Chapter 17

THE MACROECONOMICS OF GOVERNMENT FINANCE

MICHAEL HALIASSOS*
University of Maryland

JAMES TOBIN
Cowles Foundation, Yale University

Contents

0. Summary 890
1. Introduction 890
   1.1 The central questions 891
   1.2 Some doctrinal history 892
2. Government finance in traditional macroeconomics 896
   2.1 The IS/LM framework 896
   2.2 Fiscal multiplier theory 897
   2.3 Cumulative effects of deficits 899
   2.4 Dynamics of deficits and debt 901
   2.5 Multi-asset models 904
3. "Microfoundations" and parables 906
4. Financing deficits with non-monetary debt 910
   4.1 Crowding out and the interest elasticity of saving 911
   4.2 The theory of debt neutrality and the optimum debt level 912
   4.3 The conditions for debt neutrality 916
   4.4 Empirical tests of debt neutrality 924
5. Financing deficits with money 930
   5.1 Superneutrality in the steady state 930
   5.2 Superneutrality along the transition path 934
   5.3 Empirical tests of superneutrality 936
6. Monetary policies 941
   6.1 Asset exchanges 941
   6.2 Consequences of persistent deficits 948
7. Concluding remarks 952

References 954

* Haliassos' research was partially supported by the Graduate Research Board of the University of Maryland, College Park

Handbook of Monetary Economics, Volume II, Edited by B M Friedman and F H Hahn
0. Summary

This chapter is a critical survey of literature on the implications of government financial policy for economic activity. The central question is whether the mode of financing of a given path of real government purchases—by taxes, non-monetary debt issue, or money creation—has real effects, in particular real effects of macroeconomic consequence.

In Section 1, the Introduction, we define the issues with greater detail and precision. We briefly review economists' views, over the past fifty years, of the burden of public debt and the neutrality of money. Section 2 is a review of the 1960s vintage mainstream macroeconomics of fiscal and monetary policies, often called the "neoclassical synthesis". We review its implications for both short-run fluctuations and long-run trends. We include this review because the earlier tradition covers some problems and issues now neglected, because its analyses and results may still have some validity, and because they did set the stage for—one might say they provoked or inspired—the recent literature surveyed in Sections 4–6. The earlier tradition and the recent literature differ in methodology, and in Section 3 we discuss the "microfoundations" methodology that dominates contemporary macroeconomics.

Sections 4–6 are a selective critical survey of recent contributions, theoretical and empirical, designed to summarize the current state of play on the central issues: Section 4, the debt neutrality hypothesis of Robert Barro; Section 5, the effects of financing government expenditures by printing money rather than taxing, monetary superneutrality, and the "Fisher effect" of inflation on interest rates; Section 6, open market operations, and shifts between bond- and money-financing of government expenditures induced by the adoption of financial policies which are unsustainable over the longer run. In each section we first set forth the neutrality theorems purporting to show the irrelevance of the financing choice. Then we discuss articles elaborating or criticizing the theorems. In each case we conclude with a review and evaluation of some empirical tests. We conclude in Section 7 with short summary remarks.

1. Introduction

In the years 1945–1970 fiscal and monetary policies were widely regarded, both by economists and by policy-makers, as important instruments of macroeconomic stabilization. Counter-cyclical variation of budget deficits (counting surpluses as negative deficits) was accepted as stabilizing, whether it arose as the passive result of built-in responses of revenues and outlays to economic
activity or as the result of active tax and expenditure adjustments. Likewise, reactive counter-cyclical monetary policy was regarded, both by economists and by central bankers, as a useful, even an essential, stabilizer.

Subsequently, considerable disillusionment with "discretionary" policies, scornfully labeled "fine-tuning", swept professional and lay opinion. Government's ability and willingness to offset destabilizing non-policy shocks were questioned, and variability of policy instruments themselves became the favorite culprit for cyclical fluctuations. This shift of sentiment persuaded leading central banks to heed the advice of monetarists to adopt and adhere to money stock targets independent of the state of the economy, although never and nowhere to the degree the monetarist theorists prescribed.

In fiscal policy, the same shift revived the popularity of the norm of annual budget balance. In the United States in 1986 thirty-two state legislatures, the President, and majorities of both Houses of Congress supported a Constitutional Amendment mandating a balanced budget. Many economists now favor binding the central government to a fiscal rule limiting deficits. Others, however, advocate rules limiting the size and growth of government expenditures. Whether public claims on national resources are financed by taxes, by borrowing, or even by printing money is, in their view, of secondary importance.

The neoclassical counter-revolution in macroeconomic theory has lent powerful support to these intellectual trends. The two crucial strands in the New Classical Macroeconomics are *rational expectations* and continuous competitive *market clearing*. They reinforce the trained instincts of economists that unfettered markets yield optimal outcomes, that in particular there are no macroeconomic market failures which government need or can correct. They emphasize expectations of future policies as determinants of private behavior, a point especially important for macroeconomics. The famous "Lucas critique" of macroeconometrics [Lucas (1976)] says that the structures of behavior allegedly detected in econometric models will collapse if governments change policies to try to exploit them. Moreover, rational expectations models limit or obliterate the scope of monetary and other financial policies: if they are systematic, private agents will understand them and, by altering their behavior, undermine them; if they are random, the confusion of the public will make the outcomes inefficiently volatile.

1.1. The central questions

The expenditures of a sovereign central government can be financed by tax receipts, printing money, or borrowing. Does it matter in what proportions these three sources of finance are used to finance a given program of
expenditures? Do deficits (excesses of expenditures over tax revenues) affect important macroeconomic variables: real output and its growth, unemployment, prices and inflation rates, nominal and real interest rates? Given the size of a deficit, does it make a difference whether it is covered by money or borrowing? These are the central questions to which this chapter is addressed.

Those questions are most significantly posed on several assumptions: the money is fiat money, inconvertible paper; it is not commodity money; it is not even convertible into gold or foreign currency at a guaranteed rate of exchange. Public borrowing usually takes the form of the government's promises to pay amounts of its money at future dates. Most models assume that the government does not promise to deliver commodities or foreign currencies; its debt is not indexed to money prices of commodities.

In practice, of course, many governments do borrow foreign currencies or index their debts to prices of commodities or foreign currencies in their own moneys. Those means of financing raise quite different issues, easier to analyze if not easier for the borrowing governments to manage.

A closed economy is a convenient assumption for most of the issues and literature of concern in this paper. Financial policies in open economies are specifically treated elsewhere in this Handbook (Dornbusch and Giovannini, Chapter 23).

Not all governments have the tripartite financing choice described above. In some countries there is no internal way to finance deficits other than printing money. Essentially fiscal and monetary policy are the same thing. Deficits create money, and there is no other way to create government (as opposed to bank) money. The literature of economic theory contains many models in which deficits are 100 percent monetized and fiscal and monetary policy are indistinguishable. It also contains models in which government money is the only store of value for the whole society. Land and capital are assumed away. These models not only exclude analysis of non-monetized deficits but also evade the issue of "crowding out", the alleged displacement of capital formation by the absorption of saving in government debt.

Our survey will cover some models of all these varieties. But the most relevant, in our view, allow governments both monetary and non-monetary alternatives to taxes in financing current outlays and allow private savers stores of value other than government obligations.

1.2. Some doctrinal history

The burden of public debt. The major question has long been whether borrowing, rather than taxing, to finance a given program of public expenditures can shift the burden of those expenditures to future generations. Laymen
 instinctively answer "Yes". For example, President Eisenhower solemnly preached against deficit spending, on behalf of the grandchildren and later descendants of his contemporaries. Among economists, however, the long-standing orthodox view was that internal debt could not shift the burden. "We owe it to ourselves." The resources used for government expenditures, for example war, are in a closed economy with full employment necessarily drawn from other current uses of resources. The reduction in resources available for current non-government use is the same regardless of the method of financing. Subsequent payments from taxpayers to contemporaneous bondholders redistribute income between members of future generations. But these transfers involve no draft on resources in aggregate, other than the deadweight loss of non-lump-sum taxes. That may have been deferred, but not the full real burden of the expenditures themselves.

This view was stated forcefully by Lerner (1943, 1944) in his theory of functional finance. He attacks as myths the balanced-budget norm and the fear that deficits burden future generations. As a Keynesian, he does not take full employment for granted. Given the government's expenditure program, the function of taxation is to control private spending; taxes should be set and varied as necessary to equate spending to full employment output. The debt should be monetized to the degree that will bring about the real interest rate needed to sustain the optimal rate of capital investment. Given the government's expenditure program and the desired rate of capital accumulation, adjustment of taxes will make sure that consumption demand equals the remainder of full employment output.

Lerner saw clearly that fiscal and monetary stimuli are substitutes in demand management. Various mixes of the two policies can achieve full employment. Which mix to choose depends on other objectives, specifically the desired division of output between consumption and investment. Lerner did not see that this degree of freedom in macroeconomic policy undermined his assertion that the burden of government could not be shifted to future generations.

After World War II the theory of the burden of public debt excited lively controversy among economists. See the symposium edited by Ferguson (1964), reviewed by Tobin (1965b). In the symposium and elsewhere, James Buchanan attacked the orthodox "no-shift" view on the ground that deficit spending clearly defers what he regarded as the true burden, compulsory payment of taxes. He argued that lending to the government, in contrast, is voluntary and therefore burdenless.

However, the more successful revision of doctrine was what may be called the capital stock criterion. This was the prevailing view of public debt burden among American Keynesian architects of the "neoclassical synthesis" and "neoclassical" growth models in the period 1946–65. Franco Modigliani was a forceful advocate of this criterion, also in the Ferguson volume. Government
financial policies can, after all, affect the consumption opportunities of future generations. An important test is whether we would endow our descendants with a smaller or larger stock of capital, public and private, human and non-human, foreign and domestic. The neoclassical synthesis placed great emphasis on the monetary–fiscal policy mix, which Lerner had recognized as important. The criterion is quite consistent with Lerner’s strictures on the policy mix. The same short-run path of real output can be achieved with tight money and high deficits or with tight budgets and easy money. The former combination will have higher real interest rates and lower capital accumulation and growth. On a flow basis, the deficit “crowds out” productive investment; on a stock basis, the debt displaces capital in the wealth portfolios of private savers.

The next move in this contest revived the older “no-shift” orthodoxy. Robert Barro’s extraordinarily influential article, “Are Government Bonds Net Wealth?” (1974), rejected the theory of consumption and saving behavior underlying the arguments of Modigliani and other mainstream neoclassical Keynesians. According to Barro, private citizens will rationally anticipate the future taxes that any current deficit will impose on themselves or their descendants. The present value of these tax liabilities will be exactly the same as those of the debt instruments issued to finance the deficit. Internalizing the circumstances and utilities of their dynasties, current citizens will regard their net wealth as independent of current taxes and deficits. If their taxes are cut, they will not consume more; they will save the full amount of the tax cuts to enable themselves or their heirs to pay the postponed taxes. Buying the debt instruments themselves would be the perfect hedge. Much of this chapter, especially Section 4, will be devoted to discussion of this doctrine, which Barro called the Ricardian Equivalence Theorem [Ricardo (1817; 1951)].

Debt neutrality, as asserted by the Equivalence Theorem, is to be distinguished from monetary neutrality. The debt burden controversy concerns real variations in tax revenues, current and anticipated, balanced by equal and opposite variations in real amounts of net public borrowing (or repayment and lending) at market interest rates. Debt neutrality means that these variations

1The point was not wholly absent from the previous discussion. For example, in Tobin’s review (1953b) of the Ferguson book, he offers a comment which “questions the consumption-savings behavior assumed in the Modigliani notion of the burden. Is it not based on some asymmetrical illusion? Society fools itself into consuming more, thinking that possession of government paper provides for its future. Why don’t those who will have to pay taxes to service the debt consider themselves poorer and save more accordingly?” This observation threatens not only Modigliani’s concept of debt burden but equally the belief that the government can influence investment and growth by varying the fiscal-monetary mix. Indeed it comes dangerously close to denying that any internal financial and monetary arrangements are of any real consequence.” Nevertheless, the author goes on to suggest that the government as a financial intermediary can “dampen some of the needs which generate saving”, and to claim that the weight of empirical evidence is that “the private income and wealth corresponding to government deficit and debt stimulate consumption”.

have no real effects. Monetary neutrality, on the other hand, means that variations in *nominal* quantities of government-issued fiat money have no real effects.

Patinkin (1956) called this ancient orthodoxy the *classical dichotomy*. Nominal variables have nominal effects. They do not alter real quantities. They may change the general price level but they do not affect relative prices. They may alter nominal interest rates but not real rates. The quantity theory of money asserts that variations of money stocks are wholly absorbed in nominal prices. This is what happens when government expenditures are financed by issuing money rather than by taxes. The nominal interest rate on base money is fixed, usually at zero. The public is induced to hold the new money by the price increases that raise the nominal values of income, wealth, and transactions. The conclusion is that only the real size, nature, and distributional impacts of government activities affect the behaviors of economic agents and real outcomes. Debt neutrality takes this conclusion further, excluding the timing of taxes from policies of real consequence.

A quantity-theory proposition stressed by Friedman (1970), for example, is that the consequences of increasing the money stock are the same no matter how it is increased. Printing money to finance deficits has the same effect as printing money to buy back government debt obligations. Critics objected that deficit financing increases private wealth, and therefore private demand for goods and services, more than open market operations of equivalent size. The reason is that in the latter case wealth-owners sell assets to acquire money. But if government bonds are not net wealth, the two ways of increasing money stock are identical in effect, as Friedman asserted.

Wealth effects on real consumption and other demands for goods and services have played an important role in macroeconomic theory. They are the source of the real balance effect, resulting from deviations of price levels from expectations [Pigou (1943, 1947), Patinkin (1948)]. These deviations alter the real values of nominal assets – and debts. Private assets exceed private debts to the extent that government debts do not give rise in private calculations to future tax obligations. Ricardian equivalence confines the real balance effect to monetized government debt. Perhaps it obliterates the effect altogether. Wealth-owners may not regard an increase in their holdings of government money as net wealth if they expect it to be eroded by future inflation. Seignorage in modern times is not literally the sovereign’s take from coinage. The government can finance part of its budget by meeting the growth in public demands for its fiat money. These demands grow faster the higher the inflation rate, and the higher the economy’s real growth. Seignorage is a kind of tax, alternative to explicit taxation.

The real balance effect came into macroeconomics as an answer to Keynes’ contention that the market economy lacked an effective mechanism for restor-
ing full employment, once a demand shock had jarred it out of equilibrium. Keynes challenged neoclassical economists’ reliance on nominal price adjustments, on the neoclassical ground that real demand should be independent of nominal prices. In particular, Keynes said workers were incapable of lowering real wages by lowering nominal wages, although lower real wages were essential to remedy unemployment. Latter-day skepticism of the real balance effect strengthens Keynes’ challenge. Today’s new classical macroeconomists finesse the problem of equilibrating adjustment by assuming continuous equilibrium, to which nominal price developments are irrelevant.

The co-existence of monetary and non-monetary government debt continues to be a source of problems and puzzles in macroeconomic theory. Price changes affect the values of both the same way. Currency always bears zero interest; obligations to pay currency in future pay a variable market-determined rate. How come they differ in yields? The answer is that money provides implicit yields, services in kind, that make up for its lack of explicit return. These non-pecuniary yields, and therefore the explicit yield advantages of debt instruments, depend inversely on the real quantity of money. Recognition of this relationship and of its macroeconomic importance is one of Keynes’ major contributions to monetary theory.

The fixed nominal interest on money has another major implication. Neutrality with respect to nominal price levels does not imply neutrality with respect to temporal changes in prices. Those changes – inflations and deflations – alter the real rate of return on money, and possibly on other nominal assets and debts as well. Variations in real interest rates are real events and are likely to have real effects. Since money and other assets are substitutes, albeit imperfect, variations in real rates on money will be transmitted to other assets, nominal and real, at least in the short run. This effect, often bearing the name of Mundell (1963b) or Tobin (1965a), will be discussed in Section 5 below. Models that somehow deny these effects are termed “superneutral”.

2. Government finance in traditional macroeconomics

2.1. The IS/LM framework

Although Keynes eloquently stressed the heretical implications of his General Theory (1936) for fiscal and monetary policies, he did not actually model them explicitly. Hicks’ (1937) algebraic formulation of the Keynesian model proved to be a fruitful and durable framework for macroeconomic policy analysis. Generations of students learned the “IS/LM” calculus of aggregate demand, and it is still the tool of instinctive first recourse of many macroeconomists.
Indeed, it is still the basic framework of macroeconomic models, large and elaborate as they have become because of disaggregations and lags.

Fiscal policy shifts the IS curve, monetary policy the LM curve. What could be simpler? The behavioral and structural parameters on which the effects of the policies on aggregate demand depend are easily discerned: the impotence of monetary policy in the liquidity trap (flat LM) or in the event both investment I and saving S are insensitive to interest rates (vertical IS); the impotence of fiscal policy if money demand is insensitive to interest rates (vertical LM) or if investment or saving or both are perfectly interest-elastic (flat IS); and all the cases in between. The analysis can be extended to derive aggregate real demand as a function of price level, allowing for the effects of prices on real supplies of money and wealth, and this "AD" function can confront an aggregate supply "AS" relation between the same variables.

As was Hicks' original intention, the same apparatus can show the effects of financial policies in the "classical" world of market-sustained full employment. Indeed, IS/LM graphs generally look qualitatively the same when the horizontal axis denotes price level, $p$, rather than real output, $Y$. The effects of inflation expectations – Mundell–Tobin effects – can be shown very simply, given the approximation that IS involves the real interest rate but LM the nominal rate.

Early in its life the Hicksian model was extended to open economies [Fleming (1962), Mundell (1963a)] and to the analysis of macro policies with and without international capital mobility and under fixed and floating exchange regimes.

Policy analysis with the framework was not confined to shifts in policy-determined variables; the policy-makers could be modeled as following rules, and the rules built in to IS and LM relations themselves. Welfare judgments among alternative rules would depend on the stochastic environment, for example, as in Poole's classic paper (1970) on the variances and covariances of real demand shocks and financial shocks.

2.2. Fiscal multiplier theory

Keynesian analysis of fiscal policy began with the "multiplier", invented by Kahn (1931) to formalize and quantify the instinctive beliefs of Keynes and others that deficit spending for public works during the Depression would create not only direct jobs but also indirect jobs, probably much more numerous. At the same time, the analysis showed why the chain of spending and job creation was not endless, why the multiplier was finite. Kahn's discovery paved the way for the General Theory itself, which could not have been written without the consumption function. Although the original idea of
the multiplier referred to increases in government purchases, it was later realized that the same logic applies to private expenditures induced by tax reductions or transfer payments.

The explicit implications of multiplier theory for fiscal policy were developed in the late 1930s and the 1940s, largely by Alvin Hansen and his followers, mainly young colleagues and students at Harvard [Hansen (1941)]. Hansen (1953) was also an exponent of the IS/LM framework. As he realized, the multiplier tells the amount of horizontal shift of the IS curve resulting from a fiscal stimulus or other real demand shock, but this is not the end of the story. In any case, fiscal multiplier theory yielded some important elaborations and clarifications:

1. Government budgets are partly endogenous. Congresses or parliaments enact tax codes, but the revenues they realize depend on their tax bases, and these vary with the state of the economy. The same is true of some budget outlays, notably entitlements to transfers like unemployment compensation, welfare, and farm subsidies. Legislation specifies rules for determining eligibilities and amounts, but the outlays then are open-ended and depend on the economy.

2. In consequence, budget deficits are partly endogenous. In practice, they are, for given tax and entitlement codes and given expenditure programs, inversely related to economic activity, i.e. to national product. It is important to distinguish these endogenous variations in deficits (or surpluses) from those arising from exogenous policy changes in the codes and programs. The former have no multiplier consequences; they are already built into the multiplier and into the IS curve. The latter are multiplicands; they do have the consequences described by the multiplier and by shifts of IS. The distinction is the reason for calculating cyclically corrected budget deficits from year to year, eliminating endogenous cyclical variations and isolating programmatic changes. Deficit measures calculated for this purpose have, over a history that goes back to the late 1940s, been dubbed “full employment”, or “high employment”, or “structural”.

3. Taxes and transfers, as well as earned factor incomes, affect consumption spending. A common assumption has been that consumption and saving both depend positively on after-tax, “disposable” incomes. This implies, given the procyclicality of taxes minus transfers, that the multiplier is smaller the larger are the parameters that link tax liabilities positively, and transfer entitlements negatively, to income. A small multiplier limits fluctuations caused by non-policy shocks, and for that reason the elements of fiscal structure that lead to a small multiplier are called “built-in stabilizers”. The same stabilizing structure makes deficits more sensitive to macroeconomic conditions.

4. Not all structural budget changes of equivalent dollar amount have the
same effects on aggregate demand. One reason is that they differ as to "multiplicand". For example, part of a lump-sum tax cut or transfer increase will be saved, so that the multiplicand per dollar will be less than a dollar; in contrast, a dollar government purchase of goods and services increases aggregate demand for national product by a dollar even if no indirect multiplier effects occur. This distinction is the source of the celebrated "balanced-budget multiplier theorem", which says that a fully tax-financed one-dollar increase in government purchases raises aggregate demand by one dollar.

2.3. Cumulative effects of deficits

Keynes stated explicitly that his model applied to a short run in which the flow of investment determined by its solution alters negligibly the stock of capital. Indeed, that is his definition of "short run". The IS/LM version has this same limitation. It is necessary because changes in the stock of capital would alter the investment function, perhaps also the saving function and other behavioral relationships, and therefore alter also the equilibrium solution. The same is true of other stocks and flows. Saving accumulates wealth. Government budget deficits add to the stock of public debt. Current deficits on international accounts diminish the nation's net foreign claims.

IS/LM analysis usually ignores these stock-flow identities. Strictly speaking, fiscal multipliers apply to the effects of larger or smaller values of policy instruments, expenditures or tax rates, at an instant of time. No matter how large the deficit, a flow, it cannot change the stock of debt in zero time. Thus, the IS/LM snapshot cannot answer questions about alternative means of financing continuing deficits.

Of course, the LM relation, unlike IS, is equality of demand and supply of a stock, the stock of money. The snapshot model allows this stock to vary, presumably by open market exchanges for other assets, notably government securities. These exchanges, however, take no time and involve no saving or dissaving, although they may alter the value of those assets and of total wealth. Note that the IS/LM model contains a consumption/saving function describing the rate at which the public desires to accumulate wealth. But it contains no function telling in what forms, monetary and non-monetary, the public wishes to accumulate it. The money-demand function tells how, given pre-existing asset stocks as altered by instantaneous open market operations, the public chooses to divide its wealth between money and other assets.

These properties of the short-run Keynesian model were well understood by its architects. But in the 1960s and 1970s a number of critics "discovered" what they called the government budget "restraint" (surely a misnomer) or "constraint", and accused standard short-run macroeconomics of erroneously ignor-
ing it or even violating it [Christ (1968), Currie (1978)]. What they meant but mislabeled is simply the budget flow identity: outlays equal tax revenues plus security issues plus new money creation. No error is involved in confining the analysis to a short enough time period so that the resulting stock changes do not matter. The question is, however, whether the comparative-static solutions of such a snapshot model with respect to policy variations are misleading. In particular, the standard model says that, except for extreme shapes if IS and/or LM, pure fiscal stimulus, i.e. 100 percent bond-financed increases of expenditures or reductions of taxes, raises aggregate demand. The critics said that this conclusion might be reversed as the time period is extended and the deficits raise the stocks of securities relative to those of money. Some econometric equations showed fiscal multipliers dying out and even reversing sign with the passage of time [Anderson and Carlson (1970), Modigliani and Ando (1976)].

It is hard to see how this reversal would occur in an under-employed economy with constant real money supply if the two-asset framework of the IS/LM model is maintained. In that framework, government debt securities and capital are perfect substitutes in portfolios, each bearing "the" real rate of interest, and the portfolio choice between that composite asset and money depends on income, wealth, and the nominal interest rate, as in the money-demand function. Could upward shifts of LM due to the rising stock of government securities, i.e. to rising private wealth, more than offset the upward IS shift and consign the economy to a higher interest rate at lower income? Both of those outcomes would diminish the demand for money. Only an increase in wealth could do the reverse. A decline in income would diminish the demand for wealth; only the rise in real interest could do the reverse. It could achieve the reversal only if it induced portfolio managers to hold more wealth not only in the assets whose yields had increased but also in the other one, money. So the story apparently depends on an unlikely combination of effects.

This technical issue – the permanence or transience of positive effects of pure fiscal stimulus on aggregate demand – inspired formal dynamic analysis. Blinder and Solow (1974) showed how the economy would gravitate to an equilibrium of budget balance: incomes would rise until they generated enough revenue flow to offset the continuing fiscal stimulus, inclusive of the rising debt interest outlays. Tobin and Buitert (1976) obtained essentially the same result, with more elaborate modeling of the effects of wealth and capital stock on saving and investment and of the dynamics and stability conditions of the model.

The upshot of these papers is to maintain over the longer run the conclusion of short-run analysis: pure fiscal policy is expansionary unless the interest elasticity of money demand is zero. However, the endogeneity of interest rates makes budget outlays for debt service endogenous also. Just as in short-run
IS/LM analysis there can be perverse comparative statics – expansionary policies cannot balance the budget except at lower output and interest rates. Just as in short-run analysis, these equilibria are unstable. A surprising conclusion is that the long-run multiplier for fiscal stimulus may well be higher if deficits are not monetized than if they are. The equilibrium interest rate is lower if they are monetized, and it takes less income to raise enough revenues to balance the budget.

These scenarios are meant to extend the calculus of aggregate demand to situations in which stocks are allowed to vary endogenously with the passage of time. To focus on the direction and magnitude of stock effects, the authors assume that output is demand-determined. The scenarios cannot be taken as realistic simulations, because supply limitations and price movements will intervene.

Tobin and Buijer also analyze the full employment case where prices are flexible. Output is supply-determined but varies with the capital stock. The comparative statics of balanced-budget equilibrium says that fiscal stimulus is expansionary and lowers interest rates, because it takes more income and thus more capital to generate the necessary revenues. The stability of the equilibrium requires that price expectations adapt fairly quickly to price experience. As the authors noted at the time, the long-run comparative statics of models of this type do not depend on counting government debt as net wealth. The budget-balance condition does the trick. Although these papers assume stationary labor force and productivity, they could be formulated as growth models, where the natural counterpart of budget balance is growth of debt (in real terms) at the economy’s natural growth rate. Possibly more serious matters are the assumptions that all non-monetary assets are perfect substitutes, that saving and wealth demand are interest-inelastic, and that stochastic disturbances are absent.

2.4. Dynamics of deficits and debt

Some fifty years ago Roy Harrod in England and Evsey Domar in the United States independently originated modern macroeconomic growth dynamics. Domar (1944) applied his model to the growth of public debt. He saw the importance of comparing the debt interest rate (net of taxes paid on such interest) and the growth rate of output and income. Specifically, assume that enough taxes are levied to pay debt interest but not to pay for all other expenditures. The debt will grow. Will tax rates have to be raised without limit? Yes, if and only if the after-tax interest rate on the government’s debt obligations exceeds the growth rate. The insight is important in interpreting the Ricardian Equivalence Theorem. In a growing economy, the “government
budget constraint” that rational taxpayers will internalize is not that the public debt cannot grow, not even that it cannot grow indefinitely, but that it cannot grow faster than the economy itself.

Thus, interest on the outstanding debt differs significantly from other items in the budget. Interest outlays depend on the size of the debt accumulated in the past, on interest rates, on the degree of monetization, and on the composition of the interest-bearing debt by maturities and other characteristics of securities. Like taxes and transfers, interest expenses are, given prior fiscal history, largely endogenous. However, they do not depend directly on economic activity, but on interest rates as determined by the interactions of the economy, the financial markets, and monetary policy. Legislators have even less control over interest outlays in the short run than over tax revenues and entitlement transfers.

The nominal deficit, in dollars per year, is at any moment the rate \( \dot{D} \) at which the nominal debt \( D \) is growing. Let \( X \) be the nominal primary deficit, the excess at that moment of the rate per year of expenditures, i.e. purchases and transfers, over the rate of accrual of tax revenues, excluding interest outlays from transfers and excluding tax liabilities on such interest from revenues. The primary deficit is what the deficit would be, other things equal, if there were no outstanding debt. Let \( \bar{t} \) be the average nominal interest rate, after taxes, on non-monetary debt, and let \( H \) be the nominal amount of monetary debt, the monetary base. Then the deficit is:

\[
\dot{D} = X + \bar{t}(D - H) .
\]  

(2.1)

This is the deficit as conventionally calculated and debated.

However, there are good economic reasons to prefer other measures. A real deficit would convert \( D \), \( H \), and \( X \) to real quantities, “deflating” them by a price index, \( p \).

Still another measure would charge as the cost of debt service the real interest rate rather than the nominal rate. Let the real after-tax interest on the non-monetary debt be \( r \), equal to \( \bar{t} - \pi \), where \( \pi \) is the rate of inflation. The real inflation-corrected deficit is:

\[
\frac{\partial}{\partial t} \frac{(D/p)}{p} = \dot{D}/p - \pi D/p = X/p + rD/p - (r + \pi)H/p .
\]  

(2.2)

The logic of this concept is “inflation accounting”. The government’s primary budget is geared to real programs, which cost more dollars when prices are higher; if the tax law is indexed, either formally or by periodic legislative adjustment, real tax revenues are neutral with respect to prices. The deficit is government dissaving, its claim on the nation’s private saving in this sense: the
amount of investment and other non-governmental claims on private saving that the government either crowds out (when aggregate supply limits output) or makes unnecessary for the maintenance of aggregate demand (when supply is not a constraint).

Private savers are presumably interested in real wealth and real streams of present and future consumption. If so, they will save extra in dollars to offset depreciation by inflation of their net real holdings of nominal assets, including government debt. They will understand that part of their nominal interest receipts is return of principal rather than income. Eisner (1985) has shown how conventional accounting overstated deficits in inflationary years of the 1970s; inflation accounting converts those deficits into real surpluses. However, saving behavior may not be inflation-neutral in the short run, even if savers eventually adapt.

The logic of inflation accounting can be pushed a step further to allow for real growth of output and income as well as inflation. Suppose that savers are interested in maintaining the ratio of their wealth to their income. Let \( d, h, \) and \( x \) be the ratios of \( D, H, \) and \( X, \) respectively, to nominal income \( pY. \) Let the growth rate of real income \( Y \) be \( g. \) Then the growth in the debt/income ratio \( d \) is:

\[
\frac{\partial}{\partial t} \left( \frac{D}{pY} \right) = \frac{\partial}{\partial t} \left( \frac{\pi}{pY} \right) = x + (r - g)d - (r + \pi)h .
\]

The first two terms by themselves say that unless the primary deficit is negative, the debt will grow faster than national income if the net real interest rate on the debt exceeds the growth rate of the economy. The last term is equivalent to a reduction of the primary deficit. It is the seignorage “income” of the government, the interest savings due to its power to print interest-free money. To the extent that this term is augmented by inflation, it represents the “inflation tax”. In the United States \( h \) is small relative to \( d, \) of the order of 10–12 percent. Seignorage on base money is small at low or moderate rates of inflation, but there may be other ways in which inflation collects real tax revenue.

According to equation (2.3), the steady-state value of the ratio is:

\[
d^* = \frac{(x - (r + \pi)h)}{(g - r)} .
\]

Equation (2.3) itself says that the ratio \( d \) will be increasing if the primary deficit \( x \) less the seignorage term \( (r + \pi)h \) is non-negative and the real interest rate \( r \) exceeds the growth rate \( g. \) Indeed,

\[
\dot{d} = (d - d^*)(r - g) .
\]
The steady-state debt/GNP ratio will be unstable, ceteris paribus, if and only if the interest rate on government debt exceeds the growth rate of the economy. A positive $d^*$, implying positive steady-state deficits, is compatible with $r - g > 0$ provided the primary deficit less seignorage is negative, but any initial deviation from $d^*$ will be magnified.

Crowding out of capital accumulation (in open economies of net claims on the rest of the world as well as domestic capital) is the social cost of deficit finance that does not create its own saving. Suppose that the central bank's monetary policy keeps the economy at full employment with a constant inflation rate. Having no reason to expect future tax increases, consumers buy government debt issues with funds that otherwise would have financed domestic (and foreign) investment. The deficits finance private and public consumption. In a stable scenario ($r < g$) the shift from one steady state to another produces a slow and undramatic reduction in capital and consumption per capita. In an unstable scenario, developments are more spectacular and frightening. To the dynamic of equation (2.5) is added another destabilizing feedback as capital becomes more and more scarce and its marginal productivity rises, the debt interest rate rises too. Eventually the deficit absorbs all private saving, investment ceases, the capital stock gradually wears out, and equity values plummet. The story [Tobin (1986a)] is meant not as a forecast but as a cautionary tale.

2.5. Multi-asset models

Macroeconomic models, both theoretical and empirical, are aggregative in several dimensions: over agents, time periods, commodities, markets, and assets. Index number problems, essentially insoluble, abound. Mostly we ignore them and try to muddle through. In monetary and financial economics, aggregation of assets is especially problematical. Some theoretical monetary models actually contain only one store of value, “money”, thus managing to confuse the theory of money with the theory of saving. Consider the “islands” parables supporting the “Lucas supply curve”, in which intertemporal consumption and leisure choices also determine demands for money. Some other models allow only two assets, base money and capital. As observed above, these models can illuminate some points, but their asset menu allows no distinction between fiscal and monetary policies.

The standard IS/LM model distinguishes two assets, money and everything else. The constituents of the second category, capital and non-monetary government obligations, are taken to be perfect substitutes, all bearing essentially the same market interest rate. Because of money's use in transactions and its other characteristics, the other assets are not perfect substitutes for money,
whose fixed (usually zero) nominal yield is generally smaller than the yield of alternative holdings. As noted above, this model does allow for distinct treatment of monetary and fiscal policies and shocks.

Nevertheless, it has several shortcomings. Equities in real capital and obligations to pay fixed amounts of money are affected differently by various shocks to inflation, productivity, taxes, and other variables. They are portfolio substitutes, but by no means perfect. Their sources of supply are quite different. A three-asset menu—capital, base money, and government non-monetary debt—allows both monetary and fiscal policies to be more faithfully modeled. Open market operations between money and Treasury obligations, which may be bills maturing tomorrow or next week, are not the same as operations in equity or capital goods markets.

The three-asset model sheds light on the problem of Subsection 2.3. What happens when the stock of non-monetary debt is increased relative to the stocks of base money and capital, as the result of a period of bond-financed deficit spending? In particular, what happens to the rate of return wealth-owners require in order to hold the existing capital stock? If this rate goes down—that is, the prices of equities go up—the increase in stock of government debt is favorable to capital investment and to aggregate demand. If the rate goes up, the growth in debt is contractionary. The answer depends, speaking loosely, on whether money and debt are better substitutes than capital and debt. If so, the outcome is expansionary; if not, contractionary [Tobin (1961, 1963)].

Since government debts are futures contracts in government currency, they might be expected to be close substitutes for currency. This is especially true of Treasury bills maturing tomorrow, next week, or next month. By bank reserve requirements and other legal differentiae, governments seek to reduce the substitutability of money and short bills. Long-term government bonds, while connected to short-term obligations by a chain of maturity substitutions, also share many of the sources of risk of ownership of capital or private equities and bonds. It takes more than a three-asset model to handle the complexities of the situation. Tobin (1963) distinguishes long and short debts and considers the effects of debt management, i.e. variations in the maturity structure of non-monetary government debt.

B. Friedman (1977, 1978) has investigated both theoretically and empirically the basic issue: whether, as he puts it, debt accumulation "crowds in" or "crowds out" private capital formation. His conclusion is that the substitution elasticities are such that debt crowds in. This must be understood as a demand-side effect, which will prevail only if unemployed resources are available and the central bank does not contract the money stock.

Multi-asset models can be extended to "inside" assets and debts (deposits and other intermediary liabilities, and private loans and securities). In the
process, the money markets and reserve requirements through which central bank operations affect banks, financial markets, and the economy can be explicitly shown.

Banks and other financial intermediaries hold, involuntarily and voluntarily, a considerable amount of public debt, monetary and non-monetary. This enables their depositors, the indirect owners of that debt, to hold more convenient and more liquid assets than the debt itself. A much more important macroeconomic function of financial intermediaries is to monetize or to make more liquid the ownership of private debts, the obligations of businesses, entrepreneurs, home-owners, and consumers. Their investments are financed at lower rates and on easier terms than if they had to be financed directly by conservative savers, who accept intermediary liabilities as close substitutes for government money and securities.

A short-run macro model should be regarded as referring to a slice of time or a finite period, with inputs from the past, modified by the solutions of the model in the present, passed on as data for the immediate future. At each time demands for additional amounts of all the assets are equated to the new supplies. New supplies of equities come from capital investment; new supplies of government bonds and base money, in proportions determined by monetary policy, come from the government deficit; new supplies of foreign-currency assets come from net exports. Rates of return, and output and/or prices as well, adjust to induce wealth-owners to absorb these additional supplies in their portfolios. The sums, both of demands and of supplies, equal the additions to wealth. Equations for inside assets and debts can be added; their net new supplies are zero. The non-government sector can be disaggregated, for example into households, non-financial businesses, various financial intermediaries, and the rest of the world. The model is the natural framework for giving life to the Flow of Funds statistics (published in the United States by the Federal Reserve).

The standard short-run comparative statics experiments can be run. For the most part, the qualitative conclusions of IS/LM stand up, but other questions can be asked and answered. Likewise the model's long-run, steady-state properties can be investigated. For details see Tobin (1982) and B. Friedman (1980).

3. "Microfoundations" and parables

In the last twenty years the traditional macroeconomics of government finance has come under heavy theoretical and empirical fire, especially within the profession. The attackers begin with the strong prior belief that strictly financial policies and institutions cannot have real consequences. Some enter-
tain the hypothesis that even the government's real purchases and activities make no difference, because they simply displace private purchases and activities for which they are perfect substitutes. Otherwise, the government can alter real outcomes by doing real things, including the incentives and disincentives of its taxes, subsidies, and transfers. Given these characteristics, the real economy will be the same whether government outlays are financed by taxing or by borrowing or by printing money.

The philosophy underlying this approach is that government financial policy can do nothing that individuals could not have done on their own. There is no role for government as an intermediary between parties who could not have interacted otherwise; or as a creator of missing markets; or as a risk-pooler across agents subject to different sources of uncertainty; or as an agent with superior information utilizing this advantage to influence the economy in beneficial ways or simply conveying this information to others. Indeed, in most models aspects such as these are regarded as unnecessary complications or as arbitrary frictions. In subsequent sections we review this literature and its critics, both theoretical contributions and empirical findings.

Many contemporary theorists adhere to classical propositions more loyally than their original proponents: money is a veil, which may confuse unwary and myopic observers but does not affect the reality behind. Competitive markets work for the best, allocating resources among alternative uses, among different households and agents, among present and future times, and among possible contingent states of nature. Financial variables, nominal magnitudes, do not enter anyone's utility functions, resource constraints, or production technologies. Government financial interventions in markets can be undone by private agents, and optimizing agents surely will undo any real consequences. The neutrality of real outcomes with respect to nominal and financial disturbances and policies is the central message. The burden of proof, in the contemporary intellectual climate, is on those who would question these renascent classical propositions.

This trend is one facet of the New Classical Macroeconomics, which assumes continuous market-clearing and rational expectations; denies the existence of Keynesian unemployment; denies both the efficacy of and the need for policies of demand stabilization; and explains business fluctuations as optimal inter-temporal substitutions in production and consumption.

Much recent literature expounds, tests, and criticizes the newly popular neutrality propositions. By and large, the exponents and the critics share the same methodologies. These involve models with explicit "microfoundations"; that is, they base economy-wide general equilibrium results on the optimizing behaviors of agents interacting in specified markets. However, critics also incorporate institutional constraints, missing markets, market imperfections, informational asymmetries, and externalities. The neutrality of money, the
irrelevance of government finance, the ineffectiveness of macro policies, and the optimality of market outcomes are propositions that flow easily from and only from models that approximate or mimic Arrow–Debreu general equilibrium specifications. In these models, commodities are defined not just by their physical characteristics but by the times and contingencies of their deliveries. The microfoundations, the deep parameters, are the preferences of individuals, their endowments of commodities, and the technologies that transform commodity inputs into outputs. When government activities and policies are introduced, the question is whether any of these “deep” parameters are altered. If not, if, in particular, agents’ opportunity sets remain the same, the conclusion is that the interventions under study make no real difference. Otherwise, the interventions matter. Many papers, some of which we review below, are variations on this theme.

The common methodology makes this literature, even more than most economic theory, a collection, indeed a battle, of parables. These fables are quite abstract and economically primitive. Each is designed to fit some “stylized” facts, particularly qualitative generalizations from macroeconomic time series. Each is intended to have a moral relevant to some realistic institutional observation of the world. The methodology limits the scope and realism of the parables; it is very difficult to draw any “big picture” inferences from this literature.

Walrasian and Arrow–Debreu models, in their full generality, yield important theorems concerning the existence and optimality of competitive market-clearing equilibria. They are silent on the comparative-static effects of exogenous shocks and interventions on individual and aggregate outcomes, even on the signs of those effects. The permissible scope of tastes, endowments, and technologies, especially of their heterogeneity among agents, is too great. Yet those effects are the central agenda of macroeconomics, which has always relied on simplifications and specializations of general equilibrium theory.

How can short cuts be made while adhering to microfoundations methodology? One way is to model the whole economy as a single “representative agent”, a Robinson Crusoe. Then society’s economic choices are those of a single optimizer, whose tastes and opportunity sets are just microcosmic versions of those of the whole society. It is easy to derive the basic propositions of the “Modigliani–Miller” theorem for government – neutrality, irrelevance, ineffectiveness – from models of this kind. It is hard to see why Robinson Crusoe has a government, or why that projection of himself has different objectives from those of private citizen Crusoe.

The “representative agent” model has other implausible implications: there are plenty of markets and market prices, but no transactions take place. There are, in particular, zero “inside” assets and debts. The problems of coordination
emphasized by Keynes and other macroeconomists – between investors and savers, borrowers and lenders, capitalists and workers – are finessed. However, recent literature begins to recognize the shortcomings of this approach and to introduce two or more kinds of agents: rich and poor, liquidity-constrained and liquidity-unconstrained, even-period and odd-period bank customers.

The most usual differentiation among agents in current models is by birth date. The overlapping generations model with two-period lifetimes is heavily used, itself a parable with many instructive morals. Everyone is essentially the same, but in each period old and young coexist and trade. Government has a potential role as an immortal institution somehow embodying a compact among the generations. In contrast, a single representative agent model must make Crusoe himself ageless and immortal if it is to handle the intertemporal allocations – saving and capital accumulation – now properly regarded as central to macroeconomics.

On these allocations there is in principle a big difference between agents with finite horizons (because of mortality, generational selfishness, illiquidity, and/or myopia) and agents with infinite horizons (because of immortality, generational altruism, liquidity, and/or foresight). For the issues of theory and policy at stake, this difference is crucial. Infinite-horizon agents in steady states will typically accumulate wealth to the point at which assets yield returns equal to the agents' constant marginal rates of intertemporal substitution in consumption. Their demand for wealth will be infinitely elastic with respect to asset yields. Finite-horizon agents, on the other hand, will have finite demands for wealth at any rate of return; their savings may or may not respond positively to asset returns, but they will not respond infinitely. The rate of interest is determined by “thrift” and (marginal) “productivity” in the finite-horizon case, but by thrift alone in the infinite-horizon case.

One important issue to which the distinction is relevant is that of “superneutrality” in an economy with fiat money. Fiat currency has no intrinsic utility in consumption or production. And – this is crucial – its own yield is exogenously constant, usually zero. A dollar greenback today is a dollar tomorrow and tomorrow. Imagine that the stocks of fiat money and of promises to pay fiat money in future, and indeed the values of all predetermined and exogenous present and future nominal variables, are known. Suppose that all those values are scaled up or down in a common proportion. Neutrality means that no real variables – quantities or relative prices or interest rates – are different, while all nominal prices are scaled up or down in that same proportion. The common sense of this conclusion is that no one cares whether the unit of account is a dollar or a dime, and the contrived mental experiment is equivalent to a units change. The theoretical sense is simply that quantities of fiat money and other nominal variables do not enter anyone’s utility or production functions.
Superneutrality, on the other hand, relates to a different experiment. All the nominal quantities are scaled up or down, not by a common factor but in proportion to their distance in time from the present, i.e. by \( \exp(\pi t) \). It is not obvious that the paths of real variables are invariant to \( \pi \), which can be identified with the rate of inflation. Such invariance is necessary and sufficient for superneutrality. But it certainly cannot be deduced simply from the absence of money illusion, the common sense of neutrality. Clearly, at least one real, relative price is immediately altered by the experiment; the real rate of interest on fiat money depends inversely on \( \pi \). This alteration is bound to have repercussions throughout the system.

Superneutrality is usually defined, most precisely in relation to Solow-type aggregate growth models, as invariance of paths of capital stock and consumption with respect to monetary growth and price inflation. This invariance can arise from infinite elasticities of desired wealth with respect to yields on saving, as described above [Sidrauski (1967)]. Independently of how much capital they accumulate, these consumers will hold quantities of money, in real purchasing power, such that its real return too equals their constant supply-price of saving. The real return on money has two parts, the negative of the inflation rate and the implicit service return of money as means of payment; the latter depends inversely on the size of average real balances held. In Tobin (1965a, 1968, 1982), however, consumer-savers have finite horizons and finite demands for wealth, money and capital compete for room in their portfolios, and higher inflation increases the relative attractiveness of capital. This “Tobin effect” tends to raise capital stock and consumption. Against it is the resource cost (sometimes called “shoeleather cost”) of handling transactions with smaller real cash balances, which tends to reduce real incomes, consumption, and desired wealth [see Tobin (1986b)].

4. Financing deficits with non-monetary debt

This section addresses the question of whether a shift from tax finance to bond\(^2\) finance of given real government expenditures on goods and services with unchanged monetary policy absorbs any private saving. If it does not, “debt neutrality” prevails, i.e. investment, aggregate demand, realized income, employment, real interest rates, inflation, and the long-run capital intensity of the economy are not influenced by the choice between tax- and debt-financing. After a brief discussion of crowding out and of the interest sensitivity of saving,

\(^2\)For simplicity, we use the term “bond” to refer to all non-monetary government debt of whatever maturity and form. We are not distinguishing bonds from bills or notes.
we investigate the theoretical conditions for debt neutrality. Then we present the implications of relaxing each of these conditions. We conclude with a review of available empirical tests of the debt neutrality proposition.

4.1. Crowding out and the interest elasticity of saving

A bond-financed decrease in taxation may lead to "crowding out" of two kinds: lower investment expenditure flows; and a lower desired ratio of the capital stock to real GNP in a long-run steady state. There are two main sources of stock and flow crowding out: supply constraints and monetary constraints. In standard discussions, these cause interest rates to rise, thus lowering both investment expenditure and the desired capital stock. For a given tax cut, the extent of crowding out would be smaller, the larger the interest-sensitivity of desired private saving and of the desired wealth-to-income ratio. These are, of course, standard results in traditional IS/LM and growth models.

A recently popular framework for exhibiting interest sensitivity is the two-period life-cycle model in which a representative agent chooses consumption in each period so as to maximize his utility $U(C_1, C_2)$ subject to the constraint that the present discounted value of lifetime consumption is equal to the labor income received in the first period. In this model, saving responds positively (negatively) to the interest rate if the elasticity of substitution between present and future consumption is greater than (less than) one. For a "Cobb–Douglas" utility function, the elasticity of substitution is unity and saving is interest insensitive. The convenience of the Cobb–Douglas specification, the dependence of the sign of response on the choice of utility function, and the failure of a significant number of empirical studies to identify interest-rate effects on desired saving, have created a presumption in favor of postulating interest insensitivity.

The validity of this postulate has been challenged by Boskin (1978) and Summers (1981, 1982). Summers argues that the choice of a Cobb–Douglas utility function does not allow the marginal propensity to consume to depend on after-tax interest rates, and that the use of a two-period model in which all income is received in the first period obscures the negative effects of increases

---

1In an open economy, international capital mobility causes foreign capital to flow into the country experiencing the tax cut, resulting in an appreciation of its currency, a fall in net exports, and hence a total or partial displacement of foreign investment by the bond-financed budget deficit.

2The life-cycle model is due to Fisher (1930) and to Modigliani and Brumberg (1954). An extensive discussion of the two-period consumption decision can be found in Feldstein and Tsiang (1968).
in interest rates on the present discounted value of lifetime labor income, human wealth. He incorporates both effects by considering a multi-period life-cycle model with an additively separable, constant-elasticity-of-substitution utility function and the constraint that the present value of lifetime consumption equal the sum of assets and of human wealth. Summers reports simulations showing that consumption responds more negatively to the rate of return when the elasticity of substitution is greater. Permanent changes in the after-tax interest rate reduce consumption by more or increase it by less than do transitory changes.

Among empirical studies, Boskin's estimate of an interest elasticity of U.S. saving of 0.4 is widely regarded as too high, and has been criticized by Howrey and Hymans (1978) as being sensitive to the choice of sample period and to arbitrary details of data construction. Summers criticizes statistical estimates of consumption functions for using current disposable income as a proxy for human wealth, for errors in measuring interest rates and expected inflation, and for simultaneity bias. His own estimates, based on first-order conditions of an agent's intertemporal optimization problem, also suggest that the elasticity is quite high. These estimates are derived from an optimization model, and the overidentifying restrictions imposed by the first-order condition are often rejected by the data. Whether the interest elasticity of saving is positive is still an open question, especially in view of the stylized fact of a trendless U.S. wealth-to-GNP ratio over long periods of time despite the upward trend in real rates.

Whether desired saving and wealth-to-income ratios are interest sensitive or not, life-cycle models do not imply infinite interest elasticities. When they are amended to incorporate people's concern for their descendants by allowing the size of bequests to enter the utility function, they still do not imply infinite planning horizons, in contrast to the alternative assumption that the utilities of descendants are what matters. The importance of bequests [as documented, for example, by Kotlikoff and Summers (1981), and Mirer (1979)], is not necessarily inconsistent with crowding out. The choice of how to incorporate the bequest motive is a matter of empirical plausibility.

4.2. The theory of debt neutrality and the optimum debt level

4.2.1. Debt neutrality

In 1974 Robert Barro presented the case for debt neutrality. According to this doctrine, also known as the "Ricardian neutrality" or "Ricardian equivalence" or "tax discounting" hypothesis, a bond-financed tax decrease does not affect agents' consumption demand, despite its effect on contemporaneous disposable income. When the economy is in a short-run, full-employment equilibrium,
debt finance affects neither the path of prices nor that of interest rates. Finally, capital intensity in a long-run steady state is similarly unaffected. Clearly, one of the strongest implications of this doctrine is that bond-financed tax decreases and the associated deficits do not lead to any crowding out of private capital formation. Similar conclusions apply to variations of social security taxes, which transfer income between the young and the old.

A striking insight of the theory is that finite lives do not necessarily imply finite planning horizons: if the utility of every generation enters that of its predecessor, every agent incorporates the utility and budget constraints of the whole dynasty of successors into the decision concerning consumption and bequest levels.

A second key element in standard formulations of the theory is the assumption that government will not be able to service the debt by issuing more bonds forever, and that it will, therefore, eventually have to raise taxes. When this is so, it is inappropriate to analyze the effects of the bond-financed tax cuts without considering the implications of the associated future tax liabilities. According to Barro, taxpayers will not alter consumption in response to changes in tax law because they regard them merely as changes in the timing of present and future tax liabilities of unaltered total present value.

While both issues will be dealt with extensively below, it is instructive to include a few comments on the latter one here. Practical macroeconomists and macroeconometricians have long factored into their policy analyses and forecasts consumers’ and investors’ anticipations of tax changes actually scheduled or under serious political and legislative consideration. An implication of life-cycle or permanent income theories of consumption is that current consumption will respond less to tax changes consumers regard as temporary than to those they regard as permanent. Financial market variables also have been used as indicators of expectations. If crowding out is expected to occur through higher interest rates in future, long-term rates should stand above short rates in the present. If deficits are expected to lead to inflation in future, this too should increase nominal longer-term interest rates.

“Ricardian equivalence” raises two interrelated questions. One is what constraint the government actually is under, or will behave as if it is under. The second is what private taxpayers and other agents believe to be the constraints and policies of the government. What degree of rationality must individuals possess in order to know not only how the economy works under given government policies, but also how policy decisions of the entire sequence of

\[\text{This sensible idea had been noted by Ricardo and by others before Barro, but it is not sufficient by itself to establish the equivalence of taxation and of bond financing. Indeed, Ricardo in 1817 argued against the practice of not repaying government debt accumulated during wars. His reason was that either people would be fooled into thinking that they are richer or they would try to shift the burden onto others, possibly by emigrating. In either of these two cases, their current real consumption would go up [see Ricardo (1951, esp. pp. 244-249)].}\]
future governments are arrived at? In the absence of any schedule for increasing taxes and in a political and ideological climate against taxation, would a citizen of the United States in the 1980s rationally expect future fiscal corrections?

The first question is easier to tackle analytically. Suppose that the interest rate exceeds the economy's growth rate. Then, if a limitless increase in the public debt relative to national product is to be ruled out, the present value of future tax revenues – discounting by the difference between interest and growth rates – must exceed the similarly discounted present value of expenditures by the amount of the current debt. This can be shown through recursive substitutions into the single-period government budget identity, and it is usually termed the “intertemporal government budget constraint”. Although this constraint is incorporated in all standard discussions of debt neutrality, it is also possible for the real after-tax interest rate on bonds to be lower than the rate of growth of real GNP. In this case, the debt-to-GNP ratio is stable and the government can continuously issue new bonds to finance interest payments without ever having to increase taxes.

The argument usually employed in support of debt neutrality is that the condition for a stable steady-state debt-to-GNP ratio conflicts with the Phelps “golden rule” condition for dynamic efficiency. According to that condition, an economy with \( g > R \) is overcapitalized in the sense that disinvestment could make consumption per capita higher in at least one year without lowering it in any year.

However, the efficiency condition refers to the net marginal productivity of capital, \( R \). There are two respects in which \( R \) is not the same as \( r \), the net interest rate on government debt. One is that the relevant marginal contribution of capital to social product is pre-tax, while the cost of debt service is after-tax. Since the Treasury cannot in fact rely on lump-sum taxes, the after-tax rate is lower. The second is that the marginal product of capital exceeds the interest rate on government debt by a risk premium. In the United States safe debt interest rates have generally been below the economy's growth rate, while mean returns on capital have been above. This is indicated by the fact that gross profits, \( RK \), chronically exceed gross investment, \( gK \) [Abel et al. (1986)]. It is, therefore, possible to meet the debt stability condition without violating the dynamic efficiency condition for the economy. However, Ricardo Equivalence depends on the debt interest rate's exceeding the growth rate, and on the public's believing correctly that the government will not allow the ratio of its debt to the economy's output and income to increase without limit.

We now turn to the debt neutrality theorem. Tobin (1980) listed the restrictive conditions necessary for it to hold. A complete set of such conditions is the following:
First, agents are either immortal or, if finitely lived, linked by an unbroken infinite dynastic chain of intergenerational gifts and bequests. This condition is needed for agents to have infinite planning horizons and to be able and willing to undo any government financial policy that would redistribute real income intertemporally.

Second, capital markets are perfect, without liquidity constraints or credit rationing. Thus, private agents can lend and borrow on the same terms as the government, which has no role to play as a financial intermediary.

Third, all taxes, transfers, and subsidies are lump sum. This ensures that their effects on present and future agents are independent of agents' behavior. Hence agents can neutralize them only by compensating changes in their own intergenerational transfers.

Fourth, debt servicing through issuing more debt forever, is infeasible. Future tax increases, equal in present value to today's tax cuts, are required. Therefore, given rational expectations, these increases are foreseen by all living agents. This condition has been discussed above.

Formal proofs of the debt neutrality proposition usually employ an overlapping generations model where the utility of each generation is a function of its own consumptions in the two periods of its life, as well as of the maximum attainable utility level of the subsequent generation (and possibly of the immediately preceding one). It is shown that the effect of a bond-financed tax cut on the budget constraints of each generation can be offset by an appropriate change in voluntary intergenerational transfers. Thus, the maximum attainable utility levels of all generations are unaffected by the deficit, and optimal consumption levels do not change.

The intuitive reason for bond neutrality is that under the set of conditions presented above, the introduction of government bonds does not provide agents with opportunities they did not have under tax financing of government spending. Government bonds are simply a means whereby the current generation can undertake expenditure for which future generations will have to pay through increased taxes. If the optimal sizes of bequests both prior to and after the deficit are interior, then the current generation already had the option of transferring wealth from subsequent generations to itself by reducing the size of its bequest, but did not choose to do so. A similar logic applies to intergenerational gifts, and to retirement of outstanding debt by new taxes.

4.2.2. Tax smoothing and optimal debt

Suppose now that all the conditions necessary for debt neutrality do hold. If the choice between bond- and tax-financing does not entail any of the usually
assumed costs (such as shifting the burden of the debt on to future generations and crowding out capital), then what determines the optimal path of government debt for a given path of government expenditures?

In his 1979 paper, Barro considers an economy in which debt neutrality holds, the paths of government spending and of real GNP are given, and (consistent with debt neutrality) the real rate of return on bonds relative to that on private debt is not affected by the amount of government debt outstanding. He assumes that perpetual bond finance is not possible. For a given initial bond stock and present value of government spending, the intertemporal government budget constraint determines the present value of tax revenues. The optimal type and timing of taxation remain to be determined with reference to their social costs.

Having ruled out most of the traditionally assumed costs through the postulate of debt neutrality, Barro postulates costs associated with the collection of government revenues. These costs are assumed to be a function of contemporaneous tax revenue and tax base. Abstracting from the choice of tax composition, the optimal timing of taxes is that which minimizes the present value of the tax collection costs for the present value of taxes dictated by the intertemporal government budget constraint. The solution can be shown to imply a constant (planned) average tax rate. This constancy, along with the given present value of taxes, determines the optimal level of tax revenues at each point in time, and accordingly the path of the government’s bonded debt.

Barro shows that if the perceived duration of transitory changes in government spending and in income is constant, a temporary increase in government expenditure should have a positive (but less than one-for-one) effect on the current rate of growth of debt; whereas a temporary increase in the rate of growth of real income should have a negative effect. The debt-to-income ratio would be expected to remain constant on average, but would rise in periods of abnormally low income or high government spending. Expected inflation raises the growth rate of nominal debt by an amount equal to the inflation rate, because the optimal level of taxes is the same as under zero expected inflation, the real rate on bonds is assumed to be unaffected by changes in expected inflation, and the increase in nominal interest payments has to be financed through bonds.

4.3. The conditions for debt neutrality

We now investigate how dependent debt neutrality is on its assumptions by relaxing them one by one.
4.3.1. Infinite horizons and intergenerational transfers

Blanchard (1985) relaxes the assumption of infinitely-lived dynasties. He assumes that each agent is faced with a probability $p$ of dying, constant through the agent's life. Here an agent can be interpreted either as an individual or as a dynasty with probability $p$ of dying out.\textsuperscript{7} Insurance markets costlessly eliminate the risk of leaving unanticipated bequests. Agents contract to pay their whole wealth, $w$, to the insurance company when they die in exchange for receiving $pw$ in every period of life. Negative bequests are prohibited.

Agents are assumed to maximize expected utility of consumption over an infinite horizon with uncertainty only as to the time of death. When the instantaneous utility function is isoelastic,\textsuperscript{7} aggregate consumption is a linear function of the sum of aggregate human and non-human wealth. Human wealth is defined as the present value of future after-tax labor income accruing to those currently alive. Under the simplest distributional assumption, that after-tax labor income is equal for all agents at all times, and given the constant probability of death, all agents have the same human wealth. Assume that the number of people who die at each instant is equal to the size of the new cohort born at that instant, so that population is constant. Then\textsuperscript{8} aggregate human wealth accumulates at a rate equal to $(r + p)$. The intuitive reason for this key result is that agents discount income available tomorrow relative to income available today not only by the rate of return they could earn on it if it were available today, but also by the probability that they will not be alive so as to receive it.

Now consider a decrease in lump-sum taxes today, accompanied by an increase in taxes of equal present value $T$ periods from now. This present value is calculated using interest rates $r$ faced by the immortal government. Thus, the reallocation of taxes raises human wealth, as calculated by private agents, by an amount equal to the current tax cut times the probability that someone

\textsuperscript{7}The assumed invariance of the probability of death with age ensures that the propensity to consume out of wealth is the same for people of all ages, despite the fact that their wealth levels may differ. This allows the derivation of an aggregate consumption function for general population structures. One drawback is that it does not capture the varying behavior of agents throughout their lives.

\textsuperscript{7}An isoelastic utility function in consumption $c$ is of the form.

$$u(c) = (c^{1-\sigma})/(1-\sigma), \quad \sigma \neq 1, \quad u(c) = \log c, \quad \sigma = 1$$

The (constant) elasticity of substitution is equal to $1/\sigma$.

\textsuperscript{8}If aggregate human wealth is denoted by $H$, income by $Y$, and non-human wealth by $W$, then in the logarithmic utility case,

$$C = (p + \theta)(H + W), \quad H = (r + p)(H - Y), \quad W = rW + Y - C.$$
currently alive will have died by the time the future tax increase takes place. The longer taxes are deferred, the larger the effect on human wealth. Given that aggregate consumption is a function of human plus non-human wealth, this reallocation of taxes raises consumption. Debt neutrality fails.

One interpretation of this paper is that debt neutrality fails when agents have finite horizons, since the expected lifetime of an agent at each point in time is $1/p$. As $p$ goes to zero, neutrality is approached as a limiting case. An alternative interpretation is that agents represent dynasties with infinite horizons but with a probability that they end because of exogenous events. It is also possible, however, that the chain of intergenerational transfers is broken because some members decide against further transfers.9

Drazen (1978) shows that a key element in determining whether dynasty members decide to break the chain or not is the weight their utility functions attach to the utility of other generations. Consider a utility function for generation $i$, $U_i = U_i(c_i^t, c_i^e, U_{i+1}^*)$, where $c_i^t$ and $c_i^e$ are real consumption levels of generation $i$ in each of its periods of life and $U_{i+1}^*$ is the maximum attainable utility of the next generation. Assume that utility functions do not differ between generations and that real wage income is constant. If bequests are invested at the time they are put aside by the parent, the extra utility that a parent receives from one dollar of consumption when old is exceeded by the utility of his descendant from consumption of $1 + r$ dollars of bequest. If the father weighs the descendant's utility equally to his (i.e. if $\partial U_i / \partial U_{i+1}^* = 1$), then bequests will be positive. If he discounts his descendant's utility by a rate equal to (larger than) the market rate of interest, optimal bequests will be zero (negative). If negative bequests (i.e. indebting your children for your own current consumption) are not allowed, agents will be at a corner solution of zero bequests, and the issue of government bonds will have real effects by allowing agents to shift the burden of their current spending to future generations.10

Finally, concern about both parents and descendants can be formalized by postulating that $U_i = U_i(U_{i+1}^*, c_i^t, c_i^e, U_{i+1}^*)$. Then a bequest that raises the descendant's utility also raises the utility of the father. But in doing so, it raises the utility of the descendant further. For bequests not to lead to infinite utility,

9Blanchard also shows that for a logarithmic utility function, the steady state $r$ is between $\theta$ and $\theta + p$, and is an increasing function of $\theta$. The finiteness of horizons discourages capital accumulation, since $r = \theta$ for infinite horizons. If labor income declines through life, this tends to raise the steady-state capital stock, with an ambiguous net effect and the possibility that the steady-state $r$ will be negative (i.e. below the zero natural rate of growth of real GNP). In this case the level of the capital stock exceeds the "golden rule" level and the economy is dynamically inefficient, as in Diamond (1965). Finally, for the class of isoelastic utility functions the lower the elasticity of substitution, $1/\sigma$, the lower the steady-state capital stock.

10Considerations such as wage growth, taxation of interest, and population growth (in certain cases) further reduce the likelihood of positive bequests.
at least one generation must discount the utility of the other. If a generation weighs the utility of both adjacent generations sufficiently less than its own, then neither the gift nor the bequest motives will be operative and the introduction of government bonds will have real effects.\footnote{The discussion above assumes that bequests are in the form of non-human capital. Drazen argues that a significant share of bequests is in the form of investment in human capital, namely expenditure on the education of descendants. At least up to a certain level of education, such investment yields a higher rate of return than that on non-human wealth and could be used to enhance the father's own second-period consumption. When the father cannot enforce this liability on his descendant, the introduction of government bonds facilitates the transfer and thus has real effects. The enforceability of loans between parents and offspring would mitigate the importance of such considerations.}

Inoperative intergenerational links may exist for other reasons, for example that neither all members of a given generation nor all members of the same dynasty across generations are equal in natural endowments. Laitner (1979) postulates that there is an entire distribution of labor incomes in each generation. He assumes, somewhat artificially, that there is no correlation between the labor income of a family at a point in time and the incomes of its descendants. He then shows that a bond-financed tax cut that is repaid through tax increases $T$ periods from now can affect consumption. This is because the government can always pick a finite $T$ so large that some descendants within $0 < t < T$ have zero bequests with positive probability. Similarly, if the government makes transfers to and later levies taxes on all families (or a random cross-section of them), consumption will rise in a positive percentage of cases even with one-period bonds. This is because some families in each generation (or some generations of the same dynasty) cannot afford to leave bequests.

Not only is the ability of all future dynasty members to leave bequests uncertain, but also one's own ability to do so may not be foreseen at the beginning of one's career. Feldstein (1988) introduces uncertainty as to income in the second half of a parent's working life. Despite the presence of lump-sum taxes and the assumption that bequests are solely motivated by intergenerational altruism, current consumption rises as a result of debt finance (including the introduction of an unfunded social security system). The logic is that if the parent were sure that he would not leave a bequest, any increase in current disposable income arising from a tax cut would be divided between the two periods of his life. Since zero-bequest circumstances have positive probability, the parent will raise his current consumption to some extent. Even when the parent, blessed with good luck late in life, actually makes a bequest, the consumption of the second generation is reduced as a result of the tax cut the parent earlier enjoyed.

Although many of the papers on debt neutrality emphasize intergenerational transfers, most of them do not do full justice to human reproductive biology. Notable exceptions are two papers by Bernheim and Bagwell (1988) and Abel...
and Bernheim (1986). Bernheim and Bagwell (1988) demonstrate that in a properly specified dynastic framework, where everybody is "altruistic" towards members of the same dynasty and where the size of bequests per se is not a source of utility, redistributions of wealth leave everybody's consumption and resource allocation unaffected, prices play no role in resource allocation, and apparently distortionary taxes do not induce any change in individual behavior. Since these implications of the dynastic framework seem completely unrealistic, Bernheim and Bagwell argue that this framework should not be trusted for policy analyses. Debt neutrality is attacked through a reductio ad absurdum.

Specifically, Bernheim and Bagwell point out that when two individuals belonging to different families marry and have children, concern for common grandchildren links their two original families. The fact that these grandchildren also get married and have children means that more families of the current generation are linked by their common concern for those distant descendants, etc. Once it is established that two families of the same generation are linked, we can extend the chain further by moving up or down the family tree. Moreover, dynasties will typically be linked through multiple channels. Everybody is a part of every dynasty.

The authors assume that in each period $t$ there is a chain of operative linkages connecting any two living individuals, where each link consists of a transfer made sometime between periods $t$ and $t + T$, with $T$ an integer. They show that for each sufficiently small perturbation of deficit and tax policies, there exists an equilibrium in which factor prices, labor supplies, consumption, and purchases of physical capital are unaffected by the perturbation. The perturbation simply leads to offsetting private transfers and bond purchases. This is true not only for government financial perturbations, but for all exogenous shocks; for example in the natural endowments of different agents. These results survive the introduction of various types of uncertainty and informational asymmetries.

Can a dynastic framework with frictions generate plausible conclusions? Abel and Bernheim (1986) introduce frictions such as (a) the derivation of pleasure directly from the act of giving; (b) the existence of both selfish and altruistic parents combined with incomplete information about others' preferences; and (c) social norms dictating that parents should divide transfers equally among all their children. Their overall conclusion is that frictions tend to make redistributional policies non-neutral, but they lead to paradoxes of other kinds.

As an example, consider case (c). Suppose that no parent is selfish; that the size of transfers is not a direct source of utility; and that parent $t$ receives a transfer. In response to it, he raises his transfers to both his children by the same amount. Now consider the two sets of parents of the spouses of those children. They observe that one of their children is better off and, as a result, increase their consumption and lower their transfer. Because of the social
norm, however, this reduction of transfers applies to their other children as well. In turn, the families of the spouses of those other children raise their transfers, etc.

Thus, although exact Ricardian equivalence with respect to policies that affect children equally is maintained here, redistribution policies that affect children differentially are not neutralized by the parents due to egalitarian constraints. It would seem that the Berneim–Bagwell paradox has been resolved and the Barro theorem rescued from the reductio ad absurdum. The disturbing feature of this setup, however, is that an exogeneous increase in the wealth of any given individual is never Pareto improving, i.e. always makes some people worse off. In view of this paradoxical implication, and of similar implications for other frictions, Abel and Bernheim conclude that such frictions cannot provide Ricardian equivalence with a plausible theoretical foundation.

Gifts and bequests can be viewed not only as manifestations of altruism but also as assets transferring consumption from one period of life to the other. Buien (1979, 1980) and Carmichael (1982) suggest that, under certain conditions, government bonds are perfect substitutes for gifts and bequests in this role. For example, when gifts to parents are reduced, agents can increase their first-period consumption since they offer less to their parents, and lower their second-period consumption since they receive less from their children. Thus, a bond-financed tax cut that substitutes second- for first-period consumption may be neutralized through an appropriate general adjustment of gift and bequest levels for all generations. This result does not require that taxes be eventually raised to service the debt.\(^\text{12}\)

4.3.2. Perfect capital markets?

A second condition for debt neutrality is that there are no capital market imperfections leading to credit rationing. There is now a substantial volume of research on the sources and implications of credit rationing, starting with the seminal paper of Weiss and Stiglitz (1981). As an example, suppose that because certain individuals have relatively bad collateral, they can only borrow at a rate \(r_H\) which is higher than that for others, \(r_L\). As recognized by Barro (1974), when a government bond is bought by a low-discount-rate individual and the proceeds and taxes associated with it are both distributed in the same way among the two classes of individuals, the bond is in effect a loan from the

\(^{12}\)An exchange with Burbidge [Burbidge (1983, 1984) and Buien and Carmichael (1984)] shows that in order for this stronger neutrality proposition to hold, the same type of intergenerational transfer has to be operative both before and after the tax cut, a condition which is difficult to meet for large increases in the amount of debt introduced, since these are likely to induce movement to corners with respect to transfers. When this happens, bonds are not neutral, as argued above. By contrast, Burbidge’s formulation yields neutrality even for large additions to the stock of bonds, provided that gifts or bequests exist before the tax cut and that future tax increases are necessary.
low-discount-rate to the high-discount-rate individuals. The net wealth of the low-discount group is unaffected, but that of the other group rises since the rate \( r_H \) by which they discount the future tax liabilities is higher than the rate \( r_L \) by which those who hold the bonds discount the stream of coupon payments. Thus, one would expect the consumption of the high-discount group to go up in response to the increase in their net wealth.

This idea is imbedded in a formal model of credit rationing by Webb (1981). Webb argues that the government's ability to enforce tax repayments is superior to that of private lenders. There may be a higher default penalty for taxes than for private debts, and tax withholding ensures that the government receives payment before any other payments are made. The rate at which the government borrows and lends is therefore below that offered to individual borrowers in the private capital market. A substitution of government for private debt raises agents' net wealth due to this difference in interest rates. It may even be sufficiently large to bring the total volume of private debt below the critical level required to eliminate the occurrence of default in the private market. Debt neutrality fails.

The relative quantitative importance of finite horizons and of capital market imperfections is addressed in two recent papers. Poterba and Summers (1986) employ a life-cycle simulation model to consider the effects of debt-financed transfers of one dollar to each living person for \( K \) years. It is assumed that the debt is never repaid, but beginning in period \( K + 1 \), the government levies lump-sum taxes on working individuals to meet its interest payments and maintain a target real debt stock, aggregate or per capita. Simulations for a range of parameter values suggest that although deficit policies may transfer substantial tax burdens to future generations, they have only trivial short-run effects on consumption and saving. The intuitive reason is that for all but the oldest consumers, marginal propensities to consume out of wealth are quite small. This conclusion is only strengthened by considering realistic debt repayment periods. When the effects of current deficits are simulated on the basis of alternative scenarios for future deficits, even dramatic changes in deficit paths have only minor effects on consumption and saving.

These results suggest that in the absence of liquidity constraints and myopia, the distinction between overlapping generations and infinite-horizon models may be of little practical importance in evaluating short-run (but not necessarily long-run) effects of deficit policies.

Hubbard and Judd (1986a, 1986b) provide additional support for emphasizing liquidity constraints rather than finite horizons. In one illustration they consider agents with logarithmic utility functions and constant probabilities of death and of experiencing a wage increase from \( w_1 \) to \( w_2 \). With no capital market imperfections and for plausible parameter values, numerical simulations indicate that the marginal propensity to consume (MPC) out of a five-year tax cut financed with a twenty-year delay, during which time people
die at a rate of 2 percent per year, would only be about 0.05. Finite horizons per se are not sufficient to generate sizeable effects on consumption. By contrast, the MPC is more than quadrupled when 20 percent of the work force are liquidity constrained and consume all their wages.\(^1\)

4.3.3. **Lump-sum taxation?**

The third condition for debt neutrality is that taxes and transfers are lump-sum. It is well known that governments cannot typically rely on such taxes for the bulk of their revenues. It is thus appropriate to check the robustness of the theorem under distortionary taxation. When taxes are not lump-sum, the behavior of descendants can be adjusted so as to reduce the impact of tax increases, and this will be taken into account by the current generation experiencing the tax cut. Specifically, there is no reason for the parents to raise their bequest by the entire present value of future taxes implied by the current budget deficit. They will raise their consumption, knowing that their children will manage to reduce their future tax liabilities.

There has been some research on non-lump-sum taxes when the tax base is uncertain. When the increase in expected future taxes is combined with a current tax cut equal to its present value, the present value of total expected tax payments is unaffected. However, an increase in future income taxation reduces uncertainty surrounding disposable income. When income taxes are levied on descendants of the current generation, this logic implies that there will be downward pressure on bequests and upward pressure on current consumption, to the extent that the size of bequests is influenced by the parents' internalization of risk faced by their descendants. In fact, simulations by Barsky, Mankiw and Zeldes (1986) show that for plausible parameter values, the marginal propensity to consume out of a current tax cut associated with uncertain future tax liabilities is in the neighborhood of values implied by neo-Keynesian models that ignore the future tax liabilities.

4.3.4. **Limitless debt?**

All of the above conditions deal with whether a substitution of lower taxes and more bonds today for higher taxes in the future affects agents' opportunity sets. The question addressed in our introductory discussion of debt neutrality

\(^1\)A liquidity constraint may be regarded as a very short horizon, much shorter than a lifetime Hubbard and Judd also point to the relevance of the distribution of the tax cut for the effect on the aggregate MPC. In particular, if the tax cut is not uniform but results in greater relative relief for the high-income group, the effect of borrowing constraints on the aggregate MPC out of a temporary tax cut is dampened. This questions the practice of regarding the measured MPC as a good indicator of the proportion of liquidity-constrained individuals in economies with proportional or progressive tax systems.
above was whether bond-financed tax cuts do in fact necessitate future tax increases. This was the subject of the early debate between Barro (1976) and Feldstein (1976), who focused on whether the real after-tax interest rate on bonds is higher or lower than the rate of growth of real GNP.

Some recent research on whether real consumption is affected by tax cuts which are never paid for through higher future taxes has been conducted in a game-theoretic framework. An example is the paper by O’Connell and Zeldes (1987) who take up a suggestion by Gale (1983). Gale interprets Barro’s dynastic framework as a game played among successive generations. He shows that this game has a vast multiplicity of equilibria, leaving individual behavior indeterminate. O’Connell and Zeldes look at a refinement of this equilibrium set. They start with an overlapping generations model in which debt neutrality cannot be ruled out with respect to debt that is repaid through higher future taxes. They then ask whether current tax cuts would affect consumption if the government never increases taxes in the future. They find that the set of quasi-steady-state Nash equilibria (i.e. those involving unchanged consumption) is unaffected by this experiment. However, there also exist pairs of equilibria across which government financing produces real effects. The authors present an economy with bequests where debt neutrality fails. This is because under certain conditions dynasties behave as though they value terminal wealth, which in turn depends on the size of government debt. Per capita consumption rises, despite the fact that the set of equilibria is unaffected by financial policy.

4.4. Empirical tests of debt neutrality

There are two main approaches to empirical testing of debt neutrality. One consists of structural estimates of consumption or savings functions or of financial sector behavior, and of tests for the statistical significance of taxes, transfers, government debt, and (sometimes) social security wealth. The other is to utilize the assumption of rational expectations to derive cross-equation restrictions and jointly test debt neutrality along with rational expectations and the assumed model of behavior.

4.4.1. Structural estimation

The older approach is structural estimation. The main implications of the debt neutrality hypothesis for the consumption function are: (i) that for a given path of government purchases, variation of taxes has no effect, i.e. the coefficient on taxes in a consumption or saving function should be zero; (ii) since transfer payments are analogous to tax reductions, the same holds for the coefficient on
transfers; (iii) since a change in the stock of bonds does not generate a wealth
effect on consumption, its coefficient should also be zero;\(^{14}\) and (iv) if a proxy
for social security wealth is included, its coefficient should also be zero, since
according to Barro, households will save enough to compensate future gen-
erations for the extra tax burdens required to pay higher social security
benefits.

There are two main issues that differentiate papers in this strand. The first is
the extent to which they test and correct for simultaneity bias resulting from
possibly endogenous right-hand-side regressors, such as income, wealth, taxes,
transfers, and social security wealth. The second concerns the proxies used for
permanent income and the permanent (as opposed to transitory) levels of policy
variables.

Early papers estimating consumption or savings functions without allowing
for simultaneity bias include Barro (1978), Kochin (1974), and Tanner (1979),
who all find support for debt neutrality. On the other hand, Buiter and Tobin
(1978) reach the opposite conclusion.

Feldstein (1982) allows for endogeneity of tax revenues and of income by
using the lagged value of taxes and of income as instruments. He uses current
GNP and then current personal disposable income as proxies for the corre-
sonding permanent levels of these variables. Results do not favor debt
neutrality. Feldstein looks at U.S. data for 1930–77 with the war years 1941–46
omitted. When real per capita GNP is used in a regression with real per capita
consumer expenditure as the dependent variable, the hypothesis that the
coefficient on taxes is not negative can be rejected at approximately the 20
percent significance level, the coefficient on the debt variable does not support
the irrelevance of debt, and that on transfers is significant and positive.
However, the coefficient on social security wealth is smaller than its standard
error. Results are more strongly against debt neutrality when personal dispos-
able income per capita is used as a proxy for permanent income. Feldstein also
finds an insignificant coefficient on the government purchases variable suggest-
ing that government spending does not directly crowd out private consumption
by providing a substitute for it.

In a recent paper, Seater and Mariano (1985) estimate specifications similar
to those of Barro (1978) and Feldstein (1982) by two-stage least squares with
first-order serial correlation correction, in an attempt to remove simultaneity
bias. They argue that Feldstein’s use of only lagged taxes and income as
instruments may not be sufficient, as these instruments may still be correlated
with the error due to the high degree of serial correlation in the series for

\(^{14}\)If both the stock of bonds and that of total wealth (including bonds) are entered, then the
coefficient on the bond stock should be the negative of that on wealth, so that the total effect of
bonds is zero.
income and taxes. They use a variety of instrument sets and for each one they apply Hausman tests to determine which right-hand-side variables are endogenous. They then perform the second-stage regressions using fitted values for these variables which did not pass the Hausman test. Although the endogeneity of regressors is sensitive to the choice of instrument set, the conclusions from the second-stage regressions are fairly robust. For Barro’s specification, they find support for tax discounting. When using Feldstein’s specification, both taxes and transfers become significant for the periods 1931–40/1947–74, questioning tax discounting. However, in a specification which includes unemployment, taxes become insignificant while transfers remain significant.

In the main body of their paper, Seater and Mariano follow Barro (1983) in specifying consumption as a function of: (i) permanent income; (ii) the cost of government represented by its “permanent” real expenditure on goods and services \( G^p \); (iii) current real government expenditure on goods and services \( G \) to capture the possibility that increases in \( G \) for given \( G^p \) partially substitute for private consumption; and (iv) real after-tax interest rates since they induce intertemporal substitution. They estimate permanent income and permanent government expenditure only from these series’ own histories. The regressions they run include separately four “financing” variables, namely real tax collections, real transfers, the real market value of outstanding government debt, and social security wealth. The estimation technique is instrumental variables and the sample is 1929–75. The government financing variables, including transfers, are jointly and individually insignificant, in regressions both for total consumption and for consumption of non-durables.

Kormendi (1983) examines roughly the same period (1930–76) under the assumption that government consumption (defined as the portion of government spending that yields utility to the private sector in the current period) is substitutable for private consumption, in the limit perfectly. This is yet another equivalence, challenging the effectiveness even of variations in government expenditure on goods and services. He derives a private consumption function based on this “consolidated approach”. The counterpart of integrating private and government consumption is a “total disposable income” that includes the resource flow both from net private incomes and from taxes, since the latter effectively represent claims to government-provided goods and services. “Government” here means federal, state, and local government. Private consumption is a function of total real NNP, government consumption, wealth, and any discrepancy between the values of foregone private goods or services and those provided by government (which Kormendi calls “government dissipation”). By contrast, the standard permanent-income approach would make consumption a function of personal disposable income (defined as income net of taxes, corporate retained earnings, transfers, and government interest payments on its debt), but not of government consumption. Moreover, the market value of government bonds should not affect consumption under debt neutrality. Thus,
Kormendi obtains a set of testable restrictions, similar to Feldstein's, to differentiate the two approaches.

In his estimation, Kormendi uses current and lagged NNP as proxies for permanent NNP. Looking at differenced U.S. data between 1930 and 1976, he finds support for the "consolidated approach", with a fairly high implied substitutability between government and private consumption. With regard to government bonds, Kormendi's results show that there is less than 5 percent probability that future taxes implied by government debt are less than 90 percent discounted. However, the coefficient on transfers is significantly positive, a result which Kormendi attributes to redistribution of income among classes with different spending propensities.

Kormendi's strong results have been challenged by Barth, Iden and Russek (1986), by Modigliani and Sterling (1986), and recently by Feldstein and Elmendorf (1987). The first paper updates Kormendi's sample through 1983, distinguishes between federal and state-local government debt, and also looks at more recent postwar periods. The results raise some doubts about the robustness of Kormendi's findings. Modigliani and Sterling show that when Kormendi's separate tax, transfer, and government interest variables are combined into a single "net tax" variable, the sum of its distributed lag coefficients is significantly negative in the consumer expenditure equations and is not significantly different from the sum of the lag coefficients on NNP. The government purchases coefficient is statistically insignificant. They also point out that Kormendi's specification does not allow for the relatively long lags required to approximate permanent income, and that his practice of running regressions in first differences is inconsistent with his own estimated autoregressive coefficient, which is below unity.

Feldstein and Elmendorf argue that Kormendi's results are mainly due to the inclusion of the Second World War years. In that period, deficits were run to finance a massive increase in defense spending while rationing, patriotic appeals, and shortages were producing abnormally high saving rates. When they exclude these years, they obtain insignificant effects on private consumption of government purchases, along with significant negative effects of tax receipts.

No clear conclusions on debt neutrality emerge from the articles reporting structural estimations. Results seem quite sensitive to the choice of data, of variables to be included, and to estimation procedures. Kormendi's hypothesis that government purchases replace private consumption to significant degree remains a conjecture which requires further testing.

4.4.2. An alternative to structural estimation

A second set of tests seeks to avoid structural estimation that is subject to misspecification bias. Aschauer (1985) defines effective consumption as the
sum of private consumption plus \( \theta \) times government purchases of goods and services, where \( \theta \) represents the number of units of private consumption required to yield the same utility as one unit of government goods and services. He shows that the first-order conditions for utility maximization\(^{15}\) are (i) the intertemporal budget constraint and (ii) a condition involving consumption similar to the one estimated by Feldstein, but with the discounted values of future labor income and government purchases also entering the regression.

The procedure then is to combine this relationship with an assumed process for forecasting current government purchases and derive the cross-equation restrictions implied by the joint hypotheses of rational expectations, debt neutrality, the specific optimization model, and the postulated process for government spending. Aschauer estimates the system by FIML both with and without parameter constraints and carries out a likelihood-ratio test for the restrictions implied by the hypotheses. He finds that the restrictions cannot be rejected at the 10 percent significance level (or lower).

Plosser (1982) considers a model combining the efficient market hypothesis, a simple version of the expectations theory of the term structure,\(^{16}\) and an assumed moving-average representation of policy variables to derive and test cross-equation restrictions. The first two assumptions ensure that the surprise in the holding return of an \( n \)-period bond bought at time \( t \) and sold at time \( t + 1 \) as an \((n - 1)\)-period bond is negatively related to the unexpected movement in the current one-period rate, \( \hat{R}_{1,t+1} - E_t R_{1,t+1} \), and to the revisions in the forecasts of future one-period rates, \( (E_{t+j} R_{1,t+j} - E_t R_{1,t+j}) \), \( j = 2, \ldots, (n - 1) \). He also assumes that the reduced form for the one-period interest rate includes government spending, government debt held by the public, and government debt held by the monetary authority. The vector of exogenous\(^{17}\) variables, augmented by the error in \( R_t \) which is orthogonal to the policy variable, is assumed to have a moving-average representation. Under those assumptions, the current surprise in holding period returns is a function of the contemporaneous innovations in the exogenous variables (including policy variables) which are orthogonal to the innovations in the other exogenous variables.

Plosser jointly estimates an assumed autoregressive scheme for policy vari-

\(^{15}\)Aschauer assumes that a representative agent maximizes a quadratic utility function with a positive discount rate over an infinite horizon, subject to the constraint that the present value of "effective consumption" is equal to net wealth (excluding government bonds) plus the present discounted value of labor earnings plus (a multiple of) the present value of government expenditure.

\(^{16}\)See Chapter 13 by Shiller and McCulloch in this Handbook for a discussion of empirical tests of the expectations theory of the term structure.

\(^{17}\)The assumption that these variables are exogenous can be relaxed with minor changes in the interpretation, as shown by Plosser.
ables with one form of the equation for the innovations in holding returns. The technique is to stack the two equations to form a single regression and estimate it using a non-linear GLS procedure. He uses 1954–78 quarterly U.S. data and finds that the effects of surprises in government financing on nominal rates of return are statistically insignificant, with point estimates of the "wrong" sign. Innovations in government spending, on the other hand, do have significant effects on interest rates. Plosser tentatively attributes those to intertemporal substitutions.\(^8\)

Although this strand of literature provides an interesting alternative to structural estimation, it is also subject to the criticism that the processes assumed to be used by the public in forecasting may contaminate expectational variables with measurement error. This would bias the tests towards finding no impact of policy surprises, thus favoring debt neutrality.

4.4.3. Potential for future research

There is no clear conclusion emerging from the empirical research on debt neutrality. This is disturbing, though not very surprising. Bernheim (1987) points to some of the difficulties: picking the proper measure of debt or deficit; purging regressors of endogeneity by the choice of appropriate instruments; not allowing the limited significance of short-run effects to obscure the potentially considerable significance of long-run effects. Another problem is the distinction between the anticipated and unanticipated components of fiscal variables. Anticipated policy changes in a world without frictions would have little contemporaneous effect when actually implemented, but this does not mean the policies are irrelevant for individual behavior. To make things worse, current private behavior may be mainly a response to what current policy settings imply about future settings, specifically to the debt repayment horizon that taxpayers perceive as likely.

Econometric difficulties notwithstanding, one cannot ignore what must be the best “controlled experiment” for the hypothesis, namely the U.S. tax cuts of the early 1980s. Poterba and Summers (1986), among others, have drawn attention to the fact that private saving has declined while taxes have been cut and enormous chronic deficits have arisen. Their regressions show that when tax collections were reduced by legislation, consumption did increase, despite the fact that tax changes were anticipated. These events have to be explained away before it can be claimed that there is strong empirical support for debt neutrality.

\(^8\)The results are unchanged when the assumed information set of agents is augmented to include current and lagged short- and long-term interest rates as predictors of future values of policy variables, and when the processes for policy variables are differenced to eliminate non-stationarity.
5. Financing deficits with money

This section deals with substitutions of money for tax financing. Every dollar of deficit is a dollar increase in the stock of high-powered money. If changes in the rate of monetary growth have no real effects, then money is said to be “superneutral”, as in Section 3.

Usually, superneutrality refers to the invariance of the capital stock (or the capital–labor ratio), and of real per capita aggregate consumption with respect to changes in monetary growth. Effects on real (per capita) money holdings are not regarded as violating superneutrality. We present both models focusing on invariance of steady-state equilibrium values of real variables, and those investigating the transition path.

5.1. Superneutrality in the steady state

Tobin (1965a) made a case against monetary superneutrality based on portfolio effects of changes in the rate of monetary growth. A higher rate of nominal money growth is associated with a higher rate of steady-state inflation. Since the nominal interest rate on money is constant, usually at zero, this higher inflation lowers the real rate of return on money and makes it less attractive to hold. When the asset demand for capital depends not only on its own real rate of return, but also on how this rate compares with that on the alternative available assets, higher inflation encourages a portfolio shift away from money and into capital. Thus, money is not superneutral. This is known as the “Tobin effect” or the “Mundell effect”.

Increases in inflation may favor capital accumulation, at the expense of real money balances. But they also increase the resources needed to effect any given volume of transactions. These “shoeleather costs”, which were not explicitly recognized in Tobin (1965a), should be set against the gains from the portfolio effects in assessing consumption and welfare. Tobin (1986b) shows that there are cases in which it is optimal for the government to raise inflation so as to reap the benefits from increased capital and output, despite the resulting increase in “shoeleather costs”.

The Tobin effect is what exponents of superneutrality question. They consider variations in deficits and in monetary growth arising from changes in the size of transfers or taxes for given government purchases of goods and services. Superneutrality is proved by showing that the equilibrium condition

---

Note that when private saving is interest-sensitive, as discussed in Subsection 4.2, the magnitude of the Tobin effect is smaller, because desired saving and the desired wealth-to-income ratio are also lowered.
for the size of the capital stock in the steady state is unaffected by the changes in inflation associated with changes in money-financed deficits.

The classic defense of superneutrality is Sidrauski (1967). He considers a representative individual who maximizes the discounted sum of time-separable utility over an infinite horizon. Instantaneous utility is a function of real consumption and of the flow of services of money, assumed proportional to the real money stock. The agent chooses real consumption, as well as how to divide his saving between the two available assets, money and capital. Saving is the difference between income plus government transfers and consumption. Transfers are in money, in amounts unrelated to individuals' previous holdings. The government does not purchase goods and services, and it finances transfers through money creation at a constant rate. The population grows at a constant rate.

Sidrauski's main result is that the sizes of the long-run capital stock and of real consumption are independent of the rate of monetary growth. The reason is that an infinitely-lived agent will accumulate each asset up to the point where its net yield just compensates for the postponement of consumption. This point is where its marginal product equals the sum of the rate of time preference, the rate of population growth, and the rate of depreciation. The equilibrium condition is not affected by changes in the rate of inflation.

Of course, the opportunity cost of holding cash balances (in terms of foregone consumption) rises with inflation. Since the level of consumption remains unchanged, the increase in the marginal yield of cash balances necessary to re-establish equilibrium is brought about through a reduction in real money holdings. Thus, superneutrality does not extend to the per capita real money stock. Since utility is a positive function of both consumption and real money balances, monetary expansions that raise inflation lead to welfare losses.20

In view of the source of Sidrauski's superneutrality result, it is not surprising that superneutrality fails when real money balances affect the net marginal product of capital in terms of consumption. A simple setup is presented in Dornbusch and Frenkel (1973), where it is assumed that output available for consumption is a fraction of production (net of output used to meet capital needs arising from population growth). This fraction is in turn assumed to be increasing with real per capita money holdings. Instantaneous utility is a function only of consumption, and the agent maximizes the discounted sum of utilities over an infinite horizon. Although steady-state equilibrium again requires equality of the net marginal product of capital with the rate of time

---

20Sidrauski is careful to distinguish these steady-state results from short-run effects. He notes that in his model, increases in monetary growth imply increased disposable income and consumption in the short run.
preference, the former now depends on the size of per capita money balances. The new steady state with higher inflation will generally involve both lower per capita money holdings and lower capital stock. This is the opposite of the Tobin effect.

In the same class of models, an earlier paper by Levhari and Patinkin (1968) explicitly incorporates the real money stock in the production function and shows that superneutrality fails. It also shows that when the Golden Rule holds, i.e. the net marginal product of capital equals the natural rate of population growth, it is in general possible to raise per capita output and hence welfare by lowering inflation (or increasing deflation) so as to induce greater real money holdings, even though the real rate of interest and net marginal product of capital are above the natural growth rate. It will generally not be optimal to increase the real money stock to the point where money holders are satiated, i.e. where the rate of deflation equals the net marginal product of capital. The welfare optimum will involve neither Phelps' Golden Rule capital stock nor Friedman's optimal quantity of money.

An alternative mechanism negating superneutrality in steady states is suggested by Brock (1974), who introduces a labor–leisure choice into Sidrauski's model. Since the additional first-order condition involves the marginal utilities of consumption and of leisure, it now becomes necessary for those two marginal utilities to be independent of money if the steady-state levels of capital, labor, and consumption are to be unaffected by changes in the rate of monetary growth. Brock shows that when money affects the marginal utility of leisure, money growth affects the labor supply curve and thus the steady-state stock of capital. 21

A key feature of the Sidrauski model that allows it to exhibit superneutrality is the infinity of agents' horizons. Drazen (1981) shows that for a given individual who lives for two periods, both the substitution and the income effects on capital holdings of an increase in the return to money (lower inflation or higher deflation) are negative when consumption and money balances are normal goods. The negativity of the income effect can be explained as follows. An increase in the return to money arising from a reduction in inflation implies ceteris paribus higher income in the second period of life. Given the usual concavity conditions on utility, the agent will want to spread this increase over both periods. When consumption and real balances are "normal goods", this implies higher consumption and money balances in the first period. Since the usefulness of capital to the agent as an asset is only in transferring income from the first to the second period, a

21It is still possible to obtain invariance of the capital–labor ratio, provided that the production function exhibits constant returns to scale. Except in this limited sense, however, superneutrality is lost when the choice between labor and leisure is explicitly incorporated into the model.
redistribution of income in the opposite direction can be accomplished by lowering capital holdings. This negative income effect, combined with the unambiguously negative substitution effect, produces a negative total effect of an increase in the return to money.

Now, in view of the government budget identity, a reduction in monetary growth and inflation is associated with a reduction in transfer income, a lower deficit. The overall effect on the demand for capital of a reduction in the rate of monetary expansion is the sum of the effect (just described) of the increase in the return to money and the effect of the reduction of transfers. If transfers are distributed in proportion to first-period money holdings, then this overall effect is zero. (The transfers amount to nominal interest on money and exactly offset the effect of inflation on the real return to holding money.) If transfers are heavily weighted towards the old, it is possible that deflationary policy would increase demand for capital. This is because the reduction in the transfer income of the old would more than offset the effect described in the previous paragraph, inducing people to hold more capital in order to shift income and consumption to the second period of their lives.

Thus, the Tobin effect is observed if this does not occur (e.g. when transfers are sufficiently weighted towards the young). Under this condition, Drazen shows that the result generally extends to the economy-wide capital-labor ratio.

Halliassos (1987) shows that superneutrality is not an inescapable feature of infinite-horizon setups, but is due to the commonly used but unrealistic assumption that portfolio adjustments are simultaneous across different agents. When portfolio holding periods are staggered, each portfolio is held over a period of time during which real rates of return change. Even when these changes are deterministic, staggering can generate a determinate optimal composition of portfolios: assets become imperfect substitutes under perfect foresight. Holding period returns are still tied to the rate of time preference as in the Sidrauski model. But the variability of rates of return and of return differentials within each holding period can be influenced through policy. Changes in the rate of growth of nominally denominated assets affect inflation and the real rate of return of money. Since relative asset supplies are unchanged, the optimal composition of private portfolios is also invariant to the policy change. But interest income on the optimal portfolio is affected. If the same level of consumption out of portfolio income is to be maintained, real rates on non-monetary assets have to adjust. It is also shown that real wealth and real taxes net of transfers are affected by the policy change.

22Tobin, and other contributors to this debate, were always assuming that transfers were independent of money holdings. If they were random or equal per capita, they would in Drazen's terms be weighted to the young
Siegel (1983) shows that superneutrality will also fail in general when labor-augmenting (Harrod-neutral) technological change is incorporated into the model. This creates an asymmetry between the production function and, therefore, the budget constraint, which are stationary in efficiency units, and the utility function, which is stationary in per capita values. The first-order equilibrium condition for the marginal product of capital involves the rate of growth of real per capita consumption and money holdings. In the absence of technical progress, these terms are zero. However, when technological change is present, these rates of growth are positive: real consumption and money holdings are constant only per efficiency unit of labor, i.e. the scale of each economic unit increases at the rate of technical progress. The coefficients on those rates of growth are functions of the marginal utility of consumption and of its derivatives with respect to per capita consumption and money holdings. Superneutrality now requires that these coefficients be unaffected by changes in monetary growth. This will be true for utility functions which are separable in consumption and money holdings; and for non-separable functions which are isoelastic in an index of real per capita consumption and real money holdings, $\gamma m^\beta$, where the sum of the constants satisfies $\gamma + \beta < 1$.

5.2. Superneutrality along the transition path

We now turn to the issue of whether changes in the rate of nominal money growth are superneutral when the economy is outside its steady state, but moving towards it.

Fischer (1979) considers a model similar to Sidrauski's with perfect foresight. He examines the class of utility functions displaying constant relative risk aversion, of which logarithmic utility is a special case. The result is that although in such a model the steady-state value of the capital stock is invariant to changes in the rate of monetary expansion, such superneutrality does not prevail on the transition path to the steady state, except in the case of logarithmic utility. Fischer shows that higher rates of nominal monetary growth are associated with higher rates of capital accumulation on the transition path. Moreover, as the capital stock approaches its steady-state value from below, a larger rate of monetary expansion implies lower consumption. Fischer

---

23In similar vein, Tobin (1968) had shown that steady states with transactions requirements for money exist only if Harrod-neutral technological progress occurs at the same steady rate in both goods production and in transactions

24Asako (1983) notes that for more general utility functions, a sufficient condition for superneutrality on the transition path is that the function be separable in consumption and real money balances, i.e. $u_{\text{re}} = 0$. Deriving necessary conditions for general utility functions is more difficult
does not provide an estimate of the quantitative significance of these effects.

Cohen (1985) suggests a rationale for Fischer's results, using a model that emphasizes the distinction between consumers and producers. Nominal rates of interest vary along the transition path due to two conflicting factors: the increase in inflation pushes them upwards, while the decrease in real rates pushes them downwards. When the intertemporal elasticity of substitution is unity, the two effects cancel each other out and monetary policy cannot affect nominal interest rates. However, when the intertemporal elasticity of substitution is below (above) unity, the nominal rates on the transition path are above (below) their steady-state values. In those cases, monetary policy affects capital accumulation by influencing the rate of growth (or decline) of nominal rates towards their steady-state value.

This effect on accumulation comes about because the specification of the instantaneous utility function\footnote{25} allows consumption and real money holding to be treated as a composite commodity, the "price" of which is an increasing function of the nominal interest rate. Now when monetary policy raises all nominal rates, it also affects the equilibrium sequence of the relative price of the composite good from one period to another (which is a function of nominal interest rates). Cohen shows that the effect on interest rates has no consequence for the amount of consumption in his model. However, the effect on intertemporal relative prices of the composite good is such that an increased rate of monetary growth lowers the demand for the consumption good (unless elasticity is unity), while leaving the supply unaffected. The resulting decrease in consumption allows faster capital accumulation to take place all along the transition path.

Asako (1983) shows that it is possible for a utility function with constant relative risk aversion (CRRA) other than the logarithmic to violate superneutrality on the transition path in a direction opposite from that suggested by Fischer. In particular, this is true of the CRRA utility functions with relative risk aversion exceeding unity when consumption and real money holdings are perfect complements, i.e. their desired ratio does not depend on economic conditions. In this case, the rate of capital accumulation is slower, the higher the rate of monetary expansion. Moreover, neither consumption nor the real money stock is affected by changes in money growth in the steady state. This superneutrality is even stronger than Sidrauskis's, and it is due to the perfect complementarity of $c$ and $m$.

\footnote{25} \begin{align*}
U(c, m) &= (1/(1-s))(c^s m^{1-s})^{1-s}, \quad \text{if } s \geq 0 \text{ and } s \neq 1, \\
U(c, m) &= \log c^s m^{1-s}, \quad \text{if } s = 1
\end{align*}
5.3. Empirical tests of superneutrality

Superneutrality asserts invariance of capital stock and consumption with respect to inflation rates and monetary growth. These are propositions about long-run paths. Invariance of capital implies, even derives from, invariance in the real interest rates to which marginal productivity of capital is equal in equilibrium. These invariances are all very difficult to test. Consequently, empirical research on superneutrality has focused primarily on tests for the existence and stability over time of the famous hypothesis about interest rates put forward by Fisher (1930) [and modified by Darby (1975)]. The Fisher hypothesis is that a change in inflation will be fully reflected in an equal change in nominal interest rates, without affecting the real rate of interest. Darby's modification restates the hypothesis in terms of after-tax real and nominal interest rates. One strand of the modern literature focuses on tests for the presence of a Fisher effect, sometimes allowing for taxation. The other strand focuses on the stability of the response of nominal rates to inflationary expectations over time. Here we illustrate both.

5.3.1. The presence of the Fisher effect

Tests of the Fisher hypothesis up to the early 1970s typically involved regressing the nominal interest rate on a distributed lag of past inflation rates as a proxy for expected inflation, a constant intended to represent the invariant real rate, and an error assumed to be distributed independently of past, present, and future price levels. The extraordinarily long lags typically implied by the estimates were taken as evidence against the Fisher hypothesis (even by Fisher himself).26

Sargent (1973) was the first to utilize rational expectations in testing the Fisher hypothesis. Sargent argues that even when Fisher's theory is correct, the estimated lag functions do not necessarily represent optimal forecasts of inflation. He combines rational expectations with the natural rate hypothesis (NRH) embodied in the Lucas aggregate supply curve to construct a model in which the real rate of interest is independent of the expected part of the money supply. His test of the Fisherian hypothesis is simply to test this model. He

26The first researcher to test the Fisher hypothesis was Fisher himself. In 1930 he concluded that "when prices are rising, the (nominal) rate of interest tends to be high, but not so high as it should be to compensate for the rise" (1930, p. 43). In addition, Fisher observed a relationship between interest rates and past inflation rates. He interpreted this as supporting a modified version that allows for less than perfect foresight and consequently effects that are smaller and slower than price changes.
does so indirectly by testing NRH, defined as the idea that unemployment is independent of the systematic part of the money supply.

Sargent proposes two tests of NRH. In the first, the unemployment rate is regressed on lagged unemployment rates and on other variables included in the information set in period \( t - n - 1 \), where \( n \) is the order of the autoregressive process followed by the unemployment disturbance. NRH implies that all those other variables are statistically insignificant, i.e. that the innovation in unemployment is not affected by past values of any variables, including policy variables. The alternative test proposed by Sargent involves estimating an equation for unemployment which incorporates not only lagged unemployment and the unexpected part of inflation, but also the expected part of inflation. The null (Fishierian) hypothesis is that the coefficient on expected inflation is zero, and it is tested against the alternative that it is not.

Although Sargent finds that the evidence on NRH (and hence on Fisher) is mixed, he points to the lack of an alternative model that would outperform NRH in tracking unemployment. Sargent feels that if an investigator has priors in favor of NRH, the evidence is not sufficient to reject them.

Fama (1975) suggests using a joint test of the hypotheses (i) that the expected real rate of return on Treasury bills is constant and (ii) that agents make optimal use of their information concerning inflation over the next month when setting the nominal interest rate today. The latter is a version of the efficient markets hypothesis (EMH).\(^{27}\) If the hypotheses jointly hold, then it should not be possible to use any subset of information available as of \( t - 1 \) (e.g. the history of real rates) to come up with a better prediction of \( r \) than the constant \( E(r) \). This in turn means that the autocorrelations for \( r \) are zero for all lags, and this can be tested by checking the sample autocorrelations for \( r \).

A further set of tests can be obtained by generalizing the model of bill market equilibrium so that constancy of \( E(r) \) becomes a special case. Specifically, if estimating the regression

\[
\pi_t = \alpha_0 + \alpha_1 R_t + \varepsilon_t \tag{5.1}
\]

yields coefficient estimates that are inconsistent with the hypothesis that \( \alpha_0 = -E(r) \) and \( \alpha_1 = 1 \), the model of a constant \( E(r) \) is rejected. Market efficiency can then be tested by checking whether \( \varepsilon_t \) is autocorrelated. If EMH

\(^{27}\) The nominal one-month rate of return on a one-month Treasury bill paying \$1 at time \( t \) is \( R_t = (1 - v_{t-1}) / v_{t-1} \), where \( v_{t-1} \) is the price of the bill determined in period \( t - 1 \). Fama postulates that the market sets \( v \) so that it perceives the expected real return on the bill to be the constant \( E(r) \). This model of market equilibrium is combined with the assumption that the market makes optimal use of all available information concerning the stochastic rate of change of purchasing power \( \pi_t \) over the next month.
holds, $R_t$ summarizes all available information about $E(\pi_t | I_{t-1})$, including past values of the disturbance. These past values should be of no use in predicting $\pi_t$, which implies that $\varepsilon_t$ should be serially uncorrelated. This logic can be generalized to any piece of information in $t - 1$. An example is:

$$\pi_t = \alpha_0 + \alpha_1 R_t + \alpha_2 \pi_{t-1} + \varepsilon_t,$$

(5.2)

where EMH implies that $\alpha_2$ is zero and $\varepsilon_t$ is serially uncorrelated.

Although Fama's empirical results did not reject his joint hypothesis, subsequent research has challenged their robustness. Nelson and Schwert (1977) argue that looking at the ex post real rate may generate an "errors in variables" problem, since it is possible for the autocorrelation function of the ex post real rate to be close to zero at all lags even if the ex ante real rate fluctuates considerably and is highly autocorrelated. As for the test of whether $\alpha_2 = 0$ in (5.2), when the authors replace $\pi_{t-1}$ with optimal predictors of inflation, which make use of more observations on past inflation and of the time series properties of the inflation rate, the coefficient on the optimal predictor is large and significant, putting EMH into question.

Garbade and Wachtel (1978) focus on equation (5.1), where the null hypothesis is that $\alpha_0$ and $\alpha_1$ are the same over three subperiods. They argue that the alternative specified by Fama, namely that the coefficients changed by a discrete amount from one subperiod to the next but remained constant within each subperiod, makes it quite unlikely that the null will be rejected in its favor. Instead they allow each coefficient to follow a random walk without drift. Fama's finding that $\alpha_1$ is stable and equal to 1 cannot be rejected by their tests, either. The conclusion survives even when $\alpha_0$ is assumed to depend linearly on time within each of six specified subintervals and the nature of this dependence is different for each one. However, when such piecewise-linear time-variation of $\alpha_0$ is regarded as the alternative hypothesis, the null hypothesis of a constant real rate $-\alpha_0$ is rejected, irrespective of whether the constraint $\alpha_1 = 1$ is imposed or not.

The question whether nominal interest rates have adjusted sufficiently to compensate investors both for changes in (expected) inflation and for the

---

28Fama uses data on U.S. Treasury bills with one month to maturity over the period from January 1953 through July 1971, and for various subsamples. He also performs similar tests for bills of up to six months of maturity. In all cases, he assumes that the behavior of the Consumer Price Index is the relevant measure of inflation. The results of both autocorrelation and regression coefficient tests support EMH with respect to the history of inflation rates, and do not reject the model of constant expected real rates. Combining these two findings, Fama concludes that we cannot reject the hypothesis that all variation through time in $R$ reflects variation in correctly-assessed expected rates of change in purchasing power.

29This happens when the variance of errors in inflationary expectations is large relative to the variance of the ex ante real rate.
effects of interest taxation is addressed by Tanzi (1980). Using data on six- and twelve-month Treasury bills, he ran the following regression over the 1952–75 period:

\[ R_t = r_t + \beta E(\pi_t)/(1 - T_t), \]

where \( R \) is the lender’s required rate, \( r \) is the after-tax rate he would have received if expected inflation were zero, \( E(\pi_t) \) is expected inflation, and \( T \) is the tax rate on interest income. Tanzi rejects the hypothesis that \( \beta = 1 \). Since the estimate of \( \beta \) is below 1, he concludes that agents are not free of “fiscal illusion”, and are not sufficiently compensated for the effects of changes in taxes on their nominal interest income.

While the Fisher hypothesis postulates constancy of the net marginal productivity of capital, most empirical papers use data on CPI-corrected returns on financial assets (i.e. nominally denominated interest-bearing claims such as Treasury bills). Carmichael and Stebbing (1983) point out that the poor results of existing tests may be partly due to a higher degree of substitutability at the margin between money and such financial assets than between financial assets and capital. One crucial fact about money is that its nominal rate of return is “regulated” (usually set equal to zero) and consequently changes in inflation change its real after-tax rate of return by the same amount. If nominally denominated assets are close substitutes for money, this one-to-one sensitivity to inflation should also be (approximately) true for their real after-tax rate, even if the marginal productivity of capital is fully governed by the rate of time preference and other parameters independent of inflation. This inverted Fisher hypothesis for financial assets is not rejected by data on three-month U.S. Treasury bills and by two Australian interest-rate series (short- and long-term).³⁰

5.3.2. The Fisher effect over time

Tests of how the magnitude of the Fisher effect has behaved over time have mainly focused on two partitions of the available data sample. One is into the pre World War II and post World War II, while the other partitions the post World War II era into various subsamples. We briefly illustrate both.

Friedman and Schwartz (1976, 1982) and Summers (1983) have found essentially no evidence of the Fisher effect in the pre World War II period in either Britain or the United States. They have found much stronger correla-

³⁰The U.S sample is 1953 1–1978 IV; the Australian is early or mid-sixties to end of 1981. To test their hypothesis, Carmichael and Stebbing make use of portfolio arbitrage conditions on real after-tax rates of return among the three assets, and of the assumptions that expectations are unbiased and that individuals know their marginal tax rates.
tions between short-term nominal interest rates and either ex post inflation or other proxies for expected inflation rates in the post World War II period (especially post-1960). This led Friedman and Schwartz to suggest that perhaps financial markets "learned their Fisher" only gradually.

The findings concerning the United States were recently challenged by Barsky (1987). He argues that the observed difference in the magnitude of the Fisher effect is not due to a shift in any structural relationship but to differences in the stochastic process generating inflation. While inflation was essentially a white-noise process before World War I, it became a non-stationary ARIMA process in the post-1960 period. Barsky shows that an underlying Fisher effect is consistent with any observed correlation between current nominal interest rates and actual inflation rates in the current (or next) time period. The reason is that the latter simply reflect the persistence of inflation.

In addition, Barsky raises doubts as to whether regressions of nominal interest rates on inflation are reliable in assessing whether real rates are affected by changes in expected inflation. A case in point is that of the gold standard years prior to 1913, which look the least Fisherian in regressions involving nominal interest rates [e.g. Summers (1983)], but do not exhibit negative correlation between ex ante real rates and past inflation.31

Barsky attempts to reconcile this discrepancy by the fact that inflation in those years was virtually white noise. As a result, the variance of anticipated inflation was substantially lower than the variance of actual inflation, and OLS regression would lead to the incorrect conclusion that nominal interest rates failed to respond to inflation. Summers tries to allow for this by "band filtering" the data, i.e. considering only low-frequency components which can be thought of as easily forecastable. The hope is that then expected inflation can be proxied by actual inflation. McCallum (1984b) accepts the substantive conclusions of the Summers study, but points out that low-frequency estimation is affected by mis specification of the distinction between anticipated and unanticipated movements in the regressors.32

Barsky shows that the covariance between anticipated and ex post inflation does not increase relative to the variance of inflation as the frequency is lowered. This suggests that band filtering in this case may not have resulted in a better proxy for expected inflation.33

Cargill and Meyer (1980) focus on the post World War II period and ask

---

31 Another case is that of the postwar period, which looks Fisherian on the basis of nominal rate regressions, but exhibits a strong negative relationship between inflation and expected real returns on short-term instruments.

32 See also their exchange in Summers (1986) and McCallum (1986).

33 At any rate, given that expected inflation probably fluctuated very little in the pre-1913 period, that part of the sample is probably not very informative as to the validity of the Fisher hypothesis.
whether length of maturity is relevant for the existence and magnitude of the Fisher effect over time. They run regressions of the form

\[ R_t = \beta_0 + \beta_1 r_t + \beta_2 E(\pi_t) + u_t \]  

(5.4)

over subperiods between 1954 and 1975. Coefficients on expected inflation, \( E(\pi_t) \), are almost always positive and significant; they decline with increases in maturity. Many of the coefficients are close to or significantly above unity. However, estimates of \( \beta_k \) vary significantly between subperiods. Estimates for some maturities decline quite substantially for the period 1970–75 relative to 1960–75.

6. Monetary policies

The previous section discussed substitutions of money for tax financing of a given path of government expenditures and the effects of the corresponding changes in the rate of nominal money growth. In this section we discuss shifts between alternative modes of non-tax financing of government expenditure. In most setups, fiscal policy is taken as given. In a few, taxes are also varied so as to eliminate effects of asset market exchanges on agents’ opportunity sets.

Some of the models reviewed focus on the common type of open market operations, namely exchanges of money for nominal bonds. Others, however, discuss exchanges of money for real capital or for indexed bonds. Even when the term “capital” simply refers to stored amounts of the consumption good, money–capital exchanges should be distinguished from temporary increases in government expenditure \( G \) financed through money creation. The asset swaps considered are associated with a given \( G \). We first investigate the relevance or irrelevance of various asset exchanges for the real allocation of resources. We then discuss shifts between bond and money financing which are necessary when certain financing policies are unsustainable over the longer run.

6.1. Asset exchanges

We start with models exhibiting neutrality of asset exchanges and then consider setups in which open market operations have been shown to be non-neutral.

---

34They consider a wide range of maturities of government and commercial financial instruments, and look at the “term structure” of inflationary expectations, so as to match each instrument with the (geometric) average of one-period expected inflation rates over its time to maturity.
6.1.1. Neutral asset exchanges

Wallace (1981) investigates whether there is a class of open-market exchanges between fiat money and “capital” (in the form of a stored consumption good) that would leave the equilibrium sequences for real consumption allocations and for the price level unaffected. He considers a two-period, pure-exchange, overlapping generations model with a single consumption good that is storable via a constant returns to scale, stochastic technology. Complete markets in contingent claims on second-period consumption are assumed. Open-market operations consist of purchases and costless storage by the government of the consumption good in exchange for fiat money.\(^{35}\) Money is not dominated in return. The government sets the (possibly contingent) paths of government consumption, \(G(t)\); of the endowment vector for each generation \(t\), \(w(t)\); of the path of government storage, \(K^g(t)\); and of the nominal money supply path, \(M(t)\). An equilibrium is described by a sequence of real consumption levels, prices of consumption claims, aggregate storage, prices of the good, and nominal money.

Wallace starts with an equilibrium for a policy \(\{G(t), w(t), K^g(t) = 0\}\). He then identifies policies \(\{G(t), \hat{w}(t), \hat{K}(t)\}\) which support the same equilibrium configuration (with the exception of the money sequence). The choice among them is irrelevant for the equilibrium outcome. While such policies leave the price path unaffected, they combine open market operations with changes in taxes net of transfers which “pay out” to agents any additional net interest income the government receives as a result of the change in its portfolio. This is necessary if private opportunity sets are to remain unaffected.\(^{36}\) Wallace does point out that this analysis is probably most useful as a benchmark case for assessing the real effects of asset exchanges.

While Wallace’s policies involve auxiliary fiscal changes, Chamley and Polemarchakis (1984) consider “pure” open market operations but alter the price process so as to pay out to private agents the altered returns on the government’s portfolio. In addition, they do not assume the existence of a complete set of contingent markets. Suppose that the economy is in equilib-

---

\(^{35}\)The supply of fiat money is costlessly manipulated by the government. In period \(t\), consumers demand claims on consumption in period \(t + 1\), while firms (which are owned by members of generation \(t\)) supply those claims, storing the good and money in the process. In their roles as consumers, agents maximize their expected utility of consumption in the two periods of their life. As producers, they choose to undertake one or both of the two risky projects of storing money and of storing the good.

\(^{36}\)Wallace shows that \(\{\hat{K}(t)\}\) is any non-negative sequence bounded by \(\{K(t)\}\), and \(\{\hat{w}(t)\}\) is any endowment sequence that meets a particular set of restrictions. The point of the restrictions is to ensure (a) that the distribution of income among agents is unchanged, and (b) that the differences in net interest received by the government as a result of the change in its “portfolio” of the stored good and of money are offset by changes in taxes net of transfers. In this framework, the boundedness of \(K(t)\) from above is necessary, because otherwise no feasible value of private storage exists that is consistent with unchanged total accumulation.
rium when the quantity of money is fixed for all time periods and government holdings of capital are zero (along with government expenditures and taxes for simplicity). Then suppose that the government announces a new contingent sequence for its capital holdings. The authors show that this policy change has no effect on the time paths of the allocation of goods and of the aggregate capital stock, as long as money maintains a non-zero value in terms of the capital good. Individuals reduce their capital and increase their money holdings in a way that maintains both the value of each agent’s portfolio and the real return on it. Thus, any intertemporal program that was feasible for a private agent before the policy change is also feasible after it. Moreover, the portfolio response of agents is consistent with equilibrium in the markets for capital and for money.\textsuperscript{37}

The result extends to economies with many assets, in which case only the prices of the assets involved in the exchange are affected by it. It fails when short sales are bounded below or not feasible, and when open market exchanges involve money and assets denominated in nominal terms. The reason for the latter is the effect on the returns of a nominally denominated asset induced by the change in the path of the price of money. This effect is such that the conditions which the prices of money and of bonds have to satisfy for neutrality are not satisfied for a general distribution of total nominal returns on the nominal asset.

Peled (1985) focuses on neutrality of open market exchanges of fiat money for indexed government bonds in the context of a single-good, pure-exchange, overlapping generations model. Random endowments generate stochastic bond price and aggregate price levels and inflation. However, indexed bonds promise a given amount of the perishable consumption good. Peled considers a given path of taxation. He shows that if one starts with a financing scheme that results in money being willingly held, then one can change the path of bond issues almost arbitrarily and still be able to find offsetting changes in the path of money creation that would leave the paths of real consumption, taxes, and the price of bonds unaffected.\textsuperscript{38} These exchanges do not affect the value of the

\textsuperscript{37}A sufficient condition for a positive price of money in period \( t \) is that both the contemporaneous government capital holdings and the total return on capital be positive.

\textsuperscript{38}Wallace’s neutral policies can be seen as combinations of (1) an open market operation (given taxes net of transfers), which does affect the price of money, with (2) changes in the supply of money via taxes net of transfers that offset the first price effect without affecting the real quantity of money. Chamley and Polemarchakis show that monetary policies of this type do exist even in their more generalized framework. Sargent (1987, esp. p 322) states a general neutrality theorem which encompasses those of Wallace, and Chamley and Polemarchakis as special cases.

\textsuperscript{39}Note that not all financing policies which involve the same path of taxes and lead to the same government revenue leave the real allocation of resources unaffected. The added condition is that they also generate identical “deficit” paths for all periods \( t \geq 1 \) over which they are in effect “Deficit” here refers to the change in the real value of government monetary plus non-monetary debt. When this is true, unchanged aggregate consumption is guaranteed to old agents and aggregate private saving is also unchanged.
government portfolio. The assumed existence of intragenerational markets for money and bonds is crucial for the result.\footnote{If such markets were absent prior to the government intervention, then the government could facilitate risk-sharing asset exchanges among diverse individuals through its open market operations. Then, government intervention would have real effects. On the other hand, neutrality is not contingent on the lump-sum nature of taxes.}

Until recently it was thought that irrelevance results could be obtained only in models where money is not dominated in return. Sargent and Smith (1987) show that such results can also be obtained in a class of models where money is dominated in real rate of return. In those models, money is dominated because some fraction of the population (the “poor”) are assumed to be prevented by law from holding assets other than money. The “rich” can trade in all markets.\footnote{The empirical rationalization usually offered is that laws can exist (such as Peel’s Bank Act of 1844) which impose a restriction on the minimum size of privately issued securities.} Specific assumptions about the distribution of income ensure that the poor cannot afford to hold any asset other than money. A byproduct of these assumptions is that the rich, who are not excluded from any market, will not hold money, but only stored goods and state-contingent claims to future consumption. In other respects, the overlapping generations setup is similar to that of Wallace (1981).

When open market exchanges of currency for “capital” (i.e. goods for storage) are undertaken, irrelevance with respect both to consumption and to the price system is ensured – as in Wallace – through lump-sum transfers or taxes which distribute the change in earnings on the government portfolio.\footnote{The authors show that irrelevance can also be accomplished without variation in lump-sum taxes when the menu of assets is sufficiently enriched to allow the government to borrow from (and lend to) private agents. Again, the condition for irrelevance is that the open market operations in capital be accompanied by government exchanges in debt markets so as to leave private agents’ budget sets unaltered.} Sargent, Smith, and Wallace interpret the offsetting changes in taxes net of transfers, which are necessary for irrelevance in their models, as defining a “constant” fiscal policy. The difference here is that the rich are treated differently from the poor. Specifically, the poor must be induced to raise their saving to absorb the higher real stock of currency, while the rich must reduce their saving by the amount of capital purchased by the government. Irrelevance theorems when money is dominated due to legal restrictions seem to require that such restrictions be different across agents. For instance, if the original Wallace (1981) setup is augmented to include identical “reserve requirements” across all agents, money is dominated in return but irrelevance is not obtained.

Finally, Benninga and Protopapadakis (1984) adopt a specification for utility similar to that in Sidrauskis (1967) in a two-period, Arrow-Debreu, state-preference economy with complete markets, heterogeneous consumers and firms, one good, and fixed labor supply. They ask whether shifts from one
financing mode to another, holding the revenue raised through the third mode constant, would be "neutral". They are considered "strongly neutral" when both consumption and money holdings are unaffected.

A shift between bonds and taxes will be strongly neutral if it leaves the present value of taxes (including "inflation taxes") for each consumer unchanged, and if the interest rate is unchanged. The first condition ensures that each agent's budget constraint remains unaltered by the policy shift; the second guarantees that real money holdings will also be unaffected. This result is a direct extension of those on debt neutrality to models incorporating money and uncertainty. In order for an open market purchase to be weakly neutral, a decline in the inflation rate must occur, lowering nominal interest rates, inducing agents to hold a larger real stock of money, while keeping the cost of holding money constant and allowing the government to receive more revenue from issuing money. This holds for a restricted class of utility functions: separability of utility in consumption and money is a sufficient condition for weak neutrality.43 Finally, a shift in financing between money and taxes is weakly neutral only if the government undertakes an elaborate scheme of adjusting each individual's taxes in a way that offsets the changes in the cost to each individual of holding money.

Interestingly, Benninga and Protopapadakis show that in their model, no economy can be invariant under both policy shifts between money and taxes and policy shifts between money and debt. The reason is that if the present value of taxes is changed, then the cost of holding money and the revenue from money creation must also change for neutrality to hold. However, under debt–money neutrality government revenue from money creation is fixed.

6.1.2. Non-neutral asset exchanges

We now turn to models exhibiting non-neutrality of open market asset exchanges. Grossman and Weiss (1983) consider a pure exchange, Clower-constraint model with two types of infinitely-lived consumers, firms, and banks. Consumers of each type visit the bank once every two periods to withdraw money directly deposited into their accounts, in order to finance their perfectly foreseen consumption over the two periods.44 Consumers are assumed to be the only ones who hold money.

When there is an open market purchase, the addition to the money stock flows into the banks and has to be absorbed wholly by a subset of the

43This is more restrictive than the conditions imposed by Sidrauski, due to the inclusion of uncertainty into the model.
44Deposits are not checkable. Firms deposit into consumers' accounts, since consumers own shares to firms' profits. In addition to shares, the asset menu includes both interest-bearing deposits and government bonds which are perfect substitutes for deposits.
population. This subset consists of the people who have exhausted previous money holdings and are there to make a withdrawal. Since money is only held for consumption purposes, agents are going to withdraw more only if they decide to consume more. Real and nominal interest rates have to drop in order to induce that group to decide to consume more. Moreover, price effects will only be gradual, since the members of this group will spend their increased holdings gradually until their next trip to the bank. Consumers of the other type cannot increase their nominal spending before their predetermined time comes to visit the bank. Thus, they respond to higher prices by reducing real consumption.\footnote{In fact, it is possible to show that when utility is logarithmic (or when it is homothetic and demand for second-period consumption rises with inflation), the model generates damped oscillatory price behavior with overshooting after two periods and a fall in the current nominal and real rates. Two-period nominal interest rates also fluctuate, rising above their steady-state value on even dates and falling below it on odd dates, until the price path converges to its steady state. The oscillations are produced by the fact that in odd periods the types who exhaust their money holdings, and consequently have a propensity to spend out of them equal to one, are also the ones whose money holdings are larger than in the steady state. With output assumed fixed, such fluctuation in nominal spending generates fluctuations in the price level.}

A similar staggered-withdrawals setup is presented in Rotemberg (1984), except that his model incorporates capital, does not assume that output is constant, and postulates that money is withdrawn at the beginning of the period and is available for spending in the current and in the next period. The model is consistent with rate-of-return dominance of capital over money in the steady state. Rotemberg shows that there is no Tobin effect in his model, in the sense that the unique steady-state equilibrium size of the capital stock that involves positive consumption is independent of the rate of growth of the money stock. However, an increase in the rate of monetary growth, and hence in inflation, leads people to consume more right after they withdraw money; and less in the period in which they do not go to the bank and have to finance consumption with money withdrawn in the previous period. The extent of this distortion in the consumption path is partly determined by the assumption that the frequency of an agent's bank visits does not change with inflation.

Now consider an open market purchase of capital. If this had no effect on prices, consumption and hence withdrawals would also be unaffected: people at the bank would not want to hold the extra money made available to them. In fact, the price level has to rise. Rotemberg's numerical simulations of the unique non-explosive path show that the price increase results in a fall in consumption of those who do not visit the bank in that period, which raises capital, and hence output, in the following period, albeit by a small amount. How the magnitude of output response would be affected by allowing people to visit the bank more frequently in response to inflation is an open question.

Open market operations have also been shown to have real effects when
portfolio adjustments are staggered, even when we abstract from the transactions role of money. Haliassos (1987) presents such a setup in which even under infinite horizons and perfect foresight, an open market operation affects the level and variability of the real rates of return on non-monetary assets as well as the level of real wealth. The reason is that the open market operation affects the optimal composition of private portfolios. If a given level of consumption out of portfolio income is to be maintained, real rates of return have to adjust. Now rates of return on non-monetary assets absorb the brunt of adjustment, since the real rate on money is tied to the rate of growth of nominal government debt.

So far the models in this section postulate infinite horizons. Waldo (1985) constructs an overlapping generations, pure-exchange model in which the nominal and real interest rates fall and the price level rises less than proportionately to the money supply increase associated with an open market purchase. The measures of fiscal policy that are held constant are government spending and the total budget deficit, taking account of interest payments and inflation tax revenues. Thus, the effects on the total deficit of any variations in inflation and interest rates are offset through changes in lump-sum taxes.

In Waldo’s model the reason agents hold currency is to finance small transactions. Interest-bearing demand deposits, which are claims to government bonds, can only be transferred at a lump-sum cost, and are consequently used to effect large transactions. When an open market purchase takes place, the nominal interest rate falls to induce agents to substitute currency holdings for demand deposits. The excess demand for goods that results from the drop in nominal interest rates is then offset by an increase in the price level that reduces the real stock of wealth.

A different experiment, namely an increase in the rate of monetary growth (and of inflation), leads to a fall in currency demand and to an increase in goods demand. The price level rises, both to reduce the real currency supply and to lower goods demand. The effect on the nominal interest rate depends on the relative inflation elasticities of savings and currency. Even if the nominal interest rate rises, it does not do so by as much as inflation: the real interest rate definitely falls.

46 This allows inflation and the debt mpx to be determined independently of each other and eliminates any income effects of inflation, so that results only depend on substitution effects. It is different from Sargent and Wallace (1981) where government spending and the narrowly defined deficit are held constant in that model, open market purchases that lower interest rates imply lower inflation rates through the government budget identity.

47 If one uses Waldo’s setup but assumes that government spending and the primary budget deficit are held constant [i.e., the notion of constant policy in Sargent and Wallace (1981)], an open market purchase does not put downward pressure on interest rates and (via the budget identity) on inflation rates, but the effects on savings and the price level are ambiguous. If inflation effects dominate, savings rise and the price level falls to clear the goods market. The opposite happens when interest rate effects dominate.
6.2. Consequences of persistent deficits

We turn now to the question of whether persistent budget deficits are inflationary. There are two sets of issues here. One is whether such deficits are financed by issuing bonds for a given money path, or vice versa. The other is whether it is the primary or the total budget deficit which is kept constant over time.

Smith (1982) considers a constant total budget deficit. He compares a "monetarist" policy of maintaining a target path for money while financing deficits through bonds to a "bondist" policy of maintaining a target path for the bond stock while issuing money to finance the deficit. He shows that in a dynamic IS/LM model, a zero-inflation steady state may not be stable under the monetarist strategy, whereas it is more likely to be stable under the bondist strategy. The rationale is that fixed tax rates and government spending (including debt service) make government saving procyclical. Under a bond target, the change in money supply dictated by the government budget identity is countercyclical. By contrast, under a monetary target, the government must sell bonds to the public when a drop in output causes an increase in the budget deficit. Although the dynamic behavior of the model is complicated, Smith suggests that crowding out effects play a crucial role in generating the instability associated with the monetarist policy.

Of course, such crowding out effects are absent in a Ricardian world. The issue of whether persistent bond-financed deficits are inflationary in such a setup was addressed by Barro (1976) and by McCallum (1984a). Barro suggests that they would be inflationary if the rate of growth of the bond stock exceeded the rate of growth of output, the simplest example being that of an economy without population growth. The reason is that in this case the present value of the government's future taxing capacity is bounded. McCallum agrees that persistent primary budget deficits are inflationary, but shows that the result does not carry over to total budget deficits. He points out that a positive growth rate of bonds can be permanently maintained in a stationary Ricardian economy, provided that this growth rate is smaller than the rate of time preference. This is because a household's disposable income also includes interest payments from the government. As a result, taxes can exceed household output and yet be smaller than disposable income. (Presumably the assumption of lump-sum taxation plays a crucial role here. Otherwise the distortionary consequences of such high taxes would be considerable.) It remains true that debt cannot grow forever faster than disposable income.48

One of the most controversial recent papers investigating the inflationary implications of persistent primary budget deficits is that by Sargent and Wallace

48 These conclusions hold in per capita terms when the size of each household grows at a rate $n$ and utility is a function of per capita consumption and money holdings.
(1981). They assume that persistent bond-financed deficits will eventually have to be monetized. The reason they invoke is an upper bound on the public’s demand for bonds, but McCallum’s analysis is also pertinent here. The strategy they consider is “monetarist” up to a point and from then on it becomes “bondist”.

The fiscal authority is assumed to behave in its design of policy as a Stackelberg leader, whereas the monetary authority takes the role of a Stackelberg follower. All variations in the deficit considered are due to changes in $G$, since the size of after-tax per capita endowments is assumed constant and the economy is on its full-employment path with a real income growth equal to the constant rate of population growth, $n$. Monetary policy is determined by the choice of the rate of growth of the nominal stock of high-powered money, $H$. This rate of growth is assumed to be equal to $\theta$ up to time $t^*$, when the assumed arbitrary upper limit to private sector demand for bonds is attained. Thus, for $t < t^*$, the amount of bond financing is residually determined by the size of the primary deficit, the government budget constraint, and $\theta$. From then on the amount of bond financing is determined by the requirement that the per capita bond stock be kept constant at that maximum level, while it is the rate of growth of $H$ that is now residually determined.

The private sector in their model consists of the “rich” and the “poor”, while the government consists of a fiscal and of a monetary authority. A transactions motive for holding money is not incorporated, nor is there any element of uncertainty. Bonds and claims to capital are assumed to be of sufficiently large denominations for the poor not to be able to afford them, and legal restrictions are assumed (somewhat artificially) to prevent the poor from pooling funds and the rich from acting as financial intermediaries for the poor. Thus, money is held only by the poor, while the rich hold their wealth in the form of bonds and of capital which bear the same real rate of return, $R$. It is assumed that $R$ is fixed and that it exceeds the rate of population growth and the rate of return on money (equal to minus the inflation rate).

Buiter (1982) doubts whether this setup is appropriate for discussing issues of inflation and of government debt. First, in Sargent and Wallace’s formulation, government spending, $G(t)$, is wasteful, since it does not enter in private utility functions and reduces the amount of resources available for consumption or investment. Second, the presence of the postulated constraints on the portfolio behavior of the poor imply that Pareto-optimal policies should involve a deflation rate of $-R$, thus making available to all agents intertemporal market terms of trade equal to the technological intertemporal terms of trade for the economy as a whole. Third, to the extent that the constraints imposed on the poor imply underaccumulation of capital, the government should act as a net lender to the private sector and/or give subsidies to the rich.
These objections notwithstanding, two interesting questions have been asked in the context of this model. First, what are the effects on the time path of inflation of adopting a more restrictive monetary policy today? Second, what are the effects on inflation of fiscal actions that result in larger real per capita total deficits? This second question includes cases of "fiscal irresponsibility", where fiscal authorities allow the deficit to exceed the maximum seignorage obtainable in steady state. We discuss both questions here.

6.2.1. Effects of monetary restriction

The effects of tighter money today are the focus of the Sargent–Wallace paper. On the basis of the setup described above, the authors derive two results. The first is that when money demand is independent of the inflation rate, a lower value of $\theta$ (and thus a lower rate of inflation) for $t < t^*$ always necessitates a higher rate of growth of money and higher inflation after $t^*$. This is because the lower $\theta$ for $t < t^*$ causes a higher level of the real per capita bond stock to be attained at $t - t^*$ and maintained thereafter. Since $R$ is assumed to exceed $\alpha$, this means that the size of the real per capita debt service (corrected for inflation and growth) for $t \geq t^*$ is larger. As a result, the rate of monetary growth (and consequently inflation) is also larger beyond $t^*$ in order to finance the higher real per capita total (as opposed to primary) deficit.

The second result concerns the case where money demand decreases with expected inflation. Current inflation depends on the entire anticipated future path of the money supply. When a lower $\theta$ for $t < t^*$ implies higher rates of monetary growth after $t^*$, it becomes possible to construct examples in which these higher later rates dominate the lower earlier rates and produce higher inflation even before $t^*$. It should be noted, however, that this occurs only for specific parameter configurations.

Darby (1984) has noted that the Sargent–Wallace result in which monetary policy is fully dictated by the stance of fiscal policy requires that the real after-tax rate on bonds exceed the natural rate of growth of the economy. When the sign of this inequality is reversed, a given budget deficit can be dynamically consistent with a range of values for $\theta$. As long as $\theta$ remains within this range, changes in monetary policy can take place today without having to be reversed over the longer run. Darby points to evidence showing that even long-term, before-tax yields on government bonds have not approached corresponding growth rates of real output in the United States. This is a fortiori true of after-tax yields.\textsuperscript{49} Thus, there may be more flexibility in the choice of monetary policy than implied by the Sargent–Wallace analysis.

Butter (1983b) suggests that even when the real rate on bonds exceeds the natural rate of growth and when money demand depends on expected inflation,

\textsuperscript{49}This evidence is consistent with the analysis of Tobin (1986).
a lower value of $\theta$ for $t < t^*$ need not imply a higher rate of growth of the nominal money stock and higher inflation after $t^*$. While it is still true that the inflation rate for $t \geq t^*$ is higher the higher the real per capita bond stock, it is not clear that a lower value of $\theta$ for $t < t^*$ always implies a higher real per capita bond stock from $t^*$ onwards. Specifically, a lower value of $\theta$ implies a higher real per capita bond stock for $t \geq t^*$ if it lowers the real discounted present value of the government's new money creation between $t = 0$ and $t = t^*$.

6.2.2. Effects of larger deficits

Buiter (1985) explores the consequences for the time path of inflation of fiscal policies that result in large real per capita total deficits, including the case where they exceed the "maximum seignorage" obtainable in a steady state. Real seignorage is defined as the amount of real resources that a government acquires in a period simply by virtue of the fact that private agents will hold the currency it prints. The per capita measure of this is equal in the steady state to the product of the real money stock times the sum of the inflation rate and the natural rate of growth of output. By raising inflation, the government also lowers real money demand. The two effects go against each other and there generally exists a maximum amount of real per capita seignorage that the government can extract.

Buiter examines the properties of the differential equation derived from the government budget constraint, assuming that the real per capita bond stock is kept constant at $b$ and that nominal money demand per capita, $m$, depends inversely on the inflation rate, $\pi$:

$$\dot{m} = d + (R - n)b - (\pi + n)m,$$

$$m = c_1 - c_2\pi, \quad c_1, c_2 > 0,$$

where $d$ is the real primary deficit per capita, $n$ is the natural rate of growth of output, and $R$ is the real rate of return on bonds. The expression on the right-hand side of equation (6.1) is the per capita measure of the real, inflation-and growth-corrected budget deficit. Excessively high values of this measure are regarded as instances of "fiscal irresponsibility".

The point Buiter makes is that in the Sargent–Wallace model, the effect of fiscal irresponsibility can never be hyperinflation. In other words, fiscal policy in this model cannot generate unbounded increases in the inflation rate. Specifically, in the case considered by Sargent and Wallace, where $d + (R - n)b$ is below the maximum amount of seignorage (or inflation tax) that could be attained in a steady state of the system, the above equation of motion has two equilibria $(m, \pi)$, and there seems to be no economic criterion for
choosing between them. If we consider the locally unstable equilibrium (as Sargent and Wallace do), an increase in the deficit will indeed raise the long-run rate of inflation, but by a finite amount. Alternatively, if the locally stable equilibrium is chosen, a finite reduction in the long-run inflation rate will result. This difference in signs mirrors the opposite effects on the amount of steady-state seignorage generated by a reduction in monetary growth.

However, it is also conceivable that the real total per capita deficit (adjusted for inflation and growth) will be set by the fiscal authority so irresponsibly as to exceed the maximum seignorage obtainable in the steady state. Buitter shows that in this case enough seignorage revenue could be generated so as to cover this difference through a process involving continuously increasing $m$, and continuously falling rates of inflation and monetary growth outside the steady state. This process is one of "hyperdisinflation", and it is unsustainable in the Sargent–Wallace model, since $m$ in that model is bounded from above and cannot grow forever. The reason for this boundedness of $m$ is that money is held by the poor to finance second-period consumption and both income and the supply of consumer goods are bounded. Indeed, some have argued (not entirely convincingly) that an unsustainable process would not even get started in a rational expectations world. The conclusion is that the case against fiscal irresponsibility provided by the Sargent–Wallace model emphasizing inflation, cannot rest on fears of hyperinflation, but only (at worst) on fears of finite increases in inflation rates.

7. Concluding remarks

There are two ways to evaluate a theoretical hypothesis. One is to ask how well it survives relaxations of assumptions, especially to see how dependent it is on patently unrealistic premises. An alternative criterion is the consistency of the theory's implications with empirical observations, the validity of its predictions. This test ignores the plausibility of assumptions. According to Milton Friedman's "methodology of positive economics", the crucial question is whether the economy behaves as if the theory were valid.

Our pragmatic view, following Tjalling Koopmans, is that all opportunities for testing should be seized, from the plausibility of primitive assumptions to the congruence of ultimate implications with observations. Empirical testing is too difficult, and too often ambiguous, to permit us to rely exclusively on the "as if" criterion in choosing among hypotheses.

Let us try to evaluate the three strong neutrality propositions reviewed in Sections 4, 5, and 6 by both kinds of criteria.

Debt neutrality fails the first criterion by a wide margin. In Section 4 we reviewed a series of papers showing how sensitive the Barro–Ricardo Equivalence theorem is to relaxing any one of numerous assumptions. Few economists
can believe that all those assumptions hold in practice. But the profession, including even some of the harshest critics of debt neutrality, seems willing to overlook such unrealism and focus on the “as if” criterion instead.

We discussed the empirical literature but found the results inconclusive. More empirical research is needed and should be possible. Recent data on saving rates and deficits, especially in the United States, seem prima facie difficult to reconcile with debt neutrality. However, the inability of researchers to agree, even for identical sample periods and similar sets of variables, is not encouraging. We seem to need new tests as well as new data.

To reject debt neutrality is not to dismiss the sensible idea that expectations of future taxes, as of other future policies, affect current consumption and saving decisions, especially when there are credible signals of political and legislative intent.

We are reluctant to recommend more research on deficits, given the already huge volume of literature on the subject. However, there seems to be room for analysis of the effects of persistent budget deficits in a greater variety of circumstances. For example, what are the consequences of unsustainable policies not reversed before the situation gets “out of hand”? Both the profession and the lay public need a plausible “doomsday scenario” in debating the needs for and merits of fiscal austerity programs.

Superneutrality is the subject of a considerable literature examining critically its necessary assumptions. One necessary assumption for superneutrality, as well as for debt neutrality, is that consumer-savers’ horizons are infinite. But this is not sufficient; a number of interesting papers identify departures from superneutrality even with infinite horizons. Although the literature provides a fairly good understanding of cases in which superneutrality fails, it is more difficult to dismiss this proposition as dependent on patently unrealistic assumptions than to dismiss debt neutrality. There is room for more research both on identifying conditions implying departures from superneutrality and on evaluating the realism of such conditions.

The most obvious gap, however, is in the empirical literature on superneutrality. The Fisher hypothesis has received much attention, but few papers test superneutrality directly. Perhaps more could be done following the methodology of Sargent’s (1973) paper, specifying a model where superneutrality holds and carrying out joint tests of superneutrality and the underlying model. Sargent focused on the Fisher hypothesis, but the same methods could be used to test invariance of real per capita consumption or of the capital stock.

Testing superneutrality should not be the sole aim of research on the effects of inflation on capital accumulation. If there are such effects, it is important to identify their direction and importance, whether positive because of Mundell or Tobin effects or negative because of the real costs of economizing money balances.
Monetary policies, distinguished from fiscal policies, involve exchanges of assets between the government or its central bank and private agents. Models exhibiting irrelevance of asset exchanges show entire classes of policies that support the same equilibrium allocations. If this is correct, it does simplify the policy-maker's problem! Within this literature, there is still room for models exploring the precise conditions under which such policy classes exist. It is also interesting to examine whether the equilibrium in question is the only one supported by each policy in a given class. If there are multiple equilibria associated with each policy, irrelevance classes may be difficult to define.

Whether open market operations as typically implemented in practice are irrelevant, is a different question. A reader of this literature must keep in mind exactly what assets are being exchanged in each model. Some papers redefine the terms "open market operations" and "constant fiscal policy" and then try to justify their model-based redefinitions. It does not then follow that all feasible asset exchanges are irrelevant.

There is little empirical research related to this species of irrelevance literature. One reason may be that the asset swaps in the models do not resemble actually observed central bank operations. However, it should be possible to find historical approximations to at least some of the modeled asset exchanges and to test directly the hypothesis of neutrality. Indeed, if there are no such incidents, one may be tempted to ask whether it is the literature that is irrelevant, i.e. irrelevant to the monetary institutions and practices of real-world economies.

The foregoing remarks were stimulated by our surveys of neutrality literature in Sections 4–6. As they indicate, contemporary theory of fiscal and monetary policy is far removed from the practical concerns of policy-makers. In Section 3 we explained how current "microfoundations" methodology inevitably creates a wide gulf between theory and application. There was not always such a gulf. As our review in Section 2 of older traditions in macro-economics indicates, fiscal theory and fiscal policy were once closely linked, each contributing to the other. The same was true of monetary theory and monetary policy. The major challenge to theorists today is to model enough of the heterogeneities, institutional idiosyncrasies, and market imperfections of actual economies to make their theories useful to empirical researchers and interesting to policy practitioners.

References

Abel, A.B and B.D Bernheim (1986) 'Fiscal policy with impure intergenerational altruism', NBER.


Ch. 17: The Macroeconomics of Government Finance


Modigliani, F. (1964) ‘Long-run implications of alternative fiscal policies and the burden of the...


Nelson, C R and G W Schwert (1977) 'Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant', American Economic Review, 67: 478–488.


Ricardo, D (1817) On the principles of political economy and taxation, edited by Piero Sraffa with the collaboration of M H Dobb Cambridge: Cambridge University Press for the Royal Economic Society (1951)


Siegel, J J (1983) 'Technological change and the superneutrality of money', Journal of Money, Credit and Banking, 15: 563–567


Ch. 17 The Macroeconomics of Government Finance