CHAPTER 1
KEYNES AND THE CLASSIC

Almost 70 years have elapsed since the publication of Keynes’ *The General Theory of Employment, Interest and Money*, yet the controversies between his followers and those macroeconomists who favour a more classical approach have remained active. One purpose of this book is to examine some of these controversies, to draw attention to developments that have led to a synthesis of important ideas from both traditions, and to illustrate in some detail how this integrated approach can inform policy debates.

At the policy level, the hallmarks of Keynesian analysis are that involuntary unemployment can exist and that, without government assistance, any adjustment of the system back to the "natural" unemployment rate is likely to be slow and to involve cycles and overshoots. In its extreme form, the Keynesian view is that adjustment back to equilibrium simply does not take place without policy assistance. This view can be defended by maintaining either of the following positions: (i) the economy has multiple equilibria, only one of which involves "full" employment; or (ii) there is only one equilibrium, and it involves "full" employment, but the economic system is unstable without the assistance of policy, so it cannot reach the "full" employment equilibrium on its own.

We shall consider the issue of multiple equilibria in Chapter 9. In earlier chapters, we focus on the question of convergence to a full equilibrium. To simplify the exposition, we concentrate on stability versus outright instability, which is the extreme form of the issue. We interpret any tendency toward outright instability as analytical support for the more general proposition that adjustment between full equilibria is protracted.

In this first chapter, we examine alternative specifications of the labour market, such as perfectly flexible money wages (the textbook Classical model) and complements fixed money wages (the textbook Keynesian model), to clarify some of the causes of unemployment. We consider fixed goods prices as well (the model of generalized disequilibrium), and then we build on this background in later chapters. For example, in Chapter 2, we assume that nominal rigidities are only temporary, and we consider a dynamic analysis that has Classical properties in full equilibrium, but Keynesian features in the transitional periods on the way to full equilibrium. Fifty years ago, Paul Samuelson labelled this class of dynamic models the Neoclassical Synthesis.

In Chapter 3, we enrich this dynamic analysis by exploring alternative ways of bringing expectations into the analysis. With expectations involved, it is not obvious that an increased degree of price flexibility lessens the amount of cyclical unemployment that follows from a decrease in aggregate demand. By the end of Chapter 3, we will have identified two important considerations that make macroeconomic convergence more problematic: firms’ reactions to sticky prices and sales constraints, and expectations.

In Chapter 4, we rectify one major limitation of the analysis to that point — that formal micro-foundations have been missing. The inter-temporal optimization that is needed to overcome this limitation is explained in Chapter 4. Then, in Chapter 5, we examine the New Classical approach to business cycle analysis — the modern, more micro-based version of the market-clearing approach to macroeconomics, in which no appeal to sticky prices is involved. Finally, in Chapters 6 and 7, we examine what has been called the "Neoclassical Synthesis — a business-cycle analysis that blends the microeconomic rigour of the New Classical with the empirical applicability that has always been the focus of the Keynesian tradition and the original Neoclassical Synthesis.

For the remainder of the book (the final five chapters), the focus shifts from short-term stabilization issues to concerns about long-run living standards. In these chapters, we focus on structural unemployment and the challenge of rising productivity growth.

1.2 Criteria for Model Selection

To ensure a useful selection of macro models, economists rely on two broad criteria. First, models must be subjected to empirical tests, to see whether the predictions are consistent with actual experience. This criterion is fundamentally important. Unfortunately, however, it cannot be the only one for model selection, since empirical tests are often not definitive. Thus, while progress has been made in developing applied methods, macroeconomists have no choice but to put at least some weight on a second criterion for model evaluation.

Since the hypothesis of constrained maximization is at the core of our discipline, all modern macroeconomists agree that macro models should be evaluated as to their consistency with optimizing underpinnings. Without a microeconomic base, there is no well defined basis for arguing that either an ongoing stabilization policy, or an increase in the average growth rate, improves welfare. Increasingly, Keynesians have realized that they must acknowledge this point. Further, the challenge posed by New Classicalists has forced Keynesians to admit that it is utility and production functions that are independent of government policy; agents’ decision rules do not necessarily remain invariant to shifts in policy. A specific microeconomic base is required to derive how private decision rules may be adjusted in the face of major changes in policy. Another advantage is that a specific microeconomic rationale imposes more structure on macro models, so the corresponding empirical work involves fewer “free” parameters (parameters that are not constrained by theoretical considerations and can thus take on whatever value will maximize the fit of the model). It must be admitted that the empirical success of a model is compromised if the estimation involves many free parameters.

Despite these clear advantages of an explicit microeconomic base, those who typically stress these points — the New Classicalists — have had to make some acknowledgments too. They have had to admit that, until recently, their models have been inconsistent with several important empirical regularities. As a result, they, like Keynesians, now allow for some temporary stickiness in nominal variables. Also, since the primary goal of this school of thought is to eliminate arbitrary assumptions, its followers cannot downplay the significance of aggregation issues or the non-uniqueness problem that often plagues the solution of their models. These issues have yet to be resolved in a satisfactory manner.

During the 1970s and 1980s, controversy between New Classicalists and Keynesians was frustrating for students. Each group focused on the advantages of its own approach, and tended to ignore the legitimate criticisms offered by the “other side.” The discipline was fragmented into two schools of thought that did not interact. In the 1990s however, there was an increased willingness on the part of macroeconomists to combine the best features of the competing approaches so that now the subject is empirically applicable, has solid micro-foundations, and it allows for market failure — so economic policy can be explored in a rigorous fashion. Students can now explore models that combine the rigour of the New Classicalists with the policy concern that preoccupies Keynesians.

The purpose of any model is to provide answers to a series of if-then questions: if one assumes a specified change in the values of the exogenous variables (those determined outside of the model), what will happen to the set of endogenous variables (those determined within the model)? A high degree of simultaneity seems to exist among the main endogenous variables (for example, household behaviour makes
consumption depend on income, while the goods market-clearing condition makes income depend on consumption. To cope with this simultaneity, we define macro models in the form of systems of equations for which standard solution techniques (either algebraic or geometric) can be employed. A model comprises a set of structural equations, that are either definitions, equilibrium conditions, or behavioral reaction functions assumed on behalf of agents. The textbook Keynesian model, the textbook Keynesian model, and the "more Keynesian" model of generalized disequilibrium (all summarized graphically in later sections of this chapter) are standard examples.

In constructing these models, macroeconomists have disciplined their selection of alternative behavioral rules by appealing to microeconomic models of households and firms. In other words, their basis for choosing structural equations is constrained maximization at the individual level, without much concern for problems of aggregation. To keep the analysis manageable, macroeconomists sometimes restrict attention to particular components of the macroeconomy, considered at a time. They record the resulting decision rules (the consumption function, the investment function, the money-demand function, the Phillips curve, and so on, which are the first-order conditions of the constrained maximizations) as a list of structural equations. This series of equations is then brought together for solving as a standard set of simultaneous equations in which the unknowns are the endogenous variables.

In other words, the procedure has two stages:

**Stage 1:** Derive the structural equations, which define the macro model, by presenting a set of (sometimes unconnected) constrained maximization exercises (that is, define and solve a set of microeconomic problems).

**Stage 2:** Use the set of structural equations to derive the solution or reduced form equations (in which each endogenous variable is related explicitly to nothing but exogenous variables and parameters) and perform the counterfactual exercises (for example, derivation of the policy multipliers).

Before 1970, macroeconomics developed in a fairly orderly way, following this two-stage approach. In recent decades, however, the discipline has seen some changes in basic approaches following from the fact that macroeconomists have tried to take seriously the more consistent and complicated theories of household and firm behaviour. That is, the specification of the constrained maximizations in stage 1 of the analysis has been made more general by allowing for such things as dynamics and the fact that agents must make decisions on the basis of expectations of the future.

This expansion has led to some conceptual and methodological complications. Many analysts now regard it as unappealing to derive one component structural equation without reference at stage 1 to the properties of the overall system. For example, if agents' behavior turns out to depend on expected inflation, it is tempting to model their forecast of inflation so that it is consistent with the actual inflation process, which is determined as one of the endogenous variables within the model. From a technical point of view, such an approach means that stages 1 and 2 must be considered simultaneously. It also means that the form of the structural equations and, therefore, the overall structure of the model itself depends on the assumed time paths of the exogenous variables. Thus, it may be a bad practice for economists to use an estimated model found suitable for one data period as a mechanism for predicting what would happen in another structural period.

We shall consider this problem, which is referred to as the Lucas critique, in later chapters. Initially, however, we restrict attention to models whose structures are assumed to be independent of the behaviour of the exogenous variables. The textbook Keynesian and classical models (covered in the remainder of this chapter) are examples of such models.

### 1.3 The Textbook Classical Model: The Labour Market with Flexible Wages

The classical macro model is defined by the following equations:

\[ Y = C(Y - kY) + I + G \]

(1.1)

\[ I(Y, r) = M/P \]

(1.2)

\[ Y = F(N, K) \]

(1.3)

\[ W = PF(N, K) \]

(1.4)

\[ W(1-k) = PS(N) \]

(1.5)

### Equations (1.1) and (1.2) are the IS and LM relationships; the symbols \( Y, C, I, G, M, P, k \) and \( r \) denote real output, household consumption, firms' investment spending, government program spending, the nominal money supply, the price of goods, the proportional income tax rate, and the interest rate. Since we ignore expectations at this point, anticipated inflation is assumed to be zero, so there is no difference between the nominal and real interest rates. The standard assumption concerning the behavioural equations (with partial derivatives indicated by subscripts) are: \( f_k < 0, f_n > 0, k < 0, f_y < 0 \). The usual specification of government policy (that \( G, k \) and \( M \) are set exogenously) is also imposed. The aggregate demand for goods relationship follows from the IS and LM functions, as is explained below.

Equations (1.3), (1.4), and (1.5) are the production, labour demand, and labour supply functions, where \( W, N \) and \( K \) stand for the nominal wage rate, the level of employment of labour and the capital stock. The assumptions we make about the production function are standard (that is, the marginal products are positive and diminishing) \( f_n f_y f_k = f_n > 0, f_y f_k < 0 \). Equation (1.4) involves the assumption of profit maximization: firms hire workers up to the point that labour's marginal product equals the real wage. It is assumed that it is not optimal for firms to follow a similar optimal hiring rule for capital, since there are installation costs. The details of this constraint are explained in Chapter 4.

Here we simply follow convention and assume that firms invest more in new capital, the lower are borrowing costs. We allow for a positively sloped labour supply curve by assuming \( S_k > 0 \). Workers care about the after-tax real wage, \( W/P_k = k/P \).

In the present system, the five equations determine five endogenous variables: \( Y, N, r, P \), and \( W \). However, the system is not fully simultaneous. Equations (1.4) and (1.5) form a subset that can determine employment and the real wage \( w = W/P \). If the real wage is eliminated by substitution, equations (1.4) and (1.5) become \( f_n(N, K) = f_y(N, K) \). Since \( k \) and \( N \) are given exogenously, \( N \) is determined by this one equation, which is the labour market equilibrium condition. This equilibrium value of employment can then be substituted into the production function, equation (1.3), to determine output. Thus, this model involves what is called the Classical Dichotomy: the key real variables (output and employment) are determined solely on the basis of aggregate supply relationships (the factor market relations and the production function), while the demand considerations (the IS and LM curves) determine the other variables (\( r \) and \( P \)) residually.

The model can be pictured in terms of aggregate demand and supply curves (in price-output space), so the term "supply-side economics" can be appreciated. The aggregate demand curve comes from equations 1.1 and 1.2. Figure 1.1 gives the graphical derivation. The aggregate demand curve in the lower panel represents all those combinations of price and output that satisfy the demands for goods and assets. To check that this aggregate demand curve is negatively sloped, we take the total differential of the IS and LM equations, set the exogenous variable changes to zero, and solve for \( (dP/dY) \) after eliminating \( (dY) \) by substitution. The result is

\[ \text{Slope of the aggregate demand curve} = \frac{r}{(1+P)} \text{in } P-Y \text{ space:} \]

\[ dP/dY = -(k(1 + 1/c)) / (M/P^2) < 0 \]  

(1.6)

The aggregate supply curve is vertical, since \( P \) does not enter the equation (any supply of \( P \), along with the labour market-clearing level of \( Y \), satisfies these supply conditions). The summary picture, with shift variables listed in parentheses, is shown in Figure 1.2. The key policy implication is that the standard monetary and fiscal policy variables, \( G \) and \( M \), involve price effects only. For example, complete "crowding out" follows increases in government spending (that is, output is not affected). The reason is that higher prices shrink the real value of the money supply so that interest rates are pushed up and pre-existing private investment expenditures are reduced. Nevertheless, tax policy has a role to play in this model. A tax cut shifts both the supply and the demand curve to the right. Thus, output and employment must increase, although price may go up or down. Blinder (1973) formally derives the \( (dP/dY) \) multiplier and, considering plausible parameter values, argues that it is negative. "Supply-side" economists are those who favour applying this "textbook Classical model" to actual policy making (as was done in the United States in the 1980s).
From a graphic point of view, the "classical dichotomy" feature of this model follows from the fact that it has a vertical aggregate supply curve. But the position of this vertical line can be shifted by tax policy. A policy of balanced-budget reduction in the size of government makes some macroeconomic sense here. Cuts in G and k may largely cancel each other in terms of affecting the position of the demand curve, but the lower tax rate stimulates labour supply, and so shifts the aggregate supply curve for goods to the right. Workers are willing to offer their services at a lower before-tax wage rate, so profit-maximizing firms are willing to hire more workers. Thus, according to this model, both higher output and lower prices can follow tax cuts.

This model also suggests that significantly reduced prices can be assured (without reduced output rates) if the money supply is reduced. Such a policy shifts the aggregate demand curve to the left but does not move the vertical aggregate supply curve. In the early 1980s, several Western countries tried a policy package of tax cuts along with decrease money supply growth; the motive for this policy package was, to a large extent, the belief that the Classical macro model has some short-run policy relevance. Such policies are controversial, however, because various analysts believe that the model ignores some key questions. Is the real world supply curve approximately vertical in the short run? Are labour supply elasticities large enough to lead to a significant shift in aggregate supply? Many economists doubt that these conditions are satisfied.

Another key issue is the effect on macroeconomic convergence of the growing government debt that accompanies this combination policy of decreased reliance on taxation and money issue as methods of government finance. The textbook Classical model abstracts from this consideration. An explicit treatment of government debt is considered later in this book (in Chapter 7), and a negative verdict on the possibility of tax cuts paying for themselves is available in Mankiw and Weinzierl (2006).

Before leaving the textbook Classical model, we summarize a graphic exposition that highlights both the goods market and the labour market. In Figure 1.3, consider that the economy starts at point A. Then a decrease in government spending occurs. The initial effect is a leftward shift of the IS curve (and therefore, in the aggregate demand curve). At the initial price level, aggregate supply exceeds aggregate demand. The result is a fall in the price level, and this in turn causes two shifts in the labour market quadrant of Figure 1.3: (1) labour demand shifts down (because of the decrease in the marginal revenue product of labour); and (2) labour supply shifts down by the same proportionate amount as the decrease in the price level (because of workers' decreased money-wage claims). Both workers and firms care about real wages; had we drawn the labour market with the real wage on the vertical axis, neither the first nor the second shift would occur. These shifts occur because we must "correct" for having drawn the labour demand and supply curves with reference to the nominal wage. The final observation point for the economy is B in both bottom panels of Figure 1.3. The economy avoids ever having a recession in actual output and employment since the shock is fully absorbed by the falling wages and prices. These fixed levels of output and employment are often referred as economy's "natural rates" (denoted here by $F$ and $N$).

Many economists find this model unappealing for two reasons. First, they think do observe recessions in response to drops in aggregate demand. Second, adjustment within this model involves firms that are perfectly happy to let inventories accumulate. A series of large decreases in aggregate demand would cause a dramatic increase in inventories, yet firms apparently never want to work them down since the model shows no layoffs.

This implicit build-up of inventories will be particularly acute if the economy is characterized by the phenomenon that Keynes called a "liquidity trap". This special case can be considered by letting the interest sensitivity of money demand become very large: $L_s \to \infty$. By checking the slope expression for the aggregate demand curve (equation (1.6) above), the reader can verify that this situation involves the aggregate demand curve being so steep that it is almost vertical. Thus, when this curve shifts to the left, it may no longer intersect the aggregate supply curve anywhere in the positive quadrant. In this situation, falling wages and prices cannot eliminate the recession. Indeed, no consistent full equilibrium exists in this case. The Classical model can, however, be modified to avoid this problem by allowing the household consumption-savings decision to depend on the quantity of liquid assets available – by making the consumption function $(C(y, h), \bar{M}/\bar{P})$. The second term in this function is referred to as the Pigou effect.

Macro models focusing on inventory fluctuations were very popular many years ago (see, for example, Metzler (1941)). Space limitations preclude our reviewing these analyses, but the reader is encouraged to consult Blinder (1981). Suffice it to say here that macroeconomic stability is problematic when firms try to work off large inventory holdings since periods of excess supply must be followed by periods of excess demand. Thus, it is very difficult to avoid overstocking when inventories are explicitly modelled.

What changes are required in the Classical model to make the system consistent with the existence of recessions and unemployment? We consider the New Classical's response to this question in Chapter 5. But here we focus on the traditional responses to this question, so that we can clarify what most economists have assumed to be the major assumptions behind Keynes' analysis. Keynes considered: (1) money-wage rigidity; (2) a model of generalized disequilibrium involving both money-wage and price rigidity; and (3) expectations effects that could destabilize the economy. The first and second points can be discussed in a static framework and so are analyzed in the remainder of this introductory chapter. The third point requires a dynamic analysis, which will be undertaken in Chapters 2 and 3.
1.4 The Textbook Keynesian Model: The Labour Market with Money-Wage Rigidity

Contracts, explicit or implicit, often fix money wages for a period of time. In Chapter 5, we shall consider some of the considerations that might motivate these contracts. For the present, however, we simply presume the existence of fixed money-wage contracts and we explore their macroeconomic implications.

On the assumption that money wages are fixed by contracts for the entire relevant short run, $W$ is now taken as an exogenous variable stuck at value $\bar{W}$. Some further change in the model is required, however, since otherwise we would now have five equations in four unknowns — $Y$, $N$, $r$, and $P$.

Since the money wage does not clear the labour market in this case, we must distinguish actual employment, labour demand, and labour supply, which are all equal only in equilibrium. The standard assumption in disequilibrium analyses is to assume that firms have the "right to manage" the size of their labour force during the period after which the wage has been set. This means that labour demand is always satisfied, and that the five endogenous variables are now $Y, r, P, N, N'$, where the latter variable is desired labour supply. Since this variable occurs nowhere in the model except in equation (1.5), that equation solves residually for $N'$.

Actual employment is determined by the intersection of the labour demand curve and the given money wage line.

![Figure 1.4 Fixed Money Wages and Excess Labor Supply](image)

**Figure 1.4** is a graphic representation of the results of a decrease in government spending. As before, we start from the observation point $A$ and assume a decrease in government spending that moves the aggregate demand curve to the left. The resulting excess supply of goods causes price to decrease, with the same shifts in the labour demand and the labour supply curves as were discussed above. The observation point becomes $B$ in both panels of Figure 1.4. The unemployment rate, which was zero, is now $BD/CD$. Unemployment has two components: layoffs, $AB$, plus increased participation in the labour force, $AD$.

The short-run aggregate supply curve in the Keynesian model is positively sloped, and this is why the model does not display the classical dichotomy results (that is, why demand shocks have real effects). The reader can verify that the aggregate supply curve’s slope is positive, by taking the total differential of the key equations (1.13) and (1.4) while imposing the assumptions that wages and the capital stock are fixed in the short run (that is, by setting $d\bar{W} = dK = 0$). After eliminating the change in employment by substitution, the result is the expression for the slope of the aggregate supply curve:

$$\frac{dP}{dY} = -\frac{P_{F_S}}{\bar{W}^3} > 0.$$  
(1.7)

The entire position of this short-run aggregate supply curve is shifted up and in to the left if there is an increase in the wage rate (in symbols, an increase in $\bar{W}$). A similar change in any input price has the same effect. Thus, for an all-using economy, an increase in the price of oil causes stagnation — a simultaneous increase in both unemployment and inflation.

Additional considerations can be modelled on the demand side of the labour market as well. For example, if we assume that there is monopolistic competition, the marginal-cost-equals-marginal-revenue condition becomes slightly more complicated. Marginal cost still equals $W / F_S$, but marginal revenue becomes equal to $\frac{1}{n(l/n)e}P$, where $n$ and $e$ are the number of firms in the industry (economy) and the elasticity of the demand curve for the industry’s (whole economy’s) output. In this case, the number of firms becomes a shift influence for the position of the demand curve for labour. If the number of firms rises in good times and falls in bad times, the corresponding shifts in the position of the labour demand curve (and therefore in the position of the goods supply curve) generate a series of booms and recessions. And real wages will rise during the booms and fall during the recessions (that is, the real wage will move pro-cyclically). But this imperfect-competition extension of the standard textbook Keynesian model is rarely considered, as a result, the following summary is what has become conventional wisdom.

Unemployment occurs in the Keynesian model because of wage rigidity. This can be reduced by any of the following policies: increasing government spending, increasing the money supply, or reducing the money wage (think of an exogenous decrease in wages accomplished by policy as the static equivalent of a wage guidelines policy). These policy propositions can be proved by verifying that $d\bar{W}/dG, d\bar{W}/dM > 0$ and that $d\bar{W}/dP < 0$. Using more everyday language, the properties of the perfect-competition version of the rigid-money-wage model are:

1. Unemployment can exist only because the wage is "too high".
2. Unemployment can be lowered only if the level of real incomes of those already employed (the real wage) is reduced.
3. The level of the real wage must correlate inversely with the level of employment (that is, it must move countercyclically).

Intermediate textbooks call this model the Keynesian system. Many economists who regard themselves as Keynesians have a difficult time accepting these three propositions, however. They know that Keynes argued, in Chapter 19 of *The General Theory*, that large wage cuts might have only worsened the Depression of the 1930s. They feel that unemployment stems from some kind of market failure, so it should be possible to help unemployed workers without hurting those already employed. Finally, they have observed that there is no strong countercyclical movement to the real wage, indeed, it often increases when employment increases (see Solon et al (1993) and Huang et al (2004)).

1.5 Generalized Disequilibrium: Money-Wage and Price Rigidity

These inconsistencies between Keynesian beliefs on the one hand and the properties of the textbook (perfect competition version of the) Keynesian model on the other suggest that Keynesian economists must have developed other models that involve more fundamental departures from the classical system. One of these departures is the generalization of the notion of disequilibrium to apply beyond the labour market, a concept pioneered by Barro and Grossman (1971) and Malinvaud (1977).

If the price level is rigid in the short run, the aggregate supply curve is horizontal. There are two ways in which this specification can be defended. One becomes evident when we focus on slope expression (1.7). This expression equals zero if $F_{PS} = 0$. To put the point verbally, the marginal product of labour is constant if labour and capital must be combined in fixed proportions. This set of assumptions — rigid money wages and fixed-coefficient technology — is often appealed to in defending fixed-price models. (Note that these models are the opposite of supply-side economics since with a horizontal supply curve, output is completely demand-determined, not supply-determined.)

Another defence for price rigidity is simply the existence of long-term contracts fixing the money price of goods as well as factors. To use this interpretation, however, we must re-derive the equations in the macro model that relate to firms, since if the goods market is not clearing, it may no longer be sensible for firms to set marginal revenue equal to marginal cost. This situation is evident in Figure 1.5, which shows a perfectly competitive firm facing a sales constraint. If there were no sales constraint, the firm would operate at point $A$, with marginal revenue (which equals price) equal to marginal cost. Since marginal cost = $\bar{W}(dN/dY)$, this is the assumption we have made throughout our analysis of Keynesian models up to this point. But if the market price is fixed for a time (at $P$) and aggregate demand falls so that all firms face a sales constraint (sales fixed at $Y$), the firm will operate at point $B$. The marginal revenue schedule now has two components: $PB$ and $BD$ in Figure 1.5. Thus, marginal revenue and marginal cost diverge by amount $BC$.

![Figure 1.5 A Competitive Firm facing a Sales Constraint](image)
We derive formally the factor demand equations in Chapter 4——both those relevant for the textbook Classical and textbook Keynesian models (where there is no sales constraint), and those relevant for this generalized disequilibrium version of Keynesian economics. Here, we simply assert the results that are obtained in the sticky-goods-price case. First, the labour-demand curve becomes a vertical line (in wage-employment space). The corresponding equation is simply the production function — inverted and solved for \( N \). labour demand is whatever solves the production function after the historically determined value for the capital stock and the sales-constrained value for output have been substituted in. The revised investment function follows immediately from cost-minimization. Firms should invest more in capital whenever the excess of capital’s marginal product over labour’s marginal product is bigger than the excess of capital’s rental price over labour’s rental price. Using 5 to denote capital’s depreciation rate, the investment function that is derived in Chapter 4 is

\[
I = \delta (F_N (W / P)) / (F_Y (r + \delta)) - 1 \tag{1.8}
\]

The model now has two key differences from what we labelled the textbook Keynesian model. First, labour demand is now independent of the real wage, so any reduction in the real wage does not help in raising employment. Second, the real wage is now a shift variable for the IS curve, and therefore for the aggregate demand curve for goods, so wage cuts can decrease aggregate demand and thereby lower employment. (This second point is explained more fully below.) These properties can be verified formally by noting that the model becomes simply equations 1.1 to 1.3 but with \( W \) and \( P \) exogenous and with the revised investment function replacing \( \delta r \). The three endogenous variables are \( Y, r, \) and \( N \), with \( N \) solved residually by equation 1.3.

The model is presented graphically in Figure 1.6. The initial observation point is \( A \) in both the goods and labour markets. Assume a decrease in government expenditure. The demand for goods curve moves left so firms can only sell \( Y \); the labour demand curve becomes the \( N \) line, and the observation point moves to point \( B \) in both diagrams. Unemployment clearly exists. Can it be eliminated? Increases in \( M \) or \( G \) would shift the demand for goods back, so these policies would still work. But what about a wage cut? If the \( W \) line shifts down, all that happens is that income is redistributed from labour to capitalists (as shown by the shaded rectangle). If capitalists have a smaller marginal propensity to consume than workers, the demand for goods shifts further to the left, leading to further declines in real output and employment. The demand for goods shifts to the left in any event, however, since given the modified investment function (equation 1.8), the lower wage reduces investment. Thus, wage cuts actually make unemployment worse.

**Figure 1.6 The Effects of Falling Demand with Fixed Wages and Prices**

Some Keynesians find this generalized disequilibrium model appealing since it supports the proposition that activist aggregate demand policy can still successfully cure recessions while wage cuts cannot. Thus the unemployed can be helped without taking from workers who are already employed (that is, without having to lower the real wage).

However, the prediction that wage cuts lead to lower employment requires the assumption that prices do not fall as wages do. In Figure 1.6, the reader can verify that if both the given wage and price lines shift down (so that the real wage remains constant), output and employment must increase. The falling price allows point \( B \) to shift down the dashed aggregate demand curve, so the sales-constrained level of output rises (sales become less constrained). Further, with a less binding sales constraint, the position of the vertical labour demand curve shifts to the right.

Many economists are not comfortable with the assumption that goods prices are more sticky than money wages. This discomfort forces them to downplay the significance of the prediction that wage cuts could worsen a recession, at least as shown in generalized disequilibrium models of the sort just summarized. For this reason, we do not pursue these models further in this text. (For a brief but excellent survey of this literature, we refer the reader to Stoneham 1979.)

**1.6 Conclusions**

In this chapter we have reviewed Keynesian and Classical interpretations of the goods and labour markets. We have established, among other points, that unemployment can exist only in the presence of some stickiness in money wages. Appreciation of this fact naturally leads to the question: should we advocate increased wage flexibility so as to avoid at least some unemployment? There are two general responses. First, one can say that private agents must have adopted the institution of temporarily rigid wages (contracts) for a reason and that reason must be understood before one can be confident that increasing wage flexibility is “good.” Second, one can presume that the microeconomic costs of increased flexibility would not be large and, therefore, directly proceed with a macroeconomic analysis of whether the built-in stability of the overall economy is enhanced by wage flexibility. In section 1.4, we reviewed one argument against this possibility. The model of generalized disequilibrium supports the proposition that wage cuts could worsen a recession.

We shall consider other ways in which standard models have struggled to generate conclusions that are more in line with Keynes’s concerns. In particular, in Chapter 3, we shall consider another mechanism — expectations — whose presence can make wage cuts during recessions undesirable. In addition, we consider market failure in labour markets in Chapter 8. In this setting, involuntary unemployment emerges as a well-explained equilibrium outcome, and multiple equilibria can exist (so that the economy can become stuck at a low level equilibrium, just because agents become pessimistic and expect such an outcome). Many economists feel that these results provide a more convincing underpinning for some important Keynesian ideas that are usually ignored in textbook treatments of the “Keynesian” model.

Despite the fact that many mainstream economists have used modern tools to highlight some of Keynes’s most central concerns, some economists (known as Post Keynesians) still reject much of this “New Keynesian” analysis. One Post Keynesian concern is that mainstream analysis treats uncertainty in a way that Keynes argued was silly. Keynes followed Knight’s suggestion that risk and uncertainty were fundamentally different. Risky outcomes can be dealt with by assuming a stable probability distribution of outcomes, but some events occur so infrequently that the relevant actuarial information is not available. According to Post Keynesians, such truly uncertain outcomes simply cannot be modelled formally.

Another concern of Post Keynesians is that there are fundamental inconsistencies in defining aggregate capital stock that are ignored in standard analysis (see Cohen and Harcourt 2003)). Mainstream macroeconomists argue that the empirical success that economists have achieved in testing aggregate production relationships must mean that the empirical relevance of this capital controversy is limited. However, Shin (1974) has demonstrated that these tests have very low power.

An excellent general summary and introduction to the Post Keynesian approach is contained in Wollson (1994), and interested readers are encouraged to pursue this reference. Given our objective of providing a text that focuses on what is usually highlighted in a one-semester course, we cannot afford to consider Post Keynesian analysis further in this book. Instead, we focus on the New Classical revival, and we cover the research of “New Keynesians” in later chapters.

---

Muhammad Firman (University of Indonesia - Accounting)
CHAPTER 2
THE FIRST NEOClassical SYNTHESIS

2.1 Introduction

The traditional (that is, pre New Classical) analysis of economic cycles involved a compact structure that included the textbook Classical and Keynesian models as special cases. This simple – yet encompassing – framework was achieved by dropping any explicit treatment of the labour market (and the production function). Instead, a single summary relationship of the supply side of the goods market was specified. That one function was an expectations-augmented Phillips curve – a relationship that imposes temporary rigidity for goods prices in the short run, but Classical-dichotomy (natural-rate) features in full equilibrium. This simple, but complete, model of simultaneous fluctuations in real output and inflation consisted of two equations: a Phillips curve (the supply-side specification), and a summary of IS-LM theory – a simple reduced-form aggregate-demand function. The purpose of this chapter is to review the properties of this standard dynamic model.

2.2 A Simple Dynamic Model – Keynesian Short-Run Features and a Classical Full-Equilibrium

As just noted, traditional dynamic analysis combined a simple aggregate demand function and a Phillips curve, and expectations were ignored. The aggregate demand function was a summary of the IS-LM system. To proceed in a specific manner, we assume the following linear relationships:

\[
\begin{align*}
 y &= \alpha r + \beta g, & & \text{the IS function, and} \\
 m - p &= \gamma y - \delta r, & & \text{the LM relationship.}
\end{align*}
\]

where \( y, g, m, p, \) and \( r \) denoted the natural logarithms of real output, autonomous expenditure (sometimes assumed to be government spending), the nominal money supply, and the price level, respectively. The interest rate, \( \delta \) is the level (not the logarithm) of the interest rate, and since (for this chapter) we assume that expectations of inflation are always zero, \( r \) is both the real and nominal interest rate. The Greek letters are positive slope parameters.

These IS and LM relationships can be combined to yield the aggregate demand function (by eliminating the interest rate via simple substitution). The result is

\[
y = \Theta (m - p) + \xi g
\]

where \( \Theta = \alpha / (\alpha + \gamma) \) and \( \xi = \beta \Omega (\alpha + \gamma) \).

This aggregate demand function is combined with a standard dynamic supply function (a Phillips curve) in which the "core" inflation rate is assumed to be zero.

\[
p = \Phi (y - \bar{y}) + \pi.
\]

The new notation, \( \bar{y} \) and \( \pi \) denote the natural rate of output (the value that emerges in the textbook Classical model, and a value we take as an exogenous variable in the present chapter) and the core inflation rate. Since \( p \) is the logarithm of the price level, its absolute time change equals the percentage change in the price level. Thus, \( \pi \) is the inflation rate. Initially, the core inflation rate is assumed to be zero. Later on in the chapter, the core inflation rate is assumed to equal the full-equilibrium inflation rate, and since we assume a constant natural rate of output, this core inflation rate is simply equal to the rate of monetary expansion: \( \pi = \dot{m} \). If we assume the rate of monetary expansion to be zero, there is no difference between these specifications.

The full equilibrium properties of this system are: \( y = \bar{y}, p = \pi = \dot{m} \) and \( \pi = (\beta g - \gamma) / \gamma \), so (as already noted) macroeconomists talk in terms of the "natural" output rate, the "natural" interest rate, and the proposition that there is no lasting inflation-output trade-off. Milton Friedman went so far as to claim that inflation is "always and everywhere" a monetary phenomenon, but this assertion is supported by the model only if prices are completely flexible (that is, if parameter \( \phi \) approaches infinity). In general, the model involves simultaneous fluctuations in real output and inflation, bringing predictions such as disinflation must involve a temporary recession. Such properties imply that Friedman’s claim is accurate only when comparing full long-run equilibria. Nevertheless, the presumption that the model’s full equilibrium is, in fact, reached as time proceeds, should not be viewed as terribly controversial, since it turns out that this model’s stability condition can be violated in only rather limited circumstances.

Keynes’ approach to macroeconomics involved the concern that convergence to a classical full-equilibrium should not be presumed. Indeed, Keynes argued that a central job for macro theory was to identify those circumstances when convergence is unlikely, so that policy can be designed to ensure that real economies do not get into these circumstances. So while this traditional model involves sticky prices in the short run (and from this vantage point, at least, it is appealing to Keynesians), the fact that – when expectations are ignored – it rejects the possibility of instability as rather unlikely makes it offensive to Keynesians. How has this model been altered to avoid this offensive feature? The answer: by letting expected inflation depend on actual inflation, and by allowing these expectations to have demand-side effects. Thus far, we have limited the effects of anticipated inflation to the wages/price setting process (by allowing the core, or full-equilibrium, inflation rate to enter the Phillips curve). As an extension, we can allow the nominal and the real interest rates to differ by peoples’ expectations concerning inflation in the short run. But before we introduce this distinction, we discuss stability in this initial, more basic, model.

Mathematically, we can focus on the question of convergence to full equilibrium by taking the time derivative of the aggregate demand equation, assuming that autonomous spending and the money supply are not changing in an ongoing fashion (setting \( \gamma = m = 0 \)), and substituting out \( p \) by using the Phillips curve (with \( \pi = 0 \)). The result is

\[
\dot{y} = -\xi (\phi (m - p) - \pi).
\]

where \( \xi = \Phi \). For the convergence of actual real output to the natural rate, we require that \( y \) rise when it is “too low” and that \( y \) fall when it is “too high.” These outcomes are consistent with equation (2.3) only if parameter \( \xi \) is positive. Thus, the model’s stability condition is \( \xi > 0 \). Since summary parameter \( \xi \) is defined as \( \xi = \alpha / (\alpha + \gamma) \), we see that, in general, instability is impossible. The only problem that can develop is if the aggregate demand curve is not negatively sloped. It can be vertical if \( \xi \) is zero, and this (in turn) is possible if the economy gets into what Keynes called a “liquidity trap.” If the nominal interest rate approaches its lower bound of zero, the demand for money becomes limited only by agents’ wealth, and the interest elasticity of money demand approaches infinity (\( \Omega \rightarrow \infty \)). This situation is sufficient to make both \( \xi \) and \( \pi \) equal to zero. In short, the system does not converge to full employment in this case. Observing that interest rates were essentially zero in the United States during the 1930s, and in Japan during the 1990s, many analysts have argued that the economy’s self-correction mechanism breaks down in these cases. We certainly did observe very protracted recessions during these episodes (even the Great Depression in the 1930s case), so it may well be appropriate to interpret these periods in this manner. Some recent papers that examine the implications of a liquidity trap in some detail are Svensson (2003), Bernanke and Reinhart (2004), Eggertsson and Woodford (2004) and Coenen et al. (2006).

Figure 2.1 Short-Run and Long-Run Equilibria

Figure 2.1 illustrates the convergence to full equilibrium in the “normal” (non-liquidity-trap) case. The long-run aggregate supply curve is vertical at the natural rate of output – reflecting the fact that – in full equilibrium – this model coincides with the textbook Classical system. But the
Keynesian feature is that the price level is predetermined at each point in time, so the instantaneous, short-run aggregate supply curve is horizontal at that height. Normally, the aggregate demand curve is negatively sloped (and the shift variables are the nominal money supply and the level of autonomous spending). We consider a once-for-all drop in exogenous spending. The aggregate demand curve shifts to the left, and the economy moves from point A to B instantly. Output is completely demand-determined in the instantaneous short run. But then, as time begins to pass, prices begin to fall, and the short-run supply curve moves down to reflect this fact. The economy traces along the B-to-C path as output comes back up to the natural rate (as point C is reached). The recession is only temporary. But this benign conclusion does not emerge if the aggregate demand curve is very steep (or vertical). If it is very steep, and it moves significantly to the left, then the price level will fall to zero before the economy gets back to the natural output rate. Keynesian economists argue that we should not downplay this non-convergence-to-full-employment possibility. More classically minded economists are comfortable interpreting this possibility as just a remotely relevant pathology. In the next section of this chapter, we see how allowing inflationary expectations to play a role in this model makes instability a much more realistic possibility. This is why Keynesian economists always stress expectations.

2.3 The Correspondence Principle

We now extend our simple dynamic aggregate supply and demand model by allowing inflationary and deflationary expectations to affect aggregate demand. We continue to assume descriptive behavioural equations, leaving the consideration of formal micro-foundations until Chapter 4. We now distinguish real and nominal interest rates. The former is involved in the IS equation, as we assume that households and firms realize that it is the real interest rate that represents the true cost of postponing consumption and borrowing. But it is the nominal interest rate that belongs in the LM equation, as long as we assume that people’s portfolio choice is between non-indexed “bonds” (that involve a real return of \( r = i - \pi \)) and money (that involves a real return of \(-\pi\)). The real yield differential is, therefore, \( i - \pi \) the nominal interest rate. Notice that, to avoid saying to specify a relationship between actual and expected inflation, we have, simply assumed that they are equal. We consider alternative specifications in Chapter 3. In any event, when the nominal interest rate is determined by substitution, the IS-LM summary is

\[
y = \theta(m + \psi) + \xi g.
\]

(2.1a)

Two of the summary aggregate-demand parameters have already been defined earlier in the chapter. The coefficient on the new term is

\[
\psi = \sigma \Omega / (\gamma + \Omega).
\]
with sign ambiguities in essentially all macro models that are more complicated than this one. Macroeconomists have responded to this problem in three ways. First, on the basis of empirical work, theorists have become more confident in making quantitative assumptions about the model's parameters, not just qualitative (or sign) assumptions. But given the controversy that surrounds most econometric work, this strategy has somewhat limited appeal. The second approach is to provide more explicit micro-foundations for the model's behavioral equations. By having a more specific theory behind these relationships, we have more restrictions on the admissible magnitudes for these structural coefficients. While this approach limits the model's sign ambiguity problems, as we shall see in Chapter 3, it does not fully eliminate them. Thus, some reliance must remain on what Paul Samuelson called the correspondence principle many years ago. He assumed that the least controversial additional assumption that can be made concerning the model's parameters (other than their signs) is to assume that — given infinite time — the system will eventually converge to its full equilibrium. After all, most economists presume that we eventually get to equilibrium. To exploit this belief, Samuelson's recommendation was to derive system's dynamic-stability condition, and then to use that condition as a restriction to help sign the corresponding comparative static multipliers. Since macroeconomists are assuming eventual convergence implicitly, Samuelson felt that nothing more of substance is being assumed when that presumption is made more explicit to sign policy multipliers. This has been standard procedure in the profession for 60 years, and we will apply the correspondence principle in our analysis here. But before doing so, we note that some macroeconomists regard the use of the correspondence principle as suspect.

The dissenters can see that there is an analogy between macroeconomists using the correspondence principle and microeconomists focusing on second-order conditions. Microeconomists use the second-order conditions to resolve sign ambiguities in their analyses — that are based on agents obeying the first-order conditions. There is no controversy in this case, because the second-order conditions are an integral part of the model, and analysis are simply making implicit assumptions explicit. But the analogy between second-order conditions in micro and dynamic stability conditions in macro breaks down since, in most macro models, analysts have the freedom to specify more than one set of assumptions for the model's dynamics. Thus, there is an element of arbitrariness in macro applications of the correspondence principle that is not present when macroeconomists rely on second-order conditions for additional restrictions. One of the purposes of providing explicit macro-foundations for macroeconomics is to discipline macro model builders so that they have less opportunity for making what others might regard as arbitrary assumptions.

A more fundamental problem with the correspondence principle is that some economists (for example, Keynes) are not prepared to assume stability. Indeed, some of them can be viewed as arguing that this issue should be the fundamental focus of research (see Tobin 1975, 1980; and Hahn and Solow 1986). According to this approach, we should compare the stability conditions under alternative policy regimes, to see whether or not a particular policy is a built-in stabilizer. Policy regimes that lead to likely instability should be avoided. Thus, even though the stability conditions are not presumed to hold by all analysts, all economists must know how to derive these conditions. Thus, we now consider the stability condition for our aggregate supply and demand model.

The stability of the economy is assessed by taking the time derivative of equation (2.4) and using equation (2.2) to, once again, eliminate the inflation rate. In this derivation, we assume that there are no further (ongoing) changes in autonomous spending and that the natural rate is constant (\( g = \gamma = 0 \)). As in the simpler model which suppressed the distinction between real and nominal interest rates, the result is \( y = -\theta y - \theta ) \), but now the expression for the stability parameter is

\[ s = \frac{\theta \phi}{(1 - \phi) } . \]

As before, stability requires that \( s \) be positive; in this case, we require \( \phi < 1 \). Applying the correspondence principle, we see that this restriction is sufficient to guarantee that the expenditure multiplier has its conventional sign. This restriction is the familiar requirement that the income sensitivity of total spending not be "too large." In this case, the income sensitivity comes indirectly through the dependence of aggregate demand on the expected rate of inflation, and through the dependence of inflation on the output gap.

2.4 Can Increased Price Flexibility be De-Stabilizing?

We are now in a position to assess whether increased price flexibility leads to more desirable responses to aggregate demand shocks. There are both "good news" and "bad news" aspects to an increase in the size of parameter \( \phi \). The good news is that the economy's speed of adjustment back to full equilibrium (following all shocks) is higher (since the speed coefficient, \( s \), rises with \( \phi \)). There are two dimensions to the "bad" news. First, the size of the initial recession is larger; second, the likelihood of outright instability is increased. In terms of this standard model, it appears that Keynes expected the "bad" news effects to outweigh the "good" news effect. Should we agree?

Before proceeding, we should consider simple intuition. How is it possible that more flexible prices can enlarge a recession? The logic of this result runs as follows. Given that aggregate demand depends on both autonomous expenditure and inflationary expectations, which are perfectly anticipated, a decrease in autonomous expenditure has both a direct and an indirect effect. The direct effect leads to lower output. The indirect effect follows from the fact that agents realize that the fall-off in output will reduce inflation, other things being equal, this decrease raises the real interest rate, so firms reduce the investment component of aggregate demand. With increased price flexibility, this secondary effect is larger than it would be otherwise. So the Keynesian proposition that increased price flexibility has this "bad" aspect is supported by this analysis which stresses expectations effects.

The other "bad" aspect of higher price flexibility is that it can make the economy unstable. Some would argue that there is some support for Keynes on this question as well. It is widely presumed that we have had longer-term wage contracts since WWII, and this is captured in our model by a smaller value for coefficient \( \phi \). Since many believe that Western economies have displayed less volatility since WWII, there would seem to be some evidence to support Keynes. However, Romer (1986) has noted that — over time — there have been important changes concerning how we measure GDP. When GDP for recent years is measured in the same way as we were constrained to measure this aggregate in pre-WWII days, we find that the economy is not less volatile. Thus, the jury is still out on this question. Nevertheless, we can pursue Keynes' conjecture more fully if we presume stability, and calculate the cumulative output loss following a once-for-all drop in demand.

As noted already, increased price flexibility is not all "bad" in this model. As long as the economy remains stable, we know that the speed with which the temporary recession is eliminated is proportional to the stability and speed parameter \( s \). We have already seen that increased price flexibility must raise \( s \) and so speed up the elimination of output disturbances.

Assuming eventual convergence, the solid line in the Figure 2.2 shows the output time path following a once-for-all reduction in autonomous spending. The dashed line indicates how output responds when parameter \( \phi \) is a larger coefficient (on the assumption that the system remains stable). The size of the initial recession is bigger, but that larger recession is eliminated more quickly. The output time path is closer to the natural output line in the latter part of the adjustment period (when prices are more flexible) and Classicalists often argue that this may be the more important consideration.

**Figure 2.2 Implications of Increased Price Flexibility**

![Figure 2.2 Implications of Increased Price Flexibility](image_url)
One calculation that supports the Classicals' interpretation is the total undiscounted output loss that follows a permanent decrease in aggregate demand. From a geometric point of view, this measure is the area between the output time path and the natural rate line in Figure 2.2. It can be calculated as \( f(t - t')dt \), which, as Buitler and Miller (1982) show, equals the initial output loss (the impact multiplier) divided by the speed of adjustment. Thus, in this case, the cumulative output loss is \( \xi^t_0 \Phi \). According to this method of weighting the "bad" short-run effect and the "good" long-run effect of a larger parameter \( \Phi \), then, an increased degree of price flexibility is desirable. Of course, supporters of Keynes can argue that what matters is a discounted cumulative output loss calculation - not what we have just calculated. Once the short run is given more weight than the longer run, it is immediately apparent from Figure 2.2 that the undesirable aspects of an increased degree of price flexibility could dominate. Overall, this analysis provides at least partial analytical support for the Keynesian proposition that increased price flexibility may not help the built-in stability properties of the economy. This issue is particularly important since many policy analysts advocate that governments use taxes and/or subsidies to stimulate private firms and their workers to adopt such arrangements as profit-sharing and shorter wage contracts. One motive for encouraging these institutional changes is a desire to increase wage flexibility (which would indirectly bring increased price flexibility) and the proponents of these policies simply presume that their adoption would be "good."

The opposite presumption seems to be involved at central banks, such as the Bank of Canada. Analysts there have noted that one of the "beneficial" aspects of our reaching price stability, is that average contract length has increased dramatically over the last 20 years. With low inflation, people are less prepared to sign longer wage contracts. While this means lower industrial-dispute and negotiation costs, it also means that the size of parameter \( \Phi \) is now smaller. This development is "good" for macroeconomic stability only if Keynes was right. It appears that the Bank of Canada is comfortable siding with Keynes on this issue.

Before leaving this section, we investigate what things make the economy’s stability condition more or less likely to be satisfied. We derived above that convergence to full equilibrium occurs only if \( \Phi < 1 \). To assess the plausibility of this condition being met, we focus on the detailed determinants of the reduced-form aggregate demand parameters given earlier. Using these interpretations, the stability condition can be re-expressed as:

\[
1 + (\xi \Phi / \Omega) > (\xi \Phi)
\]

We consider this convergence condition in two separate situations: an ongoing inflation, such as that of the 1970s, and a deep depression, such as that of the 1930s. In a period of inflation, the equilibrium nominal interest rate is high and the interest elasticity of money demand is very low. This situation can be imposed in the stability condition by letting parameter \( \Omega \) approach zero, and the result is that the stability condition must be satisfied. The opposite case involves extremely low interest rates as were observed throughout the world in the deep depression of the 1930s, and in Japan during its prolonged recession since the 1990s. In this situation the nominal interest rate approaches zero, many macroeconomists believe that the interest elasticity of money demand, parameter \( \Omega \), approaches infinity. This "liquidity trap" situation can make it very difficult for the stability condition to be satisfied.

So, under monetary aggregate targeting at least, the efficacy of the economy’s self-correction mechanism very much depends on the interest elasticity of money demand. The intuition behind this fact is straightforward. Lower prices have two effects on aggregate demand. Falling actual prices stimulate demand, and (other things equal) help end a recession. But expectations of falling prices raise real interest rates, and (other things equal) dampen demand and thereby worsen a recession. The stabilizing effect of falling actual prices works through its expansionary effect on the real money supply, while the destabilizing effect of expected deflation works through interest rates and the associated increase in money demand. If real money demand increases more than real money supply, the initial short-fall in the demand for goods is increased. A liquidity trap maximizes the chance that the money-demand effect is stronger – making the economy unstable.

We conclude that this simple dynamic model represents a compact system that allows for Keynes's worry that wage and price decreases can worsen a deep recession through expectation effects. At the same time, however, it is important to realize that the model does not suggest that instability must always occur. Indeed, macroeconomists can appeal to this one simplified model, and consistently argue that government intervention was justified in the 1930s (to avoid instability or at least protracted adjustment problems) but was not required in the 1970s (to avoid hyperinflation). Since any scientist wants a single, simple model to "explain" a host of diverse situations, it is easy to see why variants of this model represented mainstream macroeconomics for many years.

Some macroeconomists interpret the economy as having a stable "corridor." The term corridor has been used to capture the notion that the economic system may well be stable in the face of small disturbances that do not push the level of activity outside its normal range of operation (the corridor). But that fact still leaves open the possibility that large shocks can push the economy out of the stable range. For example, as long as shocks are fairly small, we do not get into a liquidity trap. But sometimes a shock is big enough to make this extreme outcome relevant, and the economy is pushed out of the stable corridor. This appears to be a reasonable characterization of the Great Depression of the 1930s. At that time, individuals became convinced that bank runs would be prevalent. As a result, they developed an extraordinary preference for cash, and the corresponding liquidity trap destroyed the applicability of self-correcting mechanism. It is important for Keynesians that instability does not always occur, so that Keynesian concerns cannot be dismissed by simply observing that there have been many episodes which have not involved macroeconomic breakdown. Howitt (1979) has argued that models that do not permit a corridor feature of this sort cannot claim to truly represent Keynes's ideas. While this discussion of corridors has been instructive, it has not been completely rigorous. After all, formal modeling of corridors would require nonlinear relationships, and our basic model in this chapter has involved linear relationships.

2.5 Monetary Policy as a Substitute for Price Flexibility

Throughout the previous section of this chapter, we have reasoned as if government policy could choose the degree of price flexibility. While some countries have institutional structures that might permit this, for many others, policy options are less direct. The government can only affect price flexibility indirectly, through such measures as tax incentives that might stimulate the use of profit sharing arrangements. The question which naturally arises is whether monetary policy – an instrument over which the government has much more direct control – can be used as a substitute for price flexibility. Keynes believed that monetary policy could be used in this way. Even Milton Friedman (1953) agreed; his advocacy of flexible exchange rates was based on the presumption that this monetary policy could substitute for flexible wages and prices.

To investigate this issue formally we examine nominal income targeting – a monetary policy advocated by many economists recently. Letting \( x = p \gamma \) denote the logarithm of nominal GDP, we can specify the following monetary policy reaction function:

\[
m = \bar{m} - \chi (p - p - y)
\]

(2.5)

where the bars denote (constant) target values and parameter \( \chi \) defines alternative policies. \( \chi \rightarrow 0 \) implies a constant money supply (what we have been assuming in the previous section); \( \chi \rightarrow \infty \) implies pegging nominal income. Since the target variable is the equilibrium inflation rate, the equation is meaningful only under the condition that the inflation rate is zero for both of these monetary policies. This policy reaction function can be combined with equations (2.1a) and (2.2). Using the methods already described, the reader can verify that the impact autonomous spending multiplier is

\[
dy / dy = \xi (1 - \Phi + \xi \Phi)
\]

(2.5)

the stability and the adjustment speed parameter is

\[
s = \Phi (1 + \gamma) (1 - \Phi + \xi \Phi)
\]

(2.5)

and the cumulative output loss is \( \xi (\Phi (1 + \gamma)) \). These results imply that nominal income targeting (an increase in parameter \( \chi \)) reduces the size of the impact effect which follows an aggregate demand shock, and it has an ambiguous effect on the speed with which this temporary output deviation is eliminated. These effects are not quite as simple as we obtained for an increase in price flexibility, but the net effect on the undiscounted cumulative output outcome is similar. The overall output effect is made smaller by nominal income targeting. In this sense, then, a more active attempt to target nominal income can substitute for an attempt to vary the degree of wage/price flexibility.

Readers may regard this analysis of monetary policy as somewhat dated. After all, central banks no longer set policy in terms of targeting monetary aggregates. Instead, they adjust interest rates with a view to
targeting inflation directly. Their research branches investigate whether their interest-rate reaction function should focus on the deviation of the inflation rate from target (the current practice), on the deviation of the price level from target, or on the deviation of nominal GDP from target. For the remainder of this section, we recast the analysis so that it can apply directly to this set of questions. Also, we consider a different specification of disturbances. Thus far, we have focused on the effects of a once-for-all change in demand. Now, we focus on an ongoing cycle in autonomous spending. Since modern analysis views policy as an ongoing process, not a series of isolated events, this alternative treatment of disturbances is appealing for an analysis that highlights model-consistent expectations on the part of private agents who understand that policy involves an ongoing interaction with the economy.

The revised model involves the IS relationship, the central bank’s interest-rate setting rule, the Phillips curve, and the specification of the ongoing cycle (a sine curve) in autonomous demand:

\[ y = -ar + \bar{y} + \bar{g}t \]
\[ r + \bar{p} = T + \bar{w} + A(p - \bar{p}) + (1 - \lambda)p(0) \]
\[ p = \bar{p}(y - \bar{y}) + 0 \]
\[ g = g + \delta \sin(t) \]

The LM relationship is not used, since its only function in this specification of monetary policy is to solve for the real value of the money supply is necessary for the central bank to deliver the nominal interest rate that is chosen according to the bank’s reaction function. We focus on a central bank that is committed to price stability. This is why there is a zero core-inflation term in the Phillips curve, and a zero average inflation rate term in the rate-setting equation. This relationship states that the bank sets the current nominal interest rate above (below) its long-run average value whenever either the inflation rate is above (below) its target value, or whenever the price level is above (below) its target level. As noted above, one major point of debate among central banks today is whether they should pursue an inflation-rate target or a price-level target. We consider this debate in this model by examining alternative values for parameter \( \lambda \). Inflation targeting is involved if \( \lambda = 1 \), while price-level targeting is specified if \( \lambda = 0 \). Parkin (2000) highlights the implications of these two policy options for the real value of contracts over time. Inflation targeting permits long-run drift in the price level, while price-level targeting does not. Thus, if the goal is to preserve the purchasing power of money, price-level targeting is preferred. Here, we focus on the implications of this choice for the economy’s short-run built-in-stability properties.

We proceed by deriving the reduced form for real output, to see how the amplitude of the resulting cycle in \( y \) is affected by changes in the monetary policy parameter \( \lambda \). We choose units so \( \bar{y} = 0 \), and substitute the Phillips curve and interest-rate-setting relationships into the IS function, to eliminate inflation and the interest rate:

\[ (1 - \alpha\phi(1 - \lambda))y = \alpha(1 - \lambda)p + \bar{g}r - \lambda g \]

Next, we take the time derivative and use the Phillips curve again to eliminate inflation:

\[ (1 - \alpha\phi(1 - \lambda))\dot{y} = -\alpha(1 - \lambda)p\dot{y} + \bar{g}\dot{g} \]

We analyze (2.6) by positing a trial solution and using the undetermined coefficient solution procedure. It is worth a brief aside to explain this procedure – to clarify that it is likely to be more familiar to economics students than they realize. Consider the familiar example of compound interest. The basic relationship is to solve (residually) for what value of interest is \( x = (1 + r)x \), where \( r \) is the interest rate and \( x \) the accumulated value. We know that the solution equation for this dynamic process is \( x = (1 + r)x \), where \( x \) is the initial amount that is invested. Let us pretend that we do not know that this is the answer. To derive this solution equation, we simply posit a trial solution of the form \( x = \mu A \), where the arbitrary parameters, \( \mu \) and \( A \), are yet to be determined. Substituting the trial solution into the model, we have: \( \mu = (1 + r)\mu A \) or \( \mu = (1 + r) \). Similarily, substituting \( t = 0 \) into the trial solution, we have \( A = \mu_0 \). As a result, the initially arbitrary reduced form coefficients, \( \mu \) and \( A \), are now determined as functions of the economically meaningful parameters, \( r \) and \( x_0 \).

We use the same procedure for our macro model here. Following Chiang (1984, p. 472) the (trial) solution for output can be written as

\[ y = \bar{y} + B(\cos t) + C(\sin t) \]

\[ \dot{y} = -B\sin t + C(\cos t) \]

and the time derivative of the autonomous spending equation, \( \dot{g} = \delta(\cos t) \) is substituted into (2.6). The resulting coefficient identifying restrictions are

\[ B = [\alpha(1 - \lambda)]T[1 - \alpha(1 - \lambda)] \]
\[ C = [\beta(1 - \alpha(1 - \lambda))]^2(1 - \alpha(1 - \lambda)^2) \]

Even for this simple model, illustrative parameter values are needed to assess the resulting amplitude of the cycle in \( y \). Representative values (considering a period length of one year) are: \( \alpha = 0.8 \) and \( \phi = 0.2 \) (autonomous spending is 20% of GDP), \( \delta = 1 \), and \( \beta = 0.2 \) (see Walsh (2003a)). With these values, the amplitude of the real output cycle is equal to one-fifth of the amplitude of the autonomous spending cycle if the central bank targets the inflation rate (if \( \lambda = 1 \)). In contrast, the amplitude of the output cycle is a slightly larger fraction (0.23 instead of 0.20) of that of the autonomous spending cycle if the central bank targets the price level (if \( \lambda = 0 \)). According to our model, the effect of a change in inflation targeting to price-level targeting is not supported.

The intuition behind this result can be appreciated by considering an exogenous increase in the price level. With inflation-rate targeting, such a “bygone” outcome is simply accepted, and future inflation is resisted. But with price-level targeting, future inflation has to be less than zero to eliminate this past outcome. That is, only under price-level targeting is a policy-induced recession called for. So price-level targeting seems obviously “bad.” The reason that this consideration may not be the dominant one, however, is that the avoidance of any long-term price-level drift (that is a feature of price-level targeting) has a stabilizing effect on expectations. For the plausible parameter values considered here, it appears that the former (destabilizing) effect of price-level targeting slightly outweighs the latter (stabilizing) effect.

2.6 Conclusions

The analysis of this chapter has involved a simple model that summarizes mainstream macroeconomics before the rational-expectations and new-classical “revolutions”. This version of macroeconomics is called the Neoclassical Synthesis since it combines important elements of both Classical and Keynesian traditions. It involves both the long-run equilibrium defined by the static model of the Classical model, and the temporary nominal stickiness feature that is the hallmark of the textbook Keynesian model (as the mechanism whereby the system departs from its full equilibrium in the short run). We have used this model to explain the correspondence principle, to examine how several monetary policies might be used to make up for the fact that prices are not more flexible, and to establish whether more flexible prices are desirable or not.

We have learned how important it is to add expectations to a macro model. Initially, expectations was an issue stressed by Keynesians, since it represents a mechanism that makes convergence to full equilibrium less assured. But, more recently, macroeconomists of all persuasion highlight expectations in their analysis. This is because stabilization policy is now modelled as an ongoing operation, not an isolated one-time event, and analysts are drawn to models in which agents are fully aware of what the government has been, and will be, doing. Thus, as far as stabilization policy analysis is concerned, there has been a convergence of views in the sense that all modern analysis focuses on model-consistent expectations (as we did by equating actual and expected inflation in this chapter).

We extend our appreciation of these issues in the next two chapters, by exploring alternative treatments of expectations in the next chapter, and by providing more explicit micro-foundations for the synthesis model in Chapter 4. Once the rational-expectations (Chapter 3) and the new-classical (Chapter 5) revolutions have been explored, we will see in a position to consider an updated version of the synthesis model – so-called “New Neoclassical Synthesis” – in Chapters 6 and 7.
CHAPTER 3
MODEL-CONSISTENT EXPECTATIONS

3.1 Introduction
As noted in Chapter 1, early work in macroeconomics involved a bold simplifying assumption—that economic agents have static expectations concerning the model’s endogenous variables. This assumption facilitated the separation of the “stage 1” and “stage 2” analyses. The unappealing thing about this approach is that individuals are always surprised by the fact that these variables change. By 1970, macro theorists had come to regard it as unappealing to model households and firm managers as schizophrenic individuals. On the one hand, we assumed that individuals took great pains to pursue a detailed plan when deciding how much to consume and how to operate their firms (focusing on points of tangency between indifference maps and constraints). However, in that traditional approach, we assumed that these same individuals were quite content to just presume that many important variables that affect their decisions will never change. Why is it appealing to assume that individuals are content to be so consistently wrong on this score, while they are so careful making detailed plans in other dimensions of their decision making? And why should economists be content to build models which explain the time paths of these variables and yet involve individuals who apparently do not believe in the relevance of these models (since they do not use them to help them foresee what will happen)?

In fact, modern macroeconomists have become quite dissatisfied with this procedure that had been traditional, and they now limit their attention to models containing individuals whose expectations are consistent with the model itself. This is the rational expectations approach (which is also called the perfect foresight approach, if stochastic disturbances are not involved).

As far as the historical development of the subject is concerned, we can identify four approaches to modelling expectations:

- **Static Expectations**: individuals are always surprised by any changes, and so they make systematic forecast errors. This early approach had the advantage of simplicity (so that model solutions could be derived more easily) but it had the disadvantage that we assumed irrational agents. After all, when errors are systematic, it should be straightforward to do better.

- **Adaptive Expectations**: individuals forecast each endogenous variable by assuming that the future value will be a weighted average of past values for that variable—with the weights summing to unity and getting smaller as ever earlier time periods are considered. This hypothesis was “invented” by Cagan in 1956, and it was made popular by Friedman since it played a central role in his permanent-income theory of consumption. This approach had the advantage that it did not complicate solution procedures too much, and it involved long-run consistency (in the sense that forecasters eventually “get it right”). For example, consider a situation in which inflation doubles. Under adaptive expectations, expected inflation eventually doubles. However, all during the (infinite) time it takes for forecasters to reach this final adjustment, they underpredict inflation. In other words, this hypothesis still involves systematic forecast errors, so it does not involve short-run consistency.

- **Perfect Foresight**: individuals are assumed to be so adept at revisiting their forecasts in the light of new information concerning the economy’s evolution that they never make any forecast errors. Economists (and other scientists) often find it useful to “bucket the truth” by considering polar cases. For example, economists focus on the extremes of perfect competition and pure monopoly—knowing that most real-world firms find themselves in industries that are in between these two extremes. Nevertheless, since formal analysis of these intermediate cases is difficult, we learn a lot by understanding the polar cases. The hypothesis of perfect foresight has some appeal, since it is the opposite polar case to static expectations. This hypothesis was involved in our treatment of the first Neoclassical Synthesis analysis in Chapter 2.

- **Many economists find it appealing to adopt an hypothesis that is between the polar-case extremes of perfect foresight and static expectations, and in a sense, that is what adaptive expectations is. After all, perfect foresight coincides with adaptive expectations if the very recent past gets all the weight in the weighted average, and static expectations coincides with adaptive expectations if the infinitely distant past gets all the weight. But the adaptive expectation approach—with some weight given to all time horizons—is still an unappealing way to achieve a compromise since it involves systematic forecast error. In addition, there is no reason for agents to limit their attention to just past values of the variable that they are trying to forecast. Why wouldn’t they use their knowledge of macroeconomics when forecasting? For example, suppose individuals were told that the central bank will double the money supply next year. Most people would raise their expectations of inflation when presented with this policy announcement. But people following the adaptive expectations approach would make no such change in their forecast, because the announcement concerning the future cannot change pre-existing outcomes (and the latter are the only items that appear in a backward-looking forecasting rule).

Rational Expectations—Like adaptive expectations, this hypothesis is an attempt to have an analysis that is between perfect foresight and static expectations. It is used in models that involve the economy being hit with a series of stochastic shocks, so agents cannot know everything. But the agents in the model understand the probability distributions that are generating the shocks, so they can form expectations in a purposeful manner. Under the rational expectations hypothesis, each agent’s subjective expectation for the value of each endogenous variable is the same as what we can calculate as the mathematical expectation for that variable—as we formally solve the model. Thus, agents make forecast errors (so this hypothesis is more “realistic” than perfect foresight) but those errors are not systematic. Perhaps a better name for this hypothesis is model-consistent expectations. In any event, most modern macro-economists find rational expectations a very appealing approach.

As noted above, the perfect foresight and rational expectations hypotheses coincide if the variance of the stochastic shocks shrinks to zero. Since the policy implications of two models that differ only in this dimension are often the same, macroeconomists often rely on perfect foresight analyses—despite the fact that this approach “sounds” more “unrealistic.”

We proceed through a series of analyses in this chapter. First, we introduce a stochastic disturbance term in an otherwise familiar macro model, by reconsidering the basic model of Chapter 2 when it is defined in a discrete-time specification. It turns out that since stochastic differential equations are very difficult to analyze and since stochastic difference equations are much simpler—macroeconomists switch to discrete-time specifications when they wish to stress incomplete information. We consider three aspects of uncertainty: situations in which economic agents have incomplete information concerning exogenous variables, situations in which agents (and the stabilization policy authorities) have incomplete knowledge concerning the model’s slope parameters, and situations in which the functional forms of important macroeconomic relationships are not known with certainty. Finally, in later sections of the chapter, we move on to adaptive and rational expectations analyses.

3.2 Uncertainty in Traditional Macroeconomics
In the last chapter, we focused on the economy’s response to a single one-time shock, and the resulting time path—a one-time departure from its natural rate, and then a monotonic return to full equilibrium. Our only analysis of ongoing business cycles, involved specifying that at least one exogenous variable does not follow a deterministic time path; instead, it is a stochastic variable. We begin this chapter by reconsidering the static-expectations model of Chapter 2—this time specified in discrete time, with a central bank that cannot perfectly control the size of the nation’s money supply at each point in time.

Ignoring the autonomous spending variable and assuming a zero core inflation rate for simplicity, the model can be defined by the following equations:

\[
\begin{align*}
y_t &= \delta((m_t - p_t) \text{ aggregate demand}) \\
(m_t - m_{t-1}) &= \chi(S_t - \gamma) + \eta_t \text{ monetary policy} \\
p_t &= p_{t-1} = \phi(y_t - \gamma) = \phi y_t \\
\end{align*}
\]

The “\(\phi\)" denotes the output gap. The monetary policy reaction function involves the money growth rate adjusting to the most recent observation on the output gap (if \(y\) is positive, not zero). Otherwise the money growth rate is a random variable, since we assume that \(\eta\) is a standard “error” term, with expected value of zero, no serial correlation, and a constant
variance. Among other things, the model can be used to assess whether “leaning against the wind” is a good policy (as Keynesians have always recommended), or whether (as Milton Friedman has long advocated) a “hands off” policy is better. We can evaluate this “rules vs. discretionary policy” debate by analyzing whether a zero value for parameter \( \chi \) leads to a smaller value for output variance (the variable the central bank is assumed to care about (given the policy reaction function)) or whether a positive value for parameter \( \chi \) delivers the smaller variance. But we postpone this policy analysis for the moment, by imposing \( \chi = 0 \). By taking the first difference of the demand function, and substituting in both the policy reaction function and the Phillips curve, we have:

\[
y_t = \nu_Y y_{t-1} + \theta_{u_t}
\]  
(3.1)

where \( \nu = (1 - \delta \phi) \). Aside from the ongoing error term, there are (in general) four possible time paths that can follow from a first-order difference equation of this sort (as shown in Figure 3.3). We observe an explosive if \( \nu > 1 \), asymptotic approach to full equilibrium if \( 0 < \nu < 1 \), damped oscillations if \( -1 < \nu < 0 \), and explosive cycles if \( \nu < -1 \).

The standard way of using the model to “explain” business cycles is to assume that \( 0 < \nu < 1 \), and that the stochastic disturbance term keeps the economy from ever settling down to full equilibrium. With these assumptions, the model predicts ongoing cycles.

It is useful to compare the stability conditions for continuous-time and discrete-time macro models. In the former, we saw that overshoots were not possible, so the stability condition is just a qualitative restriction (that the sign of parameter \( \nu \) be appropriate). But in discrete time, overshoots of the full equilibrium are possible, so the stability condition involves a quantitative restriction on the model’s parameters (that the absolute value of \( \nu \) be less than unity). To maximize the generality of their analyses, macro theorists prefer to restrict their assumptions to qualitative, not quantitative, restrictions. This fact clarifies why much of modern macro theory is specified in continuous time. But, as noted above, if stochastic considerations are to form the focus of the analysis, we must put up with the more restrictive stability conditions of continuous-time analysis, since stochastic differential equations are beyond the technical abilities of many analysts.

Now let us re-introduce ongoing policy, by considering the possibility of \( \chi > 0 \). In this case, \( \nu = (1 - \theta \phi + \chi) \). Is an increase in \( \chi \) a “good” thing? The answer appears to depend on whether \( \nu \) is positive or negative before the interventionist policy is introduced. If it is positive, interventionist policy makes parameter \( \nu \) smaller, and this is desirable since it either eliminates explosive behavior, or speeds up the asymptotic approach path to full equilibrium (if \( \nu \) were initially a positive fraction). But making \( \nu \) positive could make \( \nu \) become negative, and so the well-intentioned stabilization policy could create cycles in economic activity that did not exist without policy. In this case, interventionist policy is not recommended. Further, if \( \nu \) were already a negative fraction, policy could make \( \nu \) go on to become the minus one value, so that policy would have caused damped cycles to be replaced by explosive cycles. These possibilities support Friedman’s concern that even well-intentioned policy can be quite undesirable.

**Figure 3.1 Possible Time Paths for Output**

<table>
<thead>
<tr>
<th>( Y_t )</th>
<th>1. Direct Instability (( \nu &gt; 1 ))</th>
<th>2. Direct Convergence (( \nu &lt; 0 ))</th>
<th>3. Damped Cycles (( 1 &lt; \nu &lt; 0 ))</th>
<th>4. Explosive Cycles—not shown (( \nu &lt; 1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_0 )</td>
<td>Now ( Y )</td>
<td>Initial ( Y )</td>
<td>Time, ( t )</td>
<td></td>
</tr>
</tbody>
</table>

Another way of summarizing this possibility is by focusing on the asymptotic variance of the output gap. To derive this expression, we take the expectations operator, \( E_\phi \), through equation 3.1, using the assumption that – entering each period – the expected value of each error term is zero. The result is:

\[
E(Y_t) = \nu E(Y_{t-1}).
\]

When this relationship is subtracted from equation (3.1), and the outcome is squared, we have

\[
|Y_t - E(Y_t)|^2 = \nu^2 |Y_{t-1} - E(Y_{t-1})|^2 + 2\theta \nu^2 + 2\theta \nu |Y_{t-1} - E(Y_{t-1})| - E(Y_{t-1}).
\]

(3.3)

The variance of \( Y \) is now calculated by taking the expectations operator through equation (3.3). The result is

\[
\sigma^2_Y = \theta^2 \nu^2 (|1 - \nu^2|)^2 \sigma^2_u.
\]

(3.3)

It is important to clarify the information set upon which this expectation is based. We discuss two extreme assumptions. The first is that expectations are based on (\( t-1 \)) period information. If so, \( E_\phi(Y_t) = Y_{t-1} \), in equation (3.2) and so \( \sigma^2_Y = \sigma^2_u \). But this assumption is not appealing if we wish to consider the effects on the economy of a whole series of stochastic shocks buffeting the system through time (at any moment, the shocks from many periods continue to have some effect). To capture this ongoing uncertainty in the variance calculation, we need to assume that expectations for period \( t \) are based on information from period (\( t-\delta \)), where \( \delta \) is much larger than one. The convention is to calculate the asymptotic variance by letting \( \delta \) approach infinity. In this case, both \( Y_t - E(Y_t) \) and \( [Y_t - E(Y_t)]^2 \) equal \( \sigma^2_Y \), so equation (3.2) leads to (3.3).

Another way of expressing the calculation of asymptotic variance helps us appreciate this result, and why it is important. The asymptotic variance captures the ongoing effect on output of an entire series of short-run disturbances. It can be calculated by evaluating:

\[
E[\left[Y_{t-1} - E(Y_{t-1})\right]^2].
\]

An expression for \( \sigma^2_Y \) can be had by recursive substitution:

<table>
<thead>
<tr>
<th>( Y_{t-1} )</th>
<th>( Y_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( = \nu Y_{t-1} + \theta_{u_{t-1}} )</td>
<td>( = \nu Y_{t-1} + \theta_{u_{t-1}} )</td>
</tr>
<tr>
<td>( = \nu(\nu Y_{t-2} + \theta_{u_{t-2}}) + \theta_{u_{t-1}} )</td>
<td>( = \nu^2 Y_{t-2} + \theta_{u_{t-1}} + \theta_{u_{t-1}} )</td>
</tr>
<tr>
<td>( = \nu^2 Y_{t-2} + \theta_{u_{t-1}} + \theta_{u_{t-1}} )</td>
<td></td>
</tr>
</tbody>
</table>

Following this pattern, we have:

\[
Y_{t-1} = \nu^2 Y_{t-2} + \theta_{u_{t-1}} + \theta_{u_{t-2}} + \ldots + \theta_{u_0}.
\]

where

\[
\sigma^2_Y = \theta^2 \nu^2 (|1 - \nu^2|)^2 \sigma^2_u.
\]

As long as \( \nu \) is a fraction and \( n \) approaches infinity, we have \( E(Y_{t-1}) = 0 \), and

\[
\sigma^2_Y = \theta^2 (|1 - \nu^2|)^2 \sigma^2_u.
\]

Since an increase in the degree of policy intervention can magnify the coefficient that connects the variance of national output to the variance of the temporary shock process, it can reduce the degree to which the economy is insulated from a series of ongoing disturbances. We conclude that when stochastic considerations are combined with time lags, it is difficult to define what institutional arrangements provide the economy with a “built-in stabilizer.”

We can summarize as follows. The relevant coefficient for evaluating built-in stability is the face of a series of temporary shocks is \( 1/(1 - \nu^2) \), not the deterministic full-equilibrium multiplier that is familiar from intermediate macroeconomics, \( 1/(1 - \nu) \). This latter expression is relevant only for summarizing the effects of a one-time permanent shock. The coefficient that relates the variances is the only one that refers to the extent to which random shocks cumulate through the system. In the case of our demonstration model, we see that a more Keynesian policy (a less hands-off Friedman-like monetary policy) must lower \( 1/(1 - \nu) \), whether \( \nu \) is positive or negative, but such a policy may raise or lower the more relevant reduced-form coefficient \( 1/(1 - \nu^2) \). A more Keynesian policy decreases this coefficient if \( \nu \) is positive, but increases it if \( \nu \) is negative. This result threatens the standard idea that an interventionist central bank can do better than one that follows Friedman’s hands-off dictum. This conclusion is interesting in itself because the question of stabilization policy is important. But there is also a more general message. Simple intuition can be misleading when both time lags and the consideration of uncertainty are involved. There appears to be no substitute for an explicit treatment of these factors.

We now consider uncertainty in two different ways. First, we refer to Brainard’s (1967) more general analysis which allows both the intercept and slope parameters to be uncertain. Second, we examine the implications of the policy maker being uncertain about the functional form of a structural equation.
Brainard was the first to analyze a stochastic macro model in which the policy maker did not know the structural parameters with certainty. Brainard considered a stabilization policy authority that tries to minimize the expected deviation of output from its target value, \( E(Y - \bar{Y})^2 \), subject to the fact that output, \( Y \), is related to the policy instrument \( G \) by a simple macro model: \( Y = aG + u \). Both \( a \) and \( u \) are stochastic variables with expected values equal to 0 and with constant variances equal to \( \sigma_a^2 \) and \( \sigma_u^2 \), and with no serial or cross correlation.

After substitution of the constraints into the objective function, we have:

\[
E(\alpha G + \varepsilon G + u - \bar{Y})^2 = \alpha^2 \sigma_G^2 + \sigma_a^2 + \sigma_u^2 - 2\alpha \sigma_G \sigma_a.
\]

Differentiating this objective function with respect to the decision variable \( G \), we have the optimizing rule, and the optimal value for the policy variable that emerges is:

\[
G^* = \frac{\alpha}{\sigma_G / (\sigma_G^2 + \sigma_a^2)}.
\]

It is instructive to consider the limiting cases. If the policy multiplier coefficient is known with certainty (\( \sigma_a^2 = 0 \)), then the optimal value for the policy variable is the target value for output divided by the multiplier. But as uncertainty about the multiplier rises (that is, as the variance of \( \alpha \) approaches infinity) \( G \) should be set at zero. For intermediate values of the variance, it is best to set the policy instrument at a value that is somewhere between the “certainty equivalent” optimum and the “do nothing” value.

This formal analysis of uncertainty supports Friedman’s long-standing argument that policy makers should attempt less when they are unsure of the connection between policy and the economy’s performance. In this sense, then, the analysis supports rules “over discretion”. There have been a number of extensions of this analysis in recent years. For example, Basu et al (1980) apply this method to a situation in which a policy advisor is unsure about whether her advice will be followed. Others (Brock, Durlauf and West (2003), Majumdar and Mukand (2004), Favero and Milani (2005) and Wieland (2006)) explore why it may be optimal for the authority to depart temporarily from what would otherwise be the optimal stabilization policy, just so that it can learn more about the structure of the system, and so be less burdened with uncertainty in the future.

There is an alternative to modelling uncertainty in a stochastic framework. Uncertainty can take the form of the policy maker not knowing the curvature involved in a particular structural relationship. To establish the implications of this form of uncertainty, we now compare two models that are exactly the same except for the fact that the precise degree of curvature in one relationship, the Phillips curve, is not known. The first model is defined by the following three equations:

\[
\begin{align*}
\dot{y} &= \theta(m - \bar{p}) \\
\dot{m} &= 1 - 2X(p - \bar{p}) \\
\dot{p}_t - p_t &= \phi(X_t - Y_t)
\end{align*}
\]

This system is quite similar to the model we have discussed thus far in this chapter. Indeed, the aggregate-demand and Phillips-curve relationships are exactly the same. The policy reaction function is more up-to-date in the present specification, with the central bank adjusting its monetary aggregate with a view to hitting a price-level, not an output, target. For simplicity, we assume that the value of the level of all variables with bars above them is unity (so the associated logarithms (the lower-case variables appearing in this system) are zero). The solution equation for the output gap then is

\[
X_t = X_{t-1} \phi(\beta + 1)
\]

where \( \phi = 1 - \delta(\chi + 1) \). As was the case with our earlier analysis, this model reduces to a first-order linear difference equation. Convergence to full equilibrium occurs if \( \phi \) is a fraction. The model cannot predict an ongoing cycle of constant amplitude. Is a policy of price-level targeting recommended? Not necessarily: an increase in parameter \( \chi \) can make \( \phi \) exceed unity, so more aggressive price-level targeting can be destabilizing. But this policy can never change the fact that output gravitates to its natural rate value (unity) as long as instability is avoided. We now show that this property is lost if the curvature involved in the Phillips curve is altered just slightly. Indeed, the classical dichotomy is lost in this case.

These assertions can be defended by replacing the Phillips curve in this simple model with the following:

\[
\dot{p}_t - p_t = \phi(X_t - Y_t - 1).
\]

This one change leads to a reduced form for output that is slightly different:

\[
Y_t - Y_{t-1} = \phi(\beta + 1)(X_t - 1).
\]

Since the left-hand side of this equation can be approximated by \((X_t - X_{t-1} - X_{t-1})\), and since we can define \( X_t = ((\beta - 1)\beta)I_t \), where \( \beta - 1 = \delta(\chi + 1) \), this reduced form can be re-expressed as

\[
X_t = \beta X_t - X_{t-1}.
\]

This reduced form is a first-order non-linear difference equation; it is the simple logistic function involved in basic chaos theory. One purpose of this section of the chapter is to relate standard stabilization policy analysis to the chaos literature in a simple and clear fashion. We will see that the average rate of output observed in the long run (a real variable) is affected by the policy reaction coefficient (that is, by monetary policy).

Despite its simplicity, equation (3.4) is consistent with many different time paths for output, as is summarized in Table 3.1. Two of the possible dynamic patterns are shown in Figure 3.2. The general shape of equation (3.4) that is shown in both parts of the figure can be verified by noting that both the values 0 and 1 for current \( X \) imply that next period’s \( X \) is zero, and that the slope of the relationship decreases as current \( X \) rises (and is zero when current \( X \) is 0.5).

The dashed lines in Figure 3.2 indicate the time sequence. Convergence to the natural rate (which involves \( X \) equal to 0.5 and \( T \) equal to 1) occurs with the “low” value for \( \beta \), but a limit cycle (one that maintains a constant amplitude forever) occurs with the “high” value for \( \beta \). In the right-hand panel of Figure 3.2, “limit cycle” involves the economy shifting back and forth between points \( A \) and \( D \) (every other period). The average level of economic activity is represented by point \( C \). Since point \( D \) corresponds to the natural rate of output, point \( C \) involves an average activity level for the economy that is below the natural rate. Thus, with a non-linear reduced form relationship, it is possible that output remains below the natural rate (on average) even in the long run.

<table>
<thead>
<tr>
<th>Table 3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of Time Path</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( \beta )</td>
</tr>
<tr>
<td>( \beta \approx 0 )</td>
</tr>
<tr>
<td>( 3.5 \approx 3.49 )</td>
</tr>
<tr>
<td>( 3.49 \approx 3.49 )</td>
</tr>
<tr>
<td>( 3.57 \approx 3.57 )</td>
</tr>
<tr>
<td>( X ) is subject to chaotic fluctuations</td>
</tr>
<tr>
<td>( \beta &lt; 4 )</td>
</tr>
</tbody>
</table>

This analysis is a deterministic variant of Lipsey’s (1960) non-linear aggregation hypothesis. The interesting features of this variant are that stochastic elements are not required to generate a cycle that maintains its amplitude as time passes, and that it is variations in monetary policy that can make these permanent output effects important.

In this model, a more aggressive attempt to target the price level (a larger value for parameter \( \chi \)) can lower the average rate of output, since uncertainty concerning the appropriate functional form for one of
he model’s structural equations is what leads to the possibility that something as basic as the classical dichotomy can break down (even in a natural-rate setting), it must be admitted that the standard practice of linear approximations in macroeconomics is a definite limitation.

One final point worth explaining is how Figure 3.2 is altered when the macroeconomy is chaotic. Chaotic cycles occur with larger values for β than are assumed for Figure 3.2, since a higher β makes the cubic-form relationship pictured in Figure 3.2 more non-linear. The end points are unaltered, but the height of the hump is increased. Chaotic cycles also involve an “obverse” at points 4 and 6 in Figure 3.2, but with the location and size of the box moving around the same forever, with no regular pattern ever emerging.

There are three reasons for integrating basic macro analysis and chaos theory. First, we can appreciate that ongoing changes in exogenous variables do not need to be assumed to explain business cycles. Second, we can see that intuitively appealing policies can create cycles (both regular ones and chaotic ones) – so once again built-in “stabilizers” can be destabilized. Finally, we can appreciate that the entire nature of the full equilibrium for real variables – whether it involves asymptotic approach to the natural rate or a limit cycle which averages out to zero output is less than the natural rate (to take just two of the possibilities) – can depend on the short-run targeting strategy for implementing inflation policy. Keynesians have become particularly excited about this implication of non-linearity, since it implies that they need not follow what has been the convention – conceding the full-equilibrium properties of macro models to the Classical, and limiting the Keynesian contribution to a discussion of approach paths. With non-linearities, the short and long-run properties of models cannot be so readily separated.

Two useful surveys of work on non-linear dynamics are Mullenouse and Peng (1993) and Rosser (1990). It has proved difficult for econometricians to distinguish between the traditional view of business cycles (linear systems with stochastic shocks) that maintains the amplitude of cycles but would otherwise dampen out and this alternative view (endogenous deterministic cycles in a non-linear system that never die down). Despite this inability to reject the relevance of the non-linear approach, it has not become too popular. Mainstream macroeconomists have limited their attention to linear stochastic systems, as we do for the remainder of the chapter.

3.3 Adaptive Expectations

Real world business cycles are not the two-period saw-tooth cycles that we see in Figure 3; they are more irregular. But if mainstream analysis does not rely on exogenous variables following sine-curve time paths, and it simplifies by ignoring non-linearities, how does the mainstream approach generate realistic-looking cycles in the model’s endogenous variables? The answer is by extending the linear model so that it is consistent with cycles that appear more like a sine curve. This can be done if the model is complicated to the point that the reduced-form for output deviations is a second-order (or higher) linear difference equation. One way of doing this is by allowing for adaptive expectations. Other ways are discussed in Chapter 5, where we explore New Classical macroeconomics.

Ignoring all policy variables and error terms for simplicity, the revised demand and supply functions are:

\[ y_t = \theta p_t + \psi \pi_t \]

\[ p_{t+1} = -\theta (y_t - \bar{Y}) + \pi_t \]

\[ \pi_t = -\pi_t + \lambda (p_t - p_{t-1} + \pi_{t-1}) \]

where \( \pi \) stands for expected inflation. As explained in Chapter 2, the revised demand function comes from distinguishing the nominal from the real interest rate. The revised Phillips-curve involves the assumption that the “core” inflation rate is the adaptively-formed expected inflation rate. The adaptive-expectations hypothesis is defined by the following equation:

\[ \pi_t = (1-\lambda)\pi_t + \lambda \pi_{t-1} \]

\[ \pi_{t+1} = (1-\lambda)\pi_{t+1} + (1-\lambda)\lambda \pi_t + \ldots \]

so that successive substitution leads to

\[ \hat{\pi} = \lambda (\lambda \pi_{t+1} + (1-\lambda)\lambda \pi_t + (1-\lambda)\lambda^2 \pi_{t-1} + \ldots) \]

This last formulation of the hypothesis states that \( \hat{\pi} \) is a weighted average of past actual values, with less weight given to the more distant past. The weights decline geometrically. Both the error-adjustment interpretation and the weighted-average interpretation suggest a certain plausibility for the adaptive expectations hypothesis, as does its long-run consistency property (discussed in the Introduction to this chapter). Thus, it is not surprising that this hypothesis was the mainstream way of modelling expectations in macroeconomics from the mid 1950s to the mid 1970s. The three equations of this simple model can be combined to yield a second-order difference equation \( \hat{\pi} \), as a function of both \( \hat{\pi} \) and \( \hat{\pi} \).

In particular, we proceed through the following steps. We begin by first-differencing the demand function, and using the Phillips curve and the adaptive expectations equations to eliminate the change in the price level and the change in expectations (respectively). The result is

\[ \hat{\pi}_t = (1+\phi(\psi t - \theta)) \hat{\pi}_{t-1} - \theta \hat{\pi}_{t-1} \]

This relationship itself is first-differenced, and the resulting \( (\pi_t - \pi_{t-1}) \) term is eliminated by combining lagged versions of the Phillips-curve and the expectations equations:

\[ (\pi_t - \pi_{t-1}) = \lambda (\hat{\pi} - \hat{\pi}) \]

The result is

\[ \hat{\pi}_t = (\pi_t - \pi_{t-1}) + (\hat{\pi}_t - \hat{\pi}_{t-1}) \]

where \( \psi = 1 + \phi(\psi t - \theta) \) and \( \pi_t = -(1+\phi(\psi t - \theta) - \theta) \). As in the first-order linear difference equation case, there are four possible time paths (as is explained in all standard math-for-economists texts such as Chiang (1984, p. 576)). The only difference is that, in this case, the cycles involve damped and explosive sine curves, not two-period saw-tooth cycles. With slope parameter values that lead to damped cycles, and exogenous disturbances to keep those cycles from ever dying down, we have the standard approach to business cycles.

It turns out that the condition for cycles to occur is that \( \psi^2 < 4\psi \), and the stability conditions are: \( \psi < 1, \psi + 1 > 0, \) and \( \psi - 1 < 1 \). It is left for the reader to verify that the bigger is the role that adaptive expectations play in the system (that is, the bigger is parameter \( \lambda \), the more likely it is for the system to display cycles, and for the system to display instability. It is for this reason that Keynesian economists have emphasized expectations. And, as we discovered in Chapter 2, similar conclusions emerge when a model-consistent form of expectations (perfect foresight) is assumed, instead of adaptive expectations.

While the adaptive-expectations hypothesis has some appealing features, it has an unappealing one as well. It makes little sense to analyze the effectiveness of any government policy that is intended to improve agents’ economic welfare if that analysis does not permit those agents to understand both the environment they live in and the effect of the ongoing policy intervention within that environment. Our modeling has allowed for all these considerations only in the perfect-foresight case – not in the adaptive-expectations case.

A less “unrealistic” version of model-consistent expectations (compared to perfect foresight) is the “rational expectations” hypothesis. According to this hypothesis, agents make forecast errors, but (since they are aware of the probability distributions that generate the random shocks that hit the micro economy), they make their forecasts of the endogenous variables as if they were calculating the best forecast possible from the formal model. Thus, forecast errors occur, but they are not systematic. The remaining sections of the chapter explain how we can analyze models of this sort, and what some of the policy implications are.

3.4 Rational Expectations: Basic Analysis

The basic idea behind rational expectations can be explained in an extremely basic setting – the simple fixed-price fixed-interest-rate income-expenditure model that students encounter in their introductory economics course. That model can be defined by the following relationships:

\[ Y_t = C_t + G_t \]

\[ C_t = eY_t \]

Muhammad Firman (University of Indonesia - Accounting)
and one of
\[ Y_t = Y_{t-1} \]
\[ Y_t = E_t(Y_{t+1}) \]

The first two equations define goods market clearing, and that private demand is proportional to agents’ expectations concerning what income they will receive next period. The remaining equations define two hypotheses concerning how those income expectations may be formed. The former can be thought of as either static expectations (the forecast for today is equal to what was actually observed yesterday), or a degenerate version of adaptive expectations (with the entire weight in the weighted average put on the most recent past). The second hypothesis is rational expectations. According to this hypothesis, agents’ subjective forecast is the same as we can calculate by evaluating the mathematical expectation of actual current income (as determined in the model).

In the static expectations case, the at-a-point-in-time solution equation for current GDP is:
\[ Y_t = cY_{t-1} + G_t \]

Consider a once-for-all increase in autonomous spending. \( G_t \). The solution for \( Y_t \) indicates that GDP rises by the same amount as the increase in \( G_t \) in that very period. Then, \( Y_t \) keeps rising by smaller and smaller amounts each period, with the overall increase in output (given infinite time) being:
\[ Y_t = (1-(1-c)G_t) \]

In the rational expectations case, the same substitution of the second and third equations into the first results in:
\[ Y_t = cE_t(Y_{t+1}) + G_t \]

This is not a solution equation for actual output, \( Y_t \), since there are still two endogenous variables (\( Y_t \) and its expectation) in this one equation. We need a second equation — one for the forecast of \( Y_t \). This can be had by taking the expectations operator through this “almost reduced-form”. The result is
\[ E_t(Y_{t+1}) = G_t/(1-c) \]

When this expression is substituted back into the almost-reduced-form, we end with the full reduced-form (solution equation) for actual output:
\[ Y_t = (1-(1-c)G_t) \]

By comparing the two solution equations (for static expectations and for rational expectations), we can determine how the model’s properties are affected by embracing this model-consistent forecasts hypothesis. We see that the impact effect of a once-for-all increase in autonomous spending is bigger in the rational expectations case. Indeed, the impact effect on \( Y_t \) is now equal to the eventual effect. With forward-looking agents, it makes sense that there not be a gradual increase in income. In this case, agents correctly see that this higher income is coming. They raise their consumption immediately as a result, and this (in turn), since output is demand-determined in this model, makes overall income rise immediately. In short, what used to be called the model’s long-run properties become the model’s short-run properties (when agents have rational expectations). Worked in this general fashion, this summary applies to all rational-expectations models.

We now explore the properties of a slightly more complicated rational-expectations model. The system we now consider is similar to that in the previous section in that it involves descriptive aggregate demand and supply functions. But there are several extensions here. For one thing, we allow for variable prices and interest rates. Also, we focus on monetary (not fiscal) policy. Initially, we continue to suppress any distinction between nominal and real interest rates, and we specify an arbitrary monetary policy reaction function. The initial model is defined by the following four equations,

\[ y_t = \alpha - \psi r_t + v_t \]
\[ p_t = p_{t-1} + \phi r_t + \phi' r_{t-1} + u_t \]
\[ r_t = r + \gamma (p_{t-1} - 0) \]
\[ r' = E_t(r_{t+1}) \]

\( \alpha \) and \( \delta \) stand for the natural logs of real output and the price level. Both the natural rate of output and the central bank’s target for the price level are unity, so the logs of these variables are zero (and \( \phi = \phi_1 = 0 \)). \( \gamma \) and \( \theta \) are stochastic shocks — drawn from distributions that involve zero means, constant variances and no serial correlation. \( \gamma \) is the interest rate, not its logarithm.

Equation (3.5) is a standard (descriptive) IS relationship, which is also the aggregate demand function once the central bank sets the interest rate. Since we focus exclusively on monetary policy in this discussion, fiscal variables are constant (and embedded in the intercept). Equation (3.6) summarizes the supply side of the goods market. It is a standard expectations-augmented Phillips curve. Equation (3.7) is the central bank’s reaction function. The bank raises (lowers) the interest rate above (below) its long-run average value whenever the bank’s forecast for the log of the price level is above (below) target. There is no need to specify an LM equation, since its only function is to determine what value the bank had to adjust the money supply to (in order for the public’s demand for money function to be satisfied at this interest rate). Since the money supply enters none of the other equations of the model, we can afford to ignore this consideration.

Equation (3.8) defines rational expectations; it specifies that the agents’ subjective expectation for price is equal to what we (as model manipulators) can calculate as the mathematical expectation of price. The time subscript for the expectations operator denotes that agents know all values for the stochastic shocks up to and including the previous period (time period \( t-1 \)). Agents (and the central banker) must forecast the current shock — at the end of the previous period before it is revealed — on the basis of all past information. Agents know the exact structure of the economy (the form of all equations and all slope coefficients). The only thing they do not know is what the current and all future values of the additive error terms are.

To solve the model, we first eliminate the interest rate and the expected price variables by simple substitution. The results are:

\[ y_t = \alpha - \psi E_t(g_{t+1}) + v_t \]
\[ p_t = \phi E_t(p_{t+1}) + u_t \]

These reduced forms lead to the following volatility expressions:

\[ \text{var}(y_t) = \sigma^2 \]
\[ \text{var}(p_t) = \phi^2 \sigma^2 + \sigma_u^2 \]

Since neither variance is a function of the policy maker’s parameter, \( \gamma \), monetary policy is irrelevant. This makes sense; after all, the central bank must set its instrument variable (the interest rate) before the current shocks are known — just as the private agents must commit to setting their nominal variables — before the current shocks are known. There is nothing that the central bank can do for the private agents that they cannot do for themselves.

We can remove this “policy irrelevance” result by changing the model in at least two ways. For one thing, we could have some of the private agents constrained to set their nominal variables more than one period in advance. For another, we could let the central bank wait until the current shocks are known before the interest rate is set. In either case, the central bank would be able to do more than can agents on their own, so monetary policy should matter. We choose the second alternative here, and consider multi-period private sector nominal contracts in later chapters. We verify that policy irrelevance is a very model-specific result by changing equation (3.7) to the following:

\[ r_t = r' + \gamma (p_{t-1} - 0) \]

and re-deriving the output and price variances. It is left for the reader to duplicate the earlier steps, and to verify that the revised solutions are:
These results imply that a more aggressive price-level targeting policy (a bigger value for parameter $\gamma$) is desirable since it makes demand shocks have a smaller effect on real output, but it is undesirable since it makes supply shocks have a larger effect on real output. So the monetary authority faces a permanent volatility trade-off, even though it does not face a permanent trade-off between the average level of real output and inflation (or the price level).

These results make sense at the intuitive level. The natural-rate feature (the Classical dichotomy) is a built-in feature of the Phillips-curve specification, so it is not surprising that there is no lasting trade-off between inflation and the level of real activity. But this does not mean that central bankers should ignore the effect of their policies on real variables. Even within the class of policies that deliver long-run price stability within a natural-rate model, there is a basis for preferring one specific monetary-policy rule to another, since there are lasting differences in the volatility of real activity about its natural-rate value.

When a negative demand shock hits, the aggregate demand curve shifts left. If the short-run aggregate supply curve is positively sloped, price falls. An aggressive price-targeting central bank reacts vigorously to reverse this effect – by shifting aggregate demand back to the right. Since the exogenous shock and the policy variable both affect the same curve (demand, not supply), the central bank cannot help but limit output volatility as it pursues price stability. But things are different when the shock is on the supply side. If the aggregate supply curve shifts left, there is pressure for the price level to rise, and for real activity to fall. To limit the rise in price, the central bank must shift the demand curve to the left. The problem is that this policy accentuates the output fall, and so there is a volatility trade-off.

The fact that we can appreciate these outcomes so easily at the intuitive level makes it seem as though there is no need to work things out formally. But such a conclusion is not warranted, as we will see in the next section – where we consider a more complicated model. In such settings, it is almost impossible to sort things out without the precision that accompanies a formal solution.

### 3.5 Rational Expectations: Extended Analysis

We now consider a more complicated model – one that involves both yesterday’s expectations of today’s outcomes and today’s expectations of tomorrow’s outcomes. A more involved solution procedure – known as the undetermined coefficient solution method – is required in such a setting. The following model is fully consistent with the micro-foundations that modern macroeconomists insist upon. Readers will be able to verify this when they read Chapter 4. For the remainder of this chapter, however, we simply take this model as “to be defended later.”

The model involves a proper distinction between nominal and real interest rates, $IS$ and Phillips curve equations that are based on inter-temporal optimization, and a central banker who optimizes. To simplify slightly, we set the supply shock to zero in all time periods. The revised $IS$ supply, and policy-reaction functions are:

$$y_i = \left(E_i(x, y) - E_i(x, y_i)\right) + E_i(x) - \psi p - v,
\quad p_i = (1/2)p_i + (\phi/2)y,$$

(3.11a)

(3.12a)

We now proceed with solving the model, using trial solutions and the undetermined coefficient method. Since there is only one pre-determined variable in the model (the previous period’s price) and one exogenous variable (the demand shock), there can only be these two terms in the solution equations for real output and the price level. Thus, we assume the following trial solutions:

$$y_i = \delta + b v_i
\quad p_i = c + d_i,$$

(3.13)

(3.14)

We use the undetermined coefficient solution procedure to determine how the reduced-form coefficients $(a, b, c, d)$ depend on the underlying structural parameters that have economic meaning $(\psi, \phi, \gamma)$.

We now substitute the trial solutions into (3.11a) and (3.12a). First, after substituting (3.13) into (3.12a), we have

$$p_i = ((1 + \psi d)/2) v_i + (b \phi/2) y_i,$$

(3.15)

Equations (3.16) and (3.15) are the same if and only if

$$c = (1 + \alpha b)/2,$$

and

$$d = \phi (b \phi/2).$$

We need two more identifying restrictions. These are obtained by using the trial solutions to get expressions for all terms on the right-hand side of (3.11a) except the error term. Once these are substituted in and the result is compared to (3.13), the remaining two identifying restrictions emerge:

$$a = a_0 - \psi c,
\quad b = (d - \phi d) + 1.$$

The four identifying restrictions are now solved explicitly for $a, b, c, d$ and $\gamma$, so the reduced forms contain only primary parameters:

$$a = 1/\phi,
\quad b = 2/(3 + \psi \phi),
\quad c = 0,
\quad d = \phi (3 + \psi \phi).$$

The expressions for price and output volatility (the asymptotic variance for each variable) follow directly from the trial solutions:

$$\sigma_y^2 = (d^2/\phi) \sigma^2,$$

(3.16)

$$\sigma_p^2 = a_0^2 \sigma^2 + b^2 \sigma^2.$$  

(3.17)

Illustrative values for these volatility expressions can be had if representative parameter values are inserted. As we shall see in Chapter 4, the most straightforward defense of the $IS$ relationship in this model involves assuming that households have a utility function equal to the logarithm of consumption. If this is assumed, our parameter $\psi$ must be unity, so we insert this value here. Walsh (2003a, p. 576) has used 0.2 as an estimate of the slope of the short-run (annual) Phillips curve, so for our illustrative calculations here, we follow his lead and take $\phi = 0.2$. With these representative values,

$$\sigma_y^2 = 0.5 \sigma^2$$

and

$$\sigma_p^2 = 0.004 \sigma^2.$$  

(3.18)

We now compare these outcomes to what occurs if the central bank pursues a different target. Suppose the bank’s goal is to set the interest rate so that the expected value for nominal GDP is right on target (equals zero). In this case, all the steps in the solution procedure are repeated, but with $E_i(p_i) - E_i(x), not E_i(x_i) = 0$, in this case. The derivations are messier with this monetary policy, so we simplify by imposing $\psi = 1$ from the outset. It is left for the reader to verify that the revised expressions for the identifying restrictions are:

$$a = c_1 - c,
\quad b = 1 - d,
\quad d = \phi (2 - c - \phi),
\quad c = (\phi + 2) \times \sqrt{(\phi + 2)^2 + 4})/2.$$
CHAPTER 4

THE MICROFOUNDATIONS OF MODERN MACROECONOMICS

4.1 Introduction

The traditional analysis that was reviewed in previous chapters involved some appeal to micro-foundations. For example, in the textbook classical model (in Chapter 1), it is customary to refer to a static theory of household utility maximization to "justify" the household labour supply function. (We assumed that households maximize a utility function containing two arguments—after-tax real income and leisure—subject to a simple budget constraint—that income equals the after-tax real wage times the amount of time worked.) Also, we assumed that firms maximize their profits, and this was the rationale behind the standard optimal hiring rule for labour—that workers be hired until the marginal product of labour is pushed down to the rental cost of labour (the wage). But we did not assume a similar optimal hiring rule for the other factor input—capital. This different treatment is explained later in the present chapter—in the section on the micro-foundations of the firms. But first, we consider the micro basis for the equations of a standard macro model that summarize the behaviour of households. And even before that, in the next section of the chapter, we motivate the drive for more explicit micro-foundations in modern macroeconomics.

4.2 The Lucas Critique

Before the early 1970s, virtually all macro policy analyses were inconsistent with the principles of microeconomics. The households and firms who operated within the macro model followed the same decision rules no matter how their environment was altered by changes in policy regime. Economics is often defined as the subject that explores the implications of constrained maximization. But this description did not apply to traditional macroeconomics. Since this was forcefully first pointed out by Nobel laureate Robert Lucas in 1976, and since macroeconomists have been working hard to avoid this criticism ever since, we begin this chapter by explaining what has become known as the "Lucas critique" of more traditional macroeconomics. The material in this section provides a compact summary of Lucas' argument—with an explicit applied example—by relying on a simple theory behind the Phillips curve. A fuller (inter-temporal) version of this theory of sticky prices is available in McCallum (1980) and Musso (1981). Here, for simplicity, we consider just one-period optimization.

Firms face negotiation costs when setting wages. When workers are dissatisfied with the prospect that wage changes will be too low, they work to rule or strike, and this industrial action lowers output (and raises firms' costs). When firm owners are dissatisfied with the prospect that wage changes will be too high, they lock out workers, and these actions also reduce output and raise costs. If prices are a mark-up on wages, this same reasoning implies that these adjustment costs are incurred whenever price increases are either too low or too high. Full equilibrium considerations dictate would be an "appropriate" price change. To capture these considerations, we assume that the price-setter's optimization involves balancing the costs of fast versus slow price adjustment. Fast adjustment lowers the cost of being away from full equilibrium, while slow adjustment lowers the adjustment costs. The following cost function captures these considerations:

\[
\phi(p_t - \bar{p}) + \beta(p_t - p_{t-1}) - (\bar{p} - \bar{p}_{t-1})^2
\]

\(p_t\) is the full-equilibrium price—the one that would have the firm operating at its natural rate of output (at the minimum point of its long-run average-cost curve). The first term in the cost function captures the cost of being away from this desirable long-run level of operations. The second term captures the adjustment cost—that exists whenever the actual change in wages (and therefore prices) is either above or below what is dictated by the firm's equilibrium considerations. Parameter \(\beta\) defines the relative importance of these adjustment costs. To keep this demonstration of the Lucas critique simplified, we take \(\beta\) to be a "primitive" parameter. Economics starts with a specification of tastes and technology. If one wants to explore the determinations of wages, one becomes a psychologist, not an economist, and if one is interested in understanding technology, one becomes an engineer, not an economist. So taste and technology are the primitive constructs in our discipline. Clearly, parameter \(\beta\) is not primitive in this ultimate sense. However, since the relative cost of adjustment can be assumed to be an aspect of the given "technology" faced by firms, we interpret \(\beta\) in this way, in the interests of a simplified exposition.

Since the equilibrium considerations and previous history are beyond the control of this period's price-setter, costs are minimized by differentiating the cost function with respect to the current actual price and setting the result equal to zero. After a bit of manipulation, we have

\[
(p_t - p_{t-1}) = (\bar{p} - \bar{p}_{t-1}) - \gamma(p_{t-1} - \bar{p}_{t-1})
\]

where \(\gamma = 1/(1 + \beta)\). This relationship can be written in continuous time as

\[
\dot{p} = \gamma(y - \bar{p})
\]

A standard version of the short-run Phillips curve is

\[
\dot{p} = \phi + \theta(m - p)
\]

and this form must now be related to the price-change relationship just derived, via an aggregate demand relationship. For simplicity, let us assume that transactions technology dictates that each level of output requires a certain amount of money. This is captured in a quantity-theory specification for aggregate demand (where parameter \(\theta\) is also "primitive"—that is--unaffected by changes in the policy regime):

\[
y = \Theta(m - \bar{p})
\]

We add a simple monetary policy reaction function:

\[
m = \bar{m} - \chi(p - \bar{p})
\]

According to this relationship, the central bank raises (lowers) the log (of the) nominal money supply—above (below) its long-run average value—whenever the price level is above (below) its target value. If parameter \(\chi\) is zero, the central bank is a monetary-aggregate targeter; if it approaches infinity, the bank is a price-level targeter.

Substituting the policy reaction function into the quantity theory equation, we obtain the nation's aggregate demand function:

\[
y = \Theta(m - \bar{p}) - \Theta(1 + \chi)p - \bar{p}
\]
Next, we define the full-equilibrium price as that value for \( p \) that makes output equal its natural rate. This value can be determined by setting \( y = y^* \) and \( p = \bar{p} \) into the demand function:

\[
\bar{y} = \phi(\bar{y} - \bar{p})
\]

This equation can be used in two ways. First, if the natural rate is constant, the time derivative of this equation states that \( \bar{y} = \bar{y} \); in other words, the core inflation rate is the money growth rate. The second use of the last equation is that it can be subtracted from the demand function. The result is:

\[(y - \bar{y}) = \phi(y + \bar{y})(\bar{y} - \bar{p}).\]

Substituting this last relationship into the price-change equation that was derived above, we end with:

\[\bar{p} = \bar{m} + \phi(\bar{y} - \bar{p})\]

where \( \phi = \bar{y}(\bar{p} + \bar{y}) \)

This Phillips curve has two important features. First, it states that temporary deviations of inflation from the monetary growth rate correlate with temporary deviations of real output from the natural rate. This implies that there is no inconsistency in the view that inflation is ultimately a purely monetary phenomenon, and the view that the Phillips curve is an important ingredient in a theory of the short-run interaction between inflation and real activity. As noted earlier chapters, this has long been a feature of standard macroeconomics.

Second, we conclude that the summary parameter \( \phi \) cannot be considered a primitive parameter. According to this derivation, \( \gamma \) and \( \alpha \) are technology parameters, since they define the adjustment-cost and transactions processes. Thus, these parameters are not affected by aggregate demand policy — that is, by changes in the value of monetary-policy parameter \( \chi \). But precisely because these “private sector” response coefficients are policy invariant, it must be the case that the slope of the short-run Phillips curve, \( \phi \), does depend on monetary policy. Thus, the slope of the short-run Phillips curve does not represent purely “supply-side” phenomena.

Standard practice in applied economics (in all fields— not just macroeconomics) involves estimating a model, and then using those estimated coefficients to simulate what would happen if policy were different. The Lucas critique is the warning that it may not make sense to assume that those estimated coefficients would be the same if an alternative policy regime were in place. The only way we can respond to this warning is to have some theory behind each of the model’s equations. We can then derive how (if at all) the coefficients depend on the policy regime. This derivation of the short-run Phillips curve illustrates two things. First, the Lucas critique does apply to the Phillips curve — its slope is a policy-dependent coefficient. Second, the derivation illustrates how we can react constructively to the Lucas critique. Just because the slope coefficient depends on policy, it does not mean that legitimate counterfactual experiments cannot proceed. The value of micro-foundations is evident. They do not just expose the non-primitive nature of the Phillips curve slope; they also outline precisely how to adjust that parameter to conduct theoretically defensible simulations of alternative policy rules.

One illustration of how the answers to policy questions change when the micro-foundations are respected can be had by reconsidering the results that were reported in Chapter 2 (Section 2.4). In that analysis — which involved essentially exactly this model but without the micro underpinnings — we determined that the speed of adjustment between full equilibria (parameter \( s \) in \( y = s(y - y) \)) could be derived in a straightforward fashion, and in this case we have \( s = \phi(\bar{y} + \bar{y}) \). Without considering micro underpinnings, we conclude that more aggressive policy increases the economy’s adjustment speed back to full equilibrium following disturbances. But, according to the Lucas critique, we should take account of the micro basis of the Phillips curve, and substitute out the policy-dependent Phillips curve slope parameter (by using the \( \phi = \bar{y}(\bar{p} + \bar{y}) \) result) before conducting the policy analysis. When this is done, the expression for the adjustment speed parameter becomes \( s = y \).

This result has dramatically different policy implications. It says that monetary policy has no effect on the economy’s adjustment speed. So respecting the Lucas critique does not just affect the smaller details of policy analysis; it can change the analysis in fundamental ways.

But is Lucas critique actually important in the real world? It appears so, if the history of recent decades is considered. Our micro-based analysis predicts that a move toward more direct price-level targeting on the part of the central bank can be expected to decrease the slope of estimated Phillips curves. This effect is not usually noted, so (for example) many people were surprised that the monetary-policy-induced recessions of the early 1980s and the early 1990s were as long and deep as they were. These outcomes are less surprising when this underlying theory is considered. In the 1980s and 1990s, central banks were increasingly using forward guidance to policy level-targets. With a flatter Phillips curve emerging as a result, the contraction in demand that was part of the disinflation policy had a noticeably larger effect on real output than the previous estimates of the Phillips curve had suggested. Viewed through the analysis of this section, these outcomes are not surprising after all. While this rather dramatic change in monetary policy clearly illustrates the importance of the Lucas critique, some authors (such as Rudolph (2003)) have noted that the empirical evidence concerning less bold changes in policy does not support the conclusion of structural breaks in the estimated reduced forms of macro models.

Despite the value of micro-foundations (that has been illustrated in a policy-relevant setting above), one consideration has kept some macroeconomists from working toward a more elaborate macroeconomic base for conventional macro models. This problem is aggregation — an issue which was not addressed in the Phillips curve analysis just given. The conclusion which emerges from the aggregation literature is that the conditions required for consistent aggregation are so rigid that constrained maximization at the individual level may have very few macroeconomic implications — that is, very few useful insights for aggregate analysis. This presents a problem since the only way to solve the Lucas critique is to use optimizing underpinnings to go “behind demand and supply curves” (Sargent 1982), and to treat only the ultimate taste and technology parameters as primitive (policy-invariant). If aggregation issues prevent these individual optimizations from imposing any restrictions on macroeconomic relationships, the Lucas critique cannot be faced. Yet only a few commentators (for example, Sargent 1985), emphasize that ignoring aggregation issues can be as important as ignoring the Lucas critique.

Thus, we are on the horns of a dilemma. Economists should ignore neither aggregation problems nor optimizing underpinnings. Yet the current convention to ignore aggregation issues by building macro models involving no differences between any individuals — the so-called representative-agent model. The only justification for this approach is an empirical one — that the predictions of the macro models, which are based on such a representative agent, are not rejected by the data. Thus macroeconomists have reacted to this dilemma in a pragmatic way since aggregate models seem consistent with the macroeconomic “facts,” then no matter how restrictive the aggregation requirements seem, not too much seems to be lost by assuming that the economy operates as if these restrictions are appropriate. Some macroeconomists find this pragmatic approach unconvincing, and they draw attention to inherent logical difficulties within the representative-agent methodology (see Kirman (1992) and Hartley (1997)).

In later chapters (10 and 12), we make use of an over-lapping generations macro model that involves both appealing optimization underpinnings and explicit aggregation across cohorts of different ages. Indeed, we introduce readers to this model that generalizes the single ever-lasting representative-agent framework later in the present chapter. The over-lapping generations model certainly respects the Lucas critique without ignoring the aggregation challenges. Nevertheless, since much of the modern macro literature that we need to survey in this book follows the representative-agent convention, we focus on this simpler model as well.

4.3 Household Behaviour

The standard (descriptive) IS relationship involves households that are liquidity-constrained; current consumption is limited by current income. In other words, many households cannot borrow and lend, so current consumption cannot exceed current income for many periods (as long as the household is solvent — in the sense that the present value of its debts can be covered by the present value of its assets). Many young families borrow as they purchase a home and then use future income to cover both the purchase price and the interest on the mortgage. Households do this because they are impatient; other things equal, they get higher utility if consumption of any item can occur sooner rather than later. The following model captures this behaviour:

Assume that the household utility function is:
utility = \sum_{i=1}^{n} \left( \frac{1}{(1 + \rho)^i} \right) \ln C_i

\text{i is the index of time periods, } \rho \text{ is the rate of time preference (the higher is } \rho, \text{ the more impatient people are), and the logarithmic form for the utility function at each point in time is consistent with two propositions -- that a certain minimum amount of consumption (defined to be one unit) is needed to live (to receive any positive amount of utility), and that (beyond that one unit) there is diminishing marginal utility of consumption.}

The household maximizes this utility function subject to the constraint that the present value of the entire stream of consumption be no more than the present value of its disposable income. As readers will have learned in basic micro theory, to achieve utility maximization, the household must arrange its affairs so that the ratio of the marginal utilities of any two items is equal the ratio of the prices of those two items. In this case, the two items are the levels of consumption in any two adjacent time periods. If the price of buying one unit of consumption today is unity, then the price of one unit of consumption deferred for one period is less than unity since the household’s funds can be invested at the real rate of interest for one period. Thus, the price of one-period-deferred consumption is \(1/(1+r)\). With this insight, and the knowledge that marginal utility for a logarithmic utility function is the inverse of consumption, we can write the condition for utility maximization as:

\[
\frac{MU_i}{MU_{i+1}} = ((\text{price of } C \text{ in period } i)/(\text{price of } C \text{ in period } i-1))
\]

\[
\frac{1}{(1/(1+r))} = 1/(1+r)
\]

\[
C_i = [(1+r)/(1+r)]C_{i+1}
\]

By subtracting lagged consumption from both sides of this last representation of the household’s decisions rule, and by noting that \((1+r)\) is approximately equal to unity (a reasonable value for the annual rate of time preference is something like 0.04), the consumption function can be simplified further:

\[
\Delta C = (r - \rho)C
\]

Re-expressing this relationship in a continuous-time format, we have:

\[
\dot{C} = (r - \rho)C
\]

We now show how this lifetime-wealth-based consumption function -- which is consistent with inter-temporal optimization -- is just another way of thinking about Friedman's (1957) model of permanent income. According to Friedman, consumption is proportional to broadly-defined wealth -- the sum of non-human assets, \(A\), and human wealth, \(H\):

\[
C = \rho A + H
\]

The factor of proportionality is the rate of time preference. To see that this is the same theory as the one just derived, we need to know how both human and non-human wealth change over time. In the case of non-human wealth, the specification is familiar. If each individual’s level of employment is one unit, he acquires assets when the sum of his wage income, \(w\), and interest income, \(rA\), exceeds current consumption:

\[
\dot{A} = w + rA - C
\]

Human wealth is the present value of all future after-tax wage income. Since taxation is more straightforward in a discrete-time specification, initially we write human wealth as:

\[
H_i = [w(1+r) + w_o(1+r)^2] + ...
\]

Writing this relationship forward one period in time, we have:

\[
H_{i+1} = [w_o(1+r) + w_o(1+r)^2] + ...
\]

Multiplying this last equation through by \((1/(1+r))\), and subtracting the result from the \(H_i\) equation, we have:

\[
\Delta H_i = rH_i - w_i
\]

In a continuous-time specification, then, Friedman’s model can be summarized by the following three relationships:

\[
C = \rho(A + H)
\]

\[
\dot{A} = w + rA - C
\]

\[
\dot{H} = rH - w
\]

By taking the time derivative of the first relationship, and substituting the other two equations into the result, we have:

\[
\dot{C} = (r - \rho)C
\]

This derivation proves that the permanent-income hypothesis and the inter-temporal-optimization theory are equivalent. As a result, the inter-temporal utility-maximization approach is supported by the extensive empirical work that has established the professions’ confidence in the applicability of the permanent-income model, and Friedman’s approach gains from this equivalence as well, since it is now seen as more consistent with formal optimization than had been previously thought.

Before proceeding with some extensions to this basic framework, it is useful to indicate how the consumption function can be derived more formally. We do so in two stages -- first with time treated as a discrete variable, and second with time treated as a continuous variable.

Households are assumed to maximize

\[
utility = \sum_{i=1}^{\infty} \left( \frac{1}{(1 + \rho)^i} \right) \ln C_i
\]

subject to

\[
C_i = [(1+r)/(1+r)]C_{i+1}
\]

After eliminating consumption by substitution, differentiating with respect to the its period value of \(A\), and setting to zero, we have

\[
C_i = [(1+r)/(1+r)]C_{i+1}
\]

as derived less formally above.

The same result can be derived in a continuous time setting if the calculus of variations is used. This is how Ramsey (1928) originally defined this analysis. In this case, the specification is that households maximize

\[
\int_0^t C e^{-\rho t} dt
\]

subject to

\[
C = \rho A + w - \dot{A}
\]

Since many readers will not have been taught how to deal with a situation in which the objective function is an integral and the constraint is a differential equation, a simple “cook book” rule is given here. Whenever readers confront this situation, they should work down what is known as the Hamiltonian. In this case it is

\[
\Delta = e^{-\rho - \lambda w}(t + w - A).
\]

Optimal behaviour then follows from

\[
\Delta_\lambda - \Delta_\lambda = 0,
\]

where subscripts stand for partial derivatives. It is left for the reader to verify that, in this case, following this procedure leads to exactly what was derived earlier:

\[
C = (r - \rho)C.
\]

Since the reader already knew the answer in this case, he/she can feel reassured that the “cook book” method for dealing with continuous-time specifications does “work”.

Figure 4.1 Ricardian Equivalence

Future Income and Consumption

Present Income and consumption

Indifference curves

Indifference curves
To have more intuition about this model of household behaviour, consider Figure 4.1. For simplicity, the diagram refers to a planning horizon with only two periods: the present (measured on the horizontal axis) and the future (measured on the vertical axis), and taxes on interest income are ignored. The household’s endowment point is A. Since the household can borrow and lend at rate r, the maximum amount of consumption in each of the two periods is marked on both axes. The line joining these two points is the inter-temporal budget constraint, and the household chooses the point on this boundary of its feasible set that allows it to reach the highest indifference curve (point B).

What happens if the government raises taxes today to retire some government bonds? Since this just amounts to the government substituting current for future taxes (with the present value of the household’s tax liabilities staying constant), all that happens is that the endowment point shifts in a northwest direction along a fixed budget constraint to a point such as C. The household simply borrows more, and remains consuming according to point B. Thus, the Ramsey model involves what is known as “Ricardian Equivalence” – the proposition that the size of the outstanding government debt is irrelevant.

Some OECD governments have been very proud of themselves in recent years – as they have been working down their debt-to-GDP ratios. In a similar vein, policy analysts have regularly referred to George W. Bush administration’s fiscal policy as irresponsible, since the United States government debt-to-GDP ratio has been rising dramatically. It is clear that these OECD Ministers of Finance and these commentators on US policy do not believe in Ricardian Equivalence. Perhaps this is because they know that some “real world” individuals are liquidity constraint – that is, they cannot borrow. We can focus on this situation in Figure 4.1, by realizing that such an individual faces a budget constraint given by DAE. Such an individual would be at point A – a corner solution – initially. Then, after an increase in current taxes (which retires some bonds and, therefore, cuts future taxes), the budget constraint becomes DCF. The individual moves to point C. Since current consumption is affected by the quantity of bonds outstanding, Ricardian Equivalence does not apply.

Allowing for liquidity constraints is just one way to eliminate the Ricardian Equivalence property. Another consideration is that people may discount the future since they expect to die. The Ramsey model assumes that the decision-making agent lives forever. If agents are infinitely lived dynasties (households who have children), this assumption may be quite appropriate. Nevertheless, some households have no children, so we should consider the case in which agents do not live forever. The model has been extended (by Blanchard (1985b)) by assuming that each individual faces a constant probability of death, p. With this assumption, life expectancy is 1/p, and (as derived in Blanchard and Fischer (1989), chapter 3) the aggregate consumption function becomes

\[ C = (p + \rho)(A + D). \]

Aggregation across individual agents of different ages (to obtain the aggregate consumption function from the first-order condition derived at the individual level) is messy. Nevertheless, it is feasible – given the assumption that life expectancy is independent of age. Not surprisingly, individuals who expect to die, consume a bigger proportion of their broadly defined wealth. Saving is less appealing when there is a larger probability that the individual will not live to enjoy the spoils. The overall rate of time preference becomes \((p + \rho)\) – the sum of the individual’s rate of impatience and her probability of death. As the simpler set-up, the consumption function can be re-expressed as a consumption-change relationship. The time derivative of the permanent-income representation is taken, and the aggregate version of the accumulation identities for both human and non-human wealth are substituted in. In this case, these identities are:

\[ \dot{A} = w + (r + \rho)A - C - pA \]

\[ \dot{H} = (r + \rho)H - w. \]

The new terms in the wealth-accumulation identities stem from the fact that people realize that the present value of their future wage income is smaller when death is a possibility. The most convenient way to think of the arrangements for non-human wealth is that there is a competitive annuity industry in the economy. It provides each individual with annuity income on her holdings of A throughout her lifetime, and in exchange the individual bequeaths her non-human wealth to the annuity company when she dies. Since a new person is born to replace each one that dies (so that the overall population size is constant), in aggregate, both these payments to and from the annuity companies are \(pA\) each period. When these identities are substituted into the level version of the aggregate consumption function, the result is:

\[ \dot{C} = (r - \rho)C - p(p + \rho)A. \]

This consumption function collapses to Ramsey’s when the death probability is zero. Since government bonds are part of non-human wealth (variable A), empirical workers have utilized this formulation to test Ricardian Equivalence (the proposition that \(p = 0\)). In pooled time-series cross-section regressions with future consumption regressed on current consumption and non-human wealth, empirical researchers are able to reject the null hypothesis that the A variable’s coefficient is zero. We expect these empirical results when using an overlapping-generations analysis to evaluate deficit and debt reduction, and the implications of an aging population, in Chapters 10 and 12. By allowing for \(p > 0\) in that analysis. However, since the New Neoclassical Synthesis approach to stabilization policy analysis simplifies by setting \(p = 0\), we follow this convention as we report on that literature in Chapters 6 and 7.

Thus far, our theory of households has been simplified by assuming exogenous labour supply. If households can vary the quantity of leisure they can consume, inter-temporal optimization leads to the derivation of both the consumption function and a labour-supply function. With endogenous leisure – both current goods consumption and leisure turn out to depend on a permanent income. As far as labour income is concerned, permanent income depends on both the current wage rate and the present (discounted) value of the future wage rate. Thus, current labour supply depends positively on both the current wage and the interest rate, and negatively on the future wage. This aspect of the inter-temporal model of household behaviour plays a central part in New Classical macroeconomics (Chapter 5) – the first generation of fully micro-based macro models. Also, the theory behind the “new” Phillips curve (that forms an integral part of the New Neoclassical Synthesis) relies very much on this same labour supply function. To have a micro base for the New Synthesis that is internally consistent, we want the labour supply function to emerge from the same theory of the household that lies behind the consumption function. Thus, for several reasons, we must now extend that earlier analysis to allow for a labour-leisure (labour supply) choice. As noted, for simplicity and to be able to report the New Neoclassical Synthesis literature as it is, we revert to the ever-lasting family-dynasty version of the theory of the household, for this extension. These household dynasties live forever and they choose a saving plan that is designed to smooth consumption over time. Here we assume that households maximize the discounted value of a utility function that is a function of consumption and a negative function of labour supplied.
for $N, K$ and $I$, the reader can verify that the following rules must be followed for the firm to maximize its affairs.

\[ F_y = \frac{a}{b} \]
\[ I = d(q - 1) \]
\[ r = \frac{F_y}{q} - \delta + \frac{\dot{q}}{q} \]

The first rule is familiar; it stipulates that firms must hire labour each period up to the point that its marginal product has been driven down to its rental cost (the real wage). The second rule is intuitive; it states that firms should invest in acquiring more capital whenever it is worth more than consumption goods (alternatively, whenever the stock market values capital at more than its purchase price). The third equation states that individuals should be content to own capital when its overall return equals what is available on alternative assets (interest rate $r$). The overall return is composed of a “dividend” plus a “capital gain”. When measured as a percentage of the amount invested, the former term is the gross earnings (capital’s marginal product divided by the purchase price of capital) minus the depreciation rate. The final term on the right-hand side measures the capital gain. In our specification of the firms’ investment function in Chapter 1 (that was embedded within the basic IS relationship), we assumed static expectations ($\theta = 0$). In Chapter 6 we are more general: we examine what insights are missed by assuming static expectations. But to simplify the exposition concerning installation costs for the remainder of this section, we assume static expectations.

It is useful to focus on the implications of this theory of the firm for the several models that were discussed in Chapter 1. In the textbook Classical and Keynesian models, firms were assumed to have investment and labour demand functions just like those that we have derived here. As a result, we can now appreciate that the implicit assumption behind these models is that firms maximize profits. That is, they pick factor inputs according to cost minimization, and they can simultaneously adjust employment and output to whatever values they want. With no installation costs for labour, the standard marginal-product-equal-wage condition applies. But, with adjustment costs for capital, a gap between the marginal product of capital and its rental cost (the interest rate plus capital’s rate of depreciation) exists in the short run. The optimal investment function is

\[ I = d\left((F_x/(r + \delta)) - 1\right) \]

4.4 Firms’ Behaviour: Factor Demands

The consumption function is not the only behavioural relationship that is embedded within the IS function. The investment function is an integral part of the IS-LM system as well, so (to have a micro base for this relationship) we consider the inter-temporal theory of the firm in this section. Since firm managers may not know the time preference rate of the firm owners, the best thing that managers can do for their owners (who consume according to the permanent-income hypothesis) is to deliver a flow of income that has the maximum present value. Thus, firms are assumed to maximize the present value of net revenues

\[ \int e^{-\lambda t} [F(N,K) - wN - I - b^t] dt \]

subject to the standard accumulation identity for the capital stock:

\[ I = K + bK, \quad \delta = \text{capital’s depreciation rate} \]. The final term in the objective function captures the “adjustment costs.” It is assumed that – to turn consumer goods into installed new machines – some output is used up in the installation process. The quadratic functional form is the simplest that can ensure that the installation costs for capital rise more than in proportion to the amount of investment being undertaken. This assumption is needed if investment is to adjust gradually to changing conditions, and it is necessary for the production-possibility frontier (between consumption and capital goods) to be the standard bowed-out shape.

The Hamiltonian for this optimization is

\[ \Delta = e^{-\lambda t} [F(N,K) - wN - I - b^t + q(I - K - bK)], \]

where $q$ is a Lagrange multiplier that denotes the value to the firm of a slight relaxation of the constraint – that is, the value of a bit more capital. Thus, $q$ is the relative price of capital goods in terms of consumption goods (and if equity markets function in an informed manner, $q$ also equals the value of equities). By using the “cook book” rule (applying it
This relationship states that investment is proportional to the gap between capital’s marginal product and its rental cost. Since capital’s marginal product is positively related to the employment of labour, this result "justifies" assuming that investment depends positively on output and negatively on the rate of interest. (This is standard in traditional IS-LM theory, although sometimes (as in Chapter 1 above), analysts simplify by excluding the income argument.)

If firms encounter a sales constraint (that is, if they pick factor inputs with the goal of achieving cost minimization without being able to simultaneously adjust employment and output to whatever values they want), the optimal investment function is:

\[ I = \alpha[F(w / P y)(F(w / \mu + \delta))] = 1. \]

Readers can verify this by re-specifying the Hamiltonian to

\[ \Delta = e^{-\mu[F(N,K) - s - \lambda I - \theta - \lambda K]} + \alpha[I - F(N,K)], \]

where \( \hat{\lambda} \) and \( \lambda \) represent the sales-constrained level of output and the Lagrangian multiplier attached to that constraint. The revised investment function emerges after \( \hat{\lambda} \) is eliminated by substitution — using the first-order conditions for labour. Some of the implications of this revision in the investment function were explained in our analysis of the extended Keynesian model of generalized disequilibrium in Chapter 1.

Returning to a setting with no sales constraints, it is worth drawing attention to the fact that the investment function that is consistent with profit maximization can be written in several different ways. We have already seen that it can be written as \( I = q(-\delta) \) and \( I = q[F(w / \mu + \delta)] = 1. \) A third alternative is also possible. Let \( K^* \) denote the optimal holding of capital in full equilibrium. In a no-growth setting, investment in full equilibrium is just for replacement purposes, so we have \( I = \delta K^* \). When this equation is combined with the accumulation identity for capital, \( I = \delta K^* \) we have \( \delta = (K^* - K) / K \). This way of summarizing optimal behaviour says that firms should follow a partial-adjustment rule when setting \( \mu \) investment. Net investment should equal a constant fraction of the gap between the actual and the desired capital stocks. Empirical workers have made this assumption for many years. The contribution of the formal micro-foundations is that we can appreciate that the partial adjustment coefficient should not be treated as a free parameter.

For the theory to receive empirical support, we must find that the estimated partial adjustment coefficient must equal the depreciation rate.

It was mentioned above that we could interpret the Lagrange multiplier, \( q \), as the value of stocks. In fact, it can also be interpreted as the slope of the nation’s production possibilities curve as well. We defined both interpretations now. In a well-functioning stock market, the value of equities equals the present value of the income derived from owning capital. If capital is held into the indefinite future, its per-period earnings will be national output, \( P(N,K) \), minus the wage bill, \( W \). To obtain the present value of this flow, it is discounted by the sum of the real interest rate and the depreciation rate. (Not only must this be done to calculate present value, but also the capital stock must be maintained.) Assuming constant returns to scale, we have \( F(N,K) = P(N,K) \). Using this fact and \( F(w / \mu + \delta) \), the market value of equities can be re-expressed in nominal terms as \( P(w / \mu + \delta) \). We can define \( q \) as the ratio of the market’s valuation of capital to its actual purchase price, \( P(w / \mu + \delta) \). This means \( q = F(w / \mu + \delta) \). This is most appealing.

When shares can be sold for such a price that capital can increase at the same rate as the ownership of the company is being diluted, and there are some additional funds left over, the existing owners should approve expansion. This was the intuition behind the Keynes-Tenho (1969) approach to the investment function. It is reassuring to know that we can embrace a model of investment that is consistent with both this Keynesian intuition and formal inter-temporal optimization.

Without adjustment costs, there is no difference between consumption goods and investment goods. In that case, the economy’s production possibilities curve (drawn in C-I space) is a straight (negatively sloped) 45-degree line. But with adjustment costs, this is not the case. Ignoring government spending for a simplified exposition, the amount of goods available for consumption is given by:

\[ C = F(N,K) - I - \delta. \]

The slope of the production possibilities curve is had by taking the total differential of this definition, while imposing the condition that factor supplies are constant (\( dI = d\mu = 0 \)). The result is:

\[ dc = \mu - I, \]

since we know from our earlier derivations that \( g = 1 - 2\mu \). Thus, Tobin’s valuation ratio can also be interpreted as the relative price of investment goods in terms of consumption goods. As long as macro theorists specify the resource constraint to include the installation cost term, the production possibilities curve has its normal bowed-out shape, and the model involves an aggregation of the two kinds of goods (even though it appears as simple as a one-sector model).

Other versions of the firm’s investment function emerge if we specify alternative installation-cost functions. For example, with an installation-cost specification that normalizes for the size of the firm (such as \( k = 1/K \) or \( b = b^2 / Y \)) we get slightly different investment functions:

\[ (K = a(q - 1)) \quad \text{and} \quad (Y = a(q - 1)). \]

In this last specification, since the investment process involves labour, there is also an important revision in the labour demand function:

\[ (1 + b^2 / Y) = W / P. \]

With this specification, the separation of demand and supply-side fiscal policy instruments is blurred. Anything, such as program spending, that can affect interest rates and therefore investment, is a policy variable that causes a shift in the position of the labour demand function. (With higher interest rates, fewer workers are needed for installing capital (at any real wage).) Since the labour market is what lies behind the aggregate supply curve for goods, \( G \) is a policy variable that shifts the position of both the demand and supply curves for goods. While some New Keynesian economists (such as Stiglitz (1992)) use models of equity rationing to give their model this very feature, space constraints limit our ability to pursue further these models involving interdependent aggregate supply and demand curves.

Finally, before closing this section, it is worth noting what happens if there were no adjustment costs. In this case, parameter \( b \) would be set to zero, and the first-order conditions imply that \( q \) would always be equal one, and that \( F(w / \mu + \delta) \) would hold at all times. In such a world, capital and labour would be treated in a symmetric fashion. Both factors could be adjusted costlessly at each point in time so that — even for capital — marginal product would equal rental cost. There would be no well-defined investment function in such a world, since firms would always save the optimal amount of capital. As a result, we cannot invest whatever households saved. We examine macro models with this feature, when we discuss economic growth in Chapters 10-12. But all models that focus on short-run fluctuations (that is, both Classical and Keynesian models of short-run cycles) allow for adjustment costs.

1.5 Firms’ Behaviour: Setting Prices

The purpose of this section is to explore the micro-foundations of what has come to be called the “new” Phillips curve. This relationship now forms an integral part of both New Classical macroeconomics and the New Neoclassical Synthesis. The particular micro model of sticky prices that is favored in the literature is Calvo’s (1983), so we explore that analysis here.

A full treatment would be very complicated. We would need to derive the firms’ factor demand functions (labour demand and investment) and the firms’ optimal price-setting strategy simultaneously — within one very general inter-temporal optimization. This is rarely attempted in the literature. Instead, it is assumed that there are two separate groups of firms. The first produces an “intermediate” product, and it is these firms who demand labour and capital. The second group of firms buys the intermediate product and (without incurring any costs, but subject to a constraint on how often selling prices can be changed) they sell them as final goods. This two-stage procedure is an ad hoc simplification that is intended to ”justify” our not integrating the optimal factor demands problem with the optimal-price-setting problem. Even with this separation a full treatment of the price-setting problem is more complicated than what is presented here. The fuller treatment involves product differentiation across firms. Strictly speaking, this is necessary, since each individual firm must have some monopoly power to have a price-setting decision. To ease exposition, however, the present discussion suppresses the formal treatment of monopolistic competition. By comparing this analysis to King (2000) and Goodfriend (1997), there are two ways to verify that the “bottom line” is unaffected by our following the many authors who take this short-cut.

Prices are sticky. Specifically, a proportion, \( \theta \), of firms cannot change their price each period. One way of thinking about the environment is to assume that all firms face a constant probability of being able to change price. That probability is \( (1 - \delta) \), so the average duration of each price is \( 1 / \delta \). Then, if \( p \) denotes the index of all prices ruling at each point in time and \( z \) denotes what is set by those who do adjust their price at each point in time, we have:

Muhammad Firman (University of Indonesia - Accounting)
\[ p_t = \theta p_{t-1} + (1 - \theta)z_t. \]

The objective function for each firm is to minimize a quadratic cost function – the discounted present value of the deviations between (the logarithms of) its price and its nominal marginal cost, \( mc \). The discount factor is \( 1/(1 + \rho) \), and the cost function is:

\[
\sum_{t=0}^{\infty} \frac{(1/(1 + \rho))^t}{\varphi(E_t(p_t - mc_{t,r}))^2}
\]

It is shown in the following paragraphs that the first-order condition for this problem can be approximated by the following equation (if the discount factor is set to unity):

\[(p_t - p_{t-1}) = (p_{t-1} - p_t) + \mu(mc_t - p_t),\]

where the \( e \) superscript denotes expectations. Assuming perfect foresight and switching to a continuous time specification, we can write

\[ \ddot{p} = \mu(mc_t - p_t). \]

To appreciate how we can reach these outcomes, we must first motivate the initial objective function, and then outline the derivation. Firms discount the future for two reasons: the normal rate of time preference applies, and (as time proceeds) there is an ever smaller probability that they are stuck with a fixed price. Using \( x = 1/(1 + \rho) \) as the discount factor, the cost function for the \( h \)th firm that was introduced above can be simplified to

\[
\sum_{t=0}^{\infty} \frac{(1/(1 + \rho))^t}{\varphi(E_t(p_t - mc_{t,r}))^2}
\]

\[ = \sum_{t=0}^{\infty} \varphi(E_t(p_t - mc_{t,r} + mc_{t,r} - 2p_t mc_{t,r})). \]

We differentiate this objective function with respect to the firm's choice variable, \( p_t \), and obtain

\[ p_t = \theta p_{t-1} + (1 - \theta)z_t. \]

We assume that a symmetric equilibrium holds in each period, so that \( p_t = z_t \) for all firms. To summarize the derivation thus far, we have:

\[ p_t = \theta p_{t-1} + (1 - \theta)z_t. \quad (4.1) \]

\[ z_t = (1 - \theta)\gamma E_t(mc_{t,r}) \quad (4.2) \]

We simplify (4.2) as follows:

\[ z_t = (1 - \theta)\gamma E_t(mc_{t,r}) + (1 - \theta)\partial \chi^1 E_t(mc_{t,r}) \]

\[ z_t = (1 - \theta)\gamma E_t(mc_{t,r}) + (1 - \theta)\partial \chi^1 E_t(mc_{t,r}) \]

\[ z_t = (1 - \theta)\gamma E_t(mc_{t,r}) + (1 - \theta)\partial \chi^1 E_t(z_{t-1}) \quad (4.3) \]

The next steps in the derivation are as follows. First, write (4.1) forward one period in time. Second, take the expectations operator, \( E_t \), through the result. Third, substitute the result into (4.3), then that result into (4.1). Finally, simplify what remains, using the definition of the inflation rate:

\[ \pi_t = p_t - p_{t-1}. \] We end with

\[ \pi_t = \chi E_t(z_{t-1}) + (1 - \theta)(1 - \theta)\partial \chi^1 E_t(mc_{t,r}) \quad (4.4) \]

where \( rmc \) stands for (the logarithm of) each firm's real marginal cost:

\[ rmc_t = \frac{mc_t}{\pi_t}. \]

Thus,

\[ \pi_t = \chi E_t(z_{t-1}) + \mu(rmc_t) \]

where \( \mu = ((1 - \theta)(1 - \theta)) \partial \chi. \] We approximate this relationship below by setting the discount factor, \( \chi \), equal to unity.

The final step in the derivation involves replacing the real marginal cost term with the output gap. This substitution can be explained as follows. We start with the definitions of nominal marginal cost, total product, and the marginal product of labour, assuming that the production function is Cobb-Douglas:

\[ MC = \frac{dTC}{dY} = \frac{d(WN)}{dY} = \frac{W(\partial N/\partial Y) = W/ MPL}. \]

\[ Y = N^\omega \]

\[ MPL = \sigma N / Y. \]

We combine these relationships with the household labour supply function (which was derived above, and which has been modified by replacing C with Y): \( N = (W/P)^{1/\sigma}. \) After eliminating MPL, \( N \) and \( Y \) by substitution, we have \( (MC/P) = Y^{1/\sigma} \), where \( \Omega = (1 + z)/\sigma. \) Re-writing this in logarithms, we have

\[ (mc - p) = \Omega z - \ln \sigma. \]
We pick units so that, in full equilibrium, price equals marginal cost equals unity. This implies that firms have monopolistic power only when they are out of long-run equilibrium, and it implies that the logarithm of both price and marginal cost are zero. Thus, the full-equilibrium version of this last equation is

\[(\bar{m} - \bar{p}) = \Omega y - \ln \sigma.\]

Subtracting this last relationship from the previous one, we end with

\[(\bar{m} - \bar{p}) = \Omega (y - \bar{y}).\]

When this relationship is combined with either the discrete-time or the continuous-time relationships derived above

\[(p_t - p_{t-1}) = (p_{t-1} - p_t) \gamma + \mu(m_t - p_t), \quad \bar{p} = -\mu(m_t - p_t),\]

the final result is the "new" Phillips curve:

\[(p_t - p_{t-1}) = (p_{t-1} - p_t) \gamma + \phi(y - \bar{y}), \quad \bar{p} = -\phi(y - \bar{y}),\]

where \(\phi = \mu \Omega.\) This completes the derivation of the closed-economy version of the "new" micro-based Phillips curve.

4.6 Conclusions

The hallmarks of modern macroeconomics are model-consistent expectations, and micro-based, not simply descriptive, behavioural reaction functions on the part of private agents. Chapter 3 contained the analysis that is needed to allow readers to solve macro systems involving model-consistent expectation, and the present chapter contains the analysis that is required to permit readers to understand the inter-temporal optimization that lies behind modern macro models.

The remaining task is to develop an understanding of how the policy implications of the "new" macro models differ from the more traditional descriptive models. We took a first pass at exploring this question in section 3.5 in the last chapter. The reader is now in a position to appreciate and verify that the system examined there is "new" – it involves both the "new" IS and Phillips-curve relationships, and it involves rational expectations. We now want to pursue this set of issues in more detail, and to do so, we proceed in two stages. First, we explore the work of the first group of macroeconomists to take micro-foundations seriously – the New Classics – in the next chapter. Then, we consider the extension that New Classics have embraced in recent years – sticky prices – in Chapter 6. New Classics refer to this extended model as a simple dynamic general equilibrium model with sticky prices. Others refer to it as a simple New Keynesian macro model. Still others refer to it as the "New Neoclassical Synthesis." Whatever the label, this compact structure now represents mainstream macroeconomics, and that is why we study its properties in some detail in Chapters 6 and 7.

5.1 Introduction

Our analysis thus far may have left the impression that all macroeconomists feel comfortable with models which "explain" short-run business cycles by appealing to some form of nominal rigidity. This may be true as far as macro policy-makers are concerned, but this has not been a good description of the view of many macro theorists. These theorists are concerned that (until recently) the profession has not been very successful in providing micro foundations for nominal rigidity. Even those who have pioneered the "new synthesis" of Keynesian and Classical approaches (for example, King (2000)) have expressed concern that some of its underlying assumptions concerning sticky prices can be regarded as "heroic."

The response of New Keynesian academics is to work at developing more convincing models of "menu" costs (the costs of changing nominal prices), and to elaborate how other features, such as real rigidities and strategic complementarities, make the basing of business cycle theory on seemingly small menu costs appealing after all (see Chapter 8). But there has been another reaction – on the part of New Classics. Their reaction to the proposition that menu costs "seem" too small to explain business cycles is to investigate whether cycles can be explained without any reference to nominal rigidities at all. They have seen remarkably successful in demonstrating that this is possible. Further, some revolutionary conclusions have been derived from this "equilibrium" approach to cycles. Perhaps the most central result is that he estimated benefit to society of completely eliminating business cycles may be trivial! This chapter explains this so-called real business cycle approach, and how it leads to this strong verdict regarding stabilization policy.

5.2 The Original Real Business Cycle Model

The unifying theme in New Classical work is "equilibrium" analysis. Markets always clear; agents make intelligent forward-looking plans; and expectations are rational. By adopting this framework, New Classics can rely on the existing general equilibrium analysis provided by microeconomists over more than a century. New Classics feel that when Keynesians focus on sticky prices and disequilibria, they lose the ability to benefit from this long intellectual heritage. New Classics see this as a big price to pay to make macroeconomics "relevant." This view becomes particularly firm when the New Classics feel that they may have demonstrated that the equilibrium approach may be just as relevant – even for explaining short-run cycles. One purpose of this chapter to permit the reader to form an independent decision on whether this claim of the New Classics can be sustained.

According to standard intermediate textbooks, Keynesian analysis explains the business cycle by assuming rigid money wages. Demand shocks push the price level up in booms and down in recessions. This set of outcomes makes the real wage rise in recessions and fall in booms, and the resulting variations in employment follow from the fact that wages slide back and forth along a given labour demand curve. Thus, the model predicts that the real wage moves countercyclically (and this is not observed), and because the labour market is not clearing, it makes the labour supply curve irrelevant for determining the level of employment.

New Classics prefer to assume that wages are flexible and that the labour market always clears. To their critics, a model which assumes no involuntary unemployment is an "obviously" bad idea. But if this is so, say the New Classics, it should be easy to reject their theory. While the approach has encountered some difficulties when comparing its predictions to the data, it has turned out to be much more difficult to reject this modelling strategy than had been first anticipated by Keynesians.

New Classics explain business cycles by referring to real shocks (shifts in technology) so the approach is called real business cycle analysis. The basic idea is that workers make a choice concerning the best time to work. If something happens to make working today more valuable (such as an increase in today's wage compared to tomorrow's wage), then workers make an inter-temporal substitution; they work more today and take a longer than usual vacation tomorrow. Similarly, an increase in interest rates lowers the present value of future wages, and so leads individuals to work more today.

If there is a positive technology shock today, the higher marginal product of labour implies a shift to the right of the labour demand curve. Similarly, a negative development in the technology field next period shifts the labour demand curve back to the left. Thus, according to this view, business cycles are interpreted as shifts in the position of the labour demand curve, not movements along a fixed labour demand curve. Cycles
\[
utility = F(Y, N, K) = (1 + \rho)\left[\ln(C_i) + \beta \ln(L_i)\right]
\]

Equation (5.1) is the household utility function; \( F \), \( \rho \), \( C \) and \( L \) denote expectations, the rate of time preference, consumption, and leisure.

Equation (5.2) is the time constraint (\( \mathbf{N} \) is employment). Formal utility maximization is used to derive consumption and labour supply functions. Equations (5.3), (5.4) and (5.5) define the production function and the one stochastic variable in the system – the technology shock, \( z \). An ongoing trend is involved, and the stochastic part is a normally distributed error term with zero mean and constant variance, and \( \phi \) is the coefficient of serial correlation in this econometric process. Formal profit maximization is used to derive the demands for capital and labour. Finally, equations (5.6) and (5.7) define the accumulation of the capital stock (\( \delta \) is depreciation rate and \( i \) is gross investment). New Classicals refer to the capital stock accumulation identity as a "propagation" or "persistence-generation" mechanism. Like the Keynesian assumption of staggered overlapping wage or price contracts, this mechanism implies persistence to the system's dynamics, and this gives the model a serious chance to fit the facts.

While it has now become common practice for practitioners to estimate simple aggregate models like this one, New Classicals initially preferred to pick plausible values for a few parameters, and then to use the model to generate data (by simulation). They compared the moments of the time series that were generated by stochastic simulations from the calibrated model, to the moments of the various time series from a real (usually the US) economy. If the model's data "looked like" the real world data, researchers concluded that this simple approach has been vindicated. Of course, to avoid this exercise being circular, analysts must choose the parameter values on the basis of empirical considerations that are not econometric papers involving aggregate time series data.

Hansen and Wright report quarterly simulations with the following chosen parameter values. The rate of time preference is set at 0.01 per quarter, which implies 4 percent on an annual basis. We have learned that (with infinitely lived agents) this parameter should be the same as the average real interest rate, so 4 percent is a plausible value.

Utility function parameter \( \beta \) is chosen so that the average proportion of discretionary time spent working is one-third – a value consistent with time-use studies. The production function exponent \( \alpha \) is set at 0.36, a very plausible value for capital's share of output in the United States (as observed in the national accounts). A similar reference justifies that 10 percent of the capital wears out annually (so \( \delta = 0.025 \)). Finally, the technology shock process was calibrated by assuming a standard deviation for the error term of 0.007 and a serial correlation coefficient of 0.95. These values were borrowed from one of the original contributions in this field (Kydland and Prescott (1982)) who in turn obtained these values by estimating the combination of equations (3) - (5) with time series data for the United States. By estimating the Cobb-Douglas production function with data for \( Y, N \) and \( K \) only, the estimated residuals – the so-called "below residuals" – can be taken as data for \( z \). Thus, except for the parameters which define this error process, all coefficient values are chosen from sources other than the time series that the modelers are trying to replicate. Nevertheless, since the coefficients for equation (5.5) were chosen so that the structures series properties of GDP would be simulated well, we cannot count that feature of the results as any victory at all.

But there is an achievement nonetheless, since other important stylized facts of the business cycle are well illustrated by the "data" that is generated from this very simple structure. For example, the simulations show that consumption is less variable than income, and that investment is more volatile than income. This outcome follows from the assumption of diminishing marginal utility of consumption. If a positive technology shock occurs today, people can achieve higher utility by spreading out this benefit over time. So they increase consumption by less than their income has increased initially, and as a result, some of the new output goes into capital accumulation. Since this additional capital makes labour more productive in the future, even a one-time technology shock has an impact for a number of periods into the future. It is impressive that these very simple calibrated models can mimic the actual magnitude of investment's higher volatility relative to consumption – and this has been accomplished without researchers allowing themselves the luxury of introducing "free" parameters.

Despite these encouraging results, however, this model does not generate the wide variations in employment and the very low variability in real wages (that we observe in real data) without an implausibly high value for the wage elasticity of labour supply (an unacceptable value for parameter \( \beta \)). Thus, in the next section of the chapter, we consider a number of extensions to the basic New Classical model that have been offered as mechanisms that can make the "data" that is generated from the calibrated models more representative of the actual co-movements in real wages and employment.

5.3 Extensions to the Basic Model

The first extension we consider is an attention in the utility function. The one given in equation (5.1) embodies the assumption that the marginal utility of leisure in one period is not affected by the amount of leisure enjoyed in other periods. When that function is changed so that this separability is removed, it makes agents more willing to substitute their leisure across time, and so respond more dramatically to wage changes. Keynesians regard this extension as ad hoc. They say that the Classicals are introducing "free" parameters (like the Classicals accuse the Keynesians have been doing when they assume arbitrary elements of nominal rigidity) just to make the model fit. If the whole point of the New Classical approach is to have a simple and standard market-clearing model that fits the facts, then such an adjustment made after it was found to be necessary indicates failure to some New Keynesians. In any event, the modern approach to adapting the utility function (central to much recent New Classical work) is to specify that today's utility depends on both today's level of consumption and today's level of habits. Habits evolve over time according to the following relationship:

\[ h_{t+1} = \alpha h_t - \beta h_{t+1} \]

This relationship is an additional persistence-generation mechanism that helps both calibrated and estimated models match real-world data (see, for example, Bouakez et al. (2005)). But since the habit-adjustment process is identical to the adaptive expectations formula, critics argue that New Classicals cannot simultaneously argue that rational expectations is fundamentally more appealing than adaptive expectations, and that this extension to the standard utility function is not ad hoc. Of course, as a technical matter, New Classicals can safely ignore such criticism. Since our subject starts from a specification of tastes and technology, every assumption about such matters is necessarily "arbitrary," and it seems to demonstrate a misunderstanding of the bounds of our discipline to call any such assumption "ad hoc." On the other hand, since the hypothesis of adaptive expectations concerns the relationship between actual and forecasted values of endogenous variables, not exogenous items such as the definition of tastes, it is legitimately viewed as an arbitrary (ad hoc) specification.

It is not useful for us to get bogged down in methodological dispute. At the practical level, two considerations are worth mentioning. First, since it is difficult to find any non-time-series-econometrics evidence to use as a basis for choosing an "appropriate" value for the habit-persistence parameter, \( \lambda \), it has been difficult for practitioners to
avoid some proliferation of the free-parameter problem as they implement this extension. The second point worth noting is that this generalization of the utility function does not adequately repair the real wage-employment correlations, so other changes to the basic model have become quite prevalent in the literature as well. It is to some of these other modifications that we now turn.

One such extension is indivisible labour. Some New Classical models make working an all-or-nothing choice for labour suppliers. At the macro level, then, variation in employment comes from changes in the number of people working, not from variations in average hours per worker. This means that the macro correlations are not pinned down by needing to be consistent with evidence from micro studies of the hours supplied by each individual (that show a very small elasticity). For more detail, see Hansen (1985) and Rogerson (1988).

Another extension focuses on endogenous market activity. Statistical agencies in OECD countries have estimated that, on average, households produce items for their own consumption equal in value to about one-third of measured GDP. Thus, “home” production is a very significant amount of real economic activity. Benhabib, Rogerson, and Wright (1991) have shown that when the real business cycle model involves this additional margin of adjustment, its real wage-employment correlations are much more realistic. The basic idea is that the amount of leisure consumed can remain quite stable over the cycle – even while measured employment in the market sector of the economy is changing quite dramatically – when households have the additional option of working at home (doing chores for which they would otherwise have paid others to do).

Other New Classical models, for example Christiano/Eichenbaum (1992) and McGrattan (1994) have introduced variations in government spending. This work involves adding an additional (demand) shock (in addition to technology shocks) to the model; for example,

\[ G_t = (1 - \gamma) G_{t-1} + u_t, \]  

(5.8)

is added to the system, and the market clearing condition is changed to

\[ Y_t = C_t + I_t + G_t. \]  

(5.7a)

When representative values for the additional persistence-generating parameter \( \gamma \) and for the variance of this additional error-term are included in the simulations, the resulting real wage-employment correlations look much more realistic. It is straightforward to understand why. Increases in government spending raise interest rates, and higher interest rates decrease the present value of working in the future. With the relative return to working in the current period thereby increased, the current-period labour supply curve shifts to the right. With these supply-side shifts in the model, some of the variations in employment are explained by shifts along a given labour demand curve. With this additional source of employment fluctuation, the magnitude of the technology shock does not need to be as great for the calibrated model to generate realistic changes in employment, in addition, with both labour supply and demand curves shifting back and forth, wide employment variations can easily occur with very modest changes in the real wage. That is, as long as the labour supply curve shifts sick and forth enough, its steepness is no longer a concern. Needless to say, while Keynesian models of the transmission mechanism quite differently, they are delighted to see the new school of thought relying on autonomous expenditure variations – a central concept in traditional Keynesian models – to improve the new model’s predictions.

It is instructive to depict the New Classical model in terms of aggregate demand and supply curves in price-output space. That picture appears just as in Figure 1.1 in the first chapter, but the list of shift influences for the aggregate demand and supply curves are different from that variables cause shifts in the textbook version of the classical framework. For simplicity, in the present discussion, we assume static expectations, so no separate expected future consumption and expected future wage rates need to be considered. But when the labour supply and demand equations are combined (to eliminate the real wage by substitution), we are left with a relationship that stipulates output as a positive function of the interest rate. We can use the IS relationship to replace the interest rate. The resulting sum of the labour demand, labour supply, production-function, and IS relationships is a vertical line in \( P-Y \) space with government spending and tax rates – in addition to the technology shock – as shift influences.

The real business cycle model allows no role for the money supply, so – to have the model determine nominal prices – we must add some sort of LM relationship to the system. Initially, to avoid having to re-specify the household’s optimization, this relationship was assumed to be the quantity-theory of money relationship \( L(Y) = M/P \) – justified as a specification of the nation’s trading transactions technology. This relationship is the economy’s aggregate demand function, and the nominal money supply is the only shift influence. Thus, the New Classical model still exhibits the classical dichotomy as far as monetary policy is concerned. But fiscal policy – even a change in program spending – has a supply-side effect, so it influences output, via it influences both the appropriate aggregate demand and supply diagrams to compare the effects on output and the price level of variations in autonomous expenditure across several models – those examined in Chapter 1 and the New Classical model of the present chapter. Since the traditional Keynesian model involves the idea that the price level falls during business-cycle booms, while this New Classical model has the property that the price level falls during booms, various researchers (for example, Cover and Pecorino (2004)) have tried to exploit this difference in prediction to be able to discriminate between these alternative approaches to interpreting cycles.

Care must be exercised when pursuing this strategy, however, since there is no reason to restrict our attention to an LM relationship that does not involve the interest rate. As we see in section 5.4 below, it is not difficult to extend the household-optimization part of the New Classical model to derive such a more general LM relationship from first principles. When this more general specification is involved in the model, the IS function is needed to eliminate the interest rate both from the labour-market relationships (to obtain the aggregate supply of goods function) and from the LM relationship (to obtain the aggregate demand for goods function). As a result, changes in autonomous spending shift the position of both the aggregate supply and demand curves, and the model no longer predicts that the price level must move contra-cyclically.

Other extensions to the basic New Classical model can be understood within the same labour supply and demand framework that we have just described. Any mechanism which causes the labour demand curve to shift over the cycle decreases the burden that has to be borne by technology shocks, and any mechanism which causes the labour supply curve to shift accomplishes the same thing – while at the same time decreasing the variability of real wages over the cycle. The variability of price mark-ups over the cycle is an example of a demand-shift mechanism, and households shifting between market-oriented employment and home production is an example of a supply-shift mechanism. Regarding the former, we know that the mark-up of price over marginal cost falls during booms because of the entry of new firms. This fact causes the labour demand curve to shift to the right during booms. Devereux, Head, and Lapham (1993) have shown how this mechanism can operate within a real business cycle model involving imperfect competition.

There are still other reasons for the labour demand curve to move in a way that adds persistence to employment variations. Some authors (such as Christiano and Eichenbaum (1992) introduce a payment lag. If firms have to pay their wage bill one period after receiving their sales revenue, the labour demand function becomes \( F_t = (W_t/P_t)(1+r) \) so variations in the interest rate shift the position of the labour demand curve. Further, firms may encounter adjustment costs when hiring/firing labour, and firing-by-doing business is an important mechanism (see Cass and Johni (2002)). In the latter case, tomorrow’s productivity is high if today’s employment level is high. Simulations have shown that the consistency between the output of calibrated equilibrium models and real-world time series is increased, when persistence-generation mechanisms such as these are added. It can be challenging to discriminate between some of these mechanisms. For example, there is a strong similarity between the habits extension and the learning-by-doing extension. But despite this, Bouakez and Kano (2006) conclude that the habits approach fits the facts better.

As noted above, it is interesting to note the convergence involved with parts of New Keynesian and New Classical work. Keynesianists have been taking expectations and micro foundations more seriously to improve the logical consistency of their systems, while Classicalists are embracing such things as autonomous expenditure variation, imperfect competition and payment lags to improve the empirical success of their models. Despite this continual evolution, however, there still is a notable difference in emphasis. Classical models have the property that the observed fluctuations in unemployment have been chosen by agents, so there is no obvious role for government to reduce output variations below what agents have already determined to be optimal. This presumption of social optimality is inappropriate, however, if markets fail for any reason (such as externalities, moral hazard, or imperfect competition).

Hansen and Wright (1992) have shown that when all of these extensions are combined, a simple aggregate model can generate data that reflects fairly well the main features of the real-world wage-employment correlations after all. But still the model does not fit the facts.
well enough, so even pioneers of that approach (for example, Goodfriend and King (1997) have called for the New Neoclassical Synthesis in which temporarily sticky prices are added to the real business cycle model. We consider this synthesis in some detail in Chapter 6.

It may seem surprising that New Classicals have embraced the hallmark of Keynesian analysis – sticky prices. Why has this happened? Perhaps because there is one fact that appears to support the relevance of nominal rigidities – the well-documented correlation between changes in the nominal money supply and variations in real output. Either this is evidence in favour of nominal rigidities, or it is evidence that the central bank always accommodates – increasing the money supply whenever more is wanted (during an upswing). In response to this reverse causation argument, Romer and Romer (1989) have consulted the minutes of the Federal Reserve’s committee meetings to establish seven clear episodes during which contractionary monetary policy was adopted as an unquestionably exogenous and discretionary development. The real effects that have accompanied these major shifts in policy simply cannot be put down to accommodative behaviour on the part of the central bank. Further evidence is offered in Romer and Romer (2004), and further support is provided by Ireland’s (2003) econometric results. This evidence – especially that which is derived from independent evidence of the central bank’s deliberations – removes the uncertainty that remains when only statistical causality tests are performed. One final consideration is that real and nominal exchange rates are very highly correlated. Many economists argue that this appears to be due to account for this fact other than by embracing short-run nominal rigidities.

Keynesians welcome this convergence of research approaches, yet (at the conceptual level) they remain concerned about the loss of market failure involved in the classical tradition. Also, they have some empirical concerns, and a few of these are summarized in the next few paragraphs.

The real business cycle approach is built upon the notion of inter-temporal substitution of labour supply. However, micro studies of household behaviour suggest that leisure and the consumption of goods are complements, not substitutes (as assumed in New Classical theory). Another awkward fact is that, in the United States at least, only 15 percent of actual labour market separations are quits. The rest of separations are layoffs. In addition, the data on quits indicates that they are higher in booms. The real business cycle model predicts that all separations are quits and that they are higher in recessions. Finally, it is a fact that a high proportion of unemployment involves individuals who have been out of work for a long time – an outcome that does not seem consistent with the assumption of random separations.

Other interpretation disputes stem from the fact that it is impossible to observe the technology shocks directly. In Soltow’s original work, the residual accounted for 48 percent of the variation in the output growth rate. Later work, which measured inputs more carefully, avoided some aggregation problems, and allowed for a variable utilization rate for both labour and capital, pushed the residual’s contribution down to 3 percent. It is no wonder that New Classicals can explain a lot with the original Soltow residuals; they contain a lot more than technology shocks. Incidentally, many analyses find it reassuring that the residuals are now perceived to be much smaller. Surely, if there are distortions due to negative technology shocks, disturbances at the individual firm or industry level would largely “cancel out” each other, so that, in the aggregate, there would not be large losses in technological knowledge. Possibly it would be better if New Classicals interpreted real shocks more broadly, and included such things as variations in the relative price of raw materials, as well as technology shocks, in what they consider as supply-side disturbances.

Quite apart from all those specific details, and even the general question of real-wage-employment correlations, some economists regard the inter-temporal substitution model as outrageous. In its simplest terms, it suggests that the Great Depression of the 1930s was the result of agents anticipating World War II and deciding to withhold their labour services for a decade until that high inflation period passed. Summers (1986) remarks that even if workers took such a prolonged voluntary holiday during the 1930s, how can the same strategic behaviour be posited for the machines that were also unemployed?

Given these problems, why does real business-cycle theory appeal to many of the best young minds of the profession? Blinder (1987) attributes the attraction to “Lucas’s keen intellect and profound influence,” but it also comes from the theory’s firm basis in macroeconomic principles and its ability to match significant features of real-world business cycles. Rebelo (2005) provides a clear and balanced assessment of both the successes and some of the challenges that remain for the research agenda of real business cycle theorists. It is likely that both this ongoing willingness to address these challenges, and the shift of the New Classicals from calibration to estimation, have strengthened the appeal of this approach to young researchers. Finally, perhaps another consideration is that this school of thought’s insistence on starting from a specification of utility makes it possible for straightforward normative (not just positive) analysts to be considered. The theory is so explicitly grounded in a competitive framework with optimizing agents who encounter no market failure problems, the output and employment calculations are not just “fairly realistic”; they can be viewed as optimal responses to the exogenous technology shocks that hit a particular economy.

Using this interpretation, economists have a basis for calculating the welfare gains from stabilization policy. Lucas (1987) has used data on the volatility of consumption over the cycle and an assumed degree of curvature in the utility of consumption function that appears to fit some facts to calculate how much business cycle volatility could be eliminated aggregate consumption variability entirely would be the equivalent in utility terms of an increase in average consumption of something like one or two tenths of a percentage point. Lucas’ conclusion has been influential; it is one of the reasons macroeconomists have shifted their emphasis to growth theory (Chapters 10-12) in recent years.

It is interesting to interpret the simulations produced by modern real business cycle theorists as an up-dated version of Adelman and Adelman (1959). These authors performed a similar stochastic simulation experiment with a small econometric model; their intention was to show that a standard Keynesian model (with just a few numerical parameters) could mimic the actual US data. Since both groups (Old Keynesians and New Classicals) have established that their models are consistent with significant parts of actual business-cycle data (and therefore should be taken seriously), how can any one of them argue that its preferred approach should have much more weight or research agenda? For this reason alone? Even New Keynesians, for example Ambler and Phaneuf (1992) have shown that an updating of the original Adelman/Adelman study gives the same support to the New Keynesian approach. Now that all groups have proved that their approach has passed this basic test – to be taken seriously as a serious alternative – the most interesting line of thought – it seems that either some additional criteria for choosing among the different approaches is required – or a synthesis of the New Classical and sticky-price approaches should be embraced. Indeed, as we explore in the next chapter, this synthesis has been just what has developed.

Before ending this chapter and moving on to the synthesis model, we do two things. First, we use a particular version of New Classical theory to illustrate how this school of thought can be used to contribute to policy debates. In particular, we show how it facilitates our estimating the long-term benefits of adopting a low-inflation policy. Second, we pursue Lucas’ proposition that the value of stabilization policy is trivial.

5.4 Optimal Inflation Policy

To justify a zero-inflation target, many analysts make the following argument. Since inflation is a tax on the holders of money, and since taxes create a loss of consumer surplus known as an “excess burden” – that tax (inflation) should be eliminated. The problem with this argument is that if one tax is eliminated, the government must raise another tax (and that other tax creates an excess burden of its own). Recognizing this, the standard approach in public finance is to recommend the “inverse elasticity rule” for setting taxes. Since excess burdens are bigger when taxes cause large substitution effects, the efficiency criterion for judging taxes stipulates that the largest tax rates should apply to the items that have the smallest price elasticities of demand. Since the interest rate is the (opportunity) cost of holding money, and since the estimated interest elasticity of money demand is very small, the inverse elasticity rule can be used to defend a relatively large tax on money. In other words, it supports choosing an inflation rate that is (perhaps well) above zero.

The full-equilibrium benefits of low inflation are independent of the complexities of the transition path that the economy takes to reach the long run (such as those caused by nominal rigidities). As a result, all analysts agree that a basic version of the New Classical model is the appropriate vehicle to use for estimating the benefits of that policy. It is true that a model with nominal rigidities is needed to assess the short-term costs of reducing inflation, and that is why we use such a model for addressing this issue in Chapter 7. But here, since our focus is on the long-term benefits of low inflation, we use an example of New Classical work that highlights money – Manmohan and Moktan (2004).
Households maximize a standard utility function. As usual, instantaneous utility is a weighted average of leisure and consumption \((\alpha \ln(1 - N) + (1 - \alpha)\ln C)\), and there is a constant rate of time preference, \(\rho.\) The budget constraint is

\[ C = wN + rK + r - \Delta. \]

Consumption is the sum of wage and employment income, plus the transfer payments received from the government, \(r\), minus the inflation tax incurred by holding real money balances and minus asset accumulation. Since \(A = K + m\) and \(l = r + x\), we can re-express the constraint as \(C = wN + rK + r - \Delta = A.\) Households do not focus on the fact that their individual transfer payment may depend on how much inflation tax the government collects from them individually. That is, they do not see the aggregate government budget constraint (given below) as applying as the individual level. However, there is an additional complication that confronts households. This novel feature is the "cash-in-advance" constraint: each period's consumption cannot exceed the start-of-period money holdings. Since the rate of return on bonds dominates that on money, individuals satisfy the financing constraint as an equality:

\[ C = m. \]

Thus, we replace the \(m\) term in the constraint with \(C.\) Finally, for simplicity, capital does not depreciate, and since there is no growth or government program spending, investment is zero in full equilibrium, and so (in full equilibrium) total output and \(C\) are identical. Firms have a Cobb-Douglas production function (with capital's exponent being \(\theta\)), and they hire labour and capital so that marginal products equal rental costs. The full-equilibrium version of the model is described by the following equations:

\[ r = \rho \]

\[ (N/(1 - N)) = (1 - \alpha)(1 - \theta)/(\alpha(1 + r + \pi)) \]

\[ C = K^\theta N^{\alpha} \]

\[ BC/K = r \]

The first equation is the Ramsey consumption function when consumption growth is zero (that follows from the household differentiating with respect to variable \(C\)). The second equation is what emerges when the labour supply function (that follows from the household differentiating with respect to variable \(N\)) is equated with the firms' labour demand function. The third equation is the production function, and the fourth is the other relationship that follows from profit maximization - that capital is hired to the point that its marginal product equals the interest rate. The final three equations can be used to determine how consumption, employment, and the capital stock respond to different inflation rates. As noted, the government budget constraint is

\[ r = \pi C. \]

This equation states that lump-sum transfer payments, \(r\), are paid to individuals, and in aggregate, these transfers are financed by the inflation tax.

With \(\pi\) endogenous, cutting inflation is unambiguously "good." Lower inflation eliminates a tax that distorts the household saving decision, and no other distortion is introduced by the government having to levy some other tax to acquire the missing revenue. Consumption, employment, output, and the capital stock all increase by the same percentage when the inflation rate is reduced. Specifically,

\[ \frac{(C(t)/C(0))}{(1 - \pi)} = (1 - \alpha)(1 - \theta)/(\alpha(1 + r + \pi)). \]

Maasoumi and Molinari calibrate the model with standard real business-cycle assumptions \(\alpha = 0.64, \rho = 0.042, \theta = 0.31\) and they assume an initial inflation rate of zero: \(\pi = 0.\) These assumptions allow them to evaluate the inflation multiplier. The result is that creating inflation of 2% lowers steady-state consumption by 1.37%. This outcome is an annual amount. With no growth and a discount rate of 0.042, the present value of this annual loss in consumption is 0.0137/0.042 = 32% of one year's level of consumption. Most analysts regard this magnitude as quite large. The policy implication is that even a two-percent inflation rate should not be tolerated.

While we do not use a formal model to derive an estimate of the transitional costs of lowering the inflation rate in this chapter, it is worth pointing out what the magnitude of these costs turns out to be. Experience has shown (see Ball (1984)) that we suffer an increase in the GDP gap of about two percentage points for about 3.5 years to lower the inflation rate by two percentage points. This means that we lose approximately percentage points of one year's GDP to lower steady-state inflation this amount. The benefit-cost analysis says that the present value of 1 benefit exceeds this present cost (32% exceeds 7%), so disinflation supported.

This is the standard defense for targeting zero inflation. It is an application of the neoclassical synthesis. The long-run benefits of lower inflation are estimated by appealing to our theory of the natural rate (New Classical macroeconomics). The short-run costs are estimated by appealing to more Keynesian models for how some shocks can give rise to natural rate that stem from temporary nominal rigidities (our Chapter 3 model if micro-foundations are not stressed, our Chapter 6 model if they are). In diagrammatic terms, the reasoning is illustrated in Figure 5.1, where we continue to abstract from any ongoing growth. New Classical analysis is used to estimate the shift up in the potential GDP line that accompanies disinflation, and the short-run synthesis model (in which potential GDP is exogenous) is used to calculate the temporary drop (and then later recovery) in actual GDP.

**Figure 5.1 Implications of Lower Inflation**

One unappealing aspect of this estimate of the long-term benefits of low inflation is that it involves the assumption that the monetary authority can dictate to the fiscal policy maker (and insist that the latter must cut transfer payments in the face of disinflation). How is the analysis affected if we assume that this is not possible? To explore this question, we add a tax on wage income. Assuming (as we have already) that the wage equals the marginal product of labour, we re-express the government budget constraint as

\[ r = \pi C + i((1 - \theta)/(C/N)). \]

There is one change in household budget constraint, since it must now stipulate that households receive only the after-tax wage. This results in only one change; the third equation in our list of the model's relationships becomes

\[ (N/(1 - N)) = (1 - \alpha)(1 - \theta)/(\alpha(1 + r + \pi)). \]

It is left for the reader to re-derive the effect on consumption of a change in the inflation rate, with \(r\) being the endogenous policy instrument instead of \(\pi.\) It is more difficult for disinflation to be supported in this case, since one distortion (the wage-tax) is replacing another (the inflation tax). Indeed, in this case, for the calibration assumed by Maasoumi and Molinari, it turns out that the "benefit" of lower inflation has to be negative!

Disinflation forces the fiscal authority to rely more heavily on a more distortionary revenue source than the inflation tax. Similar results in more elaborate calibrated models are reported in Cooley and Hansen (1991). It would therefore appear that the public-finance approach (that simplifies by using a New Classical model with no ongoing growth) does not lead to a solid under-pinning for a zero inflation target.

A somewhat more reliable argument for choosing a very low inflation rate (perhaps zero) as "best," concerns the effect of inflation on savings in a growth context. Most tax systems are not fully indexed for inflation. To appreciate why this is important, suppose you have a $100 bond that gives you a nominal return of 10%. Suppose that inflation is 5%, and that the interest rate on your bond would be 5% if inflation were zero. At the end of the year, the financial institution sends you a tax form indicating that you received $10 of interest earnings, and the government taxes you on the entire $10. In fact, however, only $5 of the $10 is interest earnings. The other $5 is compensation for the fact that the principal value of your investment has shrunk with inflation. An interest income tax system should tax only interest income, not the saver's depreciation expenses. An indexed tax system would do just this. Non-indexed tax systems make inflation amount to the same thing as a raising of the interest-income tax rate. Thus, inflation reduces the incentive to save, so that individuals living in the future inherit either a smaller capital stock (with which to work) or a larger foreign debt to service, or both. According to this analysis, then, to avoid a lowering of future living standards, we should pursue a "zero" inflation target. We evaluate this line of argument more fully in the economic growth chapters (10-12) later.
in this book. At this point, we simply assert what will be derived and explained here. If there are no "second-best" problems, growth theory supports the removal of all taxes on saving (such as inflation when the tax system is not indexed). But if there are second-best problems, this policy is not necessarily supported. Again, the case for zero inflation is far from complete.

5.5 Harberger Triangles vs. Okun’s Gap

Thus far, much of this book has focused on explanations of the business cycle and an evaluation of stabilization policy. It has been implicit that there would be significant gains for society if the business cycle could be eliminated. The standard defence for this presumption can be given by referring to Figure 5.2 (where we now allow for ongoing growth by drawing the log of the GDP time paths with a positive slope).

Without business cycles, actual output, \( y \), would coincide with the natural rate, \( y^* \), and both series would follow a smooth growth path such as the straight line labelled \( y^* \) in Figure 5.2. But because we observe business cycles, the actual output time path is cyclical – as is the wavy line labelled \( y \) in the figure. Traditionally, Keynesians equated the natural rate with potential GDP, and Figure 5.2 reflects this interpretation by having the actual and natural rates coinciding only at the peak of each cycle. Okun (1962) measured the area between the two time paths for the United States for a several decade long period of time, and since the average recession involved a loss of at least 5 percent of national output, the sum of the so-called Okun gaps was taken to represent a very large loss in material welfare. The payoff to be derived from a successful stabilization policy seemed immense.

![Figure 5.2 Output Gaps](image)

Before explaining how the New Classicals have taken issue with this analysis, it is useful to note how the likely payoff that can follow from successful microeconomic policy was estimated back when Okun was writing. As an example, consider a reduction in the income tax rate, which (since it applies to interest income) distorts the consumption-savings decision. In section 4.3, we derived the consumption function that follows from inter-temporal optimization on the part of an infinitely lived agent who is not liquidity constrained; the decision rule is:

\[
\frac{C}{I} = r(1 - \theta) - \rho
\]

where \( C, r, \rho, \) and \( \theta \) denote consumption, the real interest rate, the rate of time preference, and an income tax rate that does not exempt interest income (as we assumed in the previous section).

Traditional applied microeconomic analysis involved focusing on full equilibrium without growth (that is, on the \( r(1 - \theta) - \rho \) relationship, and combining this supply of savings function with the full-equilibrium demand for capital (\( F_K = r - \delta \), where \( F(K, N) = Y \) is the production function and \( \delta \) is the depreciation rate for capital). Assuming a fixed quantity of labour employed, a Cobb-Douglas function, \( Y = K^{\rho}N^{1-\rho} \), and that the rates of time preference and capital depreciation are independent of tax policy, these relationships imply

\[
\frac{dY}{dY} = [(r(1 + \delta)(0 - 1)(1 - \theta))(dr)] \frac{dY}{dt}.
\]

With representative parameter values (\( r = 0.3, \theta = 0.36, \delta = 0.03, \rho = 0.1) \) a 10 percent reduction in taxes (\( d\theta = -0.1 \) 10 percent reduction in taxes (\( d\theta = -0.1 \) involves an increase in national output of just one-half of one percent. This once-for-all gain in material welfare is just one-tenth the size of what Okun estimated to be the benefit of avoiding one recession. Since the analysis of micro distortions was often presented geometrically as a consumer surplus triangle by public finance specialists such as Harberger, Tobin (1977, page 468) concluded that "it takes a heap of Harberger triangles to fill an Okun's gap." Thus, traditional Keynesians have felt confident that stabilization policy was more important than microeconomic policy.

There has been a major change in thinking on these issues in recent years. For one thing, analysts no longer feel comfortable with equating the natural rate of output and potential GDP. A strict interpretation of real business cycle theory involves the presumption that there is no difference between the actual and the natural rate of output. There are simply variations in the level of output that are caused by stochastic elements in the production process. Some take a slightly less doctrinaire view of the New Classical approach. According to this view, there are differences between the natural and the actual output rates, and the natural rate is what can be sustained on an average basis. The economy operates below this level during downturns, and above this level during booms. Thus, a proper drawing of Figure 5.2 involves shifting the smooth natural rate line down so that it cuts through the midpoint of each up and down portion of the actual output time path. Using such a revised graph to calculate the total Okun's gap over a period of years gives a very different answer. Output losses are still incurred during recessions, but these losses are approximately made up for by the output gains during booms. A perfect stabilization policy would eliminate both the output losses and the output gains. Thus, it is possible that, on balance, the net benefit of (even a perfectly successful) stabilization policy is close to zero.

![Figure 5.3 Diminishing Marginal Utility and Risk Aversion](image)

Even if gains and losses did cancel out, there would still be some benefit to individuals as long as they are risk averse. That is, two income streams with the same present value are not evaluated as equal in utility terms if one income stream involves volatility. Figure 5.3 illustrates this issue. It shows that with risk aversion, an individual refuses a fair bet – for example, she refuses to pay $100 for the right to play a game in which there is a 50-50 chance of receiving either $150 or $50. The expected value of the game is $100, but – given the uncertainty – the utility that can be derived from this expected value is not as big as what is enjoyed when the $100 is certain. Thus, an individual with diminishing marginal utility is willing to give up an amount of utility equal to distance DB to eliminate the variability in her income stream. If the degree of risk aversion is very slight, the utility of income function is almost linear, and distance DB is very tiny. This is the reasoning that Lucas (1987) used in arriving at his estimate of the value of stabilization policy. Using a time-separable utility function with a constant coefficient of relative risk aversion (for which empirical demand systems yield an estimate), Lucas was able to quantify the benefits of eliminating variability, and as already noted, he concluded that they were trivial.

Keynesians have made three points in reacting to Lucas. The first concerns whether there is market failure. According to New Classicals, unemployment is voluntary, so when output is low it is because the value of leisure is high. What is so bad about an "output loss" if it is just another word for a "leisure gain"? But Keynesians think that a significant component of unemployment is involuntary, since it stems from some market failure such as asymmetric information, adverse selection, or externalities in the trading process. (We examine these possibilities in Chapter 8.) According to this view, smoothing is not the only result to follow from stabilization policy. Indeed, just the commitment to attempt stabilization may be sufficient to shift the economy to a Pareto-superior equilibrium in models that involve both market failure and multiple equilibria. Thus, stabilization policy can affect the mean, not just the variance, of income. In terms of Figure 5.2, stabilization policy can both reduce the wiggles in the y time line and shift up its intercept. Lucas' calculations simply assume that this second effect is not possible.
A second point concerns the distribution of the gains and losses over the business cycle. A relatively small proportion of the population bears most of the variability, so that the individuals are sliding back and forth around a much wider arc of their utility function (than Lucas assumed). Even staying within Lucas’ framework and numerical values, Pemberton (1995) has shown that this distributional consideration can raise the estimated benefit of stabilization policy by a factor of eight. An even bigger revision is called for if a different utility function is used. Pemberton notes that many experimental studies have cast fundamental doubt on the expected utility approach. Indeed, the equity-premium puzzle implies that we cannot have confidence in utility functions like the ones Lucas used. When some of the alternatives are used to rework Lucas’ calculations, it turns out that business cycles do involve significant welfare implications.

Thus far, our discussion of the relative size of Okun’s gap and Harberger triangles has ignored two things: what are the effects of tax changes before full equilibrium is reached? and what are the effects (if any) on the economy’s average rate of growth? These issues can be clarified with reference to Figure 5.4. As we have seen, a cut in the interest-income tax stimulates savings. As a result, current consumption must drop, as shown by the step down in the solid line time path in the left-hand panel of Figure 5.4. Individuals must suffer this lower standard of living for a time, before the increase in the stock of capital (made possible by the higher savings) takes place. The illustrative calculations have estimated the long-term gain (the step up in the dashed line in the left-hand panel of the figure) but not this short-term pain. Thus, the comparison of Okun gaps and Harberger triangles is not complete without a dynamic analysis of tax policy (which is provided in Chapters 10 and 12). But we must also note that a tax policy which stimulates savings may not just cause a once-for-all increase in the level of living standards. It may raise the ongoing growth rate of consumption, as shown in the right-hand panel of Figure 5.4. There is still a period of short-term pain in this case, but the effect on the present value of all future consumption can be more dramatic. Whether tax policy can have any effect on the long-run average growth rate has been much debated in recent years, and this debate is covered in the final two chapters of the book. But if it can, we would have to conclude that the size of Harberger triangles may be far bigger than earlier analysts had thought.

Does this mean that Lucas is right after all—that microeconomic policy initiatives are more important than stabilization policy? Not necessarily. As Fatas (2000), Barlevy (2004) and Blackburn and Pelloni (2005) have shown, there is a negative correlation between the variance of output growth and its mean value. It seems that more volatile business cycle is not conducive to investment, and so it contributes to a smaller long-run growth rate than would otherwise occur. Thus, endogenous growth analysis raises the size of both Harberger triangles and Okun gaps.

Figure 5.4 Effects on Consumption of Lower Interest-Income Taxes.

It seems that a prudent reaction to the Okun gap vs. Harberger triangle debate is to take the view that the profession should allocate some of its resources to investigating both stabilization policy and long-term growth policy (eliminating distortions). As we shall see in later chapters, the same analytical tools are needed to pursue both tasks.

5.6 Conclusions

Real business cycle theorists have convinced all modern macroeconomists of the value of explicit micro-foundations, and as a result they have made modern work much more rigorous than what preceded their challenge to that earlier literature. And now that New Classical have acknowledged that some form of nominal rigidity needs to be part of their model, we have a convergence of views. Over the last several years, a particular version of nominal rigidity (Calvo (1983))—that involves micro-foundations—has been added to the New Classical approach. The resulting “New Neo-Classical Synthesis” model (a real business cycle system with temporarily sticky prices) is quite similar to what New Keynesians had been developing independently. Indeed, when some other key features of New Keynesian work (for example, real wage rigidity stemming from incomplete information (discussed in Chapter 8)) are added to the real business cycle framework, the empirical applicability of the synthesis approach is enhanced even more. It is to these developments that we turn our attention in the next several chapters—before shifting our focus to long-term growth theory.

CHAPTER 6

THE NEW NEOCLASSICAL SYNTHESIS

6.1 Introduction

In this chapter, we analyze what has been called the “New Neoclassical Synthesis” in macroeconomics. As noted in earlier chapters, this chapter approaches to combine the best of two earlier schools of thought. First, it is consistent with the empirical “fact of life” that prices are sticky in the short run (the Keynesian tradition). Second, it is based on the presumption that the Lucas critique must be respected. That is, it is in keeping with the demands of the New Classicalists; both the temporary price stickiness and the determinants of the demand for goods must be based on a clearly specified inter-temporal optimization.

This synthesis involves the basic (infinitely lived representative agent) version of the inter-temporal theory of the household (derived in Chapter 4, section 3) to re-specify the IS relationship, and the similar theory of the firms (derived in Chapter 4, section 5) as a basis for a re-specified Phillips curve. The traditional (or “old”) IS relationship involves the level of aggregate demand (output) depending inversely on the interest rate, and the traditional Phillips curve involves the level of the inflation rate depending positively on the output gap. With the derived micro-based IS and Phillips curve relationships in Chapter 4, it appeared rather different: \( y = (r - \bar{r}) \) and \( \pi = \phi(y - \bar{y}) \). Thus the “new” IS relationship involves the change in aggregate demand depending positively on the interest rate, and the “new” Phillips curve involves the change in the inflation rate depending inversely on the output gap.

We investigated one aspect of monetary policy in a model involving these “new” relationships in a rational-expectations setting in Chapter 3 (section 5). Since that analysis was rather messy, in this chapter we simplify in three ways. First, we ignore stochastic shocks, so that rational expectations becomes the same thing as perfect foresight. Second, we use a continuous-time specification, so that a geometric approach—phase diagrams—can be used instead of algebra. Third, for most of the chapter, consider only one aspect of the new model at a time—initially, the new IS relationship with a traditional Phillips curve, and then a new Phillips curve with a traditional IS relationship. In each case, we wish to explore how (if at all) these changes in the model’s specification affect the answer to a standard stabilization policy question: what happens to output when the central bank embarks on a disinflation policy?

6.2 Phase Diagram Methodology

Chapter 2 focused on the first Neoclassical Synthesis—a model that combined traditional IS and Phillips curve relationships that were descriptive, not based on formal inter-temporal optimization. In somewhat modified notation, when a monetary-policy action function is added, that system can be defined by equations (6.1)–(6.3):

\[(y - \bar{y}) = -\phi(r - \bar{r}) \]  
\[
\pi = \rho\pi \]  
\[
r = \bar{r} + \kappa + \lambda(\pi - \pi) \]

The first equation (the aggregate demand function) states that output falls below its full-equilibrium value when the real interest rate rises above its full equilibrium value. The second equation (the dynamic aggregate supply function) states that inflation exceeds the authority’s target inflation rate whenever the actual rate of output exceeds the natural rate. The third equation states that the central bank raises the nominal interest rate above its full equilibrium value whenever the price level exceeds the bank’s target value for the price level. The slope parameters (the three Greek letters) are all positive.

Muhammad Firman (University of Indonesia - Accounting)
We focus on a contractionary monetary policy, the central bank lowers its target value for the price level in a once-for-all, previously unexpected, fashion. Further, we assume that – both before and after this change – the bank did maintain, and will then revert to maintaining, a constant value for that target variable (\(y\)). If we were to graph this exogenous variable – the level of \(x\) as a function of time – it would appear as a horizontal line that drops down in a one-time step fashion at a particular point in time. At that very instant, the slope of the graph is undefined, but both before and after that point in time, the slope, \(x\), is zero.

We are interested in knowing what the time graphs for real output and the price level are in the face of this one-time contractionary monetary policy. We learned how to answer this question in Chapter 2. We were able to re-write the model as a single linear differential equation in one variable, and from that compact version of the system, we could derive both the impact effect on real output, and the nature of the time paths after the policy change had occurred. Specifically, we learned that (as long as the system is stable) there is a temporary recession (which is biggest at the very instant that the target price level is cut). The output time path then starts rising asymptotically back up to the unaffected natural rate line, and the temporary recession is gradually eliminated (see Figure 2.2, p. 28). There is no jump in the price level; the Keynesian element of the synthesis is that the price level is a sticky variable. But while it cannot "jump" at a point in time, it can adjust gradually through time. In this case, it gradually falls (asymptotically approaches) the new lower value of \(x\). These properties represent the base for comparison in the present chapter. We want to know if the output and price-level time paths follow these same general patterns in a series of modified models.

The first modified model is defined by equations (6.1a), (6.2) and (6.3). The only change is that the traditional IS relation is replaced by the "new" IS function:

\[
\dot{x} = (r - \phi x)
\]

(6.1a)

To analyze this system, we are unable to use the methods of Chapter 2. This is because, with a second differential equation in the system, we cannot reduce it down to anything simpler than a set of two first-order ordinary differential equations. The purpose of this section is to explain how this system can be analyzed – first graphically (in what is known as a phase diagram) and then formally.

The first step in deriving the phase diagram is to reduce the system to just two differential equations that contain only the two endogenous variables that we most want to focus on. In this case, since we wish to highlight the output and price-level effects, we use the policy reaction function to eliminate the interest rate in the new IS function (after having used to Phillips curve to eliminate the inflation rate from the central bank’s reaction function). The result is:

\[
\dot{y} = \lambda (y - p) - (\phi y - \phi p)
\]

(6.4)

Equations (6.2) and (6.4) represent the compact version of the model. These relationships contain no endogenous variables other than the two we are focusing on – \(y\) and \(p\) – and also the time rates of change of no endogenous variables other than these same two. This is exactly the format we need, if we are to draw a phase diagram with \(y\) and \(p\) on the two axes. It is customary to put the “jump” variable – in this case, \(y\) – on the vertical axis, and the sticky-at-a-point-in-time variable – in this case, \(p\) – on the horizontal axis. We now explain how to use equations (6.2) and (6.4) to derive the phase diagram.

The goal is to draw two “no-motion” lines in a \(y-p\) space graph. The \((\dot{p} = 0)\) locus is all combinations of \(y\) and \(p\) values that involve \(p\) not changing through time. The \((\dot{y} = 0)\) locus is all combinations of \(y\) and \(p\) values that involve \(y\) not changing through time. The model’s full equilibrium involves neither variable changing, so – graphically – full equilibrium is determined by the intersection of the two no-motion loci. Only that one point involves no motion in both variables. To draw each no-motion locus, we must determine its three properties: What is the slope of the locus? What precise motion occurs when the economy is observed at a point that is not on this line? And How does this locus shift (if at all) when each exogenous variable is changed? We now proceed to answer all three questions for both no-motion loci.

The properties of the \((\dot{p} = 0)\) locus can be determined from the \(\dot{p}\) relationship – equation (6.2). When \(\dot{y} = 0\), this relationship reduces to \(\dot{y} = \dot{y}\), and this fact is graphed as the horizontal line in Figure 6.1, labelled \((\dot{p} = 0)\). So we have already answered question one; the slope of the \((\dot{p} = 0)\) locus is zero. What motion takes place when the economy is at a point off this line? The best way to answer this question is to assume that we are at such a point, say point \(A\) in Figure 6.1, and then determine what equation (6.2) implies about point \(A\). At \(A\), actual output exceeds the natural rate, and (according to equation (6.2)) \(\dot{p}\) must be positive at point \(A\) as a result. This is just a mathematical way of saying that “\(p\) is rising”, so we draw a horizontal line pointing to the right in this upper region of the diagram to show this rising price level. It may seem tempting to put an upward pointing arrow in the graph, since we are talking about “rising” prices. But we must remember that we are graphing \(p\) on the horizontal axis, so a rising value means an arrow pointing east, not north. Similar reasoning leads us to inserting a western pointing arrow below the \((\dot{p} = 0)\) locus. We have now summarized what is happening (with respect to the price level, at least) at every point in the plane. \(p\) is rising when it is observed at values above the line, falling when observed at points below the line, and not moving at all when observed at points on the line. The third (final) question of interest concerning the \((\dot{p} = 0)\) locus is: Does this line shift when the central bank lowers the value of \(x\) – its price-level target? Since this exogenous variable does not appear in equation (6.2), the answer is simply “no”; this policy does not shift the position of this no-motion locus. We now proceed to ask and answer these same three questions for the \((\dot{y} = 0)\) locus.

The properties of the \((\dot{y} = 0)\) locus follow from equation (6.4). To determine the slope of this no-motion locus, we substitute in the definition of no motion \((\dot{y} = 0)\), and solve for the variable that we are measuring on the vertical axis: our phase diagram.

\[
y = (\frac{\lambda}{\phi}) p + (\frac{1}{\lambda - \phi}) \dot{y}
\]

(6.4)

The slope expression is this equation is the coefficient on the variable that is being measured on the horizontal axis. Since this coefficient is \(\frac{\lambda}{\phi} > 0\), we know that the \((\dot{y} = 0)\) locus is positively sloped. The intercept is the term in square brackets. Within this term, the coefficient of \(x\) is \(-\frac{1}{\lambda - \phi}\), so we know that the reduction in \(x\) must increase the vertical intercept of the \((\dot{y} = 0)\) locus. Thus, we know that the contractionary monetary policy moves the positively sloped \((\dot{y} = 0)\) line up on the page. To answer the remaining question – what motion is involved when the economy is observed at a point that is not on this locus, we must revert to equation (6.4) – that does not involve our having imposed \(\dot{y} = 0\). From (6.4), we see that \(\dot{y} / \dot{x} = -\lambda / \phi\), which means that the time change in \(y\) goes from zero to a negative value as we move to a point off the line and above the line. Thus, we label all points above the line as involving \(\dot{y} < 0\), and this is what justifies the arrows pointing south in this region of Figure 6.2. Similarly, we label all points below the line as involving \(\dot{y} > 0\), and, as a result, we insert a northern pointing arrow – indicating this rising \(y\) motion for all observation points below the line in Figure 6.2.
There are four loci in Figure 6.4. The intersection of two – the \((y = 0)\) and the \((p = 0)\) lines – determines the full equilibrium of the system (the long-run outcome – point \(E\)). The intersection of the other two lines – the saddle path that cuts through the long-run equilibrium point and the pre-existing initial conditions constraint – determines the short-run outcome (point \(A\)). The saddle path is the line we need to jump on to if outright instability (which is present to the point \(E\) and is irrelevant on empirical grounds) is to be avoided, and the initial conditions line is what we can jump along. The intersection (point \(A\)) is the only point in the plane that is both feasible (given the historically determined starting price level) and desirable. While this methodology offers no discussion of a decentralized mechanism that would help individual agents coordinate to find point \(A\), it assumes – on the basis of instability not being observed – that agents somehow achieve this starting point. While this seems somewhat arbitrary, it must be realized that some additional assumption is needed to complete the model. After all, the two-equation system involves three endogenous items: \(p\), \(y\) and \(y\). With \(y\) being a jump variable, both its level and its time derivative are determined within the model. So an additional restriction is needed to close the model, and this additional restriction is that the system always jumps on to the relevant saddle path, the moment some previously unexpected event takes place. Any other assumption renders the system unstable, and therefore unable to be related to the (presumed to be stable) real world.

It is easier to appreciate how the model works by actually following through a specific event. We do just this by referring to Figure 6.5. The economy starts in full equilibrium at point \(A\) – the intersection of the initial no-motion loci. Then a once-for-all, previously unexpected, drop in \(x\) occurs, as the central bank performs this unprecedented monetary contraction. We have determined that this event shifts the \((y = 0)\) line up to the left – to what is shown as the dashed line in Figure 6.5. The new full equilibrium point is \(C\), but the economy cannot jump immediately to this point, because the price level is predetermined at a point in time. The initial conditions line (always a vertical line if we follow the convention of graphing the jump variable on the vertical axis) is the vertical line going through the initial equilibrium point \(A\). Since \(y\) is a jump variable, the economy can move – instantaneously – to any point on this line. To determine which point, we draw in the saddle path going through the new full equilibrium point \(C\). The intersection of this line with the initial conditions line determines the immediate jump point – \(B\).

The full solution is now summarized. The observation point jumps immediately from \(A\) to \(B\), then it travels gradually through time thereafter, from \(B\) to \(C\). So the contractionary monetary policy involves a temporary recession, and that recession gets ever smaller thereafter as real output asymptotically returns back to the natural rate. This is the same outcome that emerged in the first neoclassical synthesis model (Chapter 2), yet we have a new, not an old, \(A\) relationship in the present system. This similarity in results is "good news." Since many policy makers were educated several decades ago – before the new synthesis with its more thorough micro-foundations was developed – they find it difficult to shift to a new paradigm. It appears that this may not be a serious problem. The new synthesis analysis indicates that at least this one aspect of the earlier policy analysis is robust. (This is not true for all policies, but at least it is for some initiatives.)

Before analyzing related models with the phase diagram methodology, it is worthwhile noting how a more formal approach can yield slightly more precise predictions. In other words, it is useful to be able to calculate impact and adjustment-speed effects in models of this sort formally, rather than simply illustrating them geometrically. The remainder of this section is devoted to explaining how this can be done.

As with all differential equations, the form of the solutions (for the system defined by equations (6.2) and (6.4)) are:
where the $\delta$s are the characteristic roots, and the $z$s are determined by initial conditions. Since we are restricting attention to saddle path outcomes, we know that one characteristic root must be positive, and the other negative. Let us assume $\delta_1 > 0$ and $\delta_2 < 0$. By presuming a jump to the saddle path, the unstable root is being precluded from having influence, so $z_1 = z_2 = 0$ are imposed as initial conditions.

With a linear system, the equation of the saddle path must be linear as well:

$$ (y - \bar{y}) = \rho(p - \bar{x}) \tag{6.6} $$

where $\gamma$ is the slope of the saddle path. It is immediately clear that equations (6.5) and (6.6) are consistent with each other only if $z_1 = z_2$. Thus, once on the saddle path, all motion is defined by

$$ p_t = x + z_1 e^{\rho t} \quad \text{and} \quad y_t = \bar{y} + z_2 e^{\rho t} \tag{6.7} $$

We substitute equations (6.7), and their time derivatives, into (6.2) and (6.4) and obtain

$$ \delta_1 = \phi y \quad \text{and} \quad (\delta_1 + \phi) = \lambda. \tag{6.8} $$

These two equations can be solved for the stable root and the slope of the saddle path. The absolute value of the stable root defines the speed of adjustment, while the slope of the saddle path is used to calculate impact multipliers. For example, the impact effect on real output of the unanticipated reduction in the target price level follows from the saddle path equation (6.6):

$$ \frac{dfy}{d\hat{p}} = (\frac{dfy}{dx}) + \gamma \left( \frac{dfy}{dx} \right) - 1 \tag{6.9} $$

This result states that the impact effect on $y$ is a weighted average involving the full-equilibrium effect on $y$ (which we know, in this model, is zero), and the impact effect on $p$ (which is also zero in this model). Thus, the impact effect on real output simplifies to $-\gamma$, the speed of adjustment is $-\delta_1$, and the cumulative output loss is the former divided by the latter. Using equations (6.8), this cumulative output loss expression — which, for contractionary monetary policy is usually called the sacrifice ratio — can be simplified to (1.9).

It is left for the reader to verify (using the methods of Chapter 2) that when the new $IS$ relationship that is involved in this model is replaced by the old $IS$ function, the sacrifice ratio involved with this contractionary monetary policy is exactly the same. This fact adds to the assurance noted earlier that the old and new synthesis models can have — but do not always have — quite similar policy implications.

6.3 Stabilization Policy Analysis with a “New” Phillips Curve

The analysis in the previous section involved a model with a new $IS$ function and an old Phillips curve. In this section, we consider a model with the opposite dimension of novelty — one with an old $IS$ function and a new Phillips curve. Thus, the model in this section is defined by equations (6.1), (6.3) and (6.2a), where the revised equation is:

$$ \hat{p} = \phi(y - \bar{y}) \tag{6.2a} $$

In this case, it is most convenient to draw a phase diagram in $π-ρ$ space, where $π$ is the inflation rate. We put $π$ on the vertical axis (since it is a jump variable) and $p$ on the horizontal axis. The definition of $π$:

$$ \hat{p} = (π - \bar{p}) $$

is the equation we use to derive the properties of the $ρ = 0$ locus. It is a horizontal line — drawn at the point where $x = 0$ on the vertical axis in Figure 6.6. There are rightward pointing arrows above this line, and leftward pointing arrows below this line, indicating the forces of motion when the economy’s observation point is not on this line. Since $x$ does not enter this equation, this locus is not shifted when the central bank cuts its target price level.

![Figure 6.6 Properties of the $\dot{\rho} = 0$ Locus and the $\dot{\rho} = 0$ Locus](image)

Figure 6.6 Properties of the $\dot{\rho} = 0$ Locus and the $\dot{\rho} = 0$ Locus

To obtain the equation that is needed to determine the properties of the $\dot{\rho} = 0$ locus, we substitute (6.3) into (6.1) to eliminate the interest rate, and then substitute the result into (6.2a) to eliminate the output gap. After noting that $\dot{p} = \dot{x}$, the result is

$$ \dot{x} = \phi \phi \alpha (p - \bar{x}) - \psi p \sigma \tag{6.9} $$

This equation has exactly the same format as (6.4), with $π$ playing the role in (6.9) that $y$ plays in (6.4). This means that the $\dot{\rho} = 0$ locus in the present model has all the same properties as the $\dot{\rho} = 0$ locus had in the model of the previous section (it is positively sloped, there are downward pointing arrows of motion above the locus, and the position shifts left when $x$ is cut). Thus, the entire policy discussion proceeds in Figure 6.6 as an exact analogy to that which accompanied Figure 6.5 above. To save space, we leave it to the reader to review this discussion if necessary. We simply close this particular analysis with a short discussion that links the impact effect on $π$ to the impact effect on $y$. Equations (6.1) and (6.3) combine to yield

$$ (y - \bar{y}) = \psi \phi \alpha (p - \bar{x}) + \psi \sigma \tag{6.9} $$

The phase diagram (Figure 6.6) proves that $π$ falls when $x$ falls. Given this outcome, and the fact that $p$ cannot jump, this last equation implies that $y$ must jump down. Thus, the predictions of this new-Phillips-curve model align with those of the new-IS-curve model (of the last section) and of the first synthesis model (of Chapter 2): there is a temporary recession that immediately begins shrinking as time passes beyond the impact period. We conclude, as in the previous section, that the predicted effects of monetary policy are not specific to any one of these models.

6.4 Some Precursors of the New Synthesis

In this section, we outline a series of further alternative specifications for the aggregate demand ($IS$) and supply (Phillips curve) relationships. This survey allows readers to appreciate that versions of the new synthesis were present in the literature before the “New Neoclassical” label became widespread. The first variation is based on the fact that the new $IS$ function involves a fundamental limitation. The only component of aggregate demand is consumption spending by households. As a sensitivity test, we now consider a model that involves the opposite extreme specification — that aggregate demand is driven totally by firms’ investment spending. As has been our practice in earlier sections of this chapter, we compare this new specification to the original synthesis model. Thus, we specify this investment-oriented demand model by a set of four equations: the same old Phillips curve and central bank reaction functions as used already (equations (6.2) and (6.3)) along with the following set of two equations to replace equation (6.1):

$$ (y - \bar{y}) = \alpha (q - 1) \tag{6.1b} $$

$$ r = \beta (y - \bar{y}) + q \tag{6.1c} $$

These equations were explained in our derivation of firms’ investment behaviour in Chapter 4 (section 4). The first relationship states that aggregate demand depends positively on Tobin’s stock market valuation ratio, $q$, and the second relationship states that the overall yield on stocks is the sum of a dividend (the first two terms in (6.1c)) and a capital gain (the final term). In long-run equilibrium, the dividend is the “normal” real rate of interest; with cycles in the short run, dividends are higher in booms than they are in recessions.

To achieve a more compact version of this model, we proceed with the following steps. Substitute (6.2) into (6.3) to eliminate the inflation rate; substitute the result into (6.1c) to eliminate the interest rate;
take the time derivative of (6.1b) and use the result to eliminate the change in the stock price from (6.1c). We end with

$$y = a(x - p) - a(x + b)(y - \bar{y})$$  \hspace{1cm} (6.10)

The solution proceeds by drawing a phase diagram based on equations (6.2) and (6.10). As before, a straightforward comparison of this set of two equations to (6.2) and (6.4) indicates that the phase diagram is exactly as drawn in Figure (6.5). Thus, as long as we confine our attention to unanticipated monetary policy, there is no need to repeat the analysis. However, since much can be learned by considering anticipated policy initiatives, we proceed with this new use of the same phase diagram.

**Figure 6.7** Anticipated Monetary Policy

![Phase Diagram](image)

Initially the economy is as pictured in Figure 6.7 — at point A, the intersection of the original no-motion loci. While the variable that is calibrated and labelled on the vertical axis is national output, at some points in the following discussion we interpret movements up and down in the graph as increases and decreases in the stock market. This is permissible since, given equation (6.1b), y and q move one-for-one together. Assume that — starting from point A — the central bank cuts its target for the price level. As usual, the (y = 0) locus shifts left, point D appears as the new full-equilibrium outcome point, and the saddle path drawn through point D becomes relevant. If this monetary policy were unanticipated, the economy would jump to the point of intersection between this saddle path and the initial conditions line. Stock prices would drop as soon as individuals realized that a temporary recession has been created and — since this hurts profits and dividends — it is not rational to pay as much for stocks as had been previously the case. The stock market gradually recovers — as does the economy.

How does this scenario differ when the monetary contraction is pre-announced? For example, suppose the central bank states that in exactly one year’s time, it will cut the target price level. Forward-looking investors know that this will cause a drop in the stock market, and to avoid these anticipated capital losses, they sell stocks at the very moment that the future policy is announced. Each individual is trying to sell her shares before anyone else does, to (at least in principle) avoid any capital loss. But everyone is just as smart as everyone else, so this one-sided pressure in the stock market to sell causes a drop in stock prices immediately. Since the monetary policy has not yet been implemented, there is no reason for stock prices to fall as much as they do in the unanticipated policy case. Thus, stock prices (and the level of GDP) fall by some intermediate amount — say to point B in Figure 6.7 — the moment the policy is announced.

What happens during the one-year time interval between the announcement of the policy and its implementation? First, since the observation point of the economy is not on either of its no-motion loci, all endogenous variables — real output, stock prices and goods prices — will start to change. By locating point B in Figure 6.3, readers will see that the joint force for motion is in the north-west direction. As a result, the economy moves along this fashion until the observation point crosses the (y = 0) locus in Figure 6.7. Then, because Figure 6.3 involves south-west motion in this region, the economy’s true path bends down as it approaches point C in Figure 6.7. With reference to the original full equilibrium (point A) the economy is following an unstable trajectory. But, if agents have perfect foresight (as we are assuming), they will have chosen the location of point B so that the policy is implemented just as point C is reached. At that very moment, the saddle path through point D becomes relevant, and the economy begins to track along it from C to D, after the one year time interval has elapsed. To summarize: the economy moves instantly from A to B upon announcement of the policy, it moves gradually from B to C between announcement and implementation, and then gradually from C to D after implementation.

Of course, in reality, agents do not have perfect foresight. Thus, it can turn out that the economy is not at point C the moment the policy is implemented. In this case, the economy continues to follow an unstable trajectory until people realize that the stock market is departing by an
Another analysis that involves forward-looking agents and asset prices within the demand side of the model is Dorndusich’s (1976) model of overshooting exchange rates. The open-economy equivalent of the “old” IS-LM system is the Mundell (1963)-Furman (1962) model. Dorndusich extended this framework to allow for exchange-rate expectations and perfect foresight. His goal was to make the model more consistent with what had previously been regarded as surprising volatility in exchange rates. Dorndusich’s model involved the prediction that the exchange rate adjusts more in the short run than it does in full equilibrium — in response to a change in monetary policy. The following system is a version of Dorndusich’s model.

The small open economy model involves equations (6.2 and 6.3) along with

\[ y - y^* - r - T + \theta (e - p) \]
\[ r = P' = \frac{T_e}{T_e} \]

The first of these open-economy equations is an old IS relationship; in addition to the usual interest rate determinant of aggregate demand, there is a terms-of-trade effect stemming from the net export component of spending, \( e \) is the nominal exchange rate — the (logarithm of the) value of foreign currency. The other new relationship defines interest arbitrage: the domestic interest rate equals the foreign interest rate (and we have assumed zero inflation in the rest of the world) plus the (actual expected) depreciation of the domestic currency. The properties of this model can be determined via a phase diagram. The compact version of the system (needed to derive the phase diagram) is achieved as follows. Combine the interest arbitrage relationship and (6.3) to yield \( \theta = \frac{1}{e(p - \gamma)} \).

Combine the open-economy IS function with (6.2) and (6.3) to yield

\[ (1 - \gamma) (e(p - \gamma) - \gamma (p - \gamma) + \theta (e - p)) \]

Finally, take the time derivative of this last relationship, and substitute in the second last equation. The final result is

\[ \frac{\partial y}{\partial t} = \frac{1}{(1 - \gamma) (e(p - \gamma) - \gamma (p - \gamma) + \theta (e - p))} \]

The phase diagram can be constructed from (6.2) and (6.11). As we have encountered several times already, the form of (6.11) is the same as (6.4) so the phase diagram analysis is very similar to what was presented in detail in section 6.2. For this reason, we leave it for interested readers to complete this specific open-economy analysis. However, before moving on, let us note that (6.11) can be re-expressed, using (6.3), as a relationship that involves just three variables: \( y (e - \gamma) \) and \( (e - \gamma) \). Just as we concluded with the interest-rate term-structure model, then, Dorndusich’s system appears to be a forerunner of the “old” vs. “new” IS curve debate since this last relationship nests both versions.

The final study that we include in this brief survey of work that is related to the New Classical Synthesis is Taylor’s (1975b) model of multi-period overlapping wage contracts (specified in change form, not in levels as was the case in the most widely used version of Taylor’s work).

This specification yields a model that is similar to one that involves Calvo’s (1983) specification of the new Phillips curve, yet it relies on a descriptive, not a micro-based, definition of sticky wages. (For a more detailed comparison of the Taylor and Calvo versions of the Phillips curve, see Dixson and Karas (2006).)

Let \( \omega \) denote the log of all wages contracted in period \( t \), let \( \omega \) be the proportion of contracts that are of one-period duration, let \( \omega_1 \) be the proportion of contracts that are of two-period duration, and so on. These definitions imply that \( \omega_1 \), the log of the overall wage index, is

\[ \omega = \omega_1 + (1 - \omega_1) \omega_2 + (1 - \omega_1)^2 \omega_3 + \ldots \]

Writing this equation lagged once, multiplying the result through by \( (1 - \omega) \), and then subtracting the result from the original yields

\[ \omega_1 = \frac{\omega - \omega_1}{(1 - \omega)} \]

where \( 52 = r (1 - \omega) \). With constant returns to scale technology, units can be chosen so that the marginal product of labor is unity; thus \( p \) stands for both the wage index and the price level. Each contracted \( w \) is set with a view to the expected (actual) price that will obtain in the various periods in the future, and the state of the market price in all future periods (with the weight given to each period in the future depending on the number of contracts that will run for two, three, or more periods). Thus, we have
since we define the natural rate as zero. Writing this last equation forward one period, multiplying the result by (1-T1), and subtracting this last equation from that result, we have

\[=a(p, \ldots, a(t))\]

Continuous-time versions of these price and wage change relationships can be written as:

\[\Delta p = \Delta (\Pi \,\, \Pi)\]

The full model consists of these two relationships and equations (6.1) and (6.3). In continuous time, the length of one period is just an instant; thus, \( \Pi \), \( r \), and \( w \) are jump variables, while \( p \) is predetermined at each instant. Defining \( \Delta = \Delta p \) as an additional jump variable, we can re-express the system (using the by-now familiar steps) as:

\[V = \Delta \Pi \,\,\, \Pi \,\,\, \Pi = \Delta p\]

The first of these equations that define the compact system has the very same form as equation (6.9). The second is analogous to the definition of \( \bar{\Pi} \) in section 6.3. This symmetry implies that this descriptive model of overlapping wage contracts and the macro-based model of optimal price adjustment have the same macro properties.

6.5 An Integrated Analysis: a “New” IS Curve and a “New” Phillips Curve

Thus far, this chapter has proceeded in a piecemeal fashion; we have analyzed models with either a new IS function or a new Phillips curve, but not with both new relationships. If we put both new features in the same system, it involves a first-order differential equation, \( p' = r \), a second-order differential equation, \( p'' = -\Delta r + \cdots \), and one static relationship — the central bank interest-rate setting equation. If we arrange this system in the same format that we have followed in earlier sections of this chapter, we would have to deal with a system involving three first-order differential equations. Graphic analysis would then require a three-dimensional phase diagram involving two jump variables and one sticky variable. Such a graphic analysis is too cumbersome to pursue. We could use an extended version of the mathematical approach that was explained in the final paragraphs of section 6.2, but pursuing this more technical approach has been rejected as beyond the intended level of this book. Even if readers did invest in mastering this more advanced mathematical approach, the model would still have important limitations. In particular, we would still be unable to consider ongoing variations in the exogenous variables — as opposed to one-time changes.

What other options do we have? One is to switch to a discrete-time specification. We have already examined the full new synthesis model with rational expectations methods in Chapter 3 (section 5). But that analysis was very messy as well. There is one other option — to stay in continuous time and allow for ongoing changes in exogenous variables, but to rely on the undetermined-coefficient solution method that was explained in Chapter 2 (section 5) and used again in Chapter 3. We take this approach in this remaining section of this chapter.

The exogenous variable that we want to model as involving ongoing changes is autonomous spending. Thus, we must start by indicating how the new synthesis model is altered slightly — compared to what was derived in Chapter 4 — when this component of aggregate demand is considered. It is left for readers to rework the Chapter 4 analysis to verify that — with an exogenous component of aggregate demand, \( g \), the new IS function becomes \( S = \Delta (1 - a) \) and the “new” Phillips curve becomes \( \Delta \Pi = \Delta (\Pi - S) + \Delta \Pi = \Delta r \) and \( \Pi \) is the proportion of firms that cannot change their price each period.\( a \) should be a bit too low to justify assuming that there is no exogenous component to aggregate demand. (\( a = 1 \), this is the common assumption in the literature. We extend that literature in this section: first by focusing on the basic new synthesis model, and then by considering what has been called a “hybrid” model — that blends old and new approaches.

To keep this section self-contained so that readers do not have to keep referring back to equations that were defined much earlier in the chapter, the full set of the model’s relationships are listed together here.

The basic version of the model is defined by equations (6.12) through (6.15). These equations define (respectively) the “new” IS relationship (aggregate demand), the “new” Phillips curve (aggregate supply), monetary policy (the interest-rate setting rate), and the exogenous cyclical component of demand.

\[
\begin{align*}
\Delta \Pi = a\Delta q + \Delta \Pi \\
\Delta \Pi = \Pi - T &+ \Pi \\
\Delta \Pi = \Pi - \Pi \Pi &+ \Pi \Pi \\
\Pi = \Pi + \Pi \Pi &+ \Pi (1 - \Pi - \Pi)
\end{align*}
\]

The first two equations have already been discussed. Equation (6.14) is the central bank’s reaction function. We continue to focus on a bank that is committed to price stability and this is why a zero inflation-rate target term in the interest-rate-setting equation. But some of the remaining details are slightly different here compared to earlier sections of this chapter. We have chosen units so that both the (now constant) target price level and the natural rate of output are zero. This monetary-policy reaction function states that the bank sets the current nominal interest rate above (below) its long-run average value whenever either the inflation rate is above (below) its target value, or whenever the price level is above (below) its target level.

As noted in earlier chapters, one major point of debate among monetary policy analysts is whether central banks should pursue an inflation-rate target or a price-level target. We consider this debate here by examining alternative values for parameter \( k \). Inflation targeting is involved if \( k = 1 \), while price-level targeting is specified by \( k = 0 \). We focus on the implications of this choice for the economy’s short-run built-in stability properties. The existing literature has provided a thorough investigation of this policy choice in the face of supply shocks, but there has been nothing reported regarding demand shocks. It is partly to address this gap in the literature that we focus on business cycles that are caused by exogenous variations in autonomous demand, as defined by the size curve in equation (6.15). We proceed by deriving the reduced form for real output, to see how the amplitude of the resulting cycle in \( \Pi \) is affected by changes in the monetary policy parameter (\( X \)).

To analyze the model, we first substitute (6.14) into (6.12) to eliminate the real interest rate gap. Then, we take the time derivative of the result, and use (6.13) to eliminate the change in the inflation rate. Finally, we take one more time derivative and use (6.13) again, to eliminate the remaining inflation-change term. The result is:

\[
\frac{dY}{dt} = a(1 - a) + \Pi + (1 - a)(\Pi - \Pi) + \Pi - \Pi \Pi
\]

As in Chapter 2 (section 5), we analyze (6.16) by posting a trial solution and using the undetermined coefficient solution procedure. The solution for output must take the following form:

\[
y = \Pi = \Pi (e^{\Pi}) + \Pi (e^{\Pi})
\]

Equation (6.17), the time derivative of (6.17), \( y = B \sin(t) + C \cos(t) \), its third time derivative, \( \dot{y} = B \sin(t) - C \cos(t) \), and equation (6.15) and its first and third time derivatives, \( k = 0 \), \( 0 = C \cos(t) \) and \( k = -8 \cos(t) \), are all substituted into (6.16). The resulting coefficient-identifying restrictions are:

Illustrative parameter values are needed to assess the resulting amplitude of the cycle in \( y \). Representative values are: \( a = 0.8 \), \( k = 0.2 \) (which, given the restrictions noted in the third paragraph of this section, imply \( 3 = 0.032 \) and \( 8 = 1 \). With these values, the amplitude of the real output cycle is about 15% larger if the central bank targets the inflation rate (X = 1), than if it is the central bank targets the price level \( X = 0 \). According to the model, then, the contemplate move away from inflation-targeting to price-level targeting is supported. Long-run price stability is achieved in either case, and there is a small bonus with price-level targeting — there is a slight reduction in real output volatility.

The reason why this issue requires a formal analysis is that there are competing effects. These can be readily appreciated at the intuitive level (as was noted in Chapter 2). Consider an exogenous increase in the price level. With inflation-rate targeting, such a “bygone” outcome is simply accepted, and only future inflation is resisted. But with price-level targeting, future inflation has to be less than zero to eliminate this past outcome. That is, only under price-level targeting is a policy-induced recession called for. The reason that this consideration may not be the dominant one, however, is that the avoidance of any long-term price-level
We now explain a representative specification of a hybrid IS relationship. The log-linear approximation of the resource constraint is common to both the forward-looking and backward-looking components

\[ y_t = c_t + f(y_{t-1}) \]

The optimizers follow the Ramsey rule

\[ c_t = c_{t-1} - (\gamma - \beta - \gamma) \]

and the rule-of-thumb agents mimic what other agents did in the previous period

\[ c_t = c_{t-1} \]

Giving a one-half weight to each of these two decision rules, and replacing first differences with their derivative as we switch to continuous time, we arrive at the hybrid IS curve

\[ \dot{y}(\gamma - \beta) = (1 - \gamma)C(t) \]

As with the hybrid relationship on the supply side, the differences between several authors' particular models — concerning the size of the slope parameters, not the form of the equation — suggests that researchers should allow for a fairly wide set of sensitivity tests when reporting numerical simulation results.

We do not expect readers to derive the revised expressions for the reduced-form parameters (\( B \) and \( C \)) in the trial solution (that are relevant when our inflation-rate-vs-price-level targeting analysis involves both hybrid functions). While the analysis is not more involved on conceptual grounds in this hybrid case, the actual derivation is quite tedious. (With the extra time derivatives involved in both the demand and supply relationships, the analogue of equation (6.16) turns out to be a fifth-order differential equation in this case.) Suffice it to say that the analysis supports Balish's contention, that the free lunch that is supposed to accompany a switch from inflation-rate-targeting to price-level-targeting seems to disappear in the more general case of hybrid relationships.

Since empirical studies give stronger support for the hybrid relationships, it seems advisable that we put more weight on this result — that on the free lunch proposition that requires the original new synthesis model — with no backward-looking behavior involved — for analytical support.

Before ending this chapter, we briefly discuss several general issues concerning the new synthesis model. First, it may strike readers as odd that the modern analysis of monetary policy makes no reference whatsoever to the money supply. McCullum (2001) has addressed this concern. He allows money to play a role in reducing transactions costs (in a manner that is more general than the one we considered in Chapter 5 (section 4)). When the household's optimization is worked out in this setting, a money demand function emerges as one of the terms, and this creates an additional channel through which the interest rate can affect spending. Using estimates of the interest elasticity of money demand, McCullum calibrates the model and concludes that being more complete along these lines makes only a trivial difference to the model's properties and policy implications. Others who have investigated this issue are Soderstrom (2002) and Nelson (2003).

Some researchers have raised the issue that there is more than one channel for the monetary policy transmission mechanism. As noted in the last chapter, higher interest rates can have a cost-increasing effect, not just a demand-reducing effect. This can be important. A number of studies that are based on the standard hybrid new synthesis model show that nominal-GDP targeting can dominate inflation-rate targeting (for example, see Kim and Henderson (2003)). Malik (2004) and Ravenna and Walsh (2006) have shown that these policy implications can be affected when there is a direct effect of interest rates in the new Phillips curve.

Finally, readers may be surprised by the following feature of the new synthetic framework. Much attention is paid to the proposition that the analysis starts from a clear specification of the agents' utility function. Then, when policies are evaluated with the model, analysis seem to revert to the "old" habit of ranking the outcomes according to which policy delivers the smallest deviations of output and the price level from their full equilibrium values. Shouldn't an analysis that is based on an explicit utility function (involving consumption and leisure as arguments) use that same utility function to evaluate alternative policies? Indeed, this is the only approach that could claim to have internal consistency. It is interesting, therefore, that Woodford (2003a) has derived how the standard policy-maker's objective function — involving price and output deviations — can be explicitly derived as an approximation that follows directly from the private agents' utility function that underpins the analysis.
6.6 Conclusions

When the New Classical revolution began in the 1970s, strong statements were made concerning "old fashioned" macroeconomics. For example, Lucas and Sargent (1979) referred to that work as "faddishly flawed", and King (1993) argued that the IS-LM-Phillips analysis (the first Neoclassical Synthesis) was a "hazardous base on which to ... undertake policy advice". But, as Manikin (1992) has noted, that analysis has been "deconstructed" (see "new" IS-LM-Phillips analysis) and, as we have seen, it appears to be a useful framework for undertaking policy advice. As "new" macroeconomics involves roughly the same level of aggregation and abstraction as the older analysis, since this facilitates communication between actual policy makers (who were brought up in the older tradition) and modern analysts. In short, there is reason for much more optimism than there was 30 years ago.

It is not to say that there is no controversy remaining. Indeed, some macroeconomists still actively debate the relative merits of the "old" and "new" specifications of the IS and Phillips curve relationships. This is because, as noted above, the "old" relationships appear to be much more consistent with real-world data, despite the fact that the "new" relationships are more consistent with at least one specific version of new macroeconomics. It is frustrating to have to choose between the two criteria for evaluating macro models — consistency with the facts and consistency with optimization theory. It is for this reason that many researchers are focused on developing the hybrid models that share some of the features of both the "old" and "new" approaches.

One purpose of this chapter has been to make the reader aware of these developments, and to thereby impart some perspective. But since we wished to limit the technical demands on the reader, for the most part, we limited the analysis to a set of "partially new" models. We have found that some of the properties of these models are very similar. This is fortunate. It means that policy makers do not have to wait until all the controversies within modern macroeconomics are settled — before taking an initiative. It is true that the answers to some questions are model-specific. For example, the question as to whether the central bank should target the inflation rate or the price level receives different answers as we vary the model. In Chapter 3, using an "old" specification, we concluded that inflation-rate targeting is preferred. In section 5 of the present chapter, we saw two things — that a "new" model supports the opposite conclusion, and that a hybrid model swings the support back in the direction of inflation-rate targeting. Nevertheless, despite the sensitivity of our recommendations to changes in model specification, policy makers can still proceed on this issue. This because, as already noted, the hybrid model seems to score the best when evaluated according to the model-selection criteria (consistency with both constrained maximization theory and data).

The main methodological tool that is used in this chapter is the phase diagram. We will use this tool in later chapters as well. One of the most interesting things that has emerged from our use of phase diagrams is that the impact of government policies can be very different — depending on whether agents did or did not anticipate that policy. This sensitivity can sometimes make it very difficult for macroeconomists to test their models empirically.

CHAPTER 7
STABILIZATION POLICY CONTROVERSIES

7.1 Introduction

This is the second of a pair of chapters on the New Neoclassical Synthesis approach to macroeconomic policy questions. The first chapter in this sequence (Chapter 6) was concerned mostly with deriving the necessary methods of analysis. This preparation makes possible the focus in this chapter — we apply the new approach to four central questions of stabilization policy. We begin by investigating inflation policy: in particular, we consider whether disinflation should be applied in a "cold turkey" fashion, or whether a more gradual approach to lower inflation is more desirable. The second policy issue is also a monetary policy question — the central one for open economies: How can the country achieve a higher degree of built-in stability? By adopting a flexible exchange rate policy or by forming a currency union with one's major trading partners?
The loss function is illustrated in Figure 7.2 by the indifference curves. Curves that are closest to the point marked "best" represent higher utility. Suppose that the economy is at point $A$ with an undesirable level of unemployment (the natural unemployment rate) and the desirable amount of inflation (zero). The central bank will find it tempting to move to a point like $B$. This point is feasible — at least for a while — since it is on the currently relevant short-run trade-off line. This point is also desirable — since it is on a preferred indifference curve than is point $A$. So the central bank will be tempted to print money to generate the inflation rate indicated by the height of point $B$.

The trouble with this strategy is that point $B$ is not sustainable. It involves actual inflation (positive) exceeding expected inflation (zero). Individuals will be smart enough to revise upward their forecast of inflation. This will move the economy to a higher trade-off curve, and so to less desirable inflation curves. Actually, if individuals are clever, they will be able to see immediately the point where the economy will come to rest. After all, a full equilibrium requires two things. First, it must be a point of tangency between a short-run Phillips curve and an indifference curve. Second, it must be a point where the actual and expected inflation rates are equal: otherwise private agents will revise their forecast and that means the economy has yet to reach a state of rest. The vertical line at $u = u^*$ represents all points that involve actual inflation equal to expectations. The point of tangency is the point on this long-run Phillips curve that meets the central bank’s tangency requirement — that is, point $C$ in Figure 7.2.

This reasoning proves that it is not advisable for the central bank to try for point $B$ by falsifying the expectations that individuals had at point $A$. Once this attempt to move to $B$ is going, the system gravitates to point $C$ — which is on the least desirable of the three indifference curves. The central bank will make the right decision if it focuses on the vertical long-run Phillips curve. If it does so, and tries to achieve the point of tangency between this constraint and the indifference map, the bank will realize that point $A$ is the best tangency point. By focusing on the long run — and ignoring the fact that it can (but should not) exploit the fact that individuals have already signed their wage and price contracts and so are already committed to their inflationary expectations — the bank can deliver point $A$ (and this is better than point $C$). The moral of the story is that the bank should focus on the long run, and not give in to the temptation to create inflation surprises in the short run.

A formal proof of this proposition can be developed as follows. The central bank has to choose point $B$ in the $u$-$\pi$ plane (the Phillips curve). There is an endogenous and exogenous variable in the Phillips curve. The exogenous variable is the inflation rate, while the endogenous variable is the unemployment rate. The central bank has to choose a point on the Phillips curve that maximizes utility. The Phillips curve is given by $L = (u - \pi)^2 + b(T - \theta)^2$.

The first relationship is an inverted expectations-augmented Phillips curve, which stipulates that the unemployment rate falls below its natural value, $u^*$, only if actual inflation exceeds expected inflation. This relationship is illustrated in Figure 7.2. The vertical line at $u = u^*$ indicates the set of outcomes that are consistent with full equilibrium (that is, that expectations are realized). The fact that this line is vertical means that there is no trade-off between unemployment and inflation in the long run. The family of negatively sloped lines (each with a slope equal to $-1/\theta$) are the short-run trade-off curves. They indicate that for as long as individuals’ expectations of inflation are fixed — higher inflation will generate lower unemployment. As long as the central bank cares about unemployment, this short-run Phillips relationship will represent a temptation. By raising inflation, the bank can reduce unemployment. The problem is that raising inflation above what people were previously forecasting — the bank will cause problems in the longer run. Rising expectations of inflation shift the economy to a higher short-run trade-off line. In the end, citizens receive short-term gain (lower unemployment) in exchange for long-term pain (higher inflation).

A more formal analysis of this dynamic choice can be had by defining a specific objective function for the central bank. That is done in the second equation. It states that the bank suffers losses ($L$) whenever the unemployment rate differs from its "best" value ($u^*$), and whenever inflation differs from its "best" value (assumed to be zero). To ensure that losses are incurred when these targets are missed on both the high and low sides, the loss function stipulates that it is the square of each deviation that matters. Parameter $b$ represents the relative importance of citizens' attack to low inflation as opposed to sub-optimal levels of unemployment. We assume that parameter $f$ is a positive fraction. This ensures that the invisible hand has not worked. The unemployment rate that the market generates (on average, $u^*$) is higher than the value that is socially optimal ($u^*$). Only if this is assumed can we interpret the unemployment as "involuntary." The interpretation of unemployment as involuntary is discussed in much greater detail in Chapter 8. The analysis in that chapter provides a rigorous defense for the proposition that it is reasonable to specify parameter $f$ as a fraction.
which states that the bank raises aggregate demand (and therefore inflation) whenever unemployment rates above the natural rate. It is assumed that this reaction is announced in advance — as an ongoing decision rule. Thus, it is a policy that is not based on causing private sector forecasts, but it reflects the central bank’s expectations. If we substitute this policy reaction function into a standard Phillips curve,

\[ u = a + \beta (g - \pi) \]

the result is

\[ u = a + \beta (g - \pi) \]

where \( a = b + d + \mu \). The analysis proceeds just as before — using the revised loss function and this revised Phillips curve. The only thing that is new is the revised interpretation of parameter \( a \) — that is defined in this paragraph. As before, an increase in the rate of change makes \( a \) larger and \( b \) smaller, and we still assume that despite the increased policy uncertainty, the expected policy reaction function of the agents in the model.

More on central bank independence is available in Fischer (1993) and McKinnon (1995). As a result of this basic model of policy credibility, we now have an integrated analysis of disinflation which combines important insights from both Keynesian and Classical perspectives. Keynesians have emphasized rigidities in nominal wages and prices to explain the recession that seems to always accompany disinflations. New Classical have stressed the instability of the central bank to make credible policy announcements. Ball (1995) has shown that sticky prices and incomplete credibility are necessary to understand the short-term output effects of disinflation. Indeed, if incomplete credibility is not an issue, Ball shows that disinflation can cause a temporary boom, not a recession. Only a fraction of firms can adjust their prices quickly to disinflationary monetary policy. But if the policy is anticipated with full credibility, firms that do adjust know that money growth will fall considerably while their prices are in effect. Thus, it is rational for them to reduce dramatically their price increases, and this can even be enough to push the overall inflation rate below the money growth rate initially. This is what causes the predicted temporary boom. However, since we know that such booms do not occur during disinflations, we can conclude that policy credibility must be a central issue.

The general conclusion to follow from this monetary policy analysis is that the only dynamically consistent — that is, credible — policy approach is one that leaves the authorities no freedom to react in a previously unexpected fashion to developments in the future. Discretionary policy that is not part of an ongoing well-understood rule can be sub-optimal because there is no mechanism to induce future policy makers to take into consideration the effect of their policy, vis-a-vis the expectations mechanism, upon current decisions of agents” (Kydland and Prescott (1977, p. 61)).

This argument for “rules” (with or without predictable feedback) over “discretion” (unpredictable feedback) has many other applications in everyday life. For example, there is the issue of negotiators with kidnappers. In any one instance, negotiation is appealing — that the life of the hostage can be saved. But concessions create an incentive for more kidnappings. Many individuals have decided that the long-term benefits of a rule (no negotiations) exceed those associated with discretion (engaging in negotiations). Another example of this issue, that is less important but closer to home for university students, concerns the institution of final exams. Both students and professors would be better off if there were no final examinations. Yet students seem to need the incentive of the exam to work properly. Thus, one’s first thought is that if the final policy were to be a deterministic one in which the professor promises an exam and then, near the end of term, breaks that promise. With this arrangement, the students would work and learn but would avoid the trauma of the exam, and the professor would avoid the ambiguity. The problem is that students can anticipate such a broken promise (especially if there is a history of this behaviour), so we opt for a rules approach — exams no matter what. Most participants regard this policy as the best one. (Further discussion see Fischer (1989)).

We close this section by returning to disinflation policy. In this further treatment of the issue, we integrate political and macroeconomic considerations. We include this material for two reasons. First, it shows how political uncertainty can be a source of incomplete credibility in economic policy making and, second, it provides a framework for examining whether gradation in policy making is reasonable or not.

Blanchard (1985a) has presented a simple model which highlights this interplay. His analysis focuses on the possibility of two equilibria — both of which are durable since both are consistent with rational expectations. The essence of this model is that disinflation policy only
works if agents expect low money growth in the future. If a “hard-line” political party is in power and it promises disillusions, two outcomes are possible. On the one hand, agents can expect that the disillusions will succeed and that the anti-inflationary party will remain in power. Since both these expectations will be realized, this outcome is a legitimate equilibrium. On the other hand, agents can expect that the disillusions will fail, and that (perhaps as a result) the anti-inflationary party will lose power. Since the change in government means that the disillusions policy is not maintained, once again, the rates validate the initial expectation, and this outcome is another legitimate equilibrium. Blanchard formulates this argument to see what strategy is advisable for the hard-line party.

The basic version of the model is straightforward. Unemployment depends inversely on the excess of the money growth rate, g, over the inflation rate, Tu = μ(t−μ)/T−μ). The inflation rate is a weighted average of the current actual money growth rate and what growth rate is expected in the future, g: \( g = \frac{a - c}{a} \cdot g + \frac{c}{a} \cdot g_{E} \). The political opponent, the “soft” party, sets \( g = g_{E} \), where \( g > b \).

If agents expect the hard-liners to stay in power, it is: 
\[ u = \left( 1 - m + \right) \cdot \left( 1 - \right) \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \·
is to peg a linear combination of the domestic price level and domestic real output (imposed in the model by assuming equation (7.4)). Equation (7.4) encompasses two interesting cases: targeting the price level \( C_t = 0 \), and targeting nominal GDP (\( p = 1 \)).

We assume that business cycles in this small open economy are caused by exogenous variations in export demand, as defined by the sine curve in equation (7.5). We now proceed to derive the reduced form for real output, to see how the amplitude of the resulting cycle in \( y \) is affected by pegging one exchange-rate regime or the other. We also explain the derivation of the reduced form for real output as the flexible exchange rate case (and we assume that readers can use similar steps to verify the result that we simply report for fixed exchange rates).

First, we simplify by setting \( f = 1 \). Next, we take time derivatives of (7.4) and use the result to eliminate the first and second time derivatives of \( p \). Then, we substitute (7.3) into (7.1) to eliminate the interest rate, and use the result to eliminate the term involving the time derivative of the exchange rate in the time derivative of equation (7.2). The result is:

\[
A = \overline{A} + \overline{b} t + \overline{g} t + \overline{s} t + \overline{S} t
\]

As we did in sections 3.5 and 6.5, we use the underlined coefficients to solve the model for output. The final solution for output can be written as

\[
y = \overline{A} t + \overline{b} t + \overline{g} t + \overline{s} t + \overline{S} t
\]

7.4 The Feasibility of Bond-Financed Government Budget Deficits

Government debt has accumulated in many western countries — both in absolute terms and as a proportion of GDP. Many governments are trying to contain this development in recent years, as they stabilize, and sometimes reduce, their debt ratios. We explore these attempts in this section.

Government debt would not have accumulated if governments had taken the advice that follows from the standard theory of fiscal policy. That advice is to run a deficit during the recessionary phase of each business cycle and to run a surplus during the boom half of each business cycle. According to conventional wisdom, this policy would help to balance the economy over the cycle, and it would balance the budget over the interval of each full cycle. As a result, this would be sufficient to keep the debt from increasing over the long term.

For the last 30-40 years, governments have followed this advice during recessions — indeed, they have appreciated the fact that economists have condemned deficit spending during these periods. But governments have chosen to disregard the other half of the advice from economists, for they have not run surplus budgets for anything like half the time. By the 1990s, many governments had started to react to their exploding debt-to-GDP ratios. Many have become so preoccupied with stopping the growth in debt that they no longer make any attempt to help balance the economy.

Indeed, Europe’s Stability Pact has been much criticized for forcing this outcome on member governments.

We provide a micro-based estimate of the long-term benefits of deficit and debt reduction in Chapter 10. These benefits turn out to be substantial. Here, we restrict our attention to an investigation of how it has turned out that some governments have been able to re-establish control of their debt levels, while others have not. The basic issue is that short-term price level must be incurred to achieve debt reduction unless we can somehow “grow our way out” of the problem. It would seem as if this might be possible, since the denominator of the debt-to-GDP ratio is growing over time, and even with small annual deficits, the numerator may grow at a slower rate. The governments that have achieved a turn-around in their financial affairs are the ones that did not expect too much from this “less painful” approach of letting growth substitute for monetary policy. To defend this assertion, we now compare two fiscal policies — one that keeps a tight reign on the primary deficit, and the other that keeps the overall budget deficit as an exogenous target. How do such targets affect the stability of government finances?

To answer this question, we simplify by ignoring money issues, the tax revenue collected on bond interest, and short-run deviations of real output from its natural rate. If we define \( G \) and \( T \) as real program spending and real taxes, \( \pi \) as the nominal interest rate, and \( B \) as the non-nominal government bond outstanding. The nominal deficit, \( D \), is then defined by

\[
D = G - \pi B + T \]

Using lower-case letters to define proportions of GDP: \( d = D/B \), \( g = G/B \), \( \pi = \pi \), \( b = B/B \), we have

\[
d = g + t + b
\]
Under pure bond-financing, the deficit is financed through bond issue: 
\[ \Delta D \text{ is the change in the debt ratio} \]
and introducing notation for the inflation rate and the real GDP growth rate:
\[ P, \pi, Y^* \text{ is the inflation rate and the real GDP growth rate} \]
this accumulation identity can be re-expressed as
\[ \Delta D = \pi + \Delta Y \]
(Equation 7.12)

We consider two fiscal regimes — first, one that involves the government maintaining a fixed structural, or primary, deficit-to-GDP ratio. This policy means that \((g_p - 1)\) remains an exogenous constant, and the overall deficit, \(D\), is endogenous (determined in Equation 7.11). Substituting (7.11) into (Equation 7.12) to eliminate \(D\) yields
\[ g_p - 1 = \Delta r - n \]
where \(\Delta r = \pi + \Delta Y\) is the real interest rate on government bonds. Assuming Ricardian equivalence, both the interest rate and the growth rate are independent of the quantity of bonds outstanding. With this simplification, we can use this last equation to determine \(g_p\). Convergence to a constant debt ratio occurs only if this expression is negative, so the stability condition is \(n > \pi\). With a fixed primary deficit, then, a program of bond-financed deficit is feasible only if the economy’s long-run average real growth rate exceeds its long-run average real interest rate. We assess the likelihood of this condition being met for much of the remainder of this section. But first, we consider an alternative fiscal regime.

Fiscal authorities can adopt a target for the overall deficit instead of the primary deficit. This regime makes \(D\) an exogenous variable. In this case, the two-equation system defined by (7.11) and (Equation 7.12) becomes segmented. The dynamics of the debt ratio is determined by (Equation 7.12) alone. It implies that stability is assured if the \(\text{nominal growth rate of GDP} = \text{growth rate of bonds} \times (1 + \pi)\) is positive. As long as the central bank does not peg the interest rate, the actual interest rate on government bonds is equal to the nominal interest rate on bonds plus the inflation rate \((\pi + \Delta Y)\). Thus, the debt ratio converges to \(b = a/(n + \pi)\). We assume that this fiscal regime is in place when we use a calibrated model to estimate the size of the benefits of deficit reduction in Chapter 10.

The Canadian situation provides an interesting application of this analysis. For the 1968-1983 period, the Liberal government of Pierre Trudeau ran large primary deficits every year, and as this analysis suggests, the debt ratio exploded. Then, for the next ten years (1983-1993), the Conservative government of Brian Mulroney maintained an average primary deficit of zero, hoping that this contruction in fiscal policy would be sufficient to permit the country to grow its way out of its debt ratio problem. However, since interest rates exceeded the economy’s growth rate, the analysis suggests that this hope would not be realized, and indeed, it was not; the debt ratio continued to increase in dramatic fashion. Finally, the Liberals (under Jean Chretien and Paul Martin) returned to power in the mid-nineties. In thirteen years (1993-2006), government debt fell. During this period, the Liberal’s adopted a strict target for the overall deficit ratio. After bringing that target down over a five-year period, it remained constant (at a small surplus of about one half of one percentage point of GDP) thereafter. Again, just as the analysis predicts, the debt ratio has fallen (by about 35 percentage points). So with this policy, we can grow our way out of the debt problem. But this more appealing outcome required much pain initially getting the overall deficit to zero was much bigger than eliminating the primary deficit. To avoid this, we conclude that — as simple as it is — a model that contains nothing but the basic accounting identities for government budgets and debt provides an excellent guide for interpreting actual policy.

A final note about the Canadian episode is worthwhile. By the turn of the century, the focus was on another monster scenario. People were speaking of the “fiscal dividend,” which referred to the extra room in the budget created by the shrinking interest payment obligations. With the prospect of the debt ratio falling from its peak to the government’s announced long-run target by 50 percentage points, analysts expect significant interest payment savings. A drop in the debt ratio of 0.5 times an interest rate of 0.05 means a saving of 2.5 percent of GDP every year. There has been no shortage of people ready to advise the government about how to spend this fiscal dividend — some arguing for better funded programs and others advocating tax cuts. To have some idea of the magnitudes involved in this debate, we should note that an anuity of 2.5 percent of GDP would permit a permanent elimination of one-third of the federal government’s total collection of personal income taxes. With so much at stake, we can expect ongoing debate about how best to use the fiscal dividend.
tenth century, government bond yields exceeded real growth rates in many countries. In this environment, the entire distinction between safe and risky assets ceases to be important since there is no doubt that bond financing (along with an exogenous primary deficit) involves instability. The literal take-off of government deficits during much of that time, of course, consistent with this interpretation.

In the end, over long time intervals at least, it seems fair to say that there is serious doubt concerning the relative magnitude of the growth rate and the level of government bond deficits (for instance, 1993) developed models of bond-financed deficits that involve explicit modelling of uncertainty Ball et al. consider historical evidence for the United States since 1871. They estimate that this experience is consistent with being able, with a probability of 80 percent, to run temporary deficits and then roll over the resulting government debt forever. The country can likely grow its way out of a large debt problem, without the need for ever having to raise taxes. As a result, the welfare of all generations can be improved. This result is not certain. Nor does it violate the evidence in favor of dynamic efficiency — since there is still a positive probability that some generations will be worse off if growth is insufficient to lower the debt-to-GDP ratio.

Ball et al. use their result to argue that the usual metaphor concerning government deficits — that they are like tennis eating one's house — is inappropriate. The tax analogy suggests a gradual but inevitable disaster. They advocate thinking of deficits like a homeowner's decision not to buy fire insurance. Living standards can be higher as long as no fire occurs, but there is a large adverse effect if the fire does occur. They prefer the fire insurance metaphor because the occurrence of a fire is not inevitable. Of course, unreasonable people can disagree about the likelihood of the fire. Since we are not sure what caused the productivity slowdown in the mid-1970s, for example, it may be that extrapolating the more rapid average growth rates of the previous century into the future is ill advised. Similarly, with world capital markets now integrated, and with the high demand for savings stemming from the developing markets in Asia and elsewhere, world (real) interest rates may rise above historical norms for a prolonged period of time. In short, it may be quite prudent to put much weight on the 1945-1975 historical experience (which was an historical anomaly since it involved $n = 0$). If so, the tax analogy is not so bad after all.

7.5 An Evaluation of Balanced-Budget Rules and the European Stability Pact

Should the debt ratio be allowed to vary over the business cycle? One of the central lessons of the Great Depression was that adjusting annual spending and taxation with a view to maintaining a fixed budget-balanced target "come hell or high water" increases output volatility. Spending has to be cut and taxes raised as the economy slows down, which is exactly the time we do not want to happen. The Keynesian message was that it is better to help balance the economy by balancing the budget over the time horizon of a short business cycle rather than over an arbitrary shorter period such as one year. Thus, for at least half a century following the Depression, it was assumed that a rigid annually balanced budget approach was "obviously" to be avoided. But the Keynesian message has been increasingly ignored in recent years. As the "hell or high water" quetion came from the Canadian finance minister in 1994 indicates, governments have reverted to annual budget-balanced targets that permit only very small departures from a more rigid regime. Adoption of the "Growth and Stability Pact" in Europe has applied similar pressures. As The Economist put it, the euro area faces the possibility of its first recession... the stability pact must not only preclude any fiscal easing but even transmute the operational deficits. That could mean that these countries are required to increase taxes or cut public spending even as their economies slow. That smack of 1930s-style self-flagellation" (Aug. 25, 2001, p.13).

The Economist's editorial writers correct or are they putting too much stock in an "old" analysis that has not been modified to make it consistent with modern standards of rigor? Is it, or is it not, appropriate for the government to allow cyclical variation in its debt ratios by running deficits during recessions and surpluses during booms?

There is a long literature assessing the usefulness of Keynesian-style "built-in stabilizers." For example, 35 years ago, Gorton and Hellwig (1971) and Smith (1974) showed that these mechanisms can serve as de-stabilizers. Running a deficit budget during a downturn will very likely decrease the size of that initial recession. But over time the government debt must be worked back down, so the overall speed of adjustment of the economy is reduced. The initial recession is smaller, but the recovery takes longer. While this literature identifies this trade-off between a favourable initial impact effect and a unfavourable persistence effect, it is rather dated in that expectations are not forward-looking, and the behavioural equations are descriptive, not formally micro-based.

We can re-interpret some of our earlier modelling to make this analysis somewhat more up-to-date. Consider ongoing shocks, in the context of the simple model involving perfect foresight and descriptive behavioural relationships (discussed in Chapter 2, pages 32-34). If we focus on the inflation targeting (X > 1) case, we see that the amplitude of the real-output cycle is C = Pa A Keynesian fiscal policy involves taxation depending positively on output, and this (in turn) makes IS-curve parameters α and 13 larger. We can appreciate this by considering the relationship (they belong to the IS specification: $Y = C(l-α)-αg + G(1+α)$. The total differential of this relationship implies that the coefficients in $Y = -δr + c_2g$ must be interpreted as

$$a = \alpha Y Y + \alpha G \gamma (1-\alpha) + \gamma$$

As is readily seen, the steeper is the tax function (the larger is $\alpha$, the smaller are sum coefficients $a$ and 13. As a result, allowing for the "built-in stabilizers" does lower output volatility. Thus, with a somewhat more up-to-date analysis of fiscal policy (one that allows for the modern treatment of expectations), support for Keynesian stabilizers re-emerges.

But, as we have seen concerning monetary policy, the analytical under-pinning for any policy is now viewed as quite incomplete if that analysis does not involve micro-based behavioural relationships. Lum and Siroli (2006) has investigated whether the (undesired) "increased persistence" property of Keynesian debt policy is bigger or smaller than the (desirable) decreased impact-effect property, when this regime is compared to the rigid regime (involving a constant debt ratio) — in a setting that respects the requirements of the modern approach to business cycle theory. Here is a summary of that analysis.

Except for the presence of the (log of) government spending, the model's new IS and Phillips curve relationships are standardized:

$$a = \alpha Y + \alpha G \gamma (1-\alpha) + \gamma$$

$$= \alpha Y + \gamma$$

(7.13)

(7.14)

The remaining equations of the model define how monetary and fiscal policies are conducted. For monetary policy, an important consideration is simplification. There is one time derivative in the new IS relationship, two time derivatives in the new Phillips curve, and one time derivative needed to define Keynesian fiscal policy (the change in bond equals the current deficit). Thus, we need fiscal policy that can reduce the order of dynamics that is inherent in a macro model that involves both forward-looking dynamically optimizing agents, and the dynamics of bond-financed deficits. To this end, we assume that the central bank has a constant target value for the price level, $\beta$; and that the bank adjusts the interest rate by whatever it takes to ensure that the actual price level targets this according to the following relationship:

$$= -2\beta p$$

(7.15)

The bank ensures that the percentage change in the actual price is proportional to the percentage gap between the actual and target price levels. When the time derivative of this policy rule is substituted into the new Phillips curve, (7.14), we have a convenient static relationship:

$$2\beta (1-\alpha) Y = \alpha G \gamma$$

(7.16)

Another aspect of this specification of monetary policy is that it eliminates the new IS relationship from having any fundamental role to play. Its function is simply to determine (realistically) the value for the interest rate that the central bank has to set to deliver its monetary policy.

The second step in the government always balancing the budget, then with the government following Keynes' suggested approach. To be explicit about the government's accounts, we need to define government bonds in a specific way. It is convenient to assume that it are indexed consols. As a result, since each bond is a right to receive one unit of good every period forever, variable $B$ denotes both the quantity of bonds outstanding and the interest-payment obligations of the government. The market price of each bond is (14).

With a balanced-budget rule, the bond stock stays constant, so $B = B$, $\delta = 0$, and $G = T Y = (1-\alpha) Y$. If $T$ is the tax rate, and since we...
take it as fixed (in both fiscal regimes), the balanced-budget rule forces the government to allow program spending to vary over the cycle as real output changes through time. In full equilibrium, government spending is given by $G = r \beta (1 - r)B$. Expressing the spending relationship as a percentage deviation from equilibrium, we have

$$g = v(y - \bar{y})$$

(7.17)

where $v = r(G / FT) = r(1 - \alpha)$. The fact that this relationship involves government spending falling whenever the economy is in recession is what leads the analysts to reject a balanced-budget rule and to support the Keynesian approach.

We now turn to the specification of fiscal policy in a Keynesian setting. In this regime, temporary budget deficits and surpluses are permitted. But there is the possibility of instability in the debt-to-GDP ratio, since the economy has a positive interest rate, and no ongoing real growth. To avoid this problem, we specify that the Keynesian government reduces spending — whenever debt rises — by what is necessary to avoid instability. This policy is specified by $G = G - y_0 \beta (1 - r)B$. Recalling that $G = r \beta (1 - r)B$, we can specify government spending in the Keynesian regime by

$$g = -0(y - \bar{y}) - b$$

(7.18)

where $\bar{y} = 7(BI + VfG / y)$. The government covers the temporary deficits and surpluses by issuing and retiring government debt:

$$b = \frac{1}{2}y + (1 - \alpha)B - Y.$$ 

After substituting in the equations for both the actual and full-equilibrium values for spending, and using the fact that $b = B / B$, this bond-issue equation can be re-expressed as

$$[B_1 (1 - \alpha)] = \frac{1}{2} (1 - y) \beta (y - \bar{y}) (Y + B (1 - T)).$$

If a Taylor-series expansion is taken of the left-hand side of this budget identity, the coefficient of all the deviations that emerge from the square-bracket part all have as their coefficients the full-equilibrium value of which is zero. Thus, the bond-issue equation can be simplified to

$$y_0 = 0$$

(7.19)

where $x = F(\alpha)$ and $0 = F(\alpha) / B$. This relationship is part of the model when the Keynesian regime is in place. The fact that it involves government debt being issued whenever the economy is in recession is what forces the government to work that debt back down later on. This need makes the recession last for a longer time, and (other things equal) is what leads the analysts to reject the Keynesian approach.

Our strategy is to examine the output effects of a contractionary monetary policy — one-for-all unanticipated cut in the target price level, p. We calculate the undiscounted sum of the deviations of real output from its natural rate — that results from this contraction in demand (in each fiscal regime). The regime that delivers the smaller overall output effect is deemed to be the one that provides more built-in-stability.

The model involving the balanced-budget rule is defined by equations (7.15), (7.16) and (7.17), which determine $y, \alpha$ and $\beta$ at each point in time. The impact effect of the monetary policy on real output can be had by substituting (7.17) into (7.16) to eliminate the government spending variable, and then differentiating.

$$dy = \frac{dy_0}{\partial (1/2)}$$

(7.20)

We compare expression (7.20), the size of the initial recession caused by the monetary policy, to the similar multiplier derived in the Keynesian case (below). But first, we consider how rapidly this recession dissipates when the rigid fiscal regime is in place. Since the three-equation system is linear, all variables involve the same speed of adjustment. Given equation (7.15), that adjustment speed is 2. The undiscounted output loss is the impact effect divided by the adjustment speed, so the overall output deviation is

$$\text{[sum of output deviations]} = 2(1 - \alpha)$$

(7.21)

The model involving Keynesian fiscal policy is defined by equations (7.13), (7.17) and (7.19), which determine $\gamma, \alpha$ and $\beta$ at each point in time. The impact effect of the monetary policy on real output can be had by substituting (7.18) into (7.16) to eliminate the government spending variable:

$$\frac{\partial y}{\partial \bar{y}} = \frac{\partial (\bar{y} - p - b)}{\partial \bar{y} + (\alpha / y)}$$

(7.22)

and then differentiating the result (7.22), to get

$$\frac{dy}{dt} = \frac{1}{1 \alpha}.$$

(7.23)

By comparing expressions (7.20) and (7.23), we see that the contractionary monetary policy causes a bigger initial recession in the balanced-budget regime. Now we compare the adjustment speed across the two environments. To this end, we substitute (7.22) ante (7.19) to eliminate the output-gap term, and arrange the result and (7.13) in the following matrix format:

$$b = A(p - p_0 - b)$$. 

(7.24)

There are no jump variables involved in this dynamic process, so both characteristic roots must be negative for stability. Since the determinant and trace of matrix $A$ are the product and the sum of these two roots (respectively), stability requires both det$(A) > 0$ and trace$(A) < 0$. Thus, $(\alpha) < 0$ is necessary and sufficient for stability, and (as noted above) we assume that this condition is met. The adjustment speed in this case is

$$\text{trace}(A) = (2 - \alpha) / 2.$$ 

This expression is definitely smaller than $\alpha$, the adjustment speed that obtains in the rigid fiscal regime.

We can summarize as follows: The Keynesian approach has both desirable and undesirable features — it involves a smaller initial recession but a slower speed with which that recession is dissipated. Since the overall sum of the output deviations is a measure that gives weight to both these features, we compare this expression across fiscal policy regimes. In this Keynesian case, we have

$$\text{[sum of output deviations]} = 2(1 - \alpha)$$

(7.21)

By comparing expression (7.21) and (7.24), we see that reference to representative parameter values is necessary if we are to assess which fiscal regime delivers more built-in-stability. A full analysis is provided in Lavo and Scard. Here we simplify by assuming that the full-equilibrium involves no debt, and by assuming that — when in the Keynesian regime — the fiscal authority comes as close as possible to keeping its spending constant by letting parameter $p$ get close to zero. Further, from the calibration we explained in section 6.5, we know that it is safe to take parameter $\alpha$ as approximately zero. These assumptions imply that the cumulative output loss is twice as big when the Keynesian fiscal regime is in operation. For plausible parameter values, then, we can conclude that the adoption of rigid budget-balance rules may not involve any loss of built-in-stability after all — indeed this policy can improve matters. This result is consistent with the empirical results of Fund and Millo (2003), who study data for 91 countries and conclude that institutional constraints on governments actually decrease output volatility. It appears that the basic new neoclassical synthesis can rationalize these empirical findings.

As noted in earlier chapters, there is a debate in the theory of monetary policy about whether a price-level targeting strategy represents a "free lunch" — that it may provide improved outcomes in terms of both lower inflation volatility and lower output volatility. We have seen in this section that a similar debate has emerged in the theory of fiscal policy. A "free lunch" may be possible in this realm of policy as well if budget-balance rules can deliver both more stable debt-to-GDP ratios in the long term and lower output volatility in the short term. As we have seen, the aspect of the Keynesian approach that makes this counter-intuitive outcome emerge is that the Keynesian policy creates a need for the government to work the debt ratio back to its full-equilibrium value following temporary disturbances. This need slows the economy's speed of adjustment back to the natural rate. Historically, it has been difficult for macroeconomists to evaluate dynamic issues when their models have had only limited macro-foundations. With its solid grounding in inter-temporal optimization, however, the new neoclassical synthesis gives analysts more confidence in their ability to assess dynamic considerations of this sort.
CHAPTER 8

STRUCTURAL UNEMPLOYMENT

8.1 Introduction

In earlier chapters, we examined cyclical (temporary) unemployment. For example, when actual output falls below the natural rate of output, unemployment is temporary only above its natural rate. We can assume cyclical unemployment to have dropped to zero. Instead, we focus on equilibrium, or permanent, unemployment. We assume that unemployment can persist in full equilibrium for three reasons:

- Asymmetric information problems cause the equilibrium wage to exceed the market clearing level.
- The market power of unions causes the equilibrium wage to exceed the market clearing level.
- Transaction costs and friction in the labour market result in the simultaneous existence of both unemployed individuals and unfilled job vacancies.

We discuss each of these approaches to modelling the labour market in a separate section of the chapter. Much of this work has been pursued by a group of economists who are often called "New Keynesians" — economists who believe that there is market failure in labour markets, but who also believe that explicit macro-foundations are essential to rigorous macroeconomics. Some New Keynesians have shown that real rigidity — such as the phenomenon we study in this chapter — can make any given degree of nominal price rigidity more important from a quantitative point of view. When we reach this chapter, we can see that this argument is not only right, but that the models of the labour market, and we focus on identifying policy interventions that can be expected to lower the annual unemployment rate in all three settings.

8.2 Asymmetric Information in the Labour Market: Efficiency Wages

To start, we consider structural unemployment that results from the first of the three considerations listed above: asymmetric information. Workers know whether they are trying their best on the job, while employers can monitor this worker effort only incompletely. This differential access to information may lead firms to use wage policies as a mechanism for inducing higher productivity from their employees. Since a quasigold worker is a good worker, firms can have workers who put forth high effort if they make it the case that workers really want to keep their current job. But one price — the wage rate — cannot clear two markets.

In the next section, we focus on the relation between the real wage and the number of workers who are actually working in a given industry. Economists have found that the number of workers who are actually working in a given industry is positively correlated with the real wage. This result holds both for industries that are in the process of expanding and those that are in the process of contracting. Economists have found that the number of workers who are actually working in a given industry is positively correlated with the real wage. This result holds both for industries that are in the process of expanding and those that are in the process of contracting.
To see the intuition behind this wage-setting rule, we use the other first-order condition to substitute out $P_2 = \nu$ and get:

$$w = \delta / (1 - \beta).$$

According to this rule, firms must set the wage equal to their workers' outside option if there is no variability in worker effort (if $\alpha$ equals 0). This is what is assumed in the standard competitive model of the labour market. But with variable worker productivity ($\alpha > 0$), it becomes optimal to set the wage above the outside option to induce workers to work hard (to lower the probability of getting fired by shirking less).

The implications for the unemployment rate can be determined once the workers' outside option is defined. We assume:

$$b = (1 - \alpha) w + \delta b.$$ 

With standing for the unemployment rate, there is a probability equal to the employment rate ($1 - \omega$) that fixed workers will find a job with another firm (that in full equilibrium pays the same wage). There is a probability equal to the unemployment rate that the individual will be without a job. We assume a simple unemployment insurance program in which individuals receive $F/\omega$ of their wage unemployment-insurance benefit (without any waiting period or time limits) in this eventuality. The $b$ equation above defines the outside option as this, weighted average. When this relationship is substituted into the wage-setting rule, we have:

$$u = \alpha (1 - \beta).$$

This solution for the unemployment rate teaches us three things. First, unemployment is zero if there is no variability in worker effort (that is, if parameter $\alpha$ is zero). Second, increased generosity in the unemployment insurance system (a higher value for parameter, $\delta$) raises unemployment. This is because higher unemployment insurance reduces the relative payoff that individuals get by keeping their job. As a result, they choose to shirk more. Knowing this, firms raise the wage in an attempt to lessen this reaction of their employees. With higher wages, firms shift back along their (downward sloping) labour demand curve, and hire fewer workers. (By the way, this does not mean that unemployment insurance is "bad." After all, with unemployment insurance, any one unemployment spell hurts the individual less. It is just that there is a trade-off; this beneficial effect induces an increased frequency of unemployment spells.) The third implication of the solution equation for the unemployment rate follows from the fact that it does not include a productivity term, $F$. Thus, the proposition that investment in education (that raises overall productivity) would lower unemployment is not supported by this analysis. Higher productivity is desirable because it raises the wages of those who already have jobs, not because it brings more people jobs. This prediction is consistent with centuries of economic history. Vast productivity growth has led to similar increases in real wages, without any significant long-term trend in the unemployment rate.

We have focused on this model so that we have at least one rigorous framework for arguing that some unemployment is involuntary. We will use the theory to evaluate how fiscal policy might be used to lower unemployment in the next chapter. Before closing the present discussion of efficiency-wage theory, however, it is useful to extend Summers' analysis by considering optimism on the part of households (not just firms). It is preferable that the household variable-work-effort function be derived, not assumed. This can be accomplished by assuming that households maximize $(7\omega + 2 + 1 - 7v)b^{-\beta}$. It is the proportion of time that the individual is employed in her current job, and this proportion is: $7 + \alpha$. The first two terms in the objective function define the individual's income, $\text{receives} W$ if she keeps her current job, and she receives $b$ if she does not. The final term defines the duality associated with putting effort into one's job. To be compatible with the income components of the objective function, this term is scaled by $b$. Since it is preferable to specify that higher work effort increases the probability of keeping one's job, but at a decreasing rate, $\alpha$ must be less than one. Similarly, since it is appealing for higher effort to decrease utility at an increasing rate, $\beta$ must exceed one. Household behaviour follows from substituting in the constraint and differentiating the objective function with respect to $\omega$. The result is the effort function given above, if $\omega$ is interpreted as $1 / (\nu - \omega)$ and units are chosen so that $1 / (\omega - \omega) = 1$. 

The union achieves the highest indifference curve by packing the wage that corresponds to the point at which the labour demand curve is tangent to an indifference curve (point $A$ in Figure 8.1). Once the wage is set, firms are free to choose employment. But since the union has taken the firm's reaction into account, its members know that point $A$ will be chosen. We can derive what this model predicts concerning the real wage and employment by including a shift variable — such as $A$ in a revised production function, $AP(A)$). Comparative static predictions are calculated by taking the total differential of the labour demand curve and the equal wages condition. It is left for the reader to verify that the model does not predict real wage rigidity. One purpose of examining this model was to see whether the existence of unions in labour markets leads to wage rigidity and/or involuntary unemployment. Since the model contains no explicit explanation of why individuals deal with the firm exclusively through the union (are they forced to? did they choose to?), it is not possible to say whether lower employment means higher involuntary unemployment.

Let us now investigate whether wage rigidity occurs in the cooperative model of union-firm interaction. The outcome in the previous model is inefficient since there are many wage-employment outcomes — all the points within the shaded, lens-shaped region in Figure 8.1 — that can make both the firm and the union better off than they are at point $A$. The cooperative model assumes that the two parties reach an agreement and settle at one of the Pareto-efficient points that lie along the contract curve. Completing the model now requires some additional assumption that defines how the two parties divide the gains from trade. The additional assumption that is most common (see McDonald and Solow (1981)) is that the two bargainers reach a Nash equilibrium. Without specifying some rule of this sort, we cannot derive any predictions about how the wage level responds to shifts in the position of the labour demand function.

Employment effects can, however, be derived without any such additional specification. The equation of the contract curve is had by equating the slope expressions for the iso-profit and the indifference curves. With the shift variable for labour's marginal product assumed, this equal slopes condition is $(\delta^2 - \omega) / (\delta - \omega) = (\delta^2 - \omega) / (\delta - \omega)$. The
contract curve is vertical since w does not enter this equation. From this equation of the contract curve, we see that we can determine the effects on employment of changes in $\lambda$ and IT without having to specify the bargaining model that is required for the model to yield any real wage predictions. It appears that this co-operative union model does not support the hypothesis of real wage rigidity, but we can derive the employment effects that follow from this theory if we impose real wage rigidity in a macroeconomic context. For macroeconomic employment effects, it is as if real wages were fixed.

We complete our analysis of the co-operative model by using the standard Nash product to derive the conditions that determine the division of the rents between the union and the firm. The function that is delegated to the arbitrator to maximize involves the product of two terms: first, what the firm can earn in profits if co-operation is achieved (minus what it gets with no co-operation — zero), and second, the sum of the differentials for workers. This product can be written as $(w - w^*)$, where $V$ is profits, $P = A F (N) - w N$, $w^*$ is the bargaining power parameter ($0 = 1$; unions have all the power; $0 = 1$: firms have all the power), and $N$ is a union preference parameter. $(w = 1$: the union is unilateral in that it values all its members (not just the currently employed); $w = 0$: the union is seniority oriented (only the wages of those currently employed are valued).

After differentiating the arbitrator’s objective function with respect to $w$ and $N$, we have two labour market equations (which are the standard competition analogues for supply and demand curves) to determine wages and employment. McDonald and Solow call these two relationships the efficiency locus and the equity locus. The equation that defines the contract curve is the efficiency locus. It is $A F (N)$ with a union wage, while it is $A F (N)$ without a union-based union. With the seniority-based union, the unattained pay to attention to employment and the unemployment difference curves are horizontal lines. Thus, in this case, the labour demand curve and the contract curve coincide, and this version of the system essentially replicates the non-co-operative model. Finally, the equation which defines the division of the rent (the equity locus) is $N = (w - w^*) V - (w - w^*)$, whether unions are unilateral or not.

Pissarides (1980) uses a model of union—firm interaction that combines features of the two different approaches that have just been summarized. His model follows the "two-stage" aspect of the non-co-operative approach in that the firm chooses the employment level after the wage, has been determined. But the model involves a key feature of the co-operative approach as well, since the wage is not set unilaterally by the union, but rather as a result of a bargaining process involving both the union and the employer. A simplified version of Pissarides’ model involving risk neutrality on the part of the union and a Cobb-Douglas production function is explained here in the first stage. An arbitrator is appointed to choose the wage which maximizes the following Nash product function: $(w - w^*) V - (w - w^*)$. $V$ is the index of the union workers’ net benefit from the contract. Pissarides’ definition of this net benefit is $I = w N - (w - w^*) V - (w - w^*)$. As above, $N$ is employment in this firm, and $L$ is union membership. It is assumed that those who do not find employment in this firm seek employment elsewhere. These individuals face probabilities equal to the unemployment rate, and the unemployment rate concerning whether they secure another job (and are paid $w^*$) or whether they are without work (and receive employment insurance equal to IT). $I$ is what individuals receive if employment at this firm, $N$, is zero. Thus, $I = (w - w^*) V - (w - w^*)$. As far as the firm’s product is concerned, we have $V = V$ and the production function is $V = V$.

Differing the arbitrator’s objective function with respect to $w$, and then substituting in the equations that define full equilibrium ($N = w^*$) and the unemployment insurance system ($45 = w^*$), we have $u = k (1 - \beta)$, where $u = a (1 - \beta) (45)$. We note that this version of imperfect competition in the labour market yields the same equation for the natural unemployment rate as did our earlier wage model. Despite this similarity, there are some new results embedded within this alternative derivation of this reduced form. For example, in this union—firm interaction framework, we see that the natural unemployment rate is predicted to rise, the higher is the degree of union power. Thus, if lower structural unemployment is the goal, we might view the model as providing some support for legislation that is designed to limit workers’ rights.

The more general point is that many policies that are designed to lower the natural unemployment rate (some of which we stress in the next chapter) receive equivalent analytical support — whether one appeals to efficiency-wage theory or union—firm interaction theory. We can have more confidence about asking applied policy questions when the underlying rationale for that policy proposal is not dependent on just one interpretation of the labour market or the other.

5.4 Transaction Frictions in the Labour Market: Search Theory

We continue our examination of "natural" unemployment rate models by considering one more source of market failure. In this case, instead of incomplete information (and efficiency wages) or market power (and unions), we focus on friction in the labour market. This analytical framework highlights the fact that job seekers and employers meet in a series of decentralized one-to-one settings. Since matches do not occur instantaneously, this setting generates an equilibrium amount of frictional unemployment. This section follows Romer (2001, pages 444–526) and Hall (2003) very closely in outlining as simple a summary of search theory as is possible.

First, we define some notation. $E$, $U$ and $F$ stand for the number of individuals who are employed, unemployed, and in the labour force. The labour force is taken as an exogenous constant. In this section of the chapter, we use $f$ and $s$ to denote the job finding rate and the job separation rate. This notation can be used to define equilibrium. The labour market is in equilibrium when the unemployment rate is constant, and this, in turn, requires that the number of individuals leaving the unemployment pool each period be just balanced by the number who are entering that pool. In symbols: $f, s, F, U$. Since $E = L - U$ and since the unemployed pool is, $U$, this equilibrium condition can be solved for the unemployment rate:

$$u = \frac{N}{N} = \frac{N}{N}.$$

What search theorists have done is to build a model that makes the job finding rate, $f$, endogenous (and based on optimizing behaviour). The resulting solution for $f$ is then substituted into the unemployment-rate equation just presented, so that the determinants of the unemployment rate are then exposed.

We need additional notation to define this theory (and we use Romer’s): $A$ — the marginal product of each employed worker, $C$ — the fixed cost of maintaining a job, and $w$ — the wage rate. It is assumed that there is no cost of posting a vacancy. The profit associated with each filled job is given by $(A - C - \lambda)$, and the profits associated with each job vacancy is $(C - C)$. We assume static expectations, so that we can represent the present value of receiving these flows indefinitely, by simply multiplying them by (1 + r).

The technology of the matching process is specified by:

$$M = M (E, F),$$

where $M$ is the number of matches. If $y = 1$, it is said that there is congestion or crowding in the labour market (the more individuals and firms there are in the market, the less chance there is that any one match can be made). If $y = 0$, it is said that there is a "slack market" externality involved (the more individuals and firms there are in the market, the more likely it is that a good match will be found). This is the assumption made by Heyes (1989) and Diamond (1984) in the multiple-equilibria models that are discussed in the next chapter. Finally, if $y = 1$, we rule out both negative and positive externalities effects. Since the empirical evidence weakens this case, we proceed with this assumption, as below. Specifically, we detach the vacancy rate, the unemployment rate, and the ratio of these two measures, $v = \frac{N}{N}$, $u = \frac{N}{N}$, and $w = w / w$, we can define the job finding rate as

$$J = J (v, u, F) = \frac{V}{V} + (u / u) - (1 / 1).$$

Similarly, the job filling rate can be represented as

$$F = F (u, v, N) = \frac{V}{V} + (u / u) - (1 / 1).$$

We complete the model by defining the utility function of individuals. For simplicity, it is specified that individuals are risk neutral and that they do not save, so utility equals current income. Thus, the level of utility is $u$ if an individual is employed, and utility is zero if she is unemployed. Dynamic programming is used to define optimal behaviour. In this approach, we consider (for example) the "annual returns" that an individual receives from being employed. This value is denoted by $v R E$, and it is equal to the sum of a "dividend" and a "capital loss":

$$R E = w - w^* - V - (u / u).$$

This equation defines the annual dividend of having a job as the wage received, and the capital loss as the difference between the value of a job and the value of not having one (being unemployed). The probability of sustaining this capital loss is the job separation rate. The annual return that is associated with each state is defined in a similar manner.

The value to a firm of maintaining a filled job is
The firm's annual profit is the "dividend" and the difference between the value of a filled job and the value of an unfilled vacancy is the "capital loss". Again, the separation rate is the probability of sustaining this capital loss.

The annual return of being unemployed is given by

\[ rW_i = 0 + f (W_i - Y) \]  

(8.3)

since, without unemployment insurance, the dividend is zero, the capital gain is the increase in utility that comes with the possibility that the individual becomes employed, and the probability of receiving this gain is the job finding rate.

Finally, the annual return to the firm of maintaining an unfilled vacancy is

\[ rV_i = -C + (W_i - Y) \]  

(8.4)

The dividend is negative (the cost of maintaining any position), the capital gain is the extra profits that are received if the position is filled, and the probability of receiving this gain is the job filling rate.

The remaining equations define different aspects of full equilibrium. As noted above, a constant unemployment rate requires that the flows out of and into the unemployed pool must be equal:

\[ M = xE, \quad \text{or} \quad JU = zE \]  

(8.5)

Given that the cost of posting a vacancy is zero, firms must have posted a sufficient number to ensure that the marginal benefit is zero:

\[ rV_i = 0 \]  

(8.6)

Finally, the wage must be set so that the gains of the match are distributed between individuals and firms in a way that is consistent with the "market power" of these groups. As in our models of union/firm interaction, we assume a Nash equilibrium, and for simplicity here, we assume equal levels of bargaining power. This implies that the wage is set so that an individual's gain from a match is equal to the firm's gain from the relationship:

\[ (r - yU)(W_i - Y) = (rV_i - Y) \]  

(8.7)

The solution proceeds as follows. Subtracting (8.3) from (8.1), and (8.4) from (8.2), we have

\[ V_i = 1/2 [(f + z + r)(A - wo)(B - wo) + (A - wo) + z + r) \]  

(8.8)

and substituting (8.9) into (8.7), and solving for w, we have

\[ w = A(f + z + r)(A - wo)(B - wo) + (A - wo) + z + r) \]  

(8.9)

and substituting (8.9) into (8.8):

\[ V_i = -C + (A - wo)(B - wo) + z + r) \]  

(8.10)

Substituting (8.10) into (8.11) to eliminate w, we have

\[ (f + z + r)(x - 0) = 2(z + r) \]  

(8.11)

which is a nonlinear equation in x, the ratio of the unemployment rate to the vacancy rate. Hall (2003) picks values for z, r, x, and 13, and solves for x. Once x is known, Hall solves for the unemployment rate:

\[ u = (1/2)(1 - 0)(x) \]  

(8.12)

Part of Hall's calibration is that 13 = 0.3, so (8.12) becomes a quadratic equation. Hall chooses representative values for the other parameters as well, and as a result (8.12) becomes:

\[ 0.03x^2 - (0.9913)y = 0 \]  

where y = \( x \).

Of the two solutions, only one is positive (that is, economically admissible). This one solution is used in the version of the model that is extended to allow for several taxes — in the next chapter. As with the other models of market failure in the labour market, we identify fiscal policies that can be expected to have favourable effects on the natural unemployment rate.

8.5 Related Issues in New Keynesian Economics: Real vs. Nominal Rigidities

Some of the labour market models that were surveyed in the three previous sections of this chapter provide support for the hypothesis of real wage rigidity, but a fundamental problem for business-cycle theory is to explain why purely nominal shocks have real effects. The Keynesian and new neoclassical synthesis approaches to this question rely on nominal wage and/or price rigidity. Can the models of this chapter apply in any way to this question?

Some analysts have argued that the theories of real wage rigidity can apply to nominal wages in an indirect way. When wages are set bargaining is about the real wage. But the item that is actually set as a result of these decisions is the money wage. It is set at the level intended to deliver the desired real wage, given inflationary expectations. With this interpretation, we can argue that the theories apply to money-wage setting, although an additional assumption regarding indexing is required. We would expect agents to set the money wage with a full indexing clause so there would be no need to incur errors in inflationary expectations. But given that catch-up provisions can roughly replace ex ante un indexing formulae, and given that households and firms want to tie wages to different price indexes, the costs of full indexing are probably not worth the benefit (as McCallum (1986) has stressed).

Quite apart from the preceding argument, some analysts are uneasy about applying an adjustment-cost model (such as the one we explored in Chapter 4) to explain sticky prices. Negotiation costs between buyers and sellers are not a feature of reality for many commodities. Of course the sale of many commodities involves posting prices, but it does not seem compelling to rest all of sticky-price macroeconomics on this issue that seems rather trivial (such as the cost of praming new prices in catalogues — the so-called "menu" cost — and the cost of informing sales staff about price changes). The response that one can make to the change that adjustment costs for many nominal prices cannot be "that important" is simply to demonstrate that even explicitly small price-change costs can lead to large welfare losses. Akcigit and Yellen (1983) and Mankiw (1985) provide analysis that is intended to support this view.

Let us examine a brief summary of the argument (along the lines suggested by Romer (1991)). Consider a monopolist who must set her nominal price before the relevant period but who can change that price later (during the period) at a "small" cost. Assume as the price set its price, it did not guess the then-future position of its demand curve perfectly. As it enters the period analyzed in Figure 8.2, it has already posted a price equal to 0d, but the appropriate price is 0b. The firm must now decide whether making the change is worthwhile. As far as private Hicks taxes are concerned, the firm loses an amount equal to area \( FGBH \) by not lowering its price to 0b. The cost to society of not adjusting the price area \( DBHCE \) — potentially a much bigger amount. It is quite possible for even quite small adjustment costs to be larger than the area \( FGBH \) but much smaller than area \( DBHCE \). Thus, the social gains from price adjustment may far exceed the private gains.
Research continues on the microeconomics of menu costs. To some, the most appealing model of price changes at the individual level is known as the two-sided (\(S, b\)) adjustment rule. It involves the firm only incurring the fixed cost of adjustment when the gap between the desired price and the existing one exceeds a critical value \(b\) on the high side. On the low side, heterogeneity among firms can tolerate various forms such as differing initial positions within common \(S, b\) bands, or firm-specific shocks. As is usual in the aggregation literature, not all specifications lead to well-defined macro implications.

Ball and Mankiw (1994) draw the distinction between time-contingent adjustment models and state-contingent adjustment models.

The theory we covered in Chapter 4 is an example of the former. While the \(S, b\) models are examples of the latter. Ball and Mankiw note that no robust conclusions have emerged from the literature on state-contingent adjustment, but that this state of affairs is not necessarily uplifting. This is because time-contingent rules of the type \(S, b\) are gathering information about the state rather than making the actual price adjustment. Also, in economics with two groups of firms, one making each kind of adjustment it is clear that the sluggish adjustment on the part of the time-contingent firms makes it rational for those monitoring developments continuously according to the state-contingent model to behave much like the other group. Thus, it may well be that the quadratic adjustment cost model of Chapter 4 is not such a bad approximation of a theory with much more detailed structure.

Another issue that is being researched is whether it matters to specify explicitly that it is the gathering of information, not the re-setting of prices, that is costly. Mankiw and Rees (2002) have shown that a "sticky" information", or "sticky expectations", of an "opportunistic" sort may fit the facts better than a "sticker" price version does. Further, they demonstrate that this version of a new synthesis model can lead to different conclusions regarding the relative appeal of alternative monetary policies.

8.6 Conclusions

In Chapters 1 through 4, we focused on the deviations of real output from its natural level. We assumed that the natural output rate was unique, and that we could think of it as being determined by the Classical model. Further, we assumed — but had not formally shown — that if we add some friction to the specification of the labour market, we could generate a unique, non-zero value for the associated natural unemployment rate. The monetary task of the policy maker is then to provide such adjustments to the set of analyses for the labour market. We have found that market failure can occur in the labour market for several different reasons — incomplete information, imperfect competition, and transactions costs. In some of these cases, these features allow us to interpret the resulting unemployment as involuntary, and so it is reasonable to investigate whether policy can improve the outcome. We proceed to address this very question in the next chapter.

There was one other purpose in surveying the several leading models that generate real rigidities in labour markets. We have seen that real rigidities — such as those highlighted in this chapter — may be the implication of nominal rigidities. This is important since both New Keynesians, and, now all those who have adopted the new neoclassical synthesis, have relied on this proposition to "justify" their having their approach to business cycles depend on seemingly small 'menu' costs in an important way. Even New Keynesians such as Akerlof and Yellen (2004) — who have added a version of efficiency-wage theory to the real business cycle framework — are relying largely on this proposition to generate more persistence in real variables within their models.

There are three tasks that we address in the remaining chapters. First, as just noted, we use the models developed here to analyze a series of policy proposals designed to lower structural unemployment and to raise the economic position of those on low income. Second, we use some of these models to investigate the possibility of multiple equilibria. If theory leads to the possibility that there are two 'natural' unemployment rates — both a high-unemployment equilibrium and a low-unemployment equilibrium — there is an "unavoidable feature" rule for policy. It may be possible for the government to induce agents to focus on the high-activity outcome if agents know that the policy makes stands ready to push the system to that outcome if necessary. It is possible that no action — just the commitment to act — is all that may be necessary. Third, we must see whether policies that are geared to reducing structural unemployment have an undesirable long-run implication. Might these initiatives retard the productivity growth rate? We examine the first two issues in the next chapter, and then devote the final three chapters to an analysis of long-term growth.
CHAPTER 9
UNEMPLOYMENT AND LOW INCOMES: APPLYING THE THEORY

9.1 Introduction
In the last chapter, we summarized three approaches to modeling the labour market—standard analyses of what determines the level of structural unemployment. While this is often called the ‘natural’ unemployment rate, this term is unfortunate, since all these theories suggest that the full-equilibrium unemployment rate can be affected by fiscal policy. The primary task for this chapter is to investigate precisely how. We consider several fiscal policies:

- replacing the income tax with an expenditure tax in section 9.2,
- taxing physical capital owners to finance a tax break for wage earners in section 9.3,
- introducing low-income support policies, such as employment subsidies and guaranteed annual income, in models of both developed and developing economies in section 9.4, and
- investigating how policy can both create and react to multiple equilibria in section 9.5.

9.2 Tax Reform: Direct vs Indirect Taxation
We begin by illustrating the possibilities for fiscal policy within the Summers (1988) version of efficiency-wage theory. First, we add an income tax to the model, which can be interpreted in two ways. First, we could be an employee payroll tax (that is levied on wage income but not on unemployment-insurance benefits). Second, it could be part of a progressive personal income tax, that involves no tax on low levels of income (such as what individuals receive if all they have access to is the unemployment-insurance benefit). With this tax, the specifications of the efficiency index and the outside option change compared to how they were specified in Chapter 8. These relationships are now given as

\[ q = \left( \frac{1}{1 + \delta} - \frac{\beta}{1 + \delta} \right) \frac{w}{(1 - \gamma) N} \]  

and

\[ s = \frac{1}{1 - \gamma} \frac{w}{(1 + \delta)} \]  

where \( \delta \) is the wage-income tax rate. It is left for the reader to verify that these modifications change the unemployment rate solution to

\[ s = \frac{1}{1 - \gamma} \frac{w}{(1 + \delta)} \]  

This equation implies that an increase in the tax rate raises the natural unemployment rate. This occurs because higher taxes reduce the relative pay-off individuals receive from work. To lessen the resulting increase in worker skimping, firms offer a higher wage, and thus fewer workers at this higher price.

The importance of taxes can be illustrated by considering some illustrative parameter values. Realistic assumptions are \( a = 0.5, \delta = 0.50, \) and \( \gamma = 0.15 \). These representative values are consistent with this model only if \( a = 0.5 \), which we therefore assume. Now consider fixing \( a \) and \( \delta \) at 0.92 and 0.3 respectively, while higher tax rates are considered. The reader can verify that the unemployment rate rises by one percentage point (to \( s = 0.06 \)) when the tax rate rises by 10 percentage points to 0.25, and the unemployment rate rises by much more (2 and 2.5 percentage points, to 0.0675) when the tax rate rises by an additional 10 percentage points to 0.35. This thought experiment indicates that one does not need to have ultralight-weight views to be concerned about efficiency in government. Only with such efficiency can we have the many valuable services of government with the lowest possible taxes, and (as this numerical example suggests) high taxes can very much raise unemployment.

It is instructive to examine the effects that several other taxes have (or more precisely, do not have) within this basic version of the efficiency-wage model. With an employer payroll tax, \( t \), the firm's wage becomes \( w(1 + t) \), and with a sales tax, \( X \), the wage that concerns households is \( w^* = w(1 - X) \). It is left for the reader to verify that, when these changes are made in the specification of the efficiency-wage model, there is no change in the solution equation for the unemployment rate. It is useful to review the intuition behind why employee payroll taxes do, but these other taxes do not, affect unemployment. As already noted, both are more generous unemployment insurance systems and a higher employee payroll tax increase unemployment. Both these measures lower the relative return from working. To compensate for the deterioration in work effort that results, firms must raise wages, and this makes a lower level of employment optimal.

The other taxes do not change the relative return of work compared to being unemployed. For example, sales taxes must be paid simply because goods are purchased; it makes no difference how the purchaser obtained her funds. This is why the natural unemployment rate is unaffected by the sales tax. Similar reasoning applies to the employer payroll tax. A cut in this levy increases both the ability of the worker's employer to pay higher wages and the ability of all other firms to pay individual higher wages. Competition among firms for workers raises this entire increase in ability to pay to be transferred to those already working (in the form of higher wages). As a result there is no reduction in unemployment. The same outcome follows for anything that shifts the labor demand curve upward but has no direct effect within the workers' effort function. This is why we stressed in the previous chapter that increases in general productivity raise wages — and do not lower unemployment — in this efficiency-wage setting.

These results imply that we can have a lower natural unemployment rate if we rely more heavily on a sales tax, instead of an income tax. They also imply that investments in training and education lead to higher wages, but not to lower unemployment. But before we can have confidence in such strong predictions, and exhort real-world authorities to act on this advice, we need to know whether they are supported by the other theories of the natural unemployment rate.

To check the effects of various fiscal policies in our models of union-firm interaction, we add a wage-income tax (which, as above, can also be interpreted as the employee payroll tax), an employer payroll tax, and a sales tax. As in Chapter 8, the function that is delegated to the arbitrator to maximize involves the product of two items: first, the worker's firm can earn in profits if co-operations is achieved (unums what it gets with no co-operation — zero), and second, the similar differential in returns for workers. This product is \( (1 - \gamma) f(1 + t) \). This is the wage bargainining power parameter, and \( w \) is the union bargaining power parameter.

After differentiating the arbitrator's objective function with respect to \( w \) and \( \delta \) and simplifying, we have the tax-included versions of the two-labor market equations that determine wages and employment. The equation that defines the contract curve is \( (1 - \gamma) f(1 + t) = \frac{w}{(1 + \delta)} \) with a utilitarian union, while it is \( (1 - \gamma) f(1 + t) = \frac{w}{(1 + \delta)} \) with a seniority-based union. The equity relationship is \( (1 - \gamma) f(1 + t) = \frac{w}{(1 + \delta)} \) (1 - \( \gamma \)) whether unions are utilitarian or not. In both cases, the level of employment is unaffected by sales taxes, but it is affected by both the employees and the employer payroll tax (on income, or rather taxes raises unemployment).

We add the same set of taxes to the Posudios (1998) model of union-firm interaction that combines features of the co-operative and non-co-operative approaches. The arbitrator's objective function is still \( (1 - \gamma) f(1 + t) \), and the production function is still \( Y = aN \). There are several changes:

\[ \beta = \beta - \delta \gamma - \delta \gamma - \delta \gamma \]  

\[ \gamma = \gamma - \delta \gamma \]  

\[ \delta = \delta - \delta \gamma \]  

Proceeding with the same steps as we followed in Chapter 8, we arrive at the revised solution equation for the unemployment rate

\[ s = \frac{1}{1 - \gamma} \frac{w}{(1 + \delta)} \]  

where, as before, \( \gamma = 0.5, \delta = 0.50, \) and \( \gamma = 0.15 \). The policy implications are a little different from those that followed from the other models of union-firm interaction, but they are the same as those followed from the efficiency-wage model. For all the models, we have found that the natural unemployment is increased by higher employee payroll taxes, but it is not increased by a higher sales tax. It appears that there is one general conclusion that has emerged from both efficiency-wage theory and union-firm interaction theory: if a lower unemployment rate is what we should reduce employee-payroll and wage-income taxes, and finance these tax cuts by imposing a higher sales tax. It is reassuring that the support for this move toward a heavier reliance on indirect taxes receives the same analytical support from both theories about labor markets. We can have more confidence about making applied policy advice when the underlying rationale for that policy proposal is not dependent on just one interpretation of the labour market or the other.

But our examination of this policy proposal is not complete; we must derive its implications in the search model as well. To pursue this sensitivity test, the same set of taxes is added to that model. It is left for the reader to verify that equations (8.10) and (8.12) are altered:

\[ \omega = \frac{\omega + \omega}{(1 + t)} \left[ (1 + t) \frac{w}{(1 + t)} \right] + (1 + t) \frac{w}{(1 + t)} \]  

where \( \omega \) is the wage bargainining power parameter, and \( w \) is the union bargaining power parameter.
Proceeding with the solution, and Hall's (2003) calibration, we reach the several policy conclusions. Some are similar to the outcomes that we discovered in our analysis of efficiency wages and unions. For example, an increase in the employee payroll tax increases unemployment. Even the magnitude of this response is comparable to our earlier findings. (If it is raised from zero to 0.1, the unemployment rate rises by about one half of one percentage point.) This finding means that our earlier conclusion is robust across alternative specifications of the labour market. For this policy, at least, it appears not to matter that there is controversy concerning how best to model structural unemployment. Policy makers can proceed without waiting for this controversy to be resolved.

But this assurance does not apply to all policy initiatives, since some of the implications of search theory are different from the policy theorems that followed from the other models. For example, in this specification, both the lower payroll tax and a lower interest rate reduce the natural unemployment rate — predictions that are at odds with both the efficiency-wage model and Passerside's model of union-firm interaction. But not all of these differences are important. For example, while the interest rate matters in the present specification (since a higher interest rate lowers the benefit of having a job and so raises equilibrium unemployment), the practical significance of this effect is non-existent. The reader can verify that, when Hall's calibration is used, and when the annual interest rate is raised by even two or three percentage points, the effect on the unemployment rate is truly trivial. Hence, socially, differences across models of the labour market are irrelevant for policy purposes, and we can proceed with the policy prescriptions that accompanied the earlier specifications.

However, not all the differences across natural unemployment rate models can be dispensed with in this way. For example, in this search model, the unemployment rate is increased by the existence of a sales tax. Again, for Hall's calibration, we find that increasing the variable from zero to 0.1, makes the unemployment rate rise by about one half of one percentage point. This is a non-trivial effect, and it differs markedly from the zero response we discovered with efficiency wages and unions.

This different outcome is important for the general debate on whether we should follow the advice of many public-finance proponents — that we should replace our progressive personal income tax with a progressive expenditure tax. According to growth theory (models which usually involve no unemployment, which we examine in chapters 10-12), this tax substitution should increase long-run living standards. According to efficiency-wage and union theory, this tax substitution should bring the additional benefit of lowering the natural unemployment rate. But as just noted, this fortuitous outcome is not supported by search theory. However, this search model indicates that the cut in the wage income tax can be expected to lower unemployment by about the same amount as the increase in the expenditure tax can be expected to raise unemployment. Thus, even this model does not argue for rejecting the move to an expenditure tax. In this limited sense, then, the labour market models give a single message: with respect to lowering the natural unemployment rate, we either gain, or at least do not lose, by embracing a shift to expenditure-based taxation.

9.3 The Globalization Challenge: Is Mobile Capital a Bad Thing to Tax?

One of the primary concerns about the new global economy is income inequality. Compared with many low-wage countries, the developed economies (often referred to as the North) have an abundance of skilled workers and a small proportion of unskilled workers. The opposite is the case in the developing countries (the South). With increased integration of the world economies, the North specializes in the production of goods that emphasize their relatively abundant factor, skilled labour, so it is the wages of skilled workers that are bid up by increased foreign trade. The other side of this development is that Northern countries rely more on imports to supply goods that require only unskilled labour, so the demand for unskilled labour falls in the North. The result is either lower wages for the unskilled in the North (if there is no legislation that puts a floor on wages there) or rising unemployment among the unskilled in the North (if there is a floor on wage rates, such as that by a minimum wage law and welfare). In either case, unskilled Northerners can lose ground in the new global economy.

There is a second hypothesis concerning rising income inequality. It is that, during the final quarter of the twentieth century, skills-based technical change has meant that the demand for skilled workers has risen while that for unskilled workers has fallen. Technical change has increased the demand for skilled individuals to design and program in such fields as robotics, while it has decreased the demand for unskilled workers since the robots replace these individuals. Just as with the free-trade hypothesis, the effects of these shifts in demand depend on whether it is possible for wages in the unskilled sector to fall. The United States and Europe are often cited as illustrations of the different possible outcomes. The United States has only a limited welfare state so there is little to stop increased wage inequality from emerging, as indeed it has in recent decades. Europe has much more developed welfare states that maintain floors below which the wages of unskilled workers cannot fall. When technological change decreases the demand for unskilled labour, firms have to reduce their employment of these individuals. Thus, Europe has avoided large increases in wage inequality, but the unemployment rate has been high there for many years.

Most economists favour the skill-biased technical change explanation for rising income inequality. This is because inequality has increased so much within each industry and occupation, in ways that are unrelated to imports. The consensus has been that only 11% of the rising inequality in America can be attributed to the expansion of international trade. But whatever the causes, the plight of the less skilled is dire.

Even if globalization is not the cause of the low income problem for unskilled individuals in the North, it may be an important constraint on whether their governments can do anything to help them. This is the fundamental challenge posed by globalization. Citizens expect their governments to provide support for low-income individuals so that everyone shares the benefits of rising average living standards. The anti-globalization protesters fear that governments can no longer do this. The analysis in this section — which draws heavily on Mouton and Scharf (2004) — suggests that such pessimism is not warranted. To address this question specifically, let us assume that capitalists (the owners of capital) are "rich" and that they have the ability to re-locate their capital costlessly to lower-tax jurisdictions. All, we assume that labour is "poor" and that these individuals cannot migrate to other countries. Can the government help the "poor" by raising the tax it imposes on the capitalists and using the revenue to provide a tax cut for the workers? Anti-globalization protesters argue that the answer to this question is "obviously no." They expect capital to relocate to escape the higher tax, and the result will be less capital for the captive domestic labour force to work with. Labour’s living standards could well go down — even with the cut in the wage-income tax rate. It is worthwhile reviewing the standard analysis, since it is the basis for recommending that we not tax a factor that is supplied perfectly elastically (such as capital) for a small open economy. Figure 9.1 facilitates this review. The solid lines represent the initial demand and supply curves for capital. The demand curve is the diminishing marginal productivity relationship that is drawn for an assumed constant level of labour employed. The supply curve is perfectly elastic at the yield that owners of capital can receive on an after-tax basis in the rest of the world. Before the tax on capital is levied to finance a tax cut for labour, the economy is observed at the intersection of these solid-line demand and supply curves, and GDP is represented by the sum of the five regions numbered 1 to 5.

When the government raises the tax on capital, capitalists demand a higher pre-tax return — an amount that is just enough to keep the after-tax yield equal to what is available elsewhere. Thus, the higher (dashed) supply curve in Figure 9.1 becomes relevant. Domestically produced output falls by regions 1 and 2. Capital owners do not lose region 1, since they now earn this income in the rest of the world. Labour loses regions 3 and 4, but since the tax revenue is used to make an unconditional transfer to labour, their net loss is just region 3. But this is a loss, so the analysis supports the proposition that capital is a bad thing to tax, and that it is impossible to raise labour’s income.
But this standard analysis involves the assumption that the policy has no effect on the number of men and women employed. If the level of employment rises, capital can be a good thing to tax after all. If there is unemployment in the labour market, and no similar excess supply in the capital market, the economy involves a distortion before this policy is instituted. The existence of involuntary unemployment means that, before the policy, society’s use of labour is “too small,” and that (from society’s point of view) profit maximization has led firms to use “too much” capital compared to labour. A tax on capital induces firms to shift more toward employing labour and this helps lessen the initial distortion. But this does cause a deadweight effect of the policy package outweigh the traditional cost (the loss of income represented by product 3 in Figure 9.1). Figure 9.2 suggests that this possible. As long as the wage-income tax cut results in lower unemployment, each unit of capital has more labour to work with, and so it is more productive. This is shown in Figure 9.2 as a shift up in the position of the marginal product of capital curve (shown by the brighter shaded demand curve). In this case, the total income available to labour is affected in two ways. It is reduced by the shaded triangle, and it is increased by the shaded parallelogram.

If the gain exceeds the loss, the low-income support policy is effective at all. If it is less unemployment, it raises the total income of the “poor” (about) and it does not reduce the income of the “rich” (the owners of capital). This approach to low-income support is not a zero-sum game, in the sense that labour is not helped at the expense of capitalists. This is because the size of the overall economic “pie” has been increased by policy. Labour receives a bigger slice, and capitalist get the same slice as before. And all of this appear possible — despite the fact that the government faces the constraints that are stressed by the anti-globalization protesters. The same result was stressed in Koskala and Schob (2002). In their model, the unemployment results from unions, not asymmetric information, as is the case in our specification below. Related work involving search theory instead of either efficiency wages or unions, is available in Dominy (2005).

There are two crucial questions: First, is it reasonable to expect that a cut in the wage-income tax rate will lower the long-run average unemployment rate? We addressed that question in the previous version of this chapter, and we discovered that the answer is “yes.” The second question concerns whether it is reasonable to argue that the gain can be bigger than the loss. It is straightforward to answer this question by combining: one of our models of unemployment (we choose the efficiency-wage model), a production function that involves both capital and labour as inputs, a government incomes policy, and the hypothesis of perfect capital mobility. We now define just such a model, and derive the conditions that must be satisfied for this revenue-neutral tax substitution to provide the Pareto improvement that we have just discussed.

\[
\begin{align*}
Y &= g(\gamma, \beta, K) \\
q &= \gamma(1 - \beta - 1) + \beta \\
b &= 1 - \gamma(1 - \beta - 1) + \beta \\
(1 - \gamma)N = w \\
y &= \frac{N}{K} r \\
w &= \frac{N}{K}(1 - \frac{1}{\beta}) \\
N &= \frac{1}{r} - u \\
r = \frac{1}{r} - u \\
C + f(w) = \frac{N}{K} + \beta Nw
\end{align*}
\]

The equations are explained briefly as follows. The first is a Cobb-Douglas production function, which indicates that output is determined by the quantity of inputs — capital and effective labour. The labour effectiveness index is defined in the second equation — as is in our efficiency-wage model in Chapter 8. By combining the second, third and sixth equations, readers can verify that worker productivity turns out to be an exogenous constant — independent of tax and unemployment-insurance generosity policy. \((g(\gamma, \beta, K))\). This fact simplifies the derivations that are referred to below. The third equation defines the average income of a labourer (which is equivalent to the outside option in the efficiency-wage model). The next three equations are the firms’ first-order conditions for profit maximization. Firms hire each factor up to the point that the marginal product equals the rental cost. Also, when the firms’ optimal wage-setting relationship is combined with the optimal hiring rule for labour, the solution for the unemployment rate (the sixth equation) emerges.

The next two equations define factor supplies. Labour supply is inelastic (at any rate) so employment is one minus the unemployment rate. Capital is supplied completely elastically at the rate of return that this factor can earn in the rest of the world. This perfect capital mobility assumption is what imposes the globalization constraint — that capital can avoid paying any tax in this small open economy. Finally, the last equation defines a balanced government budget. The use of funds (on program spending and unemployment insurance) must equal the sources of funds (the taxes on capital and wage incomes). In this section of the chapter, it refers to the tax on the earnings of domestically employed capital, not an employer payroll tax.

The equations determine \(T, N, y, w, b, K, r, q, g\) and \(\gamma\). We use this system to derive the effects on the unemployment rate, \(u\), and the average income of a labourer, \(b\), of a cut in the wage rate tax, \(t\), that is financed by a change in (presumed to be an increase in) the tax on capital, \(r\). To accomplish this, we take the total differential of the system, and eliminate the other endogenous variable changes by substitution. The goal is to sign \(\Delta b\) and \(\Delta l\).

It turns out that the second of these policy multipliers has an ambiguous sign. Nevertheless, we can show that the average income of a labourer must rise, as long as the government does not encounter a “Laffer curve” phenomenon. What this means is that the government must raise one tax rate when the other is cut. Laffer believed that the opposite might be true — that a cut in one tax rate might increase the level of economic activity (the overall tax base) that overall revenue collected would increase — despite the fact that the tax rate was reduced. Most economists read the evidence as being against this proposition, and we have concluded that tax cuts do not increase tax revenues. That is, most analysts are comfortable assuming that the other tax rate would have to be raised. We assume that here. To make use of this non-controversial assumption, we need to work out \(\Delta b\) and to assume that the expression is negative. It is left for the reader to verify that this assumption is necessary and sufficient to sign the average-income response (to ensure that \(\Delta b\) is negative). The unemployment-rate response is unambiguous in any event.

We conclude that low-income support policy by governments in small open economies is quite feasible — despite the constraint imposed by globalization — as long as the revenue that is raised from taxing capital is used to lessen the existing distortion in the labour market. We believe that the transfer to labour that is not conditional on employment status does not meet this requirement, using that instrument (in an attempt to provide low-income support) fails. Nevertheless, the fact that a Pareto improvement is found in the wage-income tax cut policy, warrants that the anti-globalization protests have been premature in their verdict concerning the inability of governments in small open economies to raise the economic position of the low-income individuals within their countries.

Before closing this section, it is worth reviewing why a Pareto improvement is possible. For an initiative to be both efficiency-enhancing and equity-enhancing, the economy must be starting from a “second best” situation. Involuntary unemployment involves just this kind of situation. We can clarify by recalling an example introduced in the original paper on this topic (Lipsky and Lancaster (1950)). In a two-good economy, standard analysis leads the representative salaried worker to choose a “bad” with a tax on the purchase of just one good, the ratio of market prices does not reflect the ratio of marginal costs, so decentralized markets cannot replicate what a perfect planner could accomplish — achieve the most efficient use of society's scarce resources. Society is producing and consuming "too little" of the taxed good, and "too much" of the untaxed good. But this conclusion assumes that there is no pre-existing market distortion — before the tax is levied. A different verdict emerges if it is assumed that there is an initial market failure. For example, if one good is produced by a monopoly who restricts output and raises price above competitive levels, a similar inefficiency is reduced (with society consuming "too little" of this good and "too much" of the competitively supplied good). There are two policies that can "fix" this problem. One is to try to use the Competition Act to eliminate the monopoly; the other is to levy a selective excise tax on the sale of the other product. With this tax, both prices can be above their respective marginal costs by the same proportion, and society gets the efficient allocation of resources — even with the monopoly.
The results are rather messy, but it is left for the reader to take the total differential of the model to evaluate how each individual's average labour income, variable $b$, is affected by these policies. The government introduces either the employment subsidy or basic income (raising either $z$ or $p$ above an initial value of zero), and we assume that both tax rates are equal initially. Only one quantitative assumption needs be made to sign one of the outcomes. The signs of all other responses can be determined a priori. The employment subsidy definitively lowers the unemployment rate, and it raises average labour income as long as $y < 1/2$. This condition is satisfied for plausible parameter values (for example, capital's share of income equal to one-third and any tax rate up to 5%), so the model supports the introduction of employment subsidies. A rather different set of results emerges with the guaranteed annual income policy. It has no effect on the unemployment rate, and it must reduce average labour income. It is true that each individual is better off because she receives basic income, but she is worse off since her market wage falls. The increase in the tax on capital that is necessary to finance the basic income policy drives enough capital out of the country to make labour noticeably less productive. This negative effect must dominate, so the analysis does not support the introduction of a guaranteed annual income.

The overall conclusion is that the employment subsidy can be defended — even when the model highlights the "globalization constraint" (the fact that the financing of the initiative requires a higher tax rate which scares away capital). However, the basic income policy cannot be defended. The intuition behind this difference in outcomes is the same as that which applied in the previous section. The employment subsidy addresses a distortion (asymmetric information in the labour market) a source, while the guaranteed annual income does not.

We make an analysis of employment subsidies in our additional way. To motivate this further investigation, we note that our model has not involved any specialized features that would make it particularly applicable to developing economies. Developmental economists have stressed two things about production possibilities in the least developed countries that we now insert into our analysis. First, they have stressed that developing countries often have a limited supply of some crucial input — a problem that can be highlighted if we restrict our attention to the Cobb-Douglas production function (that allows firms to produce each level of output with any ratio of factor inputs). Second, they have stressed that workers can be under-nourished so that their effectiveness on the job can be compromised. The following adaptation of the earlier model allows for these considerations.

$$Y = 0.1 \delta N$$
$$V = (z + \delta) K$$
$$q = (w - 0.1) b$$
$$b = \frac{1}{2} - \frac{1}{2} \frac{y}{K} N = 1 - \frac{y}{K}$$

$$y = v N - y N$$
$$r = \frac{1}{2} - y N$$
$$G = v N - r N$$

For simplicity, we have removed the unemployment insurance program. The new policy parameters are $z$, $p$, and $b$, respectively. A subsidy of $y$ per employee, and the government makes this subsidy proportional (at rate $z$) to the average wage in the economy. Individuals receive a payment of $P$ as a guaranteed annual income, and the government makes this payment function $p$ of the average wage. Everyone receives this basic income, whether she is working or not. Because this receipt is independent of employment status, it does not affect the unemployment rate. But because firms get a bigger employment subsidy, the larger is their work force, the employment subsidy does affect the unemployment rate. The readers can verify the revision in the unemployment rate equation by re-solving the firm's profit maximization. In this case, profits equal $Y = v N - r N + SN$. After optimization, we simplify by substituting in $S = w N$. As in the previous section, it is assumed that both policy outcomes are financed by an increase in the tax on capital.
set taxes on both forms of labour to zero. The novel feature in the effort relationship is the second argument on the right-hand side. We can think of this as a "nourishment effect," with parameter \( \gamma > 0 \); it is this that — other things equal — the higher the unskilled labour wage, the more healthy, and therefore, the more productive are these individuals. There is no variable-worker-effort function for skilled labour. For one thing, it is assumed that their wage is high enough for there to be no concern about their basic health and nourishment. Further, since these individuals have "good" jobs, there is no reason for them to consider shrinking; they enjoy their work too much. Thus, only the unskilled become unemployed.

The fifth, sixth and seventh equations are the first-order conditions that follow from profit maximization. Profits are defined as

\[ Y - (w_r s^r + w_n s^n) = Y - (w_r r^r + w_n r^n) \]

The first two equations define the factor supply: unskilled labour is stuck within the country (inelastic supply), and the other two factors are perfectly mobile internationally. Skilled labour can earn wage \( v^n \) and capital can earn rent \( r^n \), in the rest of the world. The last equation is the government budget constraint. Program spending and the employment-subsidy expenses (paid to firms for hiring unskilled labour) are financed by a tax on capital.

We do not expect readers to work out the formal results of this model. Its structure is spelled out just so that readers are aware of how to adapt the analysis to a developing economy setting. We simply assert the result that emerges: the subsidy to firms for hiring unskilled labour brings both good and bad news. The good news is that the unemployment rate is reduced. The bad news is that sufficient capital is pushed out of the country (due to the higher tax levied on capital to finance the employment initiative) for the average income of an unskilled individual, \( b \), to fall.

To keep exposition straightforward, we followed Monot's specification of the 0-mg feature in the production function (which is much simpler than the standard specification, as in Kremer (1993a)).

Given this departure from the literature, it is useful to provide some sensitivity test. To this end, we report a different, less thorough-going, method of decreasing substitution possibilities within the production process. We revert to just the one (unskilled) labour and capital specification, but we switch from Cobb-Douglas to CES production function with an elasticity of factor substitution equal to one-half (not unity as with the Cobb-Douglas). The production and factor-demand functions become

\[
\begin{align*}
\delta & = \delta N_t' + (1 - \alpha - \alpha) c_t \\
\frac{\partial Y}{\partial w} & = \frac{\partial Y}{\partial r} = (1 - \alpha) \frac{1}{1 + \alpha}
\end{align*}
\]

Again, we simply report the results, without expecting readers to verify them, since the derivations are quite messy. (Easier tests of the reader's ability to perform derivations are available in the practice questions.) The results for this specification of limited factor substitution are very similar to what has been reported for the 0-mg model. This fact increases our confidence that the disappointing conclusions reached in that case is likely to be relevant for actual developing economies.

### 9.5 Multiple Equilibria

The model of efficiency wages presented in Section 8.2 is a convenient vehicle for illustrating the possibility of multiple equilibria. Sometimes, as formerly in Canada, the generosity of the unemployment insurance system is increased (up to a maximum) for regions of the country that have experienced high unemployment in previous periods. We can model this policy by specifying that the unemployment-insurance generosity parameter, \( f \), be higher if the previous period's unemployment rate is higher. \( f = \alpha w_{max} \) if the previous period's unemployment rate is below some upper limit \( w_{max} \) (or 1), while \( f = 1 \) once that maximum upper limit is reached (7). Since the solution equation for the unemployment rate is \( d = \alpha (1 - f) \), in the simpler version of the model with no taxes, the unemployment rate follows a first-order nonlinear difference equation, as long as it is below the upper bound.

This relationship is shown as the heavy line in Figure 9.3. Since full equilibrium involves \( w_{max} \), there are three values for the natural unemployment rate — given by the points A, B and C. But only points A and C represent stable outcomes. To see this, suppose the economy starts at point A. Consider a "small" shock that makes the unemployment rate rise from 1 to 2. The time path beyond that first period is shown in the diagram by the steps back to point 1. Now consider a "large" shock that makes the unemployment rate rise from 1 to 3. The time path in this case is the set of steps from 3 to 4. Thus, when near A or C, the economy converges to these points; convergence never occurs to point B. As stressed by Milbourn, Parvis and Sack (1991), models of this sort have awkward implications for disinflation policy. The "temporary" recession involved in disinflation may be permanent if the recession is a "large" shock.

Through almost all of the first seven chapters of this book, we have assumed that the economy has just one long-run equilibrium (the natural rate of output). Often we have restricted our investigation of the role for government to questions of how alternative policy rules affect the speed with which the economy approaches that full equilibrium or how these rules affect the asymptotic variance of variables about their full-equilibrium values. Thus, our analysis has followed Tobin's (1975) suggestion that Classical models be accepted as descriptions of the long-run outcome and that Keynesian models be recognized as very helpful descriptions of the adjustment paths toward that full-equilibrium outcome. Now, however, we can see why Keynesians find multiple-equilibria models so exciting. They suggest that Keynesians should no longer concede the long run to the Classics. With more than one natural rate, there is an additional role for policy — to try to steer the economy to the "preferred" full equilibrium.

One advantage of the multiple-equilibria model that we have just discussed is its simplicity. This simplicity permits explicit derivations. But one disadvantage is that the existence of multiple equilibria depends on the presence of a particular government policy. The response of the New Classics is simply to suggest that the policy maker avoid such policies. So this dependence of the multiple equilibria on policy itself has to temper the enthusiasm about rethinking some relevance concerning the economy's full equilibrium on the part of Keynesians. But as we shall see in the following brief (non-technical) review of other multiple-equilibria models, not all owe their existence to particular government policies.
One way to understand this class of search models is by focusing on the very simplified narrative suggested by Diamond. Consider a group of individuals who live on an isolated island by eating coconuts. Religious custom prescribes eating the coconuts that each individual has collected herself before deciding how much to produce (that is, how many trees to climb and coconuts to collect). Each individual must form an expectation about the probability that she will find another individual with whom to trade. To have some other traders is very valuable to the worker/trader, but once there is a reasonable number, more traders bring in a very small additional benefit (since by then she is already almost certain to find a partner). Thus, with y denoting her own output, and x her expectation of each other trader's output, she will set her own activity level according to the $y = f(x)$ relationship like that shown in Figure 9.4 — which captures the increasing returns to the trading process in the AB range of the graph.

Let expectations be adaptive: $x = d(y - a) + b$ where $a$ is the actual behavior of others. We can evaluate $d(x)$ in the region of full equilibrium (when everyone is at the same activity level: $a = y$). Evaluating at $a = y = f(x)$, we have $d(x) = -2(1-fx)$ so equilibrium is stable only if $f < 1$. Equilibrium involves $x = y = x$ and these points occur along the 45-degree line in Figure 9.4. Points A and C are the two stable equilibria. Given that individual traders do not receive the full social benefit that follows from entering the otherwise "thin" market, there is an externality problem. This market failure allows us to rank the two stable equilibria: C is preferred. In principle, government involvement could switch the outcome to the Pareto-superior equilibrium.

Hoover (1966, page 636) summarized the outcome as follows — stressing that both the more-preferred and the less-preferred equilibria involve rational expectations: "if everyone believes that markets will be inactive they will anticipate a high cost of transacting; this will discourage them from undertaking transactions, and the initial beliefs will be self-fulfilling." What Diamond and Hoover have done is to provide modern standards of analytical rigor to defend the very old summary of Keynesian economics — that, without policy intervention, the economy could remain stuck indefinitely with a sub-optimal amount of unemployment.

Wolfram (1982), Blanchard and Kydsgaard (1987), and Rowe (1987) have also constructed models involving multiple equilibria. One feature in some of these models is that firms face limited demand curves. The reason for the kink is not the traditional oligopoly consideration based on rivals' reactions. Instead, the kink is based on Stiglitz's (1979) assumption of asymmetries in the dissemination of information to customers. Customers learn immediately about any price change at the firm with which they have been trading, but they learn only slowly of other price changes. Price increases are noticed by a firm's own customers (so an elastic response can occur), but price decreases are not. Consequently, elsewhere (so a price decrease response occurs). The resulting kink in the demand curve causes a discontinuity in firms' marginal revenue curves. If marginal cost cuts through this discontinuity, almost any such point can be an equilibrium. Individual firms face a free-seller problem when adjusting prices. They can understand that when there is a general reduction in all firms' demand, it would be desirable to have all firms lower price. This would stimulate aggregate demand and avoid a recession. But if only one firm lowers price, this general benefit is not forthcoming since one firm is too small to matter. Every firm wants to keep its own price high while having all others pass on to it the "macro" benefit of lowering these prices. This is the standard "unstable cartel" or "prisoners' dilemma" problem. Expansionary policy can internalize this externality problem.

It is not just that imperfect competition can lead to multiple equilibria; it is that the equilibria can be formally ranked. The high-activity equilibrium is preferred since it lessens the standard efficiency cost associated with monopoly. Another implication of imperfect competition is noteworthy: Mankiw (1990) has focused on the unusually high returns that characterize a natural monopolist. Increasing returns can make the labour demand curve positively sloped. As a result, it can intersect a positively sloped wage-setting locus more than once, and so there are multiple equilibria. Farmer (1993) has also considered increasing returns to scale — examining how the New Classical model is affected by this extension. Multiple stable equilibria emerge. "Strategic complementarity" is a game-theoretic term which has been used to interpret many of the multiple-equilibria models. As Cooper and John (1983) have noted, there is a general reason that coordination fails in many of these New Keynesian models. The general feature is that the larger is aggregate production, the larger is the incentive for each individual to produce. They show that this feature provides a general underpinning for Keynesian multiplier effects. Oh and Waldman (1994) explain that it is a basis for slow adjustment. And Aby (1993) proves that strategic complementarity can (along with real rigidities) account for the importance of any nominal rigidities that are present in the system.

The most general notion of multiple equilibria is found in models that involve hysteresis. The simplest model of this sort — provided by Blanchard and Summers (1986) — is based on the idea that the more experienced workers of a union (the "inexperienced") are the ones who make the decisions on wages. The workers are aware of the preferences of union members who are no longer seen — having become unemployed. The insiders' power stems from median voter considerations. The wage is set equal to the value that makes the firm want to hire just as many workers as were employed in the previous period. Thus, the expected employment in time $t$, denoted as $E(M_t)$, equals the last period's employment, $E(M_{t-1})$. An expression for expected employment can be had by specifying a labour demand function. Blanchard and Summers assume a simple aggregate demand function for goods: $Y = c(M_t - P_t)$, and constant returns to scale in production: thus, if units are chosen so that labour's marginal product is unity. $Y = N$, and $P_t = W + W/M$, (where $W$ stands for output, $M$ for money supply, $F$ for price, and $W$ for the wage rate). The implied labour demand function is $N_t = c(M_t - W_t)$, if the expected operator is taken through this relationship and the resulting equation is subtracted from the original, the equation becomes $E(N_t) = M_t - E(M_{t-1})$, since wages are set so that $W = E(W_t)$. Replacing expected employment by $N_{t-1}$, the time path for employment becomes $E(M_t) = E(M_{t-1})$, $E(M_{t-1}) = M_t - E(M_{t-2})$, ... $E(M_0) = M_0 - E(M_{-1})$.

This model is consistent with both the random-walk observation concerning output and employment rates (Campbell/Mankiw 1987) and the "money surprise" literature (Barro 1977). Unexpected changes in aggregate demand affect employment, and there is nothing to average the level of employment back to any particular equilibrium (because the preferences of laid-off workers no longer matter for wage-setting). Blanchard and Summers consider several variations of this and other models to test the robustness of the hysteresis prediction. Some of these extensions allow the "considered" to exert some pressure on wage-setting, with the effect that the prediction of pure hysteresis is replaced by one of extreme persistence.

Another source of multiple equilibria is "the average opinion problem" in rational expectations. The economy has many equilibria — each fully consistent with rational expectations — and each one corresponding to a possible view of what all agents expect all the others to take as the going market price (see Frydman and Phelps (1983)).

Ultimately, models such as these lead us to the proposition that the belief structure of private agents is part of the "fundamentals" — much like prices and technologies. That is, economists should study the evolving equilibria rather than search for some rationale to treat all but one as unsustainable. (Readers saw how common this practice is when learning phase-diagram methods in Chapter 6.) This plea for further study inevitably forces analysts to explore how agents gradually achieve rational expectations. For example, consider even a very limited aspect of learning — can agents grope their way to knowing the actual values of a model's structural coefficients if all start with is knowledge of the form of the model? Peatman (1982) surveys some of the studies which pose this class of questions. Some plausible adaptive learning schemes converge to unique rational expectations equilibria in some contexts, but not always. Despite the assumption that agents incur no decision-making
costs, Pesaran (1981) has stressed that agents can become trapped in a kind of vicious circle of ignorance. If agents expect further learning is not economically worthwhile, insufficient information will be accumulated to properly test that initial belief and therefore to realize that the original decision may have been mistaken. This implies that systematic forecast errors may not be eliminated with economically rational expectations.

Most studies of multiple equilibria do not question the entire concept of rational expectations; instead, they stress how the economy might shift between them — resulting in fluctuations in aggregate demand that are ongoing due to the self-sustaining cycle of revised expectations (as in Woodford (1991)). This class of models is quite different from both traditional macroeconomics and New Classical work, where cycles are caused by exogenous shocks to fundamentals (such as autonomous spending in Keynesian models or technology in the real business cycle framework). In the standard approach, it is almost always the case that it is optimal for agents to absorb these shocks (at least partly) by permitting a business cycle to exist. After all, stochastic shocks are a fact of life. As we have seen in earlier chapters, attempts by the government to lessen these cycles can reduce welfare. But if cycles result solely from self-sustaining expectations, then it is much easier to defend the proposition that the elimination of cycles is welfare improving. Indeed, government may not need to actually do anything to eliminate the cycles other than make a commitment to intervene to stabilize if that were ever necessary. Knowledge of that commitment may be sufficient to cause agents to expect (and therefore achieve) a non-cyclical equilibrium.

Howitt and McAfee (1992) build a similar model of endogenous self-sustaining cycles. It is based on the theory of search behaviour in the labour market covered in section 8.4 — one that involves a supposedly "non-fundamental" random variable called (in deference to Keynes) "animal spirits." A particularly interesting feature of the analysis is that the equilibrium involving ongoing cycles between the optimistic and pessimistic phases is stable in a learning sense. Bayesian updating induces convergence to this equilibrium with positive probability, even though agents start with no belief that animal spirits affect the probability of successful matches in the labour market search activity. Models such as this one provide as solid modern pedigree for even the most (apparently non-scientific) of Keynesian ideas — animal spirits.

There are many more models of multiple equilibria in the literature that focus on other topics. But enough has been covered for readers to appreciate how many public economics terms — externality, incomplete information, missing markets, non-convexity, moral hazard, nested power — appear in New Keynesian analysis. The universe is vast and the challenge posed by the New Classicals — have firmed macro foundations for macro policies — that can then be motivated on the basis of some well-identified market failure (second-best initial condition). This means that the principles that underlie normative analysis in microeconomics are becoming consistent with the principles that underlie macroeconomic policy analysis — an outcome much applauded by New Classicals.

6.6 Conclusions

The purpose of this chapter has been to use some of the macro based models of the natural unemployment rate that were developed in the previous chapter to assess several policies that have been used or advocated for reducing structural unemployment and/or raising the incomes of unskilled individuals. Here, we summarize a few of the key findings.

First, there is considerable analytical support for a policy of decreasing our reliance on income taxation and increasing that on expenditure taxes. This tax substitution can be expected to lower the natural unemployment rate. Second, involuntary unemployment creates a second-best environment in the labour market. In such a setting, it can be welfare-improving to impose a disturbing tax — even one levied on capital that is supplied perfectly elastically — if the revenue can be used to reduce the pre-existing distortion in the other factor market (the labour market). This environment makes low-income support possible — even for the government of a small open economy that faces the "globalization constraint." This second-best analysis was extended so that the appeal of competing anti-poverty policies — providing employment subsidies to firms or providing basic income to individuals — could be compared.

Finally, we explored how natural-unemployment rate analyses could be modified to consider some of the additional constraints that can limit policy makers in developing economies, and to consider the possibility of multiple equilibria. The possibility of multiple equilibria suggests an "announcement effect" rationale for policy. With both a high-

employment equilibrium and a low-employment equilibrium possible — and with both involving self-satisfying rational expectations — policy can induce agents to focus on the high-activity outcome if agents know that the policy maker really wants to push the system to that outcome if necessary. It is quite possible that no action — just the commitment to act — is all that may be necessary.

The natural unemployment rate is a long-run concept. There is another long-run aspect of real economies that we have ignored thus far in the book. This feature is the fact that there is ongoing growth — a long-run trend in the natural rate of output. We focus on this issue — productivity growth — in the remaining chapters of the book.

CHAPTER 10

TRADITIONAL GROWTH THEORY

10.1 Introduction

In our discussion of stabilization policy, we focused on short-run deviations of real GDP from its long-run sustainable value. We now shift our focus to the determinants of the trend in GDP, and away from a focus on the deviations from trend. We care about the trend, since we wish to explore what policy can do to foster rising long-run average living standards. To highlight this issue, we now abstract from short-run deviations altogether and consider a longer-term analysis in which it is reasonable to assume completely flexible wages and prices. In such a world, there is no difference between the actual and the natural rates. We focus on the long-run determinants of per capita consumption.

All western governments try to stimulate saving. Some of the initiatives are: taxing consumption instead of income (partially replacing the income tax revenue with the expenditure taxes), allowing lower rates on capital gains and dividend income, allowing contributions to registered retirement saving plans to be tax-sheltered, keeping inflation low, and deficit reduction. One of the purposes of this chapter is to review the traditional economic analysis that is viewed as supporting initiatives such as these.

Growth theory is often described as "old" or "new." The "old" analysis refers to work that involves two features: (1) descriptive behavioural functions for agents (that are not explicitly based on inter-temporal optimization), and (2) productivity growth specified as an exogenous process. As a result, there are two sets of literature that qualify as "new" growth theory. The first continues to specify productivity growth as exogenous, but since micro-based decision rules are involved, the analysis is immune to the Lucas critique. The second branch of new growth theory — which is new in both senses — involves micro-based behavioural functions and endogenous productivity growth. This "new-new" analysis is also called endogenous growth theory. We examine old growth theory and the first class of new models in this chapter. We move on to endogenous productivity growth analysis in Chapter 11.

10.2 The Solow Model

The standard model of exogenous growth is due to Solow (1956) and Swan (1956). It is defined by the following relationships:

\[ Y = F(N,K) \]
\[ S = Y - I \]
\[ I = R + G \]

The first equation is the production function: output is produced by combining labour and capital. We assume that the production function is constant returns to scale. One implication of this assumption is that a doubling of both inputs results in an exact doubling of output. This assumption is necessary if we want the full equilibrium of the system to be what is referred to as a balanced growth path. This refers to a situation in which all aggregates (the effective labour supply, capital, output, consumption, and investment) grow at the same rate — the sum of the population and productivity growth rates. This outcome implies that wages grow at the productivity growth rate and that the interest rate is constant. The constant-returns-to-scale assumption is an appealing one if the economy is already large enough for all the gains of specialization to be exhausted, and if factors that are fixed in supply (such as land and non-renewable raw materials) are of limited importance. Standard growth theory abstracts from these issues, we consider them briefly in section 10.5 below.
Since convergence to $k = 0$ is assured, we know that, in full equilibrium, output and capital must grow at the same percentage rate as does labor (measured in efficiency units). But since that growth rate, $g$, is an exogenous variable, it cannot be affected by policy (which in this compact structure must be interpreted as variations in the savings rate, $s$). A tax policy which permanently raises the propensity to save pivots the $g{(b)}$ curve up in Figure 10.1, making the equilibrium point move to the right. The economy settles on a higher capital-labor ratio. Assuming that this higher level of capital intensity raises per capita consumption, the response is shown in Figure 10.2. Cross-country analysis shows that high saving people save a bigger share of a given level of income. But through time, the higher saving means that workers have more capital to work with, and there is higher output, and this is what permits both higher saving and higher consumption. Figure 10.2 shows both the short-term gain and the long-term gain. It also shows that the growth rate rises — but only in a transitional way. The lasting effect of the higher savings policy is an increase in the level of the capital/labor ratio (and therefore in the level of per capita consumption), not in the growth rate. Nevertheless, this can still represent a very significant increase in material welfare.

But can we be sure that there is long-term gain? By considering Figure 10.1, this appears to be possible when the initial equilibrium (before the pro-savings initiative is instituted) is on the left side of the diagram (as we have assumed). But it does not appear to be possible if the initial equilibrium is to the right. This ambiguity casts the question: how can we determine the optimal value for the savings rate — at least as long as we restrict our attention to full-equilibrium considerations? At low capital-labor ratios, labor does not have enough capital to work with to achieve maximum efficiency. At high capital-labor ratios, diminishing returns with respect to capital limits to such an extent that most of the extra output is not available for consumption — it is needed just to maintain the capital stock, and to keep it growing at the same rate as labor. With a fixed depreciation rate, a large capital stock requires a large investment each period.

The key to determining the “best” savings rate is to realize that per capita consumption is pictured in Figure 10.1 as the vertical distance between the output per capita curve, $r(k)$, and the required (for a constant $k$) investment per capita line, $(n + s)$. That gap is maximized at the capital/labor ratio identified as $k^*$ where the tangent at point $C$ is parallel to the $(n + s)$ line. Thus, the rule which must be obeyed to maximize steady-state consumption per head — the so-called “golden rule” — is that the marginal product of capital equal the sum of the depreciation rate and the rates at which the population and the state of technical knowledge are growing.

Is this condition likely to be met in modern, developed economies? It is clear that all policy makers assume that it is not, and that we are at a point to the left of the golden rule (as in Figure 10.1). If this were not the case, the analysis cannot support the universal drive among policy makers to stimulate savings. Thus, the presumption appears to be that — in actual economies — the marginal product of capital exceeds $(n + s)$. As long as we believe that firms maximize profits, we have to believe that — in long-run equilibrium — they must be equating the marginal product of capital to its rental cost, $(n + s)$. Thus, the presumption must be that $p = s + s$. Abel et al. (1989) have tested this presumption by comparing net profits, $s$, and net investment, $K$, by consulting the national accounts for many countries. For every country and every year, they found that profits exceeded investment. This finding has been taken as strong support for the proposition that all these economies are under-capitalized. Thus, Figure 10.2 correctly shows the time path for per-capita consumption following a pro-savings initiative.

So higher savings involves short-term gain during the time interval between points 1 and 2 in Figure 10.2, since — without the policy — per-capita consumption would have been higher (following along the lower dashed line). But after point 2 in time, there is long-term gain. Old people are inert by the pro-savings policy, since they could be within the 1- to-2- time period. But the young, especially those who are not born until after point 2 in time, are made better off. From the point of view of the elderly, higher saving is a policy of following the golden rule — doing unto others (the young) what elderly would like others to have done for them.

While the Solow model is the analytical base for pro-savings policy initiatives, it is not without its critics. For one thing, it seems incapable of explaining the vast differences in living standards that we have observed, both across time and across countries at a point in time. Roughly speaking, the challenge is to explain the fact that citizens in developed economies have a living standard that is 10 times the level that was observed 100 years ago in these same countries. Similarly, the
developed countries enjoy living standards that are roughly 10 times what the poorer countries are making do with today. Differences of these magnitudes seem beyond the Solow model. To appreciate this fact, take the total differential of the steady-state version of the model's basic equation, replace \( x \) with \( y \) (by using Cobb's production function given above), and evaluate the result at steady-state values. That result is:

\[
\frac{dy}{y} = \left( a \frac{K}{L} - \alpha \right) \frac{dx}{x}.
\]

If we consider a plausible value for capital's share of output, \( \alpha = 0.33 \), this result implies that per-capita output is increased by a mere 5% when a substantial increase in the savings ratio (10%) is undertaken and sustained indefinitely. This quantitative outcome is not in the league of what we need to explain.

A second empirical issue concerns the speed of convergence to full equilibrium. The model's speed of adjustment is the absolute value of the coefficient of the stability condition (the absolute value of the Old A expression). When this is evaluated at full-equilibrium, we see that the speed of adjustment measure is \( n + \alpha \frac{K}{L} \). Taking one year as the period of analysis, plausible values are: \( n = 0.02 \), \( 8 < 0.04 \) and \( \alpha = 0.33 \), so the speed coefficient is 0.04. From the “rule of 72”, it takes 72/0.04 = 18 years for the model to get half way from an initial steady state to a new steady state. Most empirical works argue that adjustment speed in the real world is much faster than this, so there is concern about the applicability of the Solow model on this score as well.

Relating to this, a great deal of empirical work has been done to test the “convergence hypothesis” — an implication of the Solow production model when it is combined with several assumptions concerning similar modes of behaviour across countries. Consider two countries for which the values of \( n, \alpha, \beta \) and \( \sigma \) are the same. The Solow model implies that these two countries must eventually have the same levels of per-capita income, no matter what the size of their initial capital/labour ratios. Initially “poor” countries will grow faster than the initially “rich” countries, and they must converge (or “catch up”) to the same standard of living. Growth rates should correlate inversely with the initial level of per-capita income. Comparable data sets for some 138 countries (annual data since 1960) have been constructed recently, and this data appeared (initially at least) to reject this convergence hypothesis. This finding was one of the things which stimulated the new theories of endogenous growth — some of which do imply convergence.

Mankiw, Romer, and Weil (1992) have shown that the Solow model is not necessarily threatened by this lack of convergence evidence. After all, countries do have different savings rates and population growth rates, so they are approaching different steady states. After accounting for this fact, and for the fact that countries have invested in human capital to different degrees, Mankiw, Romer and Weil find stronger evidence of convergence. A number of studies have questioned the robustness of the Mankiw, Romer, and Weil conclusions. Thus, while a new approach to growth modeling may not be necessary to rationalize the cross-sectional evidence, many economists remain dissatisfied with the fact that the Solow model does not attempt to endogenize, and therefore explain, the steady-state growth rate. We explore the simplest versions of some of these models in Chapter 11. Before doing so, however, we investigate how basic exogenous growth theory has been modified to respect the Lucas critique.

10.5 Exogenous Growth with Micro Foundations

The following equations define a model that has well-defined underpinnings based on constrained maximization:

\[
\begin{align*}
\min (1 - \gamma) & = \gamma - rt + rK + rP - pK \quad \text{for } r > 0 \\
k & = rK - (\gamma) - e - \gamma c - g \\
g & = rK + \gamma v
\end{align*}
\]

All variables are defined on a per-effective-unit-of-labour basis. This means, for example, that in full equilibrium, \( c \) will be constant. Per-capita consumption will be growing at the productivity growth rate, but per-capita consumption measured in efficiency units, \( e \), will not be growing.

The first equation was explained in Chapter 4; it involves intertemporal utility maximization by indefinitely lived agents (individuals who face a constant probability of death, and a life expectancy of \( 1/(1-p) \)).

The second equation follows from profit-maximizing firms that incur no adjustment costs while investing capital: capital is hired so that its marginal product equals the rental cost. For simplicity, there is no labour-leisure choice. The aggregate labour force is fixed at unity, so the population growth rate, \( z \), is zero. Total labour income, \( w \), is determined residually (and the associated equation is not listed above).

The third equation is the goods market clearing condition (often referred to as the economy's resource constraint); net investment is output minus household spending, replacement investment expenditure, and government purchases.

Finally, the fourth equation is the government budget constraint. \( t \) and \( v \) are the tax rates levied on interest and wage income (respectively). Government spending is taken as exogenous and constant. The wage income tax rate is determined residually by this equation, to balance the budget — given the permanent reduction in the interest-income tax rate, \( r \).

It is noteworthy that all the model's parameters are “parametric” in the Lucas sense. Assuming the same Cobb-Douglas production function as used above, the model involves seven parameters: \( p \) is a taste parameter, \( \phi \), \( \alpha \) and \( \omega \) technology parameters, and \( k \), \( r \) and \( g \) policy parameters.

There are no parameters in the equations defining private-sector behaviour that could be mixtures of fundamental taste and technology coefficients and the parameters that define alternative policy regimes (such as “is” in the Solow model). Thus, more legitimate policy analysis is possible in the present setting.

Before proceeding with a formal analysis of the revenue-neutral tax substitution in this model, it is useful to consider some intuition concerning its full equilibrium. Ignoring ongoing productivity growth, the cost of forgoing consumption for a period is the rate of time preference, while the benefit of forgoing that consumption is the amount of output that an additional piece of capital can generate (the marginal product of capital). As a result, the stock of capital should be expanded to the point that its marginal product equals the agents' rate of time preference. Is this condition satisfied in a decentralized economy? Firms ensure that the (net of depreciation) marginal product of capital equals \( r \), while households ensure that the rate of time preference equals \( 1 - r \). The social optimum is reached by decentralized agents only if \( r = 0 \). Thus, the interest-income tax should be zero in the steady state. But should we not pay some attention to what occurs during the transition to that full equilibrium? To answer this question, we analyze a phase diagram.

Taking a total differential of the first three equations, eliminating the change in the interest rate by substitution, and evaluating coefficients at steady-state values, we have

\[
f = \frac{\partial k}{\partial t} + \left( 1 + 1 - \gamma \right) dr
\]

where

\[
0 = \frac{\partial k}{\partial r} \quad \frac{\partial r}{\partial r} = \frac{\partial c}{\partial r} = 1
\]

\[
K(1 + 1 - \gamma) p (p + p) \frac{1 + 1 - \gamma}{1 - \gamma} - \gamma - t
\]

c can jump at a point in time, while \( k \) is predetermined at each instant, so unique convergence to full equilibrium requires a saddle path (which obtains as long as we assume that the determinant of \( s \) is negative, and this is fully consistent with representative values of the model's parameters). The reader can use the entries in the (1) matrix to pursue the methods explained in section 6.2, and verify the particulars of the phase diagram shown in Figure 10.3.

Figure 10.3 Phase Diagram

![Phase Diagram](Figure 10.3)

Figure 10.4 Dynamic Adjustment Following Lower Interest-Income Tax Rates

A cut in interest taxation shifts the \( a = 0 \) locus to the right, the economy...
moves from point 1 to point 2 immediately, and then from point 2 to point 3 gradually. The formal analysis confirms that there is short-term pain (lower c initially) followed by long-term gain (higher c in the new full equilibrium), so the result is similar to that pictured in Figure 10.2. The only difference stems from the fact that c in the present analysis is per-effective worker consumption, not per-person consumption (what was the focus in Figure 10.2). As a result, there is no positive slope to the trend lines in the version of Figure 10.2 that would apply to variable c. Nevertheless, the formal analysis accomplishes two things: it confirms that there is short-term pain followed by long-term gain outcome, and it facilitates a calculation which determines whether the short-term pain is or is not dominated by the long-term gain. To answer this question, we calculate \( dPV/dk \), where \( PV \) is the present value function:

\[
P(V) = Se^{-c}ln(c)dc
\]

and is the social discount rate. This welfare function is based on the instantaneous household utility function that was involved in the derivation of the model above. It is a consumption-oriented measure that should be no controversy about this general form of social welfare function. But there is controversy concerning what discount rate to use.

One candidate is \( 2 = r(1-r/2 - n) \), the economy’s net of tax and growth rate of interest. This is what is used in standard applied benefit-cost analysis — based as it is on the hypothetical compensation principle. Another candidate is \( 2 = r - n \), such an individual’s rate of time preference. For internal consistency, this option must be used if agents live forever (that is, \( c = 0 \)). But in the overlapping generations setting that is our focus here, the \( 2 = r \) assumption is not so obviously appealing. It is not without any appeal in this context, however, since it has been shown that this discount rate is an integral part of the only time-consistent social welfare function to be identified in the literature as consistent with this overlapping generations structure (Calvo and Obstfeld, 1988). Given the uncertainty concerning what discount rate to use in public policy analysis, it is instructive to consider both these options. Two rather different conclusions emerge (and readers can verify this by following the procedure outlined in section 6.2).

If the time preference rate of any one generation is used as the social discount rate, we find that — even allowing for the short-term pain — agents are better off if the interest-income tax is eliminated. But if the net market interest rate is used, there is less support for this initiative. This is because, with the net interest rate exceeding the time preference rate in the overlapping generations setting, this decision rule involves discounting the long-term gain more heavily. It turns out that \( dPV = 0 \) in this case, so the pro-savings policy is neither supported nor rejected when the hypothetical compensation criterion is used. This result is consistent with Gravelle (1991) who argues that such a tax substitution has more to do with distribution than efficiency.

We return to this issue — the support or lack thereof for pro-savings initiatives — in Chapter 12. In that chapter we will add interesting features to the tax-substitution analysis. First, the equilibrium growth rate will be endogenous, so fiscal policies will affect growth permanently. Second, we will consider two groups of households — one very patient and the other much less so. With each group having its own rate of time preference — one above the economy’s net interest rate, and the other below it — we can be more explicit when we compare the short-term pain and the long-term gain of fiscal policies that raise national savings. In the meantime, we extend the present exogenous-growth model with just one class of households so that we can consider a policy of government budget deficit and debt reduction, for a small open economy.

10.4.2. Benefit-Cost Analysis of Debt Reduction

Thus far, we have used the optimization-based exogenous-growth model to examine a balanced-budget tax substitution — a revenue-neutral switch in the tax system that involves the government relying more heavily on wage taxation and less heavily on interest taxation. In recent decades, however, many governments have not been balancing their budgets, and the pro-savings initiative that has become the centre of attention has been deficit and debt reduction. The purpose of this section is to bring our analyses to bear on precisely this topic. To do so, we must add the deficit and debt variables (and another differential equation — the accumulation identity for the nation’s foreign debt) to the model. This additional source of dynamics creates a problem. We do not wish to attempt the drawing of three-dimensional phase diagrams. Nor do we want to solve the set of differential equations in algebraic form. As a result, we make no attempt to analyze the entire time path of the economy’s response to fiscal retrenchment. Instead, we confine our attention to a comparison of the initial, and the final, full equilibria. Since we wish to illustrate the importance of fiscal retrenchment as a quantitative term, we calibrate by selecting empirically relevant values for the model’s slope coefficients.

The extended model is simplified slightly by our continuing to abstract from population growth, and by our now dispensing with interest taxation. The system is defined by the following five equations:

\[
a = (r - p - n)c + pg - (b - d) - a
\]

\[
f'(k) = r = 8
\]

\[
7 = -n(1 - f) \frac{c}{e} - (r + p) \frac{a}{e}
\]

\[
b - d - nb
\]

\[
a = (r - p) - (b - d) = a
\]

As above, all variables are defined as ratios — with the denominator being the quantity of labour measured in efficiency units. Units of output are chosen so that, initially, these ratios can be interpreted as ratios to GDP as well. The new variable is \( k \), the nation’s foreign debt ratio.

The first equation is the private-sector consumption function — very similar to what has been discussed above. There is one difference here; there is an additional component to non-human wealth. In addition to the domestically owned part of the physical capital stock (\( k - a \)), there is the stock of bonds issued by the government to domestic residents (\( b \)).

The second equation states that firms maximize profits and hire capital until the marginal product equals the rental cost. For this application, we focus on a small open-economy setting. As a result, the interest rate is determined from outside (it is paid down by the foreign interest rate and the assumption of perfect capital mobility). Since \( k \) is the only endogenous variable in the second equation, this optimal-lagrange multiplier approach pegs the capital stock. Thus, in analyzing domestic policy initiatives, we take \( k \) as a constant (and set \( k = 0 \)).

The third equation combines the GDP identity with the accumulation identity for foreign debt (both written in ratio form). This version of the accumulation identity states that the foreign debt-to-GDP ratio rises whenever net exports falls short of the pre-existing interest payment obligations to citizens in the rest of the world. The interest payment term reflects the fact that — even when net exports are zero — the foreign debt ratio rises if the growth in the numerator (the interest rate paid on that debt this period) exceeds the growth in the denominator (the GDP growth rate — \( n \)). Net exports are defined by the expression in square brackets in the third equation.

The final two equations define the government accounting identities (and they were discussed in Chapter 7 (section 4)). As noted, to avoid advanced mathematics, we ignore the details involved in the dynamic approach to full equilibrium, and focus on the long run. Once full equilibrium is reached, all aggregates are growing at the same rate as overall GDP (at rate \( n \)), so ongoing changes in the ratios (all the dotted terms in these five equations) are zero. The five equations then determine the equilibrium values of \( c, k, a, b \) and one policy variable — which we take to be \( t \). We take the total differential of the system, we set all exogenous variable changes except that in \( d \) to zero, and we eliminate the changes in \( a \), \( b \) and \( b \) by substitution. The result is:

\[
dc = (r - p) + pg + pb \frac{d}{e} + \frac{dc}{d} + n(1 - f) \frac{d}{e} - (r + p) \frac{a}{e}
\]

We evaluate this expression by substituting in the following representative values for the parameters: \( r = 0.05, p = 0.25, n = 0.02, \) and \( p = 0.02 \). These values ensure a rising consumption-age profile for each individual generation. Also, they involve a life expectancy (lives) once an individual reaches her initial working age of 50 years.
The illustrative calculation requires that an assumption be made concerning the amount that the full-equilibrium deficit-to-GDP ratio is reduced. We take Canada as an example of a small open economy. Since Canada's deficit-to-GDP ratio peaked in the mid-1990s, the government announced a large reduction in budget deficits, which reduced the deficit-to-GDP ratio to below 1% of GDP in 2020. This reduction in the deficit-to-GDP ratio allows us to focus on the impact of fiscal policy on income distribution and economic growth.

Finally, the illustrative calculation requires an assumption concerning the initial ratio of private consumption to GDP. We assume 0.55, since this value emerges from two sets of considerations. First, it is consistent with Canadian data. Second, it is consistent with the full-equilibrium restrictions involved in this model and other parameter value assumptions that we need to make — those noted above and the following. We have assumed that the aggregate production function is Cobb-Douglas with capital's share parameter equal to 0.3; the equilibrium capital-to-GDP ratio is 3.0; capital depreciates at rate 0.05 per year; and the size of government (as a proportion of GDP) is representative of the Canadian federal government in the early 1990s.

With all these assumptions, the percentage change in consumption turns out to be 3 percent. This means that, according to the model, the fiscal retrenchment that is about half completed as this book goes to press can be expected to raise Canadians' standard of living by 3 percent. In 2006 dollars, that amounts to an increase in material living standards of more than $4000 per year for each family of four, and this is an anomaly — an amount that would be received every year into the indefinite future. This analysis assumes that the country's risk premium is not reduced through deficit reduction — an assumption that is consistent with evidence (Filion, 1996). If deficit reduction did lead to a lower domestic interest rate, the long-term benefit of debt reduction would be even larger.

As far as it goes, this analysis supports the policy of many western governments in recent years — that of government budget deficit and debt reduction. Of course, it must be remembered that all that has been established here is that the long-term benefits are significant. Two other considerations are material parts of a complete benefit-cost analysis of debt reduction. First, there is the short-term pain that must be incurred to achieve this long-term gain. In the short-run (while some prices are sticky) there will be a temporary recession. This may not be too large, however: since a flexible exchange rate can largely insulate real output from aggregate demand shocks (see Mundell, 1956; Fleming, 1962), and one of the practice questions suggested for this chapter). Even if there were no temporary recession involved, consumption would have to fall in the short run to “finance” the higher national saving (as discussed in section 10.3). So there is a short-term pain involved in debt reduction. The second broad issue concerns the distribution of the long-term gain. We shall now see that — unless the government uses the fiscal dividend to improve labour's position — all of the gain from debt reduction can be expected to go to the owners of capital. This prediction may seem so alarming to some policymakers that it would be enough to cause them to reject debt reduction on equity grounds; they believe that only the rich have enough income to do much saving, and so they are the only ones who can benefit from these tax policies. Those who favour pro-saving-tax initiatives argue that this presumption is incorrect; indeed they argue that most of the benefits go to those with lower incomes. But since the process by which these benefits “trickle down” the income scale is indirect, they argue, many people do not understand it and thus reject these initiatives inappropriately. We now examine whether this trickle-down view is correct, first in a closed economy, and then in a small open economy.

Figure 10.5 shows the market for capital; the demand curve's negative slope shows diminishing returns (the fixed quantity of labour being shared over more workers), and the supply curve's positive slope captures the fact that savings is higher with higher returns. Equilibrium occurs at point $E$. With just two factors of production, capital and labour, the area under the marginal product of capital curve represents total output from the entire area under the demand curve up to point $E$ in Figure 10.5. Further, since each unit of capital is receiving a rate of return equal to the height of point $E$, capital's share of national income is the rectangle below the horizontal line going through point $E$. Labour gets the residual amount — the triangle above this line. A tax policy designed to stimulate savings shifts the capital supply curve to the right, as shown in Figure 10.5. Equilibrium moves from point $E$ to point $E'$, and total output increases by the additional area under the marginal product curve (that is, by an amount equal to the shaded trapezoid in Figure 10.5). So the pro-saving-tax initiatives do raise per-capita output. But how is this additional output distributed? The owners of capital get the dark shaded rectangle, and labour gets the light shaded triangle. So even if capitalists do all the saving and are the apparent beneficiaries of the tax policy, and even if workers do no saving, labour does get something.

Furthermore, labour's benefit is not just the small shaded triangle in Figure 10.5. Being more plentiful, capital's rate of return has been bid down to a lower level. Since that lower rate is being paid on all units of capital, there has been a transfer from capital to labour of the rectangle formed by the horizontal lines running through points $E$ and $E'$. So, after considering general-equilibrium effects, we see that capital owners may not gain at all, labour, on the other hand, must gain. We conclude that, in a closed economy (that can determine its own interest rate), the benefits of pro-saving-tax initiatives do trickle down.

---

**Figure 10.6: The Size and Distribution of Income: An Open Economy**

**Figure 10.7: The Size and Distribution of Income: An Open Economy**

---

**Figure 10.5: Trickle-Down Economics in a Closed Economy**

Pro-saving initiatives often take the form of the government using the tax system to stimulate private saving, instead of boosting public-sector saving via deficit reduction. Some people oppose these tax initiatives on equity grounds; they believe that only the rich have enough income to do much saving, and so they are the only ones who can benefit from these tax policies. Those who favour pro-saving-tax concessions
If \( R \) denotes the remaining stock of energy and \( E \) is what is used up each period, the differential equation that defines the resources that are lost is \( R = -E \). Micro models of optimal resource depletion often lead to the proposition that the amount used each period should be a constant proportion of the remaining stock. We assume that this policy is followed, so we assume \( 1 \leq R = 0 \). We can think of \( O \) as a policy variable. In the basic Solow model, the policy variable was a constant investment rate. In this model, the policy variable is a constant dis-investment rate. The last two relationships imply a linear differential equation \( X = R = 0 \), and the solution of this equation is \( X = OR e^{-t} \). This solution gives us an equation for \( E \) that can be substituted into the production function: \( E = OR e^{-t} \).

After substituting in, and proceeding with the same steps that were involved in the analysis of land, we get:

\[
Y = B^* - K^{*} (1 + \delta) (1 - \delta)^{1/\delta} \delta (1 - \sigma)^{1/\delta} \left( \delta^{\sigma/\delta} - 1 \right) .
\]

end with

**Growth in per capita output:** \( g = r - n \). [1 + z]

This result indicates that the drag on growth is bigger when the natural resource is shrinking. Nordhaus (1992) has estimated these parameters, and he concludes that the annual growth-reducing effect (the second term on the right-hand side) is about one-third smaller than the percentage point. The implication is that, as long as the productivity growth rate can be expected to exceed this amount, we can enjoy rising living standards — even though we are running out of non-renewable resources. A second reassuring point can be made. If resources really are running out as we have assumed in this analysis, the real price of these resources should be increasing. In many cases they are not. In other words, the rate of discovery of new supplies appears to be dominating our rate of using the resources. These observations suggest that the model may have focused on a case that is more stringent than what we have yet had to confront. On the other hand, the Cobb-Douglas function implies an elasticity of factor substitution which may very much overstate how easy it is for firms to make do with less of the dwindling resources, and there has been no mention of pollution externalities. Thus, it is worrisome that concern about non-renewable resources is left out of the standard analyses of fiscal policy and growth. Despite this concern, space limitations make it necessary for us to follow this convention in the remaining two chapters of the book.

It is worth noting that this discussion of non-renewable resources has invited our switching from exogenous-growth to endogenous-growth analysis, since in this case, the full-equilibrium growth rate is affected by a policy parameter, \( O \). We pursue a more systematic exploration of endogenous growth in the next chapter.

### 10. 6 Conclusions

The purpose of this chapter has been to position readers so that they can benefit from our exploration of "low" growth theory in the remainder of the book. Tradtional growth analysis is "old" in that (originally) it lacked formal optimization as a basis for its key behavioral relationship — the savings function, and (even today) it involves a rate of technological progress that is exogenous. In the analysis that has been covered in this chapter, we have seen that the first dimension of oldness has been removed through the addition of micro-foundations. We consider ways of endogenizing the rate of technical progress in the next chapter.

The basic policy prescription that follows from both the original and the micro-based version of traditional analysis is that pro-savings policies can be supported. These initiatives result in a temporary increase in the growth rate of consumption, and a permanent increase in the level of per capita consumption. Calibrated versions of these models support the conclusion that, even without a permanent growth-rate effect, higher saving leads to quite substantial increases in average living standards. If higher saving leads to capital accumulation (as it does in a closed economy), even a poor labourer who does not save benefits from a pro-savings fiscal policy. Such individuals benefit indirectly, since they have more capital with which to work. But if the higher saving leads only to increased demand for resources already of the same quantity of capital, the poor labourer is not made better off. So, in some circumstances, the increase in average living standards does not involve higher incomes for everyone. Our task in the next chapter is to establish whether this basic conclusion — that a pro-savings policy is supported in models that do not stress income distribution issues, but it may not be otherwise — holds in an endogenous productivity growth setting.
CHAPTER 11
NEW GROWTH THEORY

11.1 Introduction

The Solow growth model (and all the extensions we considered in Chapter 10, except the last one involving non-renewable resources) has the property that savagely/government policies cannot permanently affect the economic growth rate. This property of the model stems from the fact that the man-made item that raises labor's productivity — physical capital accumulation — involves diminishing returns. Given our standard assumption of diminishing returns, as more capital is accumulated, its marginal product must fall. This makes it ever less tempting for households to acquire more capital. As this reaction sets in over time, the temporary rise in productivity growth that initially accompanies a rise in saving must wane away. The only way we can have a model that avoids this production is by changing our assumption about the production process. We need to build a model in which there is a man-made factor of production that does not involve diminishing returns. The obvious candidate is “knowledge”. There seems to be no compelling reason why we need assume that the more knowledge we have, the less valuable more knowledge becomes. Thus, to have a theory of endogenous productivity growth, economists have built models involving constant returns to scale in the production process. We consider three of these models in this chapter.

11.2 A One-Sector Endogenous Growth Model

The most straightforward way of eliminating decreasing returns is to assume a production function that is a straight line ray emanating from the input-output origin: \( Y = AK \). For this reason, this approach is known as the "AK Model". The problem with assuming such a production function at the level of the individual firm is that — with constant returns — the size of the individual firm is indeterminate, so formal micro-foundations cannot be provided. Given our desire to keep the analysis free from the Lucas critique, this is unacceptable. This problem is usually solved by assuming that there is an externality. Each individual firm is assumed to have a "normal" production function:

\[ Y = AK \]

where the \( i \) subscript denotes individual-firm items and \( B \) is the worker-effectiveness index. Worker effectiveness is assumed to be proportional to the "state of knowledge" in the economy, and — to avoid the complications of specifying a separate "knowledge-producing" sector — it is assumed that the amount of knowledge possessed by the workers in each individual firm is simply proportional to the overall level of economic development (as measured by the aggregate physical capital stock that has been accumulated):

\[ B = AK \]

We can derive an aggregate version of the first equation by substituting in the second equation and defining \( Y = q^L \), and similar equations for the two inputs (where \( q \) is the number of firms). Ignoring population growth (assuming \( N \) is constant), the result is \( Y = AK \), where \( A = \frac{q}{N} \).

At the aggregate level, the full model consists of the following relationships:

\[ Y = AK \]

\[ \ln N = (l - a)Y \]

\[ \ln C = a(l - a) \]

\[ K = \ln C - G - SK \]

\[ G = \ln K + n \ln N \]

The second and third equations follow from profit maximization; firms hire each factor up to the point that the marginal product equals the rental cost. Combining the first and third equations, we have: \( aA = (l + 5) \).

Since three variables in this relationship are technology coefficients, this equation fixes the pre-tax rate of interest. This simplifies the interpretation of the remaining three relationships.

The fourth equation is the standard micro-based consumption function (where there is an interest-income tax, \( i \)). For simplicity we have assumed infinitely lived family dynasties (for whom the probability of death is zero). The fifth equation is the GDP identity — indicating that the capital stock grows whenever more output is produced, then is consumed. The final equation defines a balanced government budget. Spending is financed by the revenue that is collected from the interest-income tax (rate \( i \)) and the wage-income tax (rate \( t \)).

When a balanced growth path is reached, capital, output, and consumption all grow at the same rate, \( n \). This growth rate in living standards is also the labour productivity growth rate; otherwise, "effective" labour would not be growing at the same rate as all other aggregates. We rewrite the first three equations to focus on the equilibrium growth rate. In doing so, we divide the fifth equation through by \( k \), define \( x \) and \( g \) as \( C/K \) and \( G/t \) respectively, and we use the second and third equations to simplify. The results are:

\[ n = r(l - a) - p \]

\[ G = \ln C - (G/T) + n(l - a) \]

\[ g = r(\ln C) + n(l - a) \]

We assume that the government sets \( g \) and \( r \) exogenously, so the model determines the wage tax. The system is recursive. The first equation determines the growth rate. Then, given \( n \), the second equation determines \( x \). Finally, the third equation determines \( g \).

Since one of the policy variables enters the growth-rate determining equation, policy can have permanent effects on the growth rate in this model. To illustrate, we can take \( d n / dr \) to raise the growth rate. The size of this response is particularly noteworthy. With a 5% interest rate and a 10 percentage point cut in the tax rate (\( r = 0.05, dr = 0.1 \)), we conclude that the annual productivity growth rate can be increased by one half a percentage point.

This is a very big effect. This interpretation can be defended if we calculate the present value of the benefit of living in a society that involves a growth rate that is one half of one percentage point higher. The first step is to estimate how long it takes for the economy to reach its new higher-growth-rate steady state. It turns out that it takes absolutely no time at all to reach this outcome. For simplicity, we abstract from the government as we consider the transitional dynamics. In this simplified case, we can use the time derivative of the definition of \( x = C/K, dx = C/K - K \).

The model can be summarized by just two equations:

\[ C = ceb + p \]

\[ F = ceb + p \]

The total differential of this system is:

\[ d\theta = \theta \cdot dx \]

where \( \theta / \theta \) and \( \theta \) are the steady-state values of the growth rate and the consumption-capital ratio. Since both these entities are positive, both these first-order differential equations involve an unstable dynamic process. If we drew a phase diagram in the space, all arrows showing the tendency for motion would point to the right instability. There is not even a saddle path to jump on to. But since both variables are "pump" variables, the economy is capable of jumping immediately to the one and only stable point — the full equilibrium. In keeping with the convention that we limit our attention to feasible stable outcomes, we assume that the economy jumps instantly to its new steady state — the moment the exogenous shock occurs to make the steady-state growth rate higher. In short, there are no transitional dynamics in this growth model. We exploit this fact in our growth policy analysis in the final chapter. We rely on this fact here as well, as we illustrate the significance of a slightly higher growth rate. Assume that today's GDP is unity. If \( n \) and \( r \) denote the GDP growth rate and the discount rate, the present value (PV) of all future GDP is

\[ PV = f = e^{n} \cdot dt = \frac{1}{n} \cdot (n - r) \]

For illustrative parameter values \( r = 0.05 \) and \( n = 0.03 \), \( PV = 50 \). How much is that \( PV \) increased if the growth rate is increased by one half of one percentage point? When \( r = 0.05 \) and \( a = 0.03 \) are substituted in, the answer is \( PV = 66.5 \). Thus, the once-for-all equivalent of living in a world with a one-half-point higher growth rate is a gift equal to 1.67 times the starting year's GDP.
To have some feel as to whether this is “big” or “small” we compare this estimate to what was a hotly debated issue in Canada almost 20 years ago — the free trade arrangement with the United States. In that policy application, the benefits for Canada were estimated to be a per-period level effect of 3% higher living standards every year into perpetuity. The one-time lump-sum equivalent of this annuity is (0.05) (1 — n) / 0.03 = 1.5. Thus, in the past, Canadians have treated a policy that delivers this much as a big issue. The pro-savings tax cut that we are examining here involves a benefit that is more than 11 times as big. When contemplating results such as this, Robert Lucas (1988) argued that it is hard to get excited about anything else in life.

113 Two-Sector Endogenous Growth Models

Before immediately sharing Lucas’s excitement, we should check how sensitive this “big response” result is to the particular features of the AK model. To pursue such a sensitivity test, we explore the two-sector literature. In this approach, a separate knowledge-producing sector of the economy is specified. This knowledge is called human capital, Kt, and it is distinct from physical capital, C.

It is assumed that both forms of capital are important in the manufacturing sector — that part of the economy that produces consumption goods and the physical capital. A standard production function is assumed in this sector:

\[ Y = AK_t (B_t) \]  

The total stock of human capital in the economy is \( H \). Proportion \( b \) is employed in the manufacturing sector. The output \( Y \) can take the form of either consumption goods or physical capital accumulation. In other words, the usual accumulation identity for physical capital applies:

\[ Y = C_t - G_t - S \]  

Following Lucas (1988), it is assumed that physical capital is not important in the education sector. Only human capital is needed to produce knowledge, and the simplest form of a constant-returns (production function and accumulation identity) relationship is assumed.

\[ H_t = (1 - b_t) H_{t+1} \]  

(11.1)

\[ B_t = [1 - (1 - b_t)^{-1}] \]  

(11.2)

\( B \) is the gross rate of return on each unit of human capital employed in the education sector. The total amount that is employed there is proportion \( (1 - b) \) of the total. The net rate of return is \( (B - S) \), since, for simplicity, we assume that physical and human capital depreciate at the same rate.

It is not obvious how we should interpret the education sector. On the one hand, it can be thought of as a “home study” where individuals simply refrain from paid employment, and use the time to learn more. On this interpretation, these individuals are receiving no income (when away from the manufacturing sector) so there is no taxable wage income generated in the knowledge sector. This interpretation is appealing when thinking of university students. The other way of thinking about the education sector is that it is a “research institute” that employs researchers.

On this interpretation, everyone in society is automatically aware of all existing knowledge. This sector’s job is to produce new knowledge. This interpretation of the education sector is appealing when thinking about professors, since professors receive wages which the government can tax. We proceed in two stages. First, we explore the “home study” interpretation; then, later in this section, we shift to the “research institute” interpretation. Unfortunately, the policy implications of the two models are very different.

Households are assumed to optimize. One outcome is a standard consumption function (listed below). Another implication is that individuals arrange their portfolio of assets so that — in equilibrium and at the margin — they are indifferent between these three options: holding physical capital employed in the goods sector, holding human capital employed in the goods sector, and holding human capital employed in the education sector. Since there is no tax on human capital activity, the after-tax yield and the before-tax yield on human capital in that sector is the same. We denote that return by

\[ r_t = B_t - g_t \]  

(11.3)

It is clear from this relationship that, since there is no taxation in the home study sector, \( r_t \) is totally driven by two technology parameters.

The net of depreciation (but before tax) rate of return on physical capital in the goods sector is

\[ r_t = (\alpha Y) \]  

(11.4)

For portfolio equilibrium, the after-tax yield, \( r_t = 1 - t \), must equal what is available on human capital placed in the other sector.

\[ r_t = (1 - t) \]  

The model consists of the four numbered equations plus the consumption function (equation (11.5) below). Here we generalize (compared to the simple AK model) by re-introducing population growth and overlapping generations (a positive probability of death). In this case, the growth rate for aggregate consumption \( C \) is the sum of the growth rate for per capita consumption \( c \) plus the population growth rate \( C_t = c_t + y_t \). From Chapter 4 (section 4.3), we know that the consumption function is

\[ r = \left( r_t - 0.5 + 2z \right) / (1 - \tau) \]  

(11.5)

Before deriving the full-model properties, we simplify equations (11.1) and (11.2). We start by dividing (11.1) by \( K \) and (11.2) by \( N \). Next, we substitute in the full-equilibrium condition that the growth rate for both \( K \) and \( H \) is \( y = y \). Finally, we define \( x = C / K \), and we assume a common tax rate on all factor earnings in the goods sector \( t = z \). With government spending being determined endogenously, theGoods identity can be simplified further. The compact versions of equations (11.1) and (11.2) become:

\[ y = x + b + (1 - b) (1 + \tau) \]  

(11.6a)

\[ y = \frac{r}{(1 - \tau) (1 - t)} (1 - t) \]  

(11.6b)

The five equations determine \( r \), \( t \), \( b \), \( x \) and the productivity growth \( v \). In simplifying these five relationships, we have substituted in the assumption of balanced steady-state growth — that the growth rates of \( C \), \( K \) and \( N \) are all equal. Thus, we are limiting our attention to full equilibrium, and for a simplified exposition, we are ignoring the transitional dynamics (that are not degenerate in this case, as they were for the AK model).

Readers interested in an analysis of the between steady states dynamics should read Barro and Sala-i-Martin (1995, p 182).

It is worth taking stock of the model’s structure. The taste parameter \( (\alpha) \), the demographic parameters \( (p \) and \( z \), the technology parameters \( (b \) and \( S \), and the exogenous policy variable \( t \) are all given from outside. We can use the model to determine the effect on the steady-state growth rate of a shift to a smaller government. If there is an exogenous cut in the tax rate (and a corresponding drop in the level of government program spending), we would expect — on intuitive grounds — that there would be a higher growth rate. First, the lower tax rate can be expected to encourage growth since interest taxation is a disincetive to accumulate capital. Further, with lower \( G_t = n_t + \Delta K_t \), the GDP identity can be simplified further. After all, in this model, since \( G_t \) enters neither the utility function nor the production function, society loses whenever it has any government. We do not intend to argue that this is the most satisfactory way of modelling government activity. Instead, we are just pointing out that it is in this setting that the move-to-smaller-government initiative would be expected to be at its most powerful. It is alarming, therefore, that we find such tiny growth rate effects (below).

Of course, to derive quantitative results, the model has to be calibrated. For illustration, we use the following set of representative parameter values:

\[ \begin{align*}
    r & = \text{net of depreciation return on capital} \\
    t & = \text{tax rate} \\
    z & = \text{population growth rate} \\
    p & = \text{rate of time preference} \\
    \text{physical capital output share} & = \text{manufacturing} 0.33
\end{align*} \]

All these parameter values are standard and appealing on empirical applicability grounds. Further, when they are substituted into the model’s equations, they imply reasonable values for the division of market output. The (\( C_t/N_t \)) and (\( Y_t/C_t \)) ratios are: \( 0.6 \) and \( 0.23 \). We now interpret the education sector as involving market activity, with the income being taxable. The two production functions are exactly as before, but the net-of-tax return on human capital in the education sector is now:

\[ \text{net-of-tax return on human capital in the education sector} = \]
where \( I \) is the tax on wage income. (Since we are now highlighting the taxability of income in this sector, we revert to our earlier practice (followed in section 11.2 involving the AK model) of allowing for different rates of tax on the income derived from physical and human capital.)

Profit maximization in the goods sector requires:

\[ aY + \frac{1}{K}r_n - \gamma \quad \text{and} \quad (1-a)Y bH - r_n - g \]

Portfolio equilibrium requires equal yields:

\[ n(1-a)Y = aY + \frac{1}{K}r_n - \gamma \quad \text{and} \quad n(1-a)Y = aY \]

Eliminating the pre-tax rates of return from these relationships by substitution, we have

\[ (\gamma + \frac{1}{K}) [1 + \delta(t - (1-a))] [1 - \delta(t - (1-a))] - \gamma \]

These relationships (along with \( G = n + (1-a)X + \gamma(T + H + \alpha G) \), the government budget constraint) are needed to proceed through the substitutions that lead to the same compact expression of the model that was our focus in the "home study" version of the two-sector model. That compact set of equations is:

\[ r^* = (\beta - \gamma c) - \frac{\delta(y - b)}{\gamma} \]

\[ z^* = B(\beta - \gamma c) - \frac{\delta(y - b)}{\gamma} \]

\[ y^* = \frac{(1+\gamma)(1+\gamma)(1+\gamma)}{(1+\gamma)(1+\gamma)(1+\gamma)} \]

The new variable is \( c \), the sales tax rate. The five equations determine \( r^*, b, y, \gamma \) and one of the three tax rates. The taste parameter \( p \), the demographic parameters \( w, \frac{\gamma}{\gamma} \), and the technology parameters \( \beta, \frac{\gamma}{\gamma} \) and \( H \), and the exogenous policy variables \( g \) and two of the three tax rates are all given from outside. To illustrate, we consider using the model to determine the effect of the steady-state growth rate of a cut in the tax on wage income (financed by an increase in the tax on the earnings of physical capital). From the first equation, we see that the cut in \( c \) can have the same large impact on \( r^* \) that we got in the AK model. Then, focusing on the first term on the right-hand side of the fifth equation, we see that the productivity growth rate moves one-for-one with \( r^* \) (again, as in the AK setting). The presence of the other term on the right-hand side of the fifth equation means that things are a little more complicated. Nevertheless, with a long life expectancy (that is, with a small value for the death probability) this other term is not very important quantitatively. Hence, we are back to the Lucasian model of rent-seeking. With this version of the two-sector model, we expect big and permanent growth-rate effects from fiscal policy.

But there is a critical difference. In this model it is the wage-income tax that needs to be cut to stimulate growth. This correlation between these two key variables is negative. Saint-Paul's (2006) recent study is less optimistic on this score. He focuses on the relationship between wages and productivity. For a long time, wages for less-skilled individuals have increased with advances in productivity, while more recently, this correlation appears to become negative. Saint-Paul's model is designed to offer an explanation for this pattern, and it highlights the fact that people shift their demand away from the products that use low-skilled labour intensively as they become more wealthy. As a result, growth continues, there is an ever-decreasing demand for unskilled labour — in relative terms — and this growth stems from a change in the pattern of demand as the growth process continues, not from skill-based technical change. But the implication is the same — disappointment for those wanting to pursue both equity and growth.

There is an extensive multiple-equilibria literature that highlights the possibility of escaping an equity-efficiency trade-off (see Aghion et al. (1990) and Dac and Gatare (2005)). Some studies, such as Gómez and Zeno (1993) stress the possibility of "poverty traps." With nonlinearities stemming from such phenomena as increasing returns, there can be two equilibria: both a low-growth and a high-growth outcome. Quite a few development economists have become convinced of the applicability of these models, and so they argue that limited programs of foreign aid should not be wasted. According to this view, the North must offer the South enough aid to move the trapped economies all the way to the neighborhood of the higher-growth equilibrium. Some recent empirical work (Gräfen and Temple (2006)) suggests that about one-quarter of the world's economies are stuck in a low-growth equilibrium.

We turn now from policy implications to specification issues. There is controversy concerning the most appealing way of modelling the growth process. Some models (such as Romer (1990)) focus on R&D
spending by firms. Romer stresses that research output is different from human capital since it becomes common knowledge that is not lost to the production process if labour is unemployed. Despite this difference, Romer’s model has an important similarity with the human-capital model, since Romer assumes that each year’s increase in knowledge is proportional to the pre-existing level of knowledge. It turns out that the number of useful discoveries depends on the size of the population. The factor of proportionality linking the new output of knowledge to the pre-existing level of knowledge is a function of the number of people. More people mean more research centres, so there is a bigger chance of successful research emerging. This result is known as the “scale effect” — the bigger the level of the population, the higher is the productivity growth rate.

Jones (2003) is critical of Romer’s R&D model for two reasons. Despite Kerner’s (1993b) influential study of growth over the last one million years, Jones argues that the empirical evidence does not support the scale-effect prediction. Second, he argues that it is quite arbitrary to locate the necessity assumption that is necessary for non-explosive endogenous growth theory. Wouldn’t all data on individual establishments do not seem to support the assumption of absolutely constant marginal product relationships? Jones argues that it is more plausible to use a linear relationship in the population growth part of the model. Defining $b$ and $d$ as the births and death rates, and $L$ as the size of the population, we can specify $L = b - dL$, which implies $L/L = z$, and this is not arbitrary. Indeed, this relationship is true by definition. As a result, Jones argues that this equation forms a more plausible, less arbitrary, basis for endogenous growth theory. Nevertheless, there has not been a convergence of views on this controversy. For one thing, Solow (2003) has noted that there is considerable evidence in favour of continual convergence among OECD countries. These countries would have to be converging in demographic patterns for Jones’ model to explain this outcome.

Given the central nature of the issues that are being explored in endogenous growth theory, and the fact that substantial differences remain concerning the relative appeal of several modelling strategies, it is understandable that many economists are involved in this investigation — even as (though we have seen above) the quantitative relevance of policy-induced changes in the growth rate in some models can be questioned.

For the growth policy analysis in the next chapter, we rely on a human capital model that involves several features. First, we follow Barro and Sala-i-Martin (1993, pp 144-146), and collapse the two sectors (the physical and human capital production sectors) into one. This yields a system which has the simplicity of the AK model, but which retains the richer interpretation. Second, we follow (and extend slightly) Mankiw and Weil (2000) by introducing imperfect competition. This extension makes it possible for us to consider creative-destructive effects in a rudimentary fashion — while avoiding the complexity of the research-and-development class of endogenous growth models. (We do not endogenize the degree of monopoly power.) For the next chapter, we extend this framework, to allow for a subset of households to be short-sighted. Two things follow from this extension. First, these individuals never save, so they remain at a lower level of living standards. With both “rich” and “poor” in the model, we can investigate the effect of redistributive policies on the overall growth rate. The second thing is that it makes these individuals prone to shirking on the job. Employees react by paying efficiency wages, and unemployment results. Within this setting, we can investigate the effect of job-creation policies on growth.

As just suggested, compared to the literature surveyed in the last section of this chapter, the model we now outline involves a middle-of-the-road assumption concerning factor attunements. The standard AK model specifies that knowledge is proportional to the aggregate stock of physical capital, while the Lucas two-sector framework makes human capital the engine of growth. The former specification leads to the policy that interest-income taxes should not exist, while the second leads to the proposition that wage-income taxes should not exist. The specification suggested by Barro and Sala-i-Martin is an appealing intermediate specification. It involves the assumption that physical capital and human capital are used in the same proportions when producing all three items: consumer goods, new physical capital goods, and new knowledge (human capital).

To outline this framework concisely, we suppress the possibility of monopolistic competition initially. National output, $Y$, is either consumed privately. $C$, consumed in the form of a government-provided good, $G$, or used to accumulate physical or human capital, $K + H$.

$$Y = C + G + K + H.$$
The final sales of the second-stage producers constitutes the GDP. Measuring this aggregate from the expenditure side of the national accounts, we know that it equals $C + \text{I} + \text{G} + \text{F}. \text{Measuring GDP from the income side, it is the sum of primary factor incomes, } \text{wH + rK, plus the profits of the second-stage producers. it, Given the factor pricing equations, and the assumption that the production function involves constant returns to scale, the national income measure is } \text{F(K, H) = } \infty. \text{ Profits are defined as the excess of final sales } \text{F(K, H)} \text{ over what must be paid to the primary producers. F(K, H)/m = } \infty. \text{ e = F(K, H)\{m - 1\} m.}$

Substituting this definition of profits into the national income definition, we end with $C + \text{I} + \text{G} + \text{F} = \text{F(K, H). We conclude that no} \text{ specification of the economy's resource constraint is appropriate with this simple specification of imperfect competition.}$

To complete the model, we allow for the possibility that the economy-wide productivity parameter, $\theta$, is a function of the share of GDP that goes to profits. The idea is that higher profits make more innovation possible, so we specify

$$\theta = \theta(\theta(\theta))$$

where $\theta > 0$. We do not formally model the imperfect competition. Instead, we simply interpret $\theta$ as a parameter that can be affected by competition policy. Further, each individual firm takes $\theta$ as given. Thus, we consider a slightly revised production function:

$$Y = (1 - \theta)(\theta + 1) m F(K, H)$$

This is a short-cut or ‘black box’ specification of the essence of the R&D models of endogenous growth. No optimizing basis for investment in research is offered here; instead there is the straightforward assumption that an exogenous increase in the availability of profits (relative to primary factor earnings) results in a higher level of technological ability.

Removing the government for a more simplified exposition, we have the revised compact listing of the model:

$$a(A + B) = A - c$$

$$B = c - aA$$

$$A = \bar{A} \text{ for } m < 1$$

$$r = \alpha I m$$

$$n = r - p$$

This model determines $A, B, r, s$ and $c$, and we are interested in the effect of the degree of monopoly on the growth rate:

$$do \text{ dm} \equiv (a A(m - 1))(n - 1) = (m - 1)$$

If profits do not act as an incentive for innovation, as assumed by Manzini and Wenzner (2006) and Judd (2002), and which we can impose by setting $B = 0$, then a more competitive economy involves less pulling scarce resources away from accumulating capital, and therefore higher growth ($\text{do dm} > 0$). But if profits do increase innovation (which is derived as a possibility in quite elegant models such as Aghion and Howitt (1992), rather than just assumed here if $\theta = 0$), then a more competitive economy involves lower ongoing growth. In this case, the pro-growth effect of pulling resources back into the production of more capital is dominated by the anti-growth effect of these being less technical change. Most applied competition-policy discussions highlight this tension, and it is salient to the creative-destruction tension in the imperfect competition models that offer an explicit structure for modelling the decision to invest in research. In the next chapter, we use this framework to explore how the existence of imperfect competition affects the implications of endogenous growth theory for tax policy.

11.5 An Evaluation of Endogenous Growth Analysis

We have considered several endogenous growth models in this chapter. Since the results are so different, we need some sort of common-sense base for relating some of this work to the real world. Thus, in this section of the chapter, we provide a “back of the envelope” estimate of the payoff from increased investment in education — an estimate that is not based on any specific or formal macro model.

By comparing peoples’ incomes — people with and without further education — many labour economists have estimated the return to education. A typical result from these cross-section regression equations (involving earnings as a function of years of schooling) is that the annual return to education is in the 7.5% range. We use this return to estimate the highest standards of living that we might enjoy if everyone were more educated. This is a controversial thing to do, since it may be that much of the estimated return in the cross-section regressions just reflects a signal — that smarter people can make it through school and less clever people cannot. So, even if school does nothing but visibly separate people into these two groups, employers will find it useful to use this signal of native intelligence to save decision-making costs. The education process itself may not raise productivity. If this is the case, and the high private return reflects just signal value, then we should not apply it in an economy-wide thought experiment. After all, a signal has no discriminating power if everyone has the credential. Despite this controversy, we assume that education is not just a signal. This means that the following calculation is based on finding too big a pay-off from more investment in education. This bias only supports the conclusion that we reach, however, since the estimated pay-off is very small — despite the upward bias.

Consider transferring 1% of GDP out of current consumption, and into education. This reallocation is like buying an equity that pays a dividend of 7.5% of 1% of GDP forever. The present value of the stream of dividends that accompanies this year’s equity purchase is therefore

$$v = \frac{0.075(0.01)(1^n + 1^n + 1^n + \ldots)}{1 + \theta}$$

where $\theta = (1 + n)(1 + n)$, and today’s GDP is unity. This present value expression can be simplified to

$$v = \frac{0.075(0.01)}{1 + \theta}$$

The cost of this year’s equity purchase is the lost consumption, which is 0.01. The net benefit, $NB$, is therefore:

$$NB = 0.001(0.075)(1^n - 1) - 1.$$
Finally, Jones (1992) has presented some interesting projections. He used the pre-1940 growth experience of the United States as a basis for making a simple projection of how much per capita incomes would grow over the next 45 years. He implicated what someone might have predicted back then — without taking any account of later developments. He expected these projections to under-estimate actual growth, since they ignore the vast increase in human capital that has been accumulated during the later period. For example, the share of scientists and engineers in the labour force has increased by 300%. In addition, now 85% of individuals, not 25% finish high school. But despite all this, the simple projections over-estimated the actual growth in living standards. This exercise has created a challenge for new growth theory to explain.

What advice can be given to policy makers — given all this controversy? It would seem that — if a policy involves definite short-term pain and uncertain long-term growth-rate gains — it may be prudent to postpone implementing that policy until the insight from further research can be had. This cautious strategy does not mean that nothing can be recommended in the meantime. After all, a number of policies can be present-value justified if they deliver a (much less uncertain) one-time level-effect on living standards. This may not be as "exciting" as a permanent growth-rate effect, but it is still very worthwhile. For example, the increase in living standards that we estimated to follow from government debt-redaction policy (see section 10.4) is a very worthwhile outcome. This analysis assures us that there is a large scope for relevant policy analysis — even if it is too soon to base advice on an area of inquiry — endogenous growth theory — that is still in progress.

11.6 Conclusions

We are leaving our introduction to growth theory to a somewhat untidy state. For one thing, we have had to acknowledge that there are empirical issues that both endogenous-growth and endogenous-growth models have trouble explaining. For another, our coverage of the two-sector endogenous growth models has been simplified by our focusing entirely on steady-state outcomes; that is, by our ignoring what happens along the transition path between steady states. Finally, we have not attempted a detailed exposition of the much more elaborate research-and-development based endogenous-growth models. We make no attempt to rectify this latter shortcoming in the book's final chapter, for two reasons. First, there are other more advanced books (such as Aghion and Howitt (1998)) that serve this need. Second, we need the space to provide an extensive application of the more basic growth analysis to a set of policy questions: Does a pro-growth tax structure hurt those on low incomes? Does policies that reduce unemployment reduce growth? Does an aging population lead to lower growth? It is to these questions that we turn in our final chapter.

CHAPTER 12

GROWTH POLICY

12.1 Introduction

The purpose of this chapter is to assess the analytical support for several policy advocacy brought by policy advisors, using a simple version of endogenous growth theory. We proceed through the following steps. First, we consider the optimal tax question by comparing income to expenditure taxes. Without any market distortions allowed for in analysis, we show that expenditure taxes are recommended. This basic analysis has been very influential. For example, the President's Advisor Panel on Federal Tax Reform in the United States (issued in 2002) argue for a wholesale replacement of the income tax with an expenditure tax. Our analysis, indicates how the analytical underpinnings for this purpose are sensitive to the existence of other sources of market failure. For example, in section 2, we consider a second-best setting in which the government is "too big." In this situation, it is appropriate for government to levy distortions, so the income tax should be eliminated. Many policy analysts argue both that government is "too big" and that income taxes are "bad." This section of the chapter challenges these analysts to identify the analytical underpinnings of their views.

related finding follows from an extension that allows for two groups of households, one so impatient that these individuals do not accumulate physical capital. Progressive taxation is analyzed by taxing only "rich" households to finance the "poor." In the United States, the "rich" pay at a tax rate of 28%, the "poor" at 15%. This analysis does not support replacing a progressive income tax with a progressive expenditure tax.

Other second-best considerations are considered in later sections of the chapter. In section 3, we focus on consumption externalities. This consideration pushes the conclusion for tax policy in the opposite direction; it strengthens the case for the endogenous growth rate. Some negative aspects of consumption (such as Frank (2003)) argue that consumption externalities are very important from an empirical point of view. How else, asks Frank, can we explain the fact that survey measures of subjective happiness shows no increases over a half century — when this period has involved significant increases in the standard measures of economic growth (such as per capita GDP)?

Unemployment is added in section 4 of the chapter. In this second-best situation, an employment subsidy is supported as a mechanism for simultaneously lowering unemployment and raising the growth rate. This results a challenge for policy analysts who argue that we face a trade-off in the pursuit of low-income support policy (our equity objectives) and higher-growth policy (our efficiency objective). It is interesting that basic endogenous growth theory can provide examples such as this one — that illustrate the relevance of what Alan Blinder has called "reconciling" the two. In this class of models, it is relatively easy to find settings in which a fiscal policy that is designed to help those lower down on the "economic ladder" has indirect benefits for those up the ladder. This is similar in spirit, but opposite in direction, to the newly widely known "trickle down" economy. In this context of models it seems relatively easy to find settings in which a fiscal policy designed to help the rich generates indirect benefits for those further down the economic ladder.

Finally, in section 5 of the chapter, we outline what basic growth models imply about how future living standards may be affected by a major demographic event that is much discussed. We consider the increase in the old-age dependency ratio that will accompany the aging of the population that will occur as the post-war baby-boom generation moves on to retirement.

12.2 Tax Reform: Income Taxes vs. the Progressive Expenditure Tax

Many economists and tax-reform panels have called for a shift in tax policy: a decreased reliance on income taxation and an increased reliance on expenditure taxes. The standard analytical underpinnings for this reform proposal is endogenous growth theory. While less emphasis is given to equity considerations in the theory (since standard growth theory involves a single representative agent), policy analysts sometimes call for a progressive expenditure tax — to avoid the regressivity that would otherwise accompany the use of sales taxes. The purpose of this section of the chapter is to use a simple version of endogenous growth theory to review this debate.

We begin with the proposition that total output produced each period, Y, is used in two ways. First, it is purchased by households to be consumed that period. C. Second, it is purchased by households to add to their stock of capital, K. K refers to that period's increase in capital. The supply equals demand statement is

\[ Y = AK \]  

(12.2)

A = r, the rate of return on capital, is the amount that output increases as additional units of the input are hired (the average and marginal product of capital). C is the pre-tax consumption for households after subtracting taxes. K is the pre-tax capital available to the government. There are two potentially controversial aspects of this specification of the nation's input-output relation. First, there is no diminishing returns (as more capital is employed, its marginal product does not fall). As explained in chapter 11, the economy's equilibrium growth rate is independent of fiscal policy if this assumption is not made. As we have shown, that standard practice in presenting the standard policy-oriented analysis here. Second, once the output emerges from this production process in the form of consumer goods, it is costly for society to convert that new output into additional capital. Again, this assumption is made to keep the analysis consistent with standard practice. Also, for simplicity, we ignore depreciation of capital.

In this initial specification, the government has just one function: it levies a proportional income tax rate, r, on the income that households earn by employing their factor capital, and a proportional sales tax rate, s, on household consumption spending. The government uses this revenue to finance transfer payments, R, to households. The government's balanced-budget constraint is
Given that there is short-term pain to achieve long-term gain, it is not immediately clear that this tax substitution is supported. But we can address this issue by using the household utility function. It is shown later in this section that when the household utility integral is worked out, overall social welfare (SW) is

\[
SW = c_{0}e^{-(\frac{1}{2}p)}I \quad (12.5)
\]

where \( C_0 = c_{0}e^{p} \) and \( K_0 = k_{0} \) is the initial capital stock, at the time when the tax substitution takes effect. From (12.8), we can determine the effect on overall welfare of the tax substitution.

\[
dSW \quad dr = (\frac{1}{2}\sigma \frac{dI}{d} \quad (12.6)
\]

Using the one-time level and ongoing-growth-rate effects reported above, we have

\[
dSW \quad dr = r(\sigma - 1) (\frac{c_{1}p}{2})
\]

which can be simplified by using (12.7) and (12.2).

\[
dSW \quad dr = -r(\sigma - 1) (\frac{c_{1}p}{2}) < 0.
\]

The fact that this expression is negative indicates that the government can raise people’s material welfare by cutting the income tax rate — all the way to zero. This is the standard proof that we should rely on expenditure, not income, taxes to finance the transfer payments. Because the generosity of the transfer does not affect the household’s consumption-saving choice, there is no such thing as an ‘optimal’ value for transfers. Whatever level is arbitrarily chosen has to be financed by expenditure taxes if the government wishes to maximize the material welfare of its citizens.

We now consider a sensitivity test, by asking how the optimal-tax conclusion is affected by our replacing the government transfer payment with a program whereby the government buys a fraction of the GDP and distributes it free to users (as in the case of government-provided health care). We continue to assume that no resources are needed to convert newly produced consumer goods into new capital or (now) into the government service. Thus, the economy’s production function remains (12.2), and the supply-elastic-demand relationship becomes

\[
Y = c + K + G
\]

where \( G \) is the level of the government-provided good each period. The government budget constraint becomes

\[
G = rK + C
\]

Finally, we assume that households value the government-provided good, so there are two terms in each period’s utility function:

\[
utility = \int_{0}^{x} [ln C + \gamma ln T]
\]

Parameter \( y \) indicates the relative value that households attach to the government service. Since the government imposes the level of government spending, individual households still have only one choice to make: they must choose their accumulation of capital with a view to maximizing the present discounted value of private consumption. The solution to this problem is still equation (12.2). The model is now defined by equations (12.1a), (12.2), (12.3a) and (12.4). Defining \( g = G/IT \) as the ratio of the government’s spending to GDP, these relationships can be re-expressed in compact form:

\[
\frac{dx}{dt} = r + \gamma n
\]

\[
g = r(1 - \gamma) - r
\]

\[
n = r(1 - \gamma) - p
\]

and the modified overall material welfare function is

\[
SW = (C_0 + G_0 + \gamma c_{1}p)/(1 - y) p
\]

It remains true that \( dI/dt = dN/dt \), but before we make use of these outcomes, we focus on the question of the optimal level of government program spending. The criterion used to set government spending optimally is to arrange the outcome so that the following condition holds for households:

\[
\Delta G/\Delta T = \text{income consumption good} \quad \Rightarrow \quad (12.8a)
\]

Muhammad Firman (University of Indonesia - Accounting)
With the price of both goods being unity, and the utility function that has been assumed, this condition requires \( I/C = x/g \), that is, the optimal program spending-to-GDP ratio is \( g^* = y/c \). This definition is used to simplify the overall welfare effect of varying the income tax rate. The result is:

\[
dSW/\,dr = \left[ r^2 + \left(\alpha g^*\right)^2\right]\left[\left(g^* - g\right)^2\right]
\]

As before, material welfare is maximized when this expression is zero, and this requires no income tax (\( I = 0 \)) only if government is set optimally \( g^* = g \). If, however, government spending is "too big" \( g > g^* \), there should be an income tax.

What is the intention behind the result that income taxes should be used to finance some of the government when government is "too big"? The answer hinges on the fact that this resource misallocation in terms of a fixed proportion of a growing GDP. With growth, the magnitude of the distortion is being magnified. In this case, then, growth has a "bad" dimension. This is why it becomes "good" to rely, to some extent, on the anti-growth instrument for raising tax revenue (the income tax rate). Barro and Sala-i-Martin (1995, pg 144-161) and Marrero and Novados (2003) consider similar models. Their government good is infrastructure, and it enters the production function instead of the utility function. Nevertheless, the same dependence of the optimal tax question on second-best considerations — whether the government has set its expenditure "properly" — is stressed. A related result is also available in Garcia-Peñaloza and Turnovsky (2005) in a developing economy setting. The fact that the government cannot levy taxes on workers in the traditional sector creates a second-best situation, with the result that capital taxation is appropriate.

Returning to our specific analysis, we conclude that we cannot know whether the actual economy involves the "preferred" rate of income to sales tax rates or not, unless we know whether the government is "too big" or not, and by how much. Many pro-growth policy analysts defend two propositions: (i) that income taxes are too high relative to sales taxes, and (ii) that the government sector is too big. Our analysis has shown that the more correct these analysts are concerning the second proposition, the less standard growth theory supports their first proposition. Our purpose of this section of the chapter is to identify this tension that policy analysts do not appear to be aware of.

Thus far, there has been just one input in the production process, and we have interpreted that input quite broadly — as including both physical and human capital. For most of the rest of this chapter, we wish to contrast patient households who plan inter-temporally and acquire physical capital (as discussed above) with households that always spend their entire labor income and transfer payments. To make this companion explicit, we must distinguish physical capital \( (K) \) from human capital \( (H) \). As outlined in section 11.4, the supply-equals-demand equation becomes

\[
Y = C + G + K + H
\]

since output is either consumed privately, consumed in the form of a government-provided good, or used to accumulate physical and human capital.

Both forms of capital, the government service, and consumer goods, are produced via the standard Cobb-Douglas production function:

\[
Y = A \left( K^{\alpha} H^{\beta + 1} \right)
\]

By defining \( B = H/K \) and \( d = 2B \), the production function can be re-expressed as \( Y = A K^{\beta} \) — the same form as that used above. But in this setting, tax policy affects \( A \).

As explained in section 11.4, it is assumed that households real out their physical and human capital to firms (that are owned by other households). Profit maximization on the part of firms results in factors being hauled to the point that marginal products, just equal rental prices \( r \) is the rental price of physical capital, \( w \) for human capital:

\[
\alpha A = \left(1 - (\beta + 1)\right)H = w.
\]

Households maximize the same utility function. In this case, the constraint involved is: \( I + (\beta + 1)K + wI = rI \). In addition to the familiar consumption-growth rule, \( n = n(1 - r) - p \), the optimization leads to \( r = w \). In equilibrium, household must be indifferent between holding their wealth in each of the two forms of capital, and the equal-wealth relationship imposes this equilibrium condition. Finally, the government budget constraint is \( G = zK = wH + zC \), and the compact version of the full model is:

\[
A = 2B a
\]

\[
y = n(1 - n) - p
\]

\[
B = (1 - w)H
\]

\[
G = r + c l \, A
\]

The endogenous variables are \( A, B, r, n, c \) and one of the tax rates. It is left for the reader to verify that all the conclusions that we noted earlier in this section continue to hold in this slightly extended setting. The reason that the distinction between physical and human capital does not yield different conclusions is because the economy uses the two forms of capital in exactly the same proportions when producing all four items: consumer goods, the government good, new physical capital goods, and new knowledge (human capital).

Perhaps the only result that readers might need help with concerns the effect of the tax substitution on social welfare. The social welfare function is the utility function of the representative household.

\[
SW = \left[ C + yln G(\alpha / n) \right]
\]

We know that \( C = C e^{\alpha} \) and that \( G = G e^{\alpha} \), so the social welfare expression can be simplified to

\[
SW = \left[ C e^{\alpha} + yln (\alpha / n) + (1 + y)h e^{\alpha / n} \right]
\]

The integral in this last equation can be re-expressed as \( n \) times \( lie / t \), and this expression, in turn, can be simplified by using the standard formula \( f(t) = \int f(t) \) to yield

\[
(-1/p) = e^{-a} \left(1 + p / q \right)
\]

When this solution is evaluated at the extreme points in time, and the result is substituted into the expression for \( SW \), equation (12.8a) above is the result. When differentiating that expression with respect to the tax rate, we use \( c = C / K \), which implies that \( dC/dC/c \) \( dr = dC/dC/c \) and since \( K \) cannot jump (\( K_0 \) is independent of policy). Also, since \( G = gK_0 \), we know that \( dG/dr = 0 \) as well. The \( g^* = ye I + r \) result emerges from maximizing \( SW \) subject to the resource constraint, that is, by evaluating \( SW/1dK = 0 \).

We now move on to consider income-distribution issues. Mankiw (2000) has suggested that, for the sake of realism, all fiscal policy analyses should allow for roughly half of the population operating as infinitely-lived family dynasts and the other half operating hand-to-mouth. Thus far in this tax-policy analysis, we have not followed this advice. Further, since all households have been identical, progressive taxes could not be considered. We now extend our analysis so that we can follow Mankiw's suggestion — both to increase the empirical applicability of the analysis, and to make it possible to investigate progressive taxes.

Since the poor consume a higher proportion of their income than do the rich, a shift from a proportional income tax to a proportional expenditure tax makes the tax system progressive. For this reason, policy analysts are drawn to the progressive expenditure tax. It is hoped that this tax can avoid consuming an equity problem, as we take steps to eliminate what is perceived to be an efficiency problem. For the remainder of this section of the chapter, we use our otherwise standard growth model to examine shifting between income and expenditure taxes — when both taxes are strongly progressive. To impose progressivity in a stink fashion we assume that only "Group 1" households — the inter-temporal consumption smoothers who are "rich" — pay any taxes. "Group 2" households — the "poor" individuals who live hand-to-mouth — pay no taxes. In this section, we revert to our original specification concerning the expenditure side of the government budget. Once again we assume that there is no government-provided good; the taxes collected from the rich are used to make a transfer payment to the poor.

Group 2 households have a utility function just like the Group 1 utility function, except Group 2 people are more impatient. Their rate of time preference is \( \beta \), which is sufficiently larger than the other group's rate of impatience, \( \beta_2 \), that Group 2 people never save voluntarily, so they never acquire any physical capital. It is assumed that they simply have to do some saving, in the form of acquiring the human capital that is
The analysis in this section has involved numerous simplifying assumptions. With a view to providing a little balance, we consider one such assumption in the remainder of this section. By doing so, we reconcile Judd’s (2002) point that the presence of imperfect competition strengthens the case for an increased reliance on expenditure taxation. Here we add taxes to the monopolistic competition model that was explained in section 11.4. We assume that there is no government provided good, that the government returns all revenue to private agents as a lump-sum transfer, and that all forms of income (excluding monopoly profits) are taxed at the same rate. As a result, the government budget constraint is $R = z + c$. The compact listing of the model is quite similar to that discussed in earlier parts of this section, except that the monopoly mark-up parameter, $m$, appears here in the fourth equation.

$$n(1 + B) = A - c$$
$$B = (t - a - d) / A$$
$$A = f(x)$$
$$r = ca/m$$
$$n(1 - r - b - p)$$
$$z = rA$$

This model determines $A$, $B$, $n$, $r$, $a$, and $c$, and we are interested in the effect of a change in the income tax rate, $t$, on $n$, $a$, and social welfare. Thus we derive $d/n$ and $d/c$ and $dr/dt$ and substitute these results into

$$dsW / dr = (t + p)/[t + q]$$
$$dsW / dr = (t + p)/[t + q]$$

After simplifying, the result is

$$dsW / dr = (t + p)/[t + q]$$

The best value for the income tax rate is that which makes this expression zero. Since $m$ exceeds unity, that best value for $r$ is a negative number. Not only should positive income tax be eliminated; the ownership of capital should be subsidized. This outcome follows from another second-best situation. With imperfect competition, there is the opportunity for profit income. This possibility means a reduced incentive to earn income by acquiring and employing capital (compared to what exists in a competitive economy). If there is no other source of market failure, it is optimal for the government to remove this incentive — by subsidizing the non-profit source of income (that is, by subsidizing capital accumulation).

In the real economy, there are likely several market failures — for example, imperfect competition and an over-expanded government (at least as viewed from the perspective of the rich). As we have seen in this section of the chapter, the former distortion calls for a negative $r$ value, while the latter distortion calls for a positive $r$ value. Without further empirical analysis, it seems difficult to defend the proposition that the best value for the income tax rate is zero.

12.3 Economic Growth and Subjective Happiness

In this section, we address Frank’s (2009) concern that standard growth theory ignores the findings of the subjective happiness literature. Despite the dramatic increase in standards such as per-capita GDP and per-capita consumption over the last 50 years, citizens in all developed economies are telling us in surveys that they do not feel any happier. Frank’s explanation for these results is that much of our higher consumption expenditure has been on “positional goods.” If everyone acquires many of these goods, any one person’s relative position is not improved by economic growth. We offer a simple formalization of this argument in this section. As with imperfect competition, increased support for a shift toward expenditure taxation emerges.

Thus (except when we considered a government-provided good), we have assumed that only one thing affects utility — the household’s own level of consumption. We now extend this specification by allowing utility to depend on the amount of leisure that the household chooses. Further, we allow one household’s utility to depend inversely on the level of consumption of other households. This last feature is necessary if we are to explore Frank’s (2005) suggestion that positional goods be taken more seriously in otherwise standard analysis. To simplify the exposition, and to highlight these extensions, we introduce them into the most basic model that was considered in the first few paragraphs of section 12.2. In particular, there is perfect competition, only one composite form of capital, and no hand-to-mouth individuals or government-provided goods.

For the present discussion, it is best to think of capital as human capital. Letting $x$ denote the fraction of each household’s time that is spent at work, $x/H$ is employed capital, and the remainder, $t = 1 - x/H$, is what is devoted to leisure. The resource constraint, the production function, and the government budget constraint are
Masterbook of Business and Industry (MBI)

Muhammad Firman (University of Indonesia - Accounting)
condition that follows from household optimization concerning capital accumulation is that both physical and human capital must generate the same rate of return per unit, so households are indifferent between holding their wealth in each of the two forms of capital. This condition is $r = \frac{(1 + v)^{-1}}{\lambda}$.

These forward-looking households make two separate decisions. As a group, each family makes the capital-accumulation decision by following the consumption-growth relationship that was just discussed. Following Alexopoulos (2003), we can think of this decision being executed by the family matron, who takes the labour market outcomes of the various family members as exogenous. Then she chooses the optimal capital-accumulation plan, and allocates the corresponding amount of consumption each period to each family member. Each family member is free to augment that level of consumption by adjusting her labour market involvement. The workers at each firm are assumed to rely on a group representative to negotiate wages with their employer. The negotiator pursues a wage that exceeds the workers' outside option, but only to a limited degree since the negotiator values a high level of employment as well. As explained in section 8.3, unemployment emerges in this setting. Specifically, we have $w = \left(\frac{1}{(1 + v)}\right)\frac{1}{(1 + q)}$, where (as already noted) $v$ is the employment subsidy and $v$ is the exponent on employment in the labour negotiator's Cobb-Douglas objective function. $(1 - v)$ is the weight on wages. Of particular interest here is that the unemployment rate varies inversely with the level of the employment subsidy. Clearly, the growth model is not needed to arrive at this conclusion. What the model does is facilitate an examination of how an employment subsidy affects the growth rate of living standards (the productivity growth rate).

Some macroeconomists might find the separate-decisions formalism for specifying the family unappealing. After all, for some time now, the goal when providing macrofoundations for macroeconomics has been to specify one overall optimization that simultaneously yields all behavioral equations for the agents. But, in the interest of tractability, it is now common to separate certain decisions. As already noted, in her model of efficiency wages and endogenous growth, Alexopoulos (2003) adopts precisely this same separation of the asset-accumulation and labour-market-investment household decisions. Simultaneously, in the New Neoclassical Synthesis approach to stabilization policy analysis, modelers specify two separate firms: one to hire the factors and to sell "intermediate" products, and the other to buy the intermediates and to sell final products to households. These "final goods" sellers have no costs at all, other than the menu costs that are incurred when changing prices. The two-stage process involving two separate firms is adopted solely to separate the optimization for goods price adjustment from the optimization for determining factor demands. Thus, when it can be regarded as constricted, this practice of separating certain decisions appears to be accepted as a necessary way of proceeding in even the most central areas of study in modern macroeconomics. As noted, we follow this practice here.

The second group of households is important. They have such a high rate of time preference that they never save — beyond the investment in human capital that is necessary to have a job. As a result, this group simply consumes all their income — which is half the after-tax labour income in general — each period. And, if you add in the government transfers that each family receives on acquiring human capital, this group is much larger than those represented in the previous paragraph. Thus, since this group constitutes half the population, they represent half the unemploy. They are relatively poor since, by never acquiring any physical capital, they receive no "interest" income. The spending function for Group-2 households is given by $E = R + (1 + v)w + \frac{v}{\lambda}M$, where $R$ is transfer payment receipt.

Finally, the notion of government's balanced-budget constraint. It is obtained by setting $R = L - K$. In the model with government, $\gamma$ is the ratio which states that the income tax rate pays for general transfers and the employment subsidy. Balanced growth is assumed, so $C = K + I = \frac{K}{\lambda}$, and $E = \frac{K}{\lambda}$. The equations determine the responses of $n$, $c$, $r$, $v$, $w$, $w$, $d$, and $\gamma$, when the employment subsidy is introduced ($\theta$ is increased). $c$ and $d$ are defined as $C_{K}$ and $E_{K}$, and it is assumed that the government fixes the transfers-to-GDP rate, $R/I$. For the remainder of this section, we discuss four properties of this system — that $n$, $c$, and all rise, and that $u - \alpha = 0.5 \gamma$ becomes positive. It is left for the reader to use the equations to verify the results, desired.

Introducing the employment subsidy has a direct effect in the labour market — lower unemployment. As a result, physical capital has more labour to work with, and this raises physical capital's marginal product, and so raises the interest rate. Thus, there is an increased incentive to save. Raising this variable with a higher income tax rate shrinks, but does not eliminate, this increased incentive to save. The value of the formal model is that it allows us to see that between these competing effects on the after-tax return on saving — the rise in the real rate of return and the rate in the tax rate applied to that return — the former must dominate. Further, the model clarifies that there is no short-term pain involved (in either the richer or the poorer household's having to cut current consumption) in order to secure this long-term gain (higher productivity growth). Here, "good news" implies "bad news"; there are "good" productivity growth fronts: unemployment falls, the level of consumption rises, and the ongoing growth rate of living standards rises. We conclude that basic endogenous-growth analysis can support initiatives designed to reduce structural unemployment.

As we noted when reporting a similar finding in section 9.3, this conclusion will not be regarded as too surprising, if one recalls the Bhagwati (1963) theorem — a proposition which states that we have the best chance of improving economic welfare if the attempt to alleviate a distortion is introduced at the very source of that distortion.

Many prominent economists such as Phelps (1977), Solovev (1993), and Freeman (1991) have advocated employment subsidies. In practical terms, they call for a major enlargement of the earned income tax credit policy in the United States. As surprising as it may seem, given the high profile that is enjoyed by these advocates of employment subsidies, the investigation of this broad strategy within an endogenous-productivity-growth setting has not been researched at all extensively. This section of the chapter has been intended as a partial filling of this gap. Of course, much sensitivity testing is needed to see if similar results emerge in other formulations of endogenous growth. Also, it is important to note that this analysis to an open-economy setting. Van Der Ploeg (1996) and Tunovskiy (2003) provide useful starting points for pursuing this agenda.

12.5 The Aging Population and Future Living Standards

There is widespread concern about the living standards that will be available for the generation that follows the baby-boom cohort. According to conventional wisdom, when the baby-boomers are old, the much smaller number of workers in the next generation will face high tax rates (and consequently lower living standards) if the pay-as-we-go public pension and public health care programs are to be maintained. In this section of the chapter, we introduce readers to how our overlapping generations model can be adapted to make it useful for addressing these questions.

As noted, the challenge posed by the aging baby boomers is that the old-age dependency ratio is due to rise noticeably. To analyze this development within the context of the overlapping generations model that was introduced in section 4.2, we must introduce a retirement age. N vector (1994) has extended Blanchard's (1988) model in this way. He shows that — with a constant probability of death (equal to $p$) and a constant overall population (equal to $n$) — a retirement age of $r$ means that the proportion of the population that is in retirement must be $c_{r}p$ and the proportion of the population that is of working age is $(1 - c_{r})$. As long as the model is calibrated so that the retirement age is lower than each individual's expected life expectancy (1 + $p$), each individual will make a time trade-off on the assumption that she will need to finance her consumption needs during a retirement period. The old-age dependency ratio is $c_{r}p = c_{r}$. Within this framework, we can consider a higher old-age dependency ratio in the model economy by imposing a reduction in the retirement age.

It is a bit difficult to calibrate this model. For example, $p = 0.02$ and $n = 40$ are reasonable specifications in and of themselves. (Recall that, in this framework, there is no separate "youth" period, so individuals are born at working age (say 20 years), $p = 0.02$ implies a life expectancy of 20 + 50 years.) The unappealing thing is that these parameter values imply an old-age dependency ratio that is "good news" in other large (relative to the overall dependency ratio). Of course, it should not be surprising that a model which specifies that life expectancy is independent of age, and that people begin working at birth, has some difficulty fitting the facts perfectly. Why, then, do we use such a model? Since the alternatives have some unappealing features as well.

One alternative is Diamond's (1982) 1950s two-period overlapping-generations model. The problem with this specification is that each period is the length of one generation, say 35 years. This forces the modeler to specify that it takes 35 years for some output that has been produced but not consumed to appear in the production forecast as new capital. For this reason, Barro and Sala-i-Martin (1995) advise against using this specification. Despite this advice, many analyses are based on this framework. One accessible example is Scarr and Soumare (2002). A number of studies (for example, Auerbach and Kotlikoff (1987)) have overcome the limitation of Diamond's two-period specification by analyzing a multi-period, discrete-time, overlapping-generations model.

Muhammad Firman (University of Indonesia - Accounting)
The disadvantage in this case is that analytical solutions are impossible, so complete reliance on numerical simulation is required. Since it is very difficult for other analysts to undertake sensitivity tests of this work, these other analysts tend to regard the conclusions as having been generated from a ‘black box.’ As a result, meaningful debate can be limited. It is for these reasons that there is widespread interest in extending Blanchard’s continuous-time version of an economy with overlapping generations.

One such extension is by Jensen and Nielsen (1993). They allow some dependency of life expectancy on age, since they specify that the probability of death is zero for all individuals who are younger than the exogenously given retirement age. Then, there is a constant probability of death once the retirement age is reached. This specification preserves our ability to generate analytical solutions, but not quite as easily as in Nielsen (1994). Other extensions of Blanchard’s framework — designed to examine aging populations and remain analytically tractable — are Faruque (2003) and Bettendorf and Heijdra (2008). Another variation is offered by Gertler (1999). In his model, while individuals are working, they face no probability of death, but they are subject to a constant probability of being retired. Then, once retired, there is a constant probability of death. To keep his model tractable, Gertler must avoid an implication of logarithmic utility — that the degree of income risk aversion and the preference for inter-temporal substitution is pinned down by the same parameter. As a result, he assumes a more complicated preference function, and this allows him to aggregate across the different age cohorts without having to assume that wages are constant — and with only one additional state variable entering the dynamic analysis. Nevertheless, since this framework is much messier than Nielsen’s, we rely on the latter for the remainder of this section of the chapter.

Recall from section 4.2, that (ignoring taxes) the aggregate consumption function in the perpetual-youth version of the overlapping-generations model is specified by the following aggregate relationships:

\[ C = \beta + \gamma (K + H) \]
\[ A = \gamma (K + H) \]
\[ H = \gamma + \beta H - \omega \]

These relationships can be combined to yield the standard aggregate consumption function:

\[ C = \beta + \gamma (K + H) \]
\[ A = \gamma (K + H) \]
\[ H = \gamma + \beta H - \omega \]

The Ramsey (1928) specification is nested within this model: it is the special case that exists when \( \gamma = 0 \). Nielsen’s contribution is to add life-cycle features to this model, so that Blanchard’s model becomes a special case as well. In Nielsen’s set-up, the first two equations are the same as in Blanchard’s:

\[ C = \beta + \gamma (K + H) \]
\[ A = \gamma (K + H) \]
\[ H = \gamma + \beta H - \omega \]

It is the accumulation identity for human capital that is altered by the existence of retirement. By adding up over cohorts of all ages, and over the remaining years in each individual’s life, this accumulation identity becomes:

\[ H = \gamma + \beta H - \omega (\gamma + \beta \gamma) (1 - \omega) \]

The second term on the right-hand side is smaller than previously, since only a portion of the population is working (and therefore receiving wage income). The (new) third term on the right-hand side accounts for the fact that each individual receives the wage for only a portion of her remaining life. For a detailed derivation of this aggregate human capital accumulation identity, the reader must consult Nielsen (1994). Evaluation of the double integral (across cohorts of different ages and across the remaining years in any one person’s life) is only possible if wages are constant. Perfect international capital mobility and constant-returns-to-scale technology are sufficient assumptions to deliver this independence of wages from changes in demographics. As a result, it is prudent to limit our application of this particular model to small open economies. In any event, when the three relationships are combined, we have the revised aggregate consumption function:

\[ C = \beta + \gamma (K + H) - \omega \gamma + \beta H - \omega (\gamma + \beta \gamma) (1 - \omega) \]

As can be seen, the aggregate consumption function is altered when a portion of the population is retired. This is because retirement lowers the aggregate stock of human capital (since human capital is embodied within workers), and it is also because people have an incentive to save more when they have to plan for a period of lower income later in life. As asserted above, this more general consumption function nests the earlier models. For example, when retirement is eliminated (when parameter \( \chi \) goes to infinity), Nielsen’s consumption function reduces to Blanchard’s overlapping generations relationship. Then, when the probability of death falls to zero (\( \gamma = 0 \)), Blanchard’s model reduces to Ramsey’s (1928) analysis of the infinitely lived representative agent. The most general of these three frameworks is needed to consider rising old-age dependency.

As explained in Scarf (2004), a calibrated version of this model can be applied to Canada. The results indicate that Canadian need the government debt-to-GDP ratio to fall to about 20% by 2020 if material living standards are to be increased by about the same amount that the aging population can be expected to (other things equal) lower living standards. Of course, not all dimensions of the aging population are captured in this model. For example, the elderly consume a higher proportion of services, and since it is more difficult to have productivity increases in the service sector, aging can be expected to bring lower growth, and this mechanism is not included in this model. Despite this fact, the Canadian federal government has accepted the 20% value as its target for the debt rate, and officials have stated that this target is based on the desire to stabilize living standards from the coming demographic shock. So it certainly is the case that policy makers do pay attention to the models that we have examined in this book.

12.6. Conclusions

How can our tax system be designed to promote both an increase in fairness (as we compare rich and poor today) and an increase in average living standards over the years to come? How can our structural unemployment problems be addressed without jeopardizing our desire to have more rapidly rising material living standards? What is needed to limit the threat to living standards that is posed by the aging population? These questions are among the most disputed topics in public policy analysis today. It is the job of the policy-oriented economist to use the analytical structure of our discipline to help inform policy makers on these matters. It is important to identify both the trade-offs and the “free lunches” that are possible, when confronting these issues. This chapter has been designed to help readers meet this challenge, by explaining how basic growth theory can be directly applied to help us understand these topical questions.