CONTROVERSY OVER THE FEDERAL BUDGET DEFICIT: A THEORETICAL PERSPECTIVE

Michael Dotsey

I. INTRODUCTION

Currently there is a great deal of controversy surrounding the federal budget deficit. Claims have been made that the deficit is a burden on future generations and that it crowds out private investment and reduces the capital stock. For example, in testimony before the Senate Budget Committee, Rudolph Penner, director of the Congressional Budget Office testified “Certainly one consequence of persistent deficits about which there seems little disagreement is their adverse effect on future generations . . . . Slower growth of the private capital stock would result in lower productivity than would occur with smaller deficits and the income of future generations would be lower.” This crowding-out of investment is alleged to occur because budget deficits supposedly raise real rates of interest. The same theme is also present in the Congressional testimony of Federal Reserve Board Chairman Paul Volcker as well as in other public statements. “In essence, the demands of the Federal Government limit the rate of growth of other credit-absorbing sectors of the economy. The rationing device is interest rates held higher than would otherwise be the case.” The basic evils of budget deficits and the benefits of cutting the deficit are echoed in numerous market newsletters and by many private economists. For instance, Martin Feldstein in testimony before the Senate Finance Committee states “The short term effect of enacting a deficit reduction program of that magnitude would be a substantial decline in medium term and long term real interest rates . . . .”

There is a minority view that budget deficits do not have much consequence for aggregate economic activity. A major proponent of this line of argument is the University of Rochester’s Professor Robert J. Barro, who has resurrected the notion of Ricardian equivalence or neutrality of budget deficits. In a recent Wall Street Journal article, Barro states “There is no evidence that shifts in deficits lead to higher rates, especially to higher real rates that would deter investment . . . . Basically, this crowding-out view of deficits is a myth, which is reinforced mainly by repetition.” These opposing viewpoints are irreconcilable and yet each finds its adherents among respected economists. This article attempts to place these viewpoints in the context of the economic models from which they are derived. This procedure focuses attention on the assumptions that underlie the various conclusions concerning the effects of budget deficits and helps appraise the realism or degree of confidence that should be placed in these various views.

In addition, the article will investigate a fairly recent area of research concerning optimal deficits. This literature concludes that the optimal level of government debt and budget deficits is not generally zero. The implication is that given an exogenous time path for government debt and budget deficits is not generally zero. The implication is that given an exogenous time path for government spending, a policy of balancing the government budget in each and every period is costly in terms of social welfare. Further the less stringent requirement of the Balance Budget Amendment that passed the Senate on October 1, 1982 does not generally represent a policy that is consistent with the literature on optimal deficits.

* I would like to thank Marvin Goodfriend, Thomas Humphrey, and Roy Webb for their valuable comments and suggestions.


The optimal deficits literature does not consider the potential relationship between taxing and government spending. Some proponents of the Balanced Budget Amendment (see Friedman and Friedman [13]), feel that there exists a link between the level of government spending and the constraint of paying for spending as you go. However, I am unaware of any systematic discussion concerning this issue. More importantly, there are specific portions of the Balanced Budget Amendment that deal with the level of government spending directly by attempting to remove certain features of the current budgetary process that are likely to produce a nonoptimal level of government spending. These direct limitations on spending could be enacted independently of the requirement that the expected budget deficit be zero.
The organization of the article follows a historical approach to the theory of budget deficits. The departure point is the standard Keynesian model which is discussed in Section II. Section III analyzes some shortcomings of this model and the conclusions drawn from it. Section IV moves on to the notion of Ricardian equivalence, while Section V discusses criticisms concerning this theory. Section VI examines the notion of optimal budget deficits, while Section VII contains a brief summary and conclusions.

II. THE EFFECTS OF BUDGET DEFICITS IN A STANDARD KEYNESIAN MODEL

The initial analysis of the effects of budget deficits will be carried out within the framework of a standard Keynesian model. Starting the analysis in this manner will provide the theoretical background necessary for evaluating much of the current discussion involving deficits. It will also serve as a useful point of departure for examining the more recent rediscovery of Ricardian equivalence.

The Goods Market

The analysis begins with the closed economy accounting identity of equation (1) that aggregate output \( Y \) is equal to the sum of aggregate consumption \( C \), aggregate investment \( I \), and government spending \( G \).

\[ Y = C + I + G \]

In order to give this identity economic content, the behavior of consumption, investment, and government spending must be examined. Typically, in these models consumption is hypothesized to depend positively on disposable income \( Y_d \), which is merely income net of taxes \( T \). That is, \( Y_d = Y - T \) and consumption is given by an equation such as (2).

\[ C = c_0 + c_1Y_d \quad c_0 > 0 \quad 0 < c_1 < 1 \]

The coefficient \( c_1 \) is greater than zero and less than one implying that only a portion of an extra dollar of disposable income will be spent. It will be shown that the nature of the consumption function is crucial in determining the effects of a budget deficit.

For simplicity, let investment be written as a negative function of the real rate of interest. At higher real interest rates fewer projects are profitable, since the opportunity cost of funding a project is higher. Therefore, at higher real rates of interest there will be less investment. This is depicted by equation (3)

\[ I = I_0 - i_1r \]

where \( r \) is the real rate of interest.

Finally, let government spending be exogenously given at the level \( G \). The analysis concentrates on the consequences of changes in financing a given level of government expenditure and not on the effects of government expenditure itself. Holding government spending fixed at a predetermined level will allow the analysis to focus on the financing decision without introducing the complicating effects of government spending. Also, taxes are initially assumed to be lump sum in nature at the level \( T \).

That is, the level of taxes an individual pays is independent of his income. Within the context of the model currently under consideration, this assumption is of little consequence. However, as the analysis proceeds to examine the nature of budget deficits in more detail the distinction between lump sum taxes and proportional taxes will be important.

The income levels and interest rates consistent with clearing of the goods market can be derived from equations (1), (2), and (3). This is expressed as equation (4), and shows the negative relationship between income levels and interest rates that are needed to clear the goods market.

\[ r = \frac{C_0 + I_0 + G_1 - c_1T_1}{i_1} - \frac{(1 - c_1)}{i_1} Y_1 \]

To understand this negative relationship, perform the following thought experiment. Suppose that at an income level of \( Y_0 \), and a real interest rate \( r_0 \), that income equaled consumption plus investment plus government spending. Now suppose that income increased to \( Y_1 \). Since consumption only increases by a fraction of the rise in income, it is easy to see that the goods market is no longer cleared. Output, \( Y_1 \), is now greater than the sum of \( C + I + G \). If the real rate of interest were to fall (to, say, \( r \)), so that investment rose, the goods market would once again clear. Therefore, the locus of points depicting market clearing in the goods market, commonly referred to as the IS curve, is downward sloping as indicated in Figure 1.

The Money Market

As shown, the IS curve defines an infinite number of real interest rates and income pairs that are consistent with equilibrium in the goods market. In order to find the unique level of the real interest rate and income that will arise in our model economy, the money market must be analyzed. Equilibrium in the money market is given by values of income and
nominal interest rates that equate the demand for money with the supply of money. For simplicity, the analysis assumes that inflation is zero and therefore that nominal and real interest are equivalent. This locus of points, commonly called the LM curve, is upward sloping and is drawn in Figure 1. An intuitive explanation relies on the observation that the demand for money is positively related to income and negatively related to interest rates. Suppose that money demand were equal to money supply at \( r_l \) and \( Y_1 \). An increase in income to, say, \( Y_2 \) would increase the demand for money making it greater than a given money supply. Therefore, to return to equilibrium the interest rate would have to rise (to, say, \( r_2 \)) lowering the demand for money so that it once again equals money supply.

**Equilibrium and the Effects of Budget Deficits**

Equilibrium for the economy as a whole occurs when demand equals supply in both the money market and goods market simultaneously. This is depicted in Figure 1, and occurs at the point \( E_1 \) with an interest rate of \( r_l \) and an income level of \( Y_1 \). That is, the point \( E_1 \) is consistent with government spending of \( G_1 \) and taxes of \( T_1 \). Now let the government decide to increase the deficit by lowering \( T_1 \), while holding government spending fixed. This has no initial effect on the money market since taxes do not directly affect the demand or supply of money. Therefore, the LM curve does not shift. However, a decrease in taxes increases disposable income for any given level of income and therefore raises desired consumption. This means that aggregate demand is now greater than income at \( Y_1 \), and that a higher rate of interest, which lowers investment must now be associated with \( Y_1 \) in order to return to equilibrium. The result is that the IS curve will shift upward and to the right (to IS'). The economy wide equilibrium is now given by \( E_2 \), with both income and interest rates higher than at \( E_1 \). In the Keynesian system budget deficits are seen to be expansionary raising the level of both output and interest rates.

The increased budget deficit also changes the composition of output. Since the interest rate has risen, investment will have fallen. This interest reduced decline in investment is commonly called the crowding-out effect of a budget deficit. Also, since government spending has not changed, consumption must have risen substantially in order for output to rise on net. Therefore, an increase in government deficits causes the proportion of output comprised by consumption to rise while the ratio of investment to output falls. With lower investment, the future capital stock and productivity of the economy will fall. The opinions of many of the economists quoted in the introduction therefore seem to be based on the Keynesian model of the economy.

**III. SHORTCOMINGS OF THE KEYNESIAN ANALYSIS**

A major problem with the Keynesian analysis presented in Section II is the myopic nature of individual behavior embodied in the consumption function (2). Specifically, individuals are assumed to consume a constant proportion of current disposable income regardless of the future time path of income or taxes. An individual with $20,000 of disposable income in each period of his life would consume exactly the same amount as an individual whose disposable income alternated between $20,000 and $0 over his lifetime, with his current disposable income.
being $20,000. The permanent income hypothesis serves as an alternative way of modelling consumption. According to this theory consumption depends on the present value of lifetime disposable income or wealth. Making this one modification essentially negates the entire effects of a budget deficit that was presented in Section II.

To show this, assume that the world is two periods long (a period being of indeterminate length) and that current consumption depends on the present value of lifetime disposable income, \( W \), where

\[
W = Y_1 - T_1 + \frac{Y_2 - T_2}{1 + r}
\]

and where the subscripts 1 and 2 denote time periods one and two. Equation (2) is then modified to read

\[
(2') \quad C_1 = aW. \quad \text{5}
\]

Since the model has taken on dynamic considerations that were not present in Section II, government behavior must be analyzed in more detail. Specifically, a constraint requiring the government to balance its budget on average is imposed. That is, the present value of government spending must equal the present value of tax receipts, but the government budget need not balance in any particular period. Specifically,

\[
(6) \quad G_1 + \frac{G_2}{1 + r} = T_1 + \frac{T_2}{1 + r}
\]

This constraint is reasonable since no one would lend to the government without the promise of being paid back. Suppose that the government ran a deficit of \( D_1 \) in period one (i.e., \( G_1 - T_1 = D_1 \)). If the government doesn’t raise enough revenue in period two to cover its spending plus the repayment of its debt and interest on the debt, no one would lend to the government in period one. Basically, in order to borrow the government must promise to repay or no one will lend to it in the first place. Extending the model to include more than two periods does not change the underlying justification for imposing a government budget constraint.\text{6}

Using (5) and (6) it is straightforward to show that an increase in the current budget deficit will have no effect on the economy. Government spending is given and invariant and hence so is the present value of taxes. Therefore, wealth will not be changed by a change in the timing of taxes so long as the present value of taxes remains unchanged. Therefore, consumption will not be changed and the initial level of output and interest rates will be consistent with equilibrium. In essence, the decrease in taxes in period 1 will be used by consumers to purchase the additional bonds sold by the government to finance its unchanged level of spending \( G_1 \). These bonds will then be sold to pay for the increase in taxes that must occur in period 2. As a result, consumption is unchanged and the effect of the government deficit is nil.\text{7} The basic thrust of this analysis underlies the Ricardian equivalence proposition, which is discussed in more detail in the next section.

IV. RICARDIAN EQUIVALENCE

The model presented in Section III still has a number of unrealistic features. However, it should be noted that the results of Section II rest on a consumption function that has very little analytical appeal. First and foremost is the assumption that individuals and the government have the same life span or time horizon. One could alternatively view the government as having an infinite time horizon, with individuals being finite lived. In that case the government need not raise taxes in period 2 to offset the decline in taxes in period 1, but could wait to some point in the distant future before bringing its budget into balance. The absence of offsetting taxes would imply a rise in current individual wealth and an increase in current consumption. Consequently, the results of Section II would obtain.

Barro’s [2] contribution has led to the reemergence of the notion of Ricardian equivalence, or that pure government financing decisions do not matter even in the case of finite lived individuals. He

\text{6}This analysis neglects the potential effects of liquidity constraints which will be dealt with in more detail in Section IV.

\text{7}This parameter \( a \) will depend on the individual’s rate of time preference and perhaps other attributes of his utility function. For the case where the rate of time preference is equal to \((1 + r)\), \( a = \frac{1}{2} \).

\text{7}This line of argument does not exactly hold when the number of periods is infinite. However, if the economy grows at rate \( n < r \), then the present value of the future taxes that must be levied to finance the interest on the debt net of debt finance is exactly equal to the original debt issuance (see Barro [3]). McCallum [15] shows that in an economy with a fixed factor (such as land) that inefficient capital accumulation does not occur and that \( n < r \). This inequality holds (without land) if both agents and the government are infinitely lived. Barro [3, p. 345] conjectures that it holds in an overlapping generations model in which bequests are operational.

\text{7}This analysis would carry over to proportional income taxes, since in the current setting the supply of labor is fixed.
employs the reasonable assumption that members of a family care about each other's welfare. This assumption ties one generation to another and makes the problem equivalent to analyzing individuals as if they were infinitely lived. Since individuals effectively have the same time horizon as the government, any restructuring of debt and taxes does not alter individual wealth. Therefore, there are no changes in behavior due to an increase in government debt.

The nature of Ricardian equivalence can be illustrated by the following simple overlapping generations model. In this model, an individual is assumed to live for two periods. Further, he is assumed to care about his own consumption and the attainable utility or welfare of his heir. For simplicity, the analysis considers an economy with a constant population, that is, one in which there is one child per parent. An individual maximizes his lifetime utility subject to a lifetime budget constraint. His earnings include wages when young and generally a bequest from his parent that he receives in the second period of his life. Formally, an individual of generation \( i \) maximizes utility \( U_i \) subject to a lifetime budget constraint of

\[
(7a) \quad \mathbf{w} - T_i = c_i^y + A_i^o
\]

and a second period budget constraint of

\[
(7b) \quad A_i^y(1+r) + A_{i-1}^o(1+r) = c_i^o + A_i^o
\]

where \( U_i \) is the utility of generation \( i \), \( c_i \) is consumption of generation \( i \), \( T_i \) are taxes levied, and \( A_i \) are assets of generation \( i \). The superscripts \( y \) and \( o \) refer to the periods when the individual is young and old. Therefore, \( A_i^y \) refers to assets accumulated while a member of generation \( i \) is young and \( A_i^o \) are the assets accumulated by agent \( i \) when old that will serve as a bequest to his heir who is a member of generation \( i+1 \). Wage earnings are denoted by \( \mathbf{w} \), while the real rate of interest is given by \( r \).

In this framework, the consumption and asset demands of the old and young can be written as functions of their net-of-tax bequests, wages, and interest rates. This is most easily seen by combining the budget constraints of an agent when young and old (7(a) and 7(b)) into a total lifetime budget constraint given by

\[
(8) \quad w(1+r) + A_{i-1}^o(1+r) - T_i(1+r) = c_i^y + c_i^o + A_i^o
\]

This implies that the maximum utility level of an individual is indirectly determined by his wages, his bequest from his parent, and the interest rate.

Suppose that the government engages in deficit financing by lowering the taxes paid by generation \( i \) and raising the taxes of generation \( i+1 \) by enough to pay off the deficit plus accumulated interest on the additional debt. Within the framework of the previous section, since the \( i^{th} \) generation is paying less taxes over its lifetime, members would feel wealthier and increase their consumption in period \( i \). The deficit would be expansionary. However, that need not be the case when generation \( i \) is concerned about the welfare of generation \( i+1 \) as they are currently assumed to be. If generation \( i \) is sufficiently concerned to leave positive bequests to generation \( i+1 \), then there will be no wealth effect of the government deficit in period \( i \).

To see this update and reexamine (8), and recall that the attainable utility of a member of generation \( i+1 \), \( U_{i+1}^* \), depends on his bequest net of taxes. The government deficit in period \( i \) is financed by an increase in generation \( i+1 \)'s taxes. Essentially, this financing is equivalent to the government taking away money from generation \( i+1 \) and giving it to generation \( i \). However, since generation \( i \) had decided to give their descendants a bequest, this option was already available to them in terms of lowering their bequests. That generation \( i \) did not do this means that they did not find the lower bequests optimal in terms of maximizing their own level of utility. Hence, in order to undo what the government has done, members of generation \( i \) will increase their bequests by the amount of the deficit. Essentially, the government deficit forced individuals in generation \( i \) away from an optimal allocation of resources between generations by taking away income from generation \( i+1 \) and giving it to generation \( i \). Generation \( i \) can easily offset the actions of the government by increasing its bequests leaving the net bequest to its heirs unchanged. In doing so the entire profile of market equilibria is unchanged and the government deficit is neutral. The same results occur if the analysis is extended to having taxes being paid by generations further in the future.

\[8 \] For necessary conditions on the utility function see Drazen [11].
V. CRITICISMS OF RICARDIAN EQUIVALENCE

The Existence of Bequests

The first and perhaps most obvious criticism of Ricardian equivalence is the assumption of a bequest motive. As shown in Section IV, a necessary condition for the neutrality of government deficits is the existence of operational intergenerational transfers. If bequests between generations are not often observed, then the argument presented above is merely a theoretical exercise with little policy content. However, it is more difficult to dismiss the bequest motive than is often realized. For example, bequests can be of a very general form, such as money spent on a child’s education, and need not only represent money left in a will. In fact, any intergenerational transfer of resources will do. Casual empiricism seems to indicate that there does indeed exist some form of intergenerational transfers in most families. If so, deficits will have an expansionary effect only to the extent that some members of the older generation do not want to leave a bequest, and one would probably expect this effect to be small.

Also, operational intergenerational transfers need not only flow from the old to the young to derive the neutrality results of Section IV. Transfers of gifts from the young to the old would lead to the same sort of conclusions. Therefore, in a situation where wages were growing at a relatively fast rate (in excess, say, of the real rate of interest), making bequests from old to young less likely, the existence of gifts from young to old would serve to preserve Ricardian equivalence.

The Difference between Investment in Education and Bequests

As mentioned, intergenerational transfers are a necessary condition for deriving the Ricardian equivalence proposition. Also, no distinction is made between investment in an heir’s education or human capital and a monetary bequest. However, Drazen [11] points out that there may be important differences between these two types of bequests. Specifically, under the condition of unidirectional bequests from old to young that are primarily in the form of investments in human capital, he finds that government deficits may imply a change in net wealth, and therefore have real effects.

His argument is based on the potential difference between the return to human capital, \( r' \), and the real rate of interest, \( r \). Suppose it is reasonably assumed that the return on human capital is positively related to the amount invested in education, but that each additional dollar of investment yields a slightly lower return than the previous dollar invested. That is, there are diminishing marginal returns to investment.

Under these assumptions, and since people do indeed invest in education, the first dollar so invested yields a return greater than the real rate of interest. Indeed, one may suppose that the return on education is greater than the return on capital up to some critical value of educational investment, denoted by \( \bar{e} \). In this setting the optimal method of bequests would be for parents to first make investments in their child’s education up to an amount \( \bar{e} \), and if further bequests are desired to leave them in the form of a monetary transfer. In the case where the optimal level of bequests was less than \( \bar{e} \), the return on human capital would be greater than the real rate of interest. In this setting, government debt issuance can have a real effect. This real rate of government debt results from a failure of private markets to equalize rates of return among various investments. Alternatively, the outcome in which \( e < \bar{e} \) represents a situation of underinvestment in human capital.

It is this existence of a private market failure that creates the possibility for government debt to have real effects. This can be seen by reexamining the analysis of Section IV for the case where \( e < \bar{e} \). Suppose the government decreases the taxes of the parents (generation \( i \)) when they are young and finances some government spending through the creation of debt. It then raises the taxes of the young (generation \( i+1 \)) to pay off the debt as well as the interest on the debt. In Section IV, it was shown that a dollar for dollar change in bequests entirely offsets the government financing decision. However, since the return on human capital is greater than the real rate of interest, parents do not have to invest all of the tax cut in their child’s education to leave their child’s attainable utility level unaffected. The parent only has to invest a portion of the tax cut in his child’s education, and could spend the rest. A more likely occurrence would be for the parent to invest enough in the child’s education to make the child slightly better off and to spend the rest making the parent slightly better off. As long as human capital possesses a greater return than the real rate of interest both generations can be made better off by an increase in government debt, even though the increase in debt will raise the real rate of interest. In this case it is inappropriate to refer to debt as a burden on future generations.
Again it should be stressed that Drazen’s results occur because of a market failure. This market failure arises for three basic reasons. One is the reasonable assumption that there is no market in human capital. That is, one cannot generally enter into a contract with his son saying that I will invest so much in your education and force the son to repay the loan at an interest rate equivalent to the return on human capital. Another is that the son cannot invest in his own human capital. This seems less reasonable, since educational loans (at the post-high-school level) are readily available, and it would not be optimal to forego investing in one’s own education so long as the rate of return was greater than the real rate. Allowing for own investment would tend to equalize the real rate of interest and the return on human capital, making changes in government financing neutral.

The third reason is the omission of a gift motive from son to father. If one also allows the son to care about his father, then it seems that in the presence of operational transfers in both directions the situation in which \( r' > r \) would not represent an equilibrium. For instance, suppose that intergenerational transfers are operative in both directions and that \( r' > r \). The father will be investing in his son’s human capital because the return to the son is greater than the cost to the father. Further, the son gives gifts to his father because the return on his education may make him better off in real terms than his father. Now with \( r' > r \), the father could invest an additional dollar in his son’s education making his son’s attainable utility higher while decreasing his own utility. However, we have assumed that the son cares about the father and that he gives gifts. That is, with bidirectional transfers operational, a gift of a dollar from the father (at the margin) will be returned by the son leaving the proposed equilibrium unchanged. But, because the return on the investment made in his education is sufficient to both fully reimburse his father and still leave enough to make himself better off the situation with \( r' > r \) is not optimal. That is, it pays for the father to invest in the son up to the point where \( r' = r \).

Alternatively, suppose that the only form of bequest is investment in education (i.e., the case where \( e < e' \)), and that the bequests motive is operational in both directions. Suppose the government takes a dollar from the father (when young) and gives it with interest to the son. Since the son was already giving a gift to his father, he could have consumed more if he desired. That he did not implies that he will give the dollar plus interest back to his father. The father, therefore, knows that an additional bequest to the son will be returned with interest. However, since \( e < e' \), a dollar invested in education will yield the son more than the real rate of return. This means that the son will return at least a dollar plus interest to the father, while still having more for himself. Both will be better off, and the situation with \( r' < r \) is inefficient. That is the combined budget sets of both the son and the father will be maximized when investment in education is made up to the point where \( r' = r \). Having transfers operate in both directions essentially opens a market in human capital between the father and son, driving the return on human capital down to \( r \). In this situation there are no longer any real effects of government debt.

**Liquidity Constraints**

The situation in which some consumers are liquidity constrained also opens up an avenue that allows government debt to have real consequences. It is sometimes assumed that individuals cannot always borrow to the extent that they desire, since capital markets may not lend money based on future risky income streams. Or, perhaps, individuals knowing themselves better than a lending officer discount their risky income at a lower rate. In any event, if individuals are constrained to some extent from borrowing against risky future income streams, budget deficits will have real effects. The size of the effects will depend on the proportion of income earned by those who are constrained.

The effect originates because some portion of the society would like to consume more of their wealth but are constrained from doing so. A cut in taxes today coupled with an appropriate increase in taxes tomorrow is equivalent to making a loan to constrained individuals at the rate \( r \). Since these individuals were constrained they will increase their consumption by at least some portion of the tax cut. This will cause the real rate of interest to rise. However, the effect on social welfare of the government debt is unclear. If the government is providing an intermediation service that was not available due to some private market failure then social welfare could be enhanced. However, the loan to constrained individuals is made at the government’s borrowing rate, which may not be the appropriate rate. This implies that some subsidization is occurring between con-

---

For more detail see Zeldes [21], [22].
strained and unconstrained individuals making the net effect on social welfare ambiguous.

Further, the apparent credit rationing may not be due to any private market failure, but could result as a consequence of optimal lending arrangements in private markets. In that case the intermediation process provided by the government may be non-optimal.

Related to this issue is the question regarding the relative efficiency of the government in providing intermediation services. If the government is more efficient than the private sector, then social welfare would be increased and the economic effects associated with a positive wealth effect of government debt would occur. The postulated relative efficiency of the government may be important in developing countries with unsophisticated financial sectors but is unlikely to be relevant for the U.S. economy. Also, in the case where the government does have a relative advantage in performing financial intermediation it is inappropriate to speak of the burden of public debt since social welfare is enhanced.

Uncertainty over Income Streams

To this point, the analysis of the paper has dealt with lump sum taxes rather than taxes that are related to income. Doing so has been of little consequence, since income is exogenously given and an individual’s income stream has been certain. (To see that this is true the reader need only replace $T$ by $tY$ in the preceding discussion, where $t$ is the proportional tax rate.) However, in an analysis that considers uncertain future income streams and risk averse individuals, a distinction exists between lump sum taxes and proportional income taxes, even if income is taken to be exogenous.

The reason for this is straightforward. Risk averse agents do not like risk. Therefore, the usual effect of a reduction in risk is to make agents feel wealthier implying that they will consume more. Consider an agent who lives two periods earning a known and therefore certain amount of income, $Y_1$, in period 1 and an uncertain amount of income, $Y_2$, in period 2. The variance of an individual’s after-tax income in period 1 is zero, since he knows $Y_1$ with certainty, while the variance of his period 2 after-tax income is $(1-t)^2$ times the variance of $Y_2$. Reducing this after-tax variability will make individuals better off and generally increase period 1 consumption.

The government can accomplish this by lowering the first period tax rate and increasing period 2’s tax rate by enough to cover the resulting increase in the deficit. Increasing $t$ will lower the $(1-t)^2$ component of the variance of after-tax income. Thus, individuals will be better off and will wish to consume more. This will cause the real rate of interest to rise.

In essence, individuals are better off because the government is operating an income insurance scheme. This improvement in welfare can be seen by more closely examining the taxes levied on a particular agent in period 2, $T_2 = tY_2$, which can be rewritten as

$$\tau Y_2 = \tau \bar{Y} + \tau (Y_2 - \bar{Y})$$

where $\bar{Y}$ is the known average income level for individuals in period 2. If our agent has a lower than average income level he receives an insurance payment, while if his income is relatively high he pays what amounts to an insurance premium. By changing $t$, the government can affect the degree of after-tax income risk. Also, if $\bar{Y}$ were uncertain, meaning that aggregate income was uncertain, then an increase in $t$ would lower aggregate uncertainty about after-tax income, resulting in a welfare gain.

The provision of this insurance policy by the government improves welfare. If the market were able to fully insure individuals against income risk there would be no room for government financing decisions to affect the economy. As in the previous discussion, it is the presence of a market failure that allows government financing decisions to be nonneutral. Therefore, the existence of an economic effect caused by government debt in the presence of uncertain income presumes that the government has a relative advantage in providing insurance.

Uncertainty over the Incidence of Future Taxes

If future tax rates are uncertain, perhaps because aggregate income is uncertain—therefore implying that the government is unsure about the income tax rate needed to raise necessary revenue—the wealth effects of government debt could be negative. This is because the variance of future after-tax income may now be higher.

---

11 A discussion of credit rationing in perfectly competitive markets is given in Stiglitz and Weiss [18].
12 For a more thorough discussion see Barro [2] and Chan [9].
13 See Chan [9] and Barsky, Mankiw, and Zeldes [7].
To isolate the effect of future income tax uncertainty, consider the case where taxes are lump sum and the $i^{th}$ individual pays taxes equal to $T + k_e$, where $e$ is a random number and $k_e$ represents the uncertain portion of tax payments faced by individual $i$. The sum of all random taxes is equal to zero. This means that some individuals pay more than average while some pay less. Now suppose that an increase in future taxes is associated with an increase in the uncertainty regarding the incidence of future taxes (i.e., an increase in $k$). That is, the more revenue the government needs to raise the greater is the uncertainty concerning who will have to pay. In this case, the increased uncertainty caused by an increase in debt makes agents worse off, and they therefore reduce consumption demand in period 1, resulting in a fall in interest rates. Combining both income and tax rate uncertainty into the analysis will make the net effect of government debt ambiguous.

**Summary of Ricardian Equivalence**

In Sections IV and V, an analysis of Ricardian equivalence has been presented along with a number of the major criticisms that have been raised against this theory. The major criticisms involve the absence of a bequest motive, and the existence of various market failures. On net these criticisms do not seem to be capable of generating any large wealth effects of government debt. The proportion of society that does not exhibit any intergenerational altruism does not seem to be very large. Also, in a country like the United States with sophisticated financial markets, the importance of liquidity constraints resulting from private market failures may not be very great. Further, it is hard to believe that the government is more efficient at providing intermedial services. Finally, the combined effects of income uncertainty and tax incidence uncertainty are ambiguous. Therefore, when income is exogenous, the Ricardian theory of government debt appears to hold up fairly well and one would expect the wealth effects of government debt to at best be fairly small.

However, the exogeneity of income is a questionable assumption. Under this assumption the amount that individuals work is independent of their tax rate. This means that no consideration is given to substitution between labor and leisure, and that when no wealth effects result from purely government financing decisions there are no social losses imposed by the tax system. The implication is that for a given path of government spending one amount of debt is as good as another. If so, then the intensity of the current debate over the size of the deficit is hard to understand. However, it could also be that the assumption of income exogeneity is inappropriate and that relaxing this assumption will provide some useful insights into the importance of deficits. This step is taken in the next section. The importance of the Ricardian equivalence theory is that it has forced one to take this step, since it suggests that wealth effects are not a major factor in understanding the effects of government debt.

**VI. THE OPTIMAL LEVEL OF GOVERNMENT DEBT**

In this section the assumption regarding the exogeneity of income is dropped. Individuals can use their time to either work or take leisure. The decision on how to spend their time will generally depend on their after-tax wage rate. Therefore, tax rates will affect the amount of labor supplied and will create distortions in the economy. The job of the government will be to choose tax rates that minimize the amount of distortions subject to a future path of government spending and an inherited amount of debt. Choosing a path of tax rates for a given path of government spending amounts to choosing a path for government debt. This path is chosen to maximize social welfare and usually involves both positive and negative deficits. Hence a balanced budget is not generally optimal.

It is important to note that the literature on optimal government deficits assumes that government spending is exogenous. This is a fairly extreme assumption, since government spending should more appropriately be viewed as a choice variable. However, treating government spending as a choice variable with respect to some specific objective function is unlikely to change the qualitative results presented below. This is due to the reasonable belief that the ratio of government spending to income will vary optimally over time. For example, a state of war would cause this ratio to temporarily rise. Also, (with the exception of wartime spending) individuals would probably prefer a relatively smooth path of government spending for the same reasons that they
prefer a smooth consumption stream. This permanent income analogy between consumption of publicly and privately produced goods implies that the ratio of government spending to income should move counter-cyclically. Although a complete analysis that jointly determines the optimal levels of government spending and tax rates would be valuable, the article follows the treatment of the current literature by assuming that government spending is exogenous.

In analyzing the optimal level of government debt, the economy examined will be fairly simple. Individuals are assumed to live for two periods as does the government. Also, the production process involves only labor thereby omitting a discussion of capital. The relevant points concerning debt optimality can be made without considering capital.

**The Model without Taxes**

The representative consumer maximizes his utility by choosing consumption, c, and leisure, l, in each period of his life subject to a lifetime budget constraint. Specifically, he maximizes

\[
U = c_1 - \frac{1}{2}c_1^2 + l_1 - \frac{1}{2}l_1^2 + \beta[c_2 - \frac{1}{2}c_2^2 + l_2 - \frac{1}{2}l_2^2]
\]

subject to

\[
c_1 + \frac{c_2}{1+r} = 1 - l_1 + \frac{1-l_2}{1+r},
\]

For simplicity, a time separable quadratic utility function is used and it is assumed that one unit of labor yields one unit of output. This means that the real wage rate is normalized to one. Using a time separable quadratic utility function allows for easy numerical calculations that illustrate the nature of optimal taxation and budget deficits. In this case the marginal utility of consumption and leisure are 1-c and 1-l. This can be seen by taking the derivative of U with respect to c and l. For the long-run steady state case of \(\beta = 1/(1+r)\) individuals spend equal amounts on consumption and leisure in each period implying that \(c_1 = c_2 = l_1 = l_2 = \frac{1}{2}\).

These values are consistent with economy-wide market clearing in each period since \(c_1 = 1 - l_1 = \frac{1}{2}\) and \(c_2 = 1 - l_2 = \frac{1}{2}\). That is, aggregate demand equals supply. (Since c and l are consumption and leisure choices of the representative individual, it is permissible to equate individual and aggregate values without affecting the results of the analysis.) The solution to the individual’s choice problem can be depicted graphically by defining \(I = c + l\). The budget line has a slope of \((1+r)\) and for the case where \(\beta = 1/(1+r)\), the individual will choose \(I_1 = I_2 = 1.0\). This is depicted by the solid line in Figure 2.

**The Model with Government Spending and Taxation**

The analysis now turns to the problem when there is a government that spends and raises revenue through distortionary taxes. Taxes are levied against labor income and government spending is not valued.
The individual now maximizes his utility (9) subject to the constraint that the present value of consumption is equal to the present value of after-tax income. In this case one would no longer expect consumption and leisure to be equal since the relative price of leisure has fallen. Prior to the introduction of a government that raises revenue by taxing wages the foregone income resulting from additional leisure was one, while now it is $1 - \tau$. With leisure being cheaper, people will work less and produce less. This is the distorting nature of taxes.

The economy-wide market clearing conditions are also altered to take into account the presence of the government. In each period the amount of goods produced must equal consumption plus government spending. Specifically,

$$c_1 + \frac{c_2}{1 + r} = (1 - \tau_1)(1 - l_1) + (1 - \tau_2)\frac{1 - l_2}{1 + r}.$$  \hfill (10')

The government is also required to satisfy a budget constraint given by

$$c + g = 1 - l,$$  \hfill (11)

where $g$ is government spending.

The government is also required to satisfy a budget constraint given by

$$g_1 + \frac{g_2}{1 + r} = \tau_1(1 - l_1) + \tau_2\frac{1 - l_2}{1 + r}$$  \hfill (12)

which indicates that the present value of government spending equals the present value of government revenue. This condition requires the government to pay off any debt acquired in period 1 by raising revenue in period 2. Without this requirement it would be unreasonable for anyone to hold government debt and there would be no government debt issue to consider.

Given the model economy described by equations (9) (10') (11) and (12), the government chooses tax rates $\tau_1$ and $\tau_2$ to maximize the utility of its citizens, given an exogenous time path of government spending. To gain some insight into the nature of optimal deficits two simple cases are analyzed below.

Case 1. Optimal deficit with equal government spending, $g_1 = g_2$.

The first case examines the situation when government spending is constant over time. Intuitively, one would expect the government to equalize the level of distortions in each period and that $\tau_1 = \tau_2$. This implies that the government budget will be balanced in each period.

In particular let $g_1 = g_2 = .10$. In this case the government chooses tax rates $\tau_1 = \tau_2 = .2053$. Individuals therefore choose $c_1 = c_2 = .3870$, and $1 - l_1 = 1 - l_2 = .4870$. Also the government consumes a little over twenty percent of output. And, as conjectured, the government balances its budget in each period.

It is interesting to compare this case to the example of no government initially discussed in this section. Consumption and output have fallen, while leisure has risen. Because government spending is not valued by individuals, utility has also declined. In this case let $I = c + l(1 - \tau)$. This implies that $I_1 = I_2 = .795$. This situation is depicted by the dotted lines in Figure 2.

Interestingly enough, because $\tau_1$ and $\tau_2$ are chosen as the solution to an optimization problem by the government, small changes in $\tau_1$ and $\tau_2$ will be neutral. That is, given that the government has chosen the level of public debt optimally, small movements in tax rates and the deficit will have no real consequences. If changes in tax rates are large, or if the political outcome is one that doesn’t produce optimal taxation, then government financing decisions will have real consequences.

Case 2. The optimal deficit when government spending is unequal, $g_1 \neq g_2$.

This example investigates the optimal budget deficit when government spending is unequal across periods. For example, suppose that $g_1 > g_2$. This could occur if the government were financing a war or a temporary buildup in defense expenditures in period 1. Similarly, one might view period 1 as representing a cyclical downturn in which government transfer payments increase, while period 2 represents a cyclical upturn in which government transfer payments fall.

Intuitively one would no longer expect the government to balance its budget in both periods. Rather, the government would try to balance the distorting aspects of taxes over both periods, running a deficit in period 1 and a surplus in period 2.

---

\[^7\] For the more complicated case where government spending is productive and valued by consumers see Aschauer and Greenwood [1].
To highlight this behavior, consider the case where $g_1 = 0.20$ and $g_2 = 0$. Let $\beta = 0.75$. Under these conditions the tax rate in period 1 is 0.40, while the tax rate in period 2 is 0.282. Consumption in period 1 is 0.18 and consumption in period 2 is 0.418, while leisure in period 1 is 0.482 and leisure in period 2 is 0.582. The government runs a deficit in period 1 of $0.20 - 0.24(1 - 0.482) = 0.076$ and runs a surplus in period 2 equal to $0.282(1 - 0.582) = 0.118$ which has a present value of 0.076 given an equilibrium real rate of interest of 0.55.

Compare this to the nonoptimal behavior of balancing the budget period by period. In this case the first period tax rate is 0.67, consumption is 0.1, and leisure is 0.7. In the second period tax rates are zero, and consumption and leisure are 0.5. The equilibrium real rate is 1.40. Under an optimal budget deficit consumption and leisure are more effectively smoothed over time. This smoothing reflects the desire of individuals to have the burden of taxes appropriated over each period. Forcing the budget to balance in each period reduces the welfare of individuals.

**Current Political Concerns**

The material in this section has shown that under distorting taxes, individuals care about deficits in the sense that deficits help to allocate the burden of uneven levels of government spending optimally over time. The theory, therefore, implies that today’s optimal deficit depends on the perceived future time path of government spending as well as the inherited existing debt structure of the current government. The question of whether or not to reduce the current deficit through increased taxes hinges upon beliefs of future government spending patterns relative to GNP. The Reagan administration’s reluctance to increase taxes may therefore feel that government spending will exhibit secular growth relative to GNP and that some of this future burden should be optimally allocated to the current period. The theoretical analysis presented in this section indicates that much of the argument concerning the size of the deficit is directly related to beliefs over the expected size of the government.

**The Balanced Budget Amendment**

This section discusses the relationship between the Balanced Budget Amendment (BBA) passed by the Senate on October 1, 1982 and the literature on optimal budget deficits. Of particular relevance is the material contained in section 1 of the amendment (see Friedman and Friedman [13; p. 56]), which requires that actual government spending be less than or equal to planned government spending which in turn must be less than or equal to planned receipts. A formal interpretation of this section is that $\mathbb{E}_t[g_t - g_{t+1}] \leq \mathbb{E}_t[T_t - T_{t+1}]$ where $g_t$ is government spending, $T_t$ is tax revenue, and $\mathbb{E}_t$ is the conditional expectation of the variable in question.

As shown in Case 2, it is not generally optimal for the government to balance its budget in each and every period. The wording of the Balanced Budget Amendment does not require the government budget to balance ex post, since tax revenues are generally uncertain and may deviate from their planned or expected value. However, the language of section 1 requires that the budget balance on average.

Even though the BBA does not require strict budget balance in each period it still may force tax policy to deviate from the optimal policies previously outlined. This can be shown by means of a simple example. Consider an economy that wasn’t growing (growth could easily be incorporated into the analysis) and where government spending was fixed at $g$ in every period. Assume that taxes are income taxes and that income fluctuates in a predictable way around some average level $\overline{Y}$. An optimal policy under these circumstances would be to set taxes so that there was an anticipated surplus during relatively prosperous times and a deficit when income was depressed. On average the governments budget would balance. The BBA does not allow for this type of behavior which would characterize optimal tax policy over a business cycle. That is, the BBA only allows for deviations from budget balance when income movements are unanticipated, but does not allow for planned changes in the level of debt that would optimally accompany anticipated cyclical changes in economic activity. This lack of response to anticipated cyclical activity represents a welfare cost with regard to this amendment.

A similar example exhibiting the welfare costs of the proposed BBA, could be depicted for the case of temporary anticipated movements in government expenditure, such as short-term increases in defense spending. This is because the BBA does not gener-
ally allow for planned government budget deficits and surpluses that are necessary if taxes are to be optimally smoothed. However, the amendment does contain an important element which allows for departure from expected budget balance if three-fifths of the Congress approves it. Therefore, it is conceivable that optimal budget deficits and surpluses could be produced under the amendment if economic conditions clearly warrant a departure from expected budget balance. However, there is no guarantee that the provision of section 1 of this amendment will produce an optimal time path for taxes.

An Estimate of the Optimal Budget Deficit

In a recent Wall Street Journal article, Professor Robert Barro estimates $165 billion as the optimal budget deficit for 1984. The actual deficit was approximately $170 billion and very close to the optimum. Also, the size of the deficit which is in the range of 4-5 percent of GNP, is claimed to be in line with historical experience.

Barro’s method of estimation is linked with the second example presented in this section. During recessions, or periods where government spending may be temporarily high relative to output, budget deficits rise. This is part of the tax smoothing role of deficits. According to Barro, the effect of the last recession, 1980-83, has contributed approximately $60 billion to the 1984 deficit. This estimate is a result of the fact that output is still below its full employment level and therefore the business cycle is still contributing to a positive deficit.

Another major consideration is the extent to which interest on the debt should be financed by taxes. A financing of the expected inflation component of interest payments on the debt by the issue of new debt would leave the debt to GNP ratio unchanged, and place no further real burden on society. Barro estimated that $70-$75 billion of an optimal amount of debt arises from this component.

Another part of interest payments that should be financed through debt rather than taxes are those that are due to higher than average real interest rates. The reason for using debt to smooth these payments is analogous for using debt to smooth taxes over the business cycle. When real rates are higher than normal the debt will grow and when real rates are lower than normal the debt will fall. On average there would be no long-run change in government debt and therefore no reason to shift tax rates to finance this portion of interest payments. Since real rates appear to be abnormally high, Barro estimates that this factor has contributed roughly $30 billion to the 1984 deficit. Adding up the three factors produces an optimal deficit of $160-$165 billion for 1984.

The prediction for future optimal deficits is much the same as the prediction for 1984. Basically, the optimal level of the deficit will change if there is rapid future economic growth relative to changes in government spending, or if inflation and real interest rates fall. Current forecasts of these variables over the near future do not point to any major changes in the optimal level of the budget deficit. Therefore it is only to the extent that proposed deficits differ from the optimum that the deficit should be a matter of concern.

It should be mentioned that Barro’s estimates depend on assumptions regarding the normal level of the real interest rate, the extent to which the business cycle downturn of 1980-83 has contributed to the deficit, and forecasts of expected inflation. Differing views concerning these factors will produce different estimates of the optimal deficit. However, Barro’s analysis is useful in providing a framework for examining the current deficit with respect to some desirable standard.

VII. SUMMARY

This article has examined the effects that government budget deficits have on the economy given a variety of assumptions regarding the actual workings of the economy. Although the survey presented is far from exhaustive, it does attempt to highlight the major issues associated with the economic consequences of budget deficits. In that regard, the article first focused on the expansionary effects as well as the economic burdens of government deficits that are derived in a simple Keynesian framework. Criticisms of that framework were then discussed in light of the Ricardian equivalence proposition. Features of that proposition were explored, with the conclusion that wealth effects are probably not an important channel for analyzing the effects of government deficits.

The article then focuses on the notion of optimal deficits in an economy with taxes on labor income. Because distortions are introduced into the economy through the tax system, the government has an in-

10The provisions of the amendment may also be waived in any fiscal year in which there exists a declaration of war. Since wars are one of the primary examples of temporary government spending that should not be totally financed over their duration, the amendment does take declared wars explicitly into consideration.
centive to minimize these distortions given a projected path of government spending and an inherited structure of public debt. It does so by smoothing taxes over time. This process leads to movements in debt that are associated with temporary changes in government spending relative to output. Wars and business cycle phenomena are therefore primary factors in causing the optimal level of debt to vary over time.

The paper concludes with a brief analysis of the political controversy surrounding the current budget deficit in relation to the ideas concerning the optimal debt. Perhaps much of the controversy is related to differing desires or forecasts over movements in the size of government. A rough estimate of $165 billion as the optimal level of government debt is given as a yardstick on which to evaluate proposed budget deficits.

References


Although critics may dismiss it as a mere empirical correlation masquerading as a tradeoff, the Phillips curve relationship between inflation and unemployment has nevertheless been a key component of macroeconomic models for the past 25 years. In 1960 Paul Samuelson and Robert Solow [16, p. 192] named the relationship after A. W. Phillips, the New Zealand economist who in 1958 gave it its best known (but hardly its first) modern formulation (see Figure 1). Since then it has evolved through at least five successive versions as analysts sought to expand its explanatory power, its theoretical content, its policy relevancy, and its ability to fit the facts.

Phillips' [15, p. 290] initial wage-change version \( w = f(U) \) related the rate of wage inflation \( w \) via the function \( f(\cdot) \) to the excess demand for labor as measured by \( U \), the deviation of unemployment from its equilibrium or labor-market clearing rate. Transformed through the assumed markup of prices over wages into the price-change equation \( p = f(U) \), where \( p \) is the rate of price inflation, it was widely interpreted as a stable enduring tradeoff or menu of inflation-unemployment combinations from which the authorities could choose. In its shift-adjusted form \( p = f(U) + Z \), it incorporated a vector of variables \( Z \), including past price changes, trade union effects, unemployment dispersion, demographic factors and the like, to account for observed shifts in the inflation-unemployment tradeoff or menu of policy choices. In its expectations-augmented form \( p - p' = f(U) \), where \( p' \) is the expected rate of inflation, it asserted (1) that the tradeoff is between unemployment and unexpected inflation, (2) that the tradeoff vanishes when expectations are realized, and (3) that unemployment returns to its natural equilibrium rate at this point. Provided expectations adjust to actual inflation with a lag, it also implied the accelerationist notion that unemployment can be pegged permanently below its natural rate only if inflation is continually accelerated so as to always stay a step ahead of expectations. That is, while denying a permanent tradeoff between unemployment and the rate of inflation, it implied that there may be a permanent tradeoff between unemployment and the rate of acceleration of inflation.

The preceding versions reflect a non-market-clearing view of the world, expressing as they do the disequilibrium response of wages and prices to a mismatching of demand and supply in the labor market. By contrast, the alternative New Classical or market clearing version \( U = g(p-p') \) assumes that the labor market is always in equilibrium and that deviations of unemployment from its natural rate stem solely from inflation misperceptions and vanish when those misperceptions end. When com-
bined with the assumption of rational expectations (according to which actual inflation differs from expected inflation only by a random forecast error) this version says that tradeoffs are solely the result of unpredictable random shocks and cannot be exploited by systematic (predictable) policies.

The foregoing interpretations are well known. Not so well known, however, is the origin and early history of the inflation-unemployment relationship. For the most part, textbooks typically trace the idea to Phillips’ famous 1958 Economica article without saying anything about what went before. They correctly describe the five versions of the Phillips curve outlined above. But they fail to note that at least three of those versions (including the version presented by Phillips himself) had already been spelled out long before Phillips. The result is to neglect at least ten predecessors whose names deserve to be associated with the Phillips curve. In an effort to redress this oversight and to set the record straight, the paragraphs below document what Phillips’ predecessors had to say about the inflation-unemployment relationship.

John Law (1671-1729)

It is probably unrealistic to expect to find a Phillips curve in the writings of John Law, the famous eighteenth century banker and finance minister whose schemes to promote economic development via the creation of a paper currency secured by land ended with the collapse of the Mississippi Bubble in 1720. To be sure, he believed that money stimulates real activity. But he also believed that it does so at constant or even decreasing prices owing to the availability of idle resources and scale economies in production. As a result, there is either no Phillips curve inflation-unemployment relation in his analysis or it works in the wrong direction—falling unemployment being associated with falling, not rising, prices.

David Hume (1711-1776)

The prototypical Phillips curve analysis is to be found in the writings of the eighteenth century Scottish philosopher-economist David Hume. As early as 1752, he presented the essentials of a Phillips curve relationship of the form \( U = g(dP/dt) \), where \( U \) is the deviation of unemployment from its natural (equilibrium) rate and \( dP/dt \) is the change in the price level with respect to time. This relationship derived straight from his assumption that unemployment disturbances stem from price perception errors (the difference between actual and perceived prices) and that such errors persist only when prices are changing. Expressed symbolically, he assumed that

\[
U = h(P - P^e) \quad \text{and} \\
P - P^e = k \frac{dP}{dt}
\]

where \( P \) and \( P^e \) denote actual and perceived prices and \( k \) is a coefficient relating price perception errors to price level changes. Substitution of the latter equation into the former yields Hume’s version of the Phillips curve \( U = g(dP/dt) \) mentioned above. That version embodied his hypothesis that one must continually raise prices to peg unemployment at arbitrarily low levels since only by doing so can one produce the price perception errors that sustain the tradeoff. In short, Hume’s explanation stresses the employment effects of unperceived monetary-induced price changes. He [8, pp. 37-40] says:

though the high price of commodities be a necessary consequence of the increase of gold and silver, yet it follows not immediately upon that increase; but some time is required before the money circulates through the whole state and makes its effect be felt on all ranks of people. At first, no alteration is perceived; by degrees the price rises, first of one commodity, then of another; till the whole at last reaches a just proportion with the new quantity of specie . . . . In my opinion, it is only in this interval or intermediate situation, between the acquisition of money and rise of prices, that the increasing quantity of gold and silver is favourable to industry . . . . From the whole of this reasoning we may conclude, that it is of no manner of consequence, with regard to the domestic happiness of a state, whether money be in a greater or less quantity. The good policy of the magistrate consists only in keeping it, if possible, still increasing; because, by that means, he keeps alive a spirit of industry . . . . There is always an interval before matters be adjusted to the new situation; and this interval is as pernicious to industry, when gold and silver are diminishing, as it is advantageous when these metals are increasing.

Three points stand out in Hume’s analysis [10]. First, the tradeoff is between unemployment and unperceived changes in money and prices; it vanishes once perceptions fully adjust to reality. Second, price perceptions, though slow to adjust, eventually catch up to one-time changes in the level of money and prices. It follows that such changes can at best generate temporary but not permanent tradeoffs.
Third, the only way the tradeoff can be sustained is to generate a continual succession of changes in money and prices. Hume here makes the distinctly non-rational-expectations argument that such changes will, because of the lag in the adjustment of price perceptions, keep prices forever marching a step ahead of perceptions, perpetually frustrating the latter’s attempts to catch up. In this way, he claims, the gap between actual and perceived prices will be maintained thus permanently lowering unemployment. Hume notes that this process works symmetrically for price deflation — such deflation, if prolonged, producing an enduring rise in unemployment. It follows at once that a permanent tradeoff $U = g(dP/dt)$ exists between unemployment and the rate of change of money and prices. One must therefore agree with Charles R. Nelson’s [14, p. 2] recent judgment that

Hume was clearly of the opinion that the level of activity would be raised permanently by a steady increase in the quantity of money, prices, and wages. Hume was therefore a believer in a stable, long-run Phillips curve.

**Henry Thornton (1760-1815)**

Like Hume, Henry Thornton also described a Phillips curve of the form $U = g(dP/dt)$, where the variables are as defined above [10]. In his classic *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain* (1802) he [19, p. 237] says that a monetary expansion stimulates employment by raising prices:

... additional industry will be one effect of an extraordinary emission of paper, a rise in the cost [i.e., price] of articles will be another. Probably no small part of that industry which is excited by new paper is produced through the enhancement of the cost of commodities.

This same tradeoff, he [19, p. 238] notes, also holds in reverse as monetary and price deflation bring painful rises in unemployment.

If we assume the augmented paper to be brought back to its ordinary quantity, we must suppose industry to languish for a time through the ill success [of] mercantile transactions.

In his discussion of the Phillips curve, Thornton was careful to distinguish between alternative levels of money and prices and continuous changes of those variables. Only the latter, he said, can affect real activity and sustain the tradeoff. This is epitomized in his [19, p. 256] remark that “it is the progressive augmentation of bank paper, and not the magnitude of its existing amount, which gives the relief.” In other words, money and prices stimulate activity only when they are continually increasing. For, says Thornton [19, p. 238], “While paper is encreasing, and articles continue rising, mercantile speculations appear more than ordinarily profitable.” But “as soon . . . as the circulating medium ceases to encrease, the extra profit is at an end,” and the stimulus vanishes. Thus a one-time rise in the money stock and level of prices cannot sustain the tradeoff. Instead, a continuous increase or “progressive augmentation” is required. The tradeoff is between output and the rate of change of prices.

As for the tradeoff’s source, Thornton attributed it chiefly to a tendency for money wages to consistently lag behind prices. He explicitly stated (1) that inflation stimulates activity, (2) that it does so by reducing real wages and raising real profits, (3) that this output-enhancing redistribution occurs because money wages lag behind prices, and (4) that this wage lag persists as long as inflation is sustained. Like Hume, he did not explain why the lag would persist nor why wages would not eventually catch up with prices once inflationary expectations had fully adjusted to actual inflation. His analysis is largely silent about inflation anticipations; he did not incorporate them into his Phillips curve.

Finally, he disagreed with Hume over the desirability of exploiting the Phillips curve for policy purposes. Hume clearly believed that the policy authorities in the closed world economy should exploit the curve, using monetary gold inflation to stimulate employment. Hume [8, pp. 39-40] says as much in his advice to the policymaker.

The good policy of the magistrate consists only in keeping [money], if possible still encreasing; because, by that means, he keeps alive a spirit of industry in the nation, and encreases the stock of labor, in which consists all real power and riches.

In contrast, Thornton opposed the exploitation of the Phillips curve for policy purposes. Such exploitation involved inflation, which he saw as an unmitigated evil. All inflationary policy, he [19, p. 239] said, is “attended with a proportionate hardship and injustice.” True, output and employment would rise. But such gains, he thought, would be far too small to be worth the costs (uncertainty, in-
justice, social discontent) of higher inflation. In short, the Phillips curve at the economy’s normal level of operations was very steeply sloped, allowing little increase in output per unit rise in inflation. Thus while “paper possesses the faculty of enlarging the quantity of commodities by giving life to some new industry,” the unfavorable tradeoff ensures that “the increase of industry will by no means keep pace with the augmentation of paper.” Moreover, because the economy normally operates close to its absolute full capacity ceiling, stimulative policy will quickly reach the point where it is obvious that the antecedently idle persons to whom we may suppose the [monetary inflation] to give employ, are limited in number; and that, therefore, if the increased issue is indefinite, it will set to work labourers, of whom a part will be drawn from other, and, perhaps, no less useful occupations.

On these grounds he [19, p. 236] concluded that there exist narrow “bounds to the benefit which is to be derived from an augmentation of paper; and, also, that a liberal, or, at most, a large increase of it, will have all the advantageous effects of the most extravagant emission.”

The Attwood-Mill Debate

The Phillips curve concept continued to flourish in the hands of more than one British classical writer after Henry Thornton. That this is so is evident from a glance at the celebrated interchange between Thomas Attwood (1783-1856) and John Stuart Mill (1806-1873) in the 1820s. Attwood, an inflationist proponent of inconvertible paper currency regimes and full employment at any cost, believed in a stable long-run tradeoff relationship of the form U=g(P) where U and P denote unemployment and the price level, both taken relative to their normal (base period) values. Attwood used this relation, in which the inflation variable enters as a price level rather than its Hume-Thornton rate of change, to argue (1) that high unemployment stems from low prices, (2) that low unemployment emanates from high prices, and (3) that the government can and should achieve a zero target rate of unemployment with inflationary monetary expansion. For him nothing short of absolute full employment would suffice. Said he [3, p. 467], “so long as any number of industrious honest workmen in the Kingdom are out of employ-

ment, supposing such deficiency of employment not to be local but general, I should think it the duty, and certainly the interest, of Government, to continue the depreciation of the currency until full employment is obtained and general prosperity.” “Restore the depreciated state of the currency,” he [2, p. 66] declared, and “you restore everything that constitutes the commercial prosperity of the nation.”

Opposing him was John Stuart Mill who reasoned in terms of the relationship U=g(P-P^E) where U is the discrepancy between unemployment and its natural steady-state level, P is the price level, and P^E is its expected or perceived level. Using this relationship, Mill argued (1) that tradeoffs are temporary, (2) that they stem from unexpected price changes and vanish once perceptions adjust to reality, and (3) that, contrary to Attwood, one cannot peg real activity at arbitrarily low levels simply by pegging a nominal price (or inflation) variable since the two variables are independent of each other in steady-state equilibrium [9].

To be sure, Mill admitted that a temporary inflationary stimulus is possible. It is true, he [13, p. 79] said, that an unexpected inflation, if misperceived as a rise in relative prices, “may create a false opinion of an increase of demand; which false opinion leads, as the reality would do, to an increase of production.” But it is also true that the real expansion is “followed . . . by a fatal revulsion as soon as the delusion ceases.” In other words, once producers correctly perceive price increases as nominal rather than real, economic activity reverts to its steady-state level, but only after undergoing a temporary recession to correct for the excesses of the inflationary boom.

In Mill’s view, the steady-state Phillips curve is a vertical line at the economy’s natural rate of unemployment. To assert otherwise (as Attwood did), he thought, was to argue that people can be fooled perpetually into believing that nominal gains are real and that commodities can be created from paper money expansion. But according to Mill, one cannot fool all the people all the time. Money illusion, he contended, is not permanent. Attempts to peg real activity are therefore bound to be futile. Inflation cannot permanently stimulate activity. Mill’s reply to Attwood dispels the notion that expectations-augmented Phillips curves and the natural rate hypothesis are of recent origin.
Irving Fisher (1867-1947)

As noted above, Hume and Thornton helped lay the theoretical foundations of the particular Phillips curve relationship \( U = g(dP/dt) \). It was Irving Fisher, however, who provided the first statistical evidence of that relationship [7]. In his 1926 *International Labour Review* article, "A Statistical Relationship Between Unemployment and Price Changes," he investigated the correlation between unemployment \( U \) and lagged price changes \( (dP/dt)_L \), where the subscript \( L \) denotes a linear distributed lag (Fisher himself being the inventor of the lag distribution concept) on the price-change variable. Using monthly U.S. data for the period 1915-1925, he obtained correlation coefficients as high as 90 percent between the two variables. Likewise, his time series chart displayed a similar strong correspondence between lagged price changes and employment (see Figure 2). From this evidence he concluded that there was indeed a strong relationship between them. He [7, p. 502] also concluded that the relationship was causal as well as empirical, that causality runs unidirectionally from price changes to unemployment, and that there are good theoretical reasons for this being so. His theory of price-to-unemployment causality relies on fixed contracts, the inertia of custom, and other inhibiting factors that prevent costs from adjusting as fast as prices when prices change. Owing to the lag of costs behind prices, changes in the latter affect profits and thereby the level of real activity and employment. Via this linkage, causality, he argued, runs from inflation to unemployment as confirmed by his finding that the former variable leads the latter.

Jan Tinbergen

Although he presented no formal econometric equations, Fisher was the first to offer empirical corroboration of the Phillips curve's market clearing version \( U = g(dP/dt) \) according to which causality runs from inflation to unemployment. By contrast, Jan Tinbergen [4] in 1936 was the first to estimate the alternative shift-augmented wage-change version \( w = f(U) + Z \) in which causality runs from unemployment or some equivalent measure of demand pressure in the labor market to the wage inflation rate and a vector of shift variables enters to affect the wage-unemployment tradeoff. More precisely, his equation was of the form \( dW = F(E, dP^{-1}) \) where \( dW \) is the change in money wages, \( E \) is employment relative to its normal (i.e., trend) level, and the lagged price-change variable \( dP^{-1} \), represents catch-up or cost-of-living wage adjustment factors thought capable of shifting the curve. Thus in his "An Economic Policy for 1936" he presents the expression \( dW = 0.16 E + 0.27 dP^{-1} \) in which the numerical coefficients are estimated from the Netherlands data for the period 1923-1933.

About this equation three things must be said. It was the first econometric Phillips curve equation ever to appear in print. It also was the first to explain the tradeoff in terms of the law of supply and demand according to which the price of any good or service (including labor) varies in proportion to the excess demand for it. In other words, for the first time the Phillips curve was interpreted as a wage-reaction function relating the disequilibrium response of wages to demand pressure in the labor market, this pressure being measured by employment relative to trend. Finally, as mentioned above, Tinbergen's equation was the first to include a price change shift variable to account for observed movements in the wage-employment relationship. In these respects, it
foreshadowed 1960s-vintage wage equations that likewise represented the Phillips curve as a demand-pressure wage-response function subject to shifts owing to changes in the cost of living.

Tinbergen returned to the Phillips curve issue once again in his Business Cycles in the United Kingdom 1870-1914, published in 1951 fully seven years before Phillips' contribution. There, using \( W \) to denote wages and \( E \) to denote employment, he [21, p. 50] writes the Phillips curve equation as

\[
\frac{dW}{dt} = \dot{W} = f(E) = \lambda E
\]

and gives it the excess-demand wage-reaction interpretation. "The theory expressed" in the equation, he says, "may be given the well-known formulation that a high unemployment figure 'exerts a pressure on' the wage rate and that, on the other hand, a small unemployment figure causes wages to go up." He also notes that the equation's empirical fit might be improved if the demand-pressure variable were entered nonlinearly and that this could be accomplished by replacing the employment variable \( E \) with the inverse of the unemployment rate \( U^{-1} \). Finally, he suggested adding variables representing cost-of-living changes and the degree of unionization of the labor force to the equation to improve its statistical fit. On all of these innovations he pioneered the practice of fitting econometric Phillips curve equations.

**Klein and Goldberger**

Lawrence Klein and Arthur Goldberger also estimated econometric inflation-unemployment equations before Phillips. In their famous 1955 study *An Econometric Model of the United States, 1929-1952*, they [11, p. 19] presented a wage-change Phillips curve equation of the form \( dW = F(U, dP_{-1}) \). More precisely, their equation was

\[
\Delta W = 4.11 - 0.74 U + 0.52 \Delta P_{-1} + 0.54 t
\]

where \( U \) is total unemployment, \( t \) is a time trend in years (\( t=1 \) in 1929), and the other variables are as defined above.

Like Tinbergen, Klein and Goldberger expressed the wage inflation variable in first difference rather than percentage rate of change form. Besides including a time trend variable, they also entered the unemployment variable linearly rather than nonlinearly into their equation. Except for these minor differences, their equation is virtually the same as the later formulations of Phillips and R. G. Lipsey, who clarified and extended Phillips' work. And like those latter writers, Klein and Goldberger interpreted their equation as a wage-reaction function in which money wages change in response to excess labor demand in an effort to clear the market. According to them [11, p. 18]

the main reasoning behind this equation is that of the law of supply and demand. Money wage rates move in response to excess supply or excess demand in the labor market. High unemployment represents high excess supply, and low unemployment below customary frictional levels represents excess demand.

Here is the essence of the Phillips-Lipsey interpretation, an interpretation that also runs in terms of the law of supply and demand.

**A. J. Brown and Paul Sultan**

As documented above, the theoretical, empirical, and econometric foundations of the Phillips curve had been thoroughly established by the mid-1950s several years in advance of Phillips' own contribution. It remained, however, for someone to present a Phillips-type relationship on a statistical scatter diagram and then to draw the familiar downward-sloping convex tradeoff curve that bears his name. Credit for being the first to accomplish these tasks goes not to Phillips himself but rather to two other economists, A. J. Brown and Paul Sultan.

The former, in his 1955 volume *The Great Inflation 1939-1951*, presented scatter diagrams similar to Phillips' (see Figure 3) that plotted annual wage inflation rates against unemployment rates for the United Kingdom for the periods 1880-1914 and 1920-1951, and for the United States for the period 1921-1948. From these charts Brown [5, pp. 91-101] concluded (1) that the two variables are inversely related, and (2) that the relationship between them is nonlinear since wages change at faster rates at low than at high rates of unemployment. He also used his charts to estimate the critical noninflationary level of unemployment below which wage inflation exceeds productivity growth so that prices rise. He did not, however, fit a curve to his data. Thus, although he presented a Phillips-type graph, he failed to draw the eye-catching curve made famous by
Phillips. For this reason, one must reject A. P. Thirlwall’s [18] contention that the curve should bear Brown’s name rather than Phillips’.

Priority for drawing the Phillips curve goes to Paul Sultan, whose contribution predates Phillips’ by one year. Thus, in his 1957 textbook Labor Economics, Sultan presents the curve in a diagram (see Figure 4) described by him [17, p. 555] as follows:

the vertical scale measures the annual changes in the price level expressed as a percentage, while the horizontal scale measures the percentage of the work force unemployed. The line relating unemployment to inflation . . . is strictly hypothetical, but it suggests that the tighter the employment situation the greater the hazard of inflation . . . . Assuming that a fairly precise functional relationship exists between inflation and the level of employment, it is possible to determine the “safe” degree of full employment. In our hypothetical case, we are assuming that when unemployment is less than 2 percent of the work force, we face the dangers of inflation. And when unemployment is larger than 6 percent, we face the problem of serious deflation.

Here is the first diagrammatic representation of the price-change Phillips curve as a stable tradeoff relationship $p=f(U)$ between inflation and unemployment. On the basis of this diagram, three writers [1] recently have suggested that the Phillips curve could with equal justification be called the Sultan schedule.

Concluding Comments

Given the evidence presented in the preceding paragraphs, the label “Phillips curve tradeoff” must be judged both misleading and incomplete. For, as documented above, Phillips was far from the first to postulate an inflation-unemployment tradeoff or to draw the curve bearing his name. Even the econometric wage-price equations employed in modern Phillips curve analysis together with their excess demand and alternative market clearing interpretations long predate Phillips. In short, Phillips and his successors inherited (albeit unknowingly) these concepts; they did not invent them. In this sense at least, their work may be said to constitute the continuation rather than the origin of Phillips curve analysis.

Still, it was Phillips’ formulation and not those of his predecessors that captured the attention of the economics profession. One must ask why this was so. Certainly it cannot be explained by the novelty of his
curve or its empirical derivation; these were hardly innovations at the time he presented them. Nor can it be attributed to any originality in his explanation of his curve. His theory was simply the law of supply and demand according to which the price of any commodity or service (including labor) changes at a rate proportional to the excess demand for it. This explanation of course had been advanced by Tinbergen years before Phillips. Rather his phenomenal success probably stemmed from three factors. First was his striking finding of the apparent near 100-year empirical stability of his curve, a stability not suspected before. Second was the persuasive early expositions of his work provided by such influential economists as Lipsey [12], and Samuelson and Solow [16]. Especially important was the Samuelson-Solow interpretation of Phillips’ curve as a menu of policy choices, a menu from which the authorities could select the best (or least undesirable) inflation-unemployment combination and then use their policy instruments to attain it. By providing a ready-made justification for discretionary intervention and activist fine tuning, this interpretation helped make the Phillips curve immensely popular among Keynesian policy advisors. Third was Phillips’ presentation of his curve at just the right time to satisfy the Keynesians’ search for an explanation of how changes in nominal income divide into price and quantity components. Whatever the reason, his name alone was attached to the tradeoff concept even though at least ten predecessors over a period of roughly 250 years also shared in its formulation.

References


