.12}$$

Zero-profit condition for sector 3:

$$X_3 P_3 = (1 + T_s^3)(W L_3 + \rho K_b^3) \ . \tag{7.13}$$

Labor market equilibrium condition:

$$L = L_a^P + L_a^I + L_b^I + L_1 + L_2 + L_3 \ . \tag{7.26}$$

Capital B market equilibrium condition:

$$K_b = K_b^1 + K_b^2 + K_b^3 \ . \tag{7.27}$$

Consumption:

$$C = WL(1 - T^w) + (D_a + D_b)(1 - T^d) + LS \ . \qquad (7.14)$$

Government spending:

$$G = T^d(D_a + D_b) - T^s(P_3(I_a + I_b) + W(\theta_a I_a^2 + \theta_b I_b^2))$$
$$+ T_s^a P_a X_a + T_s^1 P_1 X_1 + T_s^2 X_2 P_2 + T_s^3 P_3 X_3$$
$$+ T^w WL - LS \ . \qquad (7.18)$$

Market equilibrium for good i, $i \in \{A, 1, 2\}$:

$$P_i X_i (1 + T_s^i) = \alpha_C^i C + \alpha_G^i G \ . \qquad (7.22\text{-}7.24)$$

Market equilibrium for good 3:

$$X_3 = I_a + I_b \ . \qquad (7.25)$$

Price deflator:

$$\zeta = \frac{X_a P_a (1 + T_s^a) + X_1 P_1 (1 + T_s^1) + X_2 P_2 (1 + T_s^2)}{X_a [P_a (1 + T_s^a)]_b + X_1 [P_1 (1 + T_s^1)]_b + X_2 [P_2 (1 + T_s^2)]_b} \ . \qquad (7.28)$$

A2.4. Expectations

Next, there are 6 equations which determine investors' expectations. These hold at each grid point, so on a grid of N intervals, there will be $6(N + 1)$ expectations equations in all. The equations are as follows:

$$W^e = (W)^{\lambda_n}(W^x)^{1-\lambda_n} \ , \qquad (8.1)$$

$$\rho^e = (\rho)^{\lambda_n}(\rho^x)^{1-\lambda_n} \ ,$$

$$P_a^e = (P_a)^{\lambda_n}(P_a^x)^{1-\lambda_n} \ ,$$

$$P_3^e = (P_3)^{\lambda_n}(P_3^x)^{1-\lambda_n} \ ,$$

$$T^{de} = (T^d)^{\lambda_x}(T^{dx})^{1-\lambda_x} \ ,$$

$$T^{se} = (T^s)^{\lambda_x}(T^{sx})^{1-\lambda_x} \ .$$

Finally, we can verify that the model is correctly identified by counting its equations and variables. From the table A1 in appendix (A1) it can be seen that the complete model has 56 variables at each grid point, or $56(N+1)$ in all. Adding up the number of dynamic, intraperiod and expectations equations gives a total of $36N+34$. Subtracting this from $56(N+1)$ gives $20(N+1)+2$, the number of variables which must be set exogenously. This is exactly the number of exogenous variables shown in table 9.1, so we conclude that the model is correctly identified.