LABOR SUPPLY AND AGGREGATE FLUCTUATIONS

Robert E. Hall*

Stanford University

and

National Bureau of Economic Research

I. Introduction

Issues of labor supply are at the heart of macroeconomic explanations of the large cyclical fluctuations of output observed in modern economies. Plainly, the population puts in more work in booms than in slumps. Economists' explanations of this phenomenon range from a pure market-clearing supply-and-demand view at one extreme to a dismissal of almost any role of supply and of market clearing at the other extreme. Disagreement is intense because failure of labor markets to clear may create a strong case favoring activist macroeconomic policy.

This paper starts with a serious empirical examination of the view that the labor market is always in balance–that every observed combination of employment and compensation is a point of intersection of the relevant supply and demand curves. I will call this the "intertemporal substitution" model of fluctuations. According to this model, workers are willing to shift their hours of work from one year to another in response to modest shifts in relative wages. This model stands up reasonably well on its own ground. The elasticity of labor supply inferred from the response of employment and relative wages to exogenous shifts in demand is around one half. This figure is quite consistent with microeconomic evidence about the responses of individual workers to temporary increases in their own wages; this evidence is available from longitudinal panel studies and from the various negative income tax experiments.

The paper goes on to point out a strong implication of the pure intertemporal substitution model, namely, the irrelevance of changes in the money supply for the labor supply function. A model where markets clear instantly ought to obey full monetary neutrality. The data refute this implication absolutely unambiguously. The money stock unambiguously shifts the labor supply function. The pure substitution model seems untenable in the light of this evidence.

7

The paper then turns to explanations of the nonneutrality of money in the short run. According to the most carefully worked out line of thought, monetary shocks cause workers to make inappropriate intertemporal shifts in labor supply, because they lack complete information about the source of aggregate shocks and are forced to respond in the same way to real and nominal disturbances. The paper is skeptical about the relevance of these models in view of the extensive, timely information that is available to modern economies about a wide variety of economic aggregates. The paper is equally skeptical about the explanatory value of the line of thought based on contracts in the labor market—if contracts are efficient, they should generate virtually the same level of employment as a labor market that clears instantaneously. If contracts are inhibited by lack of information, then employment should behave in the way described by the theory of markets with incomplete information, a theory which has already been found wanting. Contract theory does not seem to make any independent contribution to the explanation of the sensitivity of employment to movements in the money supply.

Finally, the paper turns to the view that, in the short run, labor supply is largely irrelevant to the determination of aggregate employment. According to a view prevailing among the majority of practical economists and well summarized in a textbook IS-LM model with an empirical Phillips curve, employers determine employment unilaterally by equating the marginal value product of labor to a predetermined nominal wage. Short-run movements in labor demand push workers above or below their microeconomic labor supply functions. Only very gradually, as wages adjust sluggishly, does the level of employment come to respect labor supply. The evidence does not refute this view. But supporters of this line of thought have not, so far, been very successful in explaining why large movements away from the labor supply function are consistent with basic economic principles.

II. The Pure Intertemporal Substitution Hypothesis

According to the intertemporal substitution model of fluctuations, people work more hours in some years than in others, because the market rewards them for this pattern. People are not tricked into extra work in a boom; they find the work desirable, because the return, in the relevant intertemporal sense, is unusually high. Slumps are just periods when a lower level of effort is economically efficient.

Before expounding the intertemporal substitution model itself, I think it may be helpful to clear up some potential sources of confusion. The history of the model has been bound up with the hypothesis of rational expecta-

tions. Robert Lucas, the dean of the rational expectations school, was one of the authors of the most famous exposition of the intertemporal model.[1] That paper is probably even more famous for introducing one of the first formalizations of the role of expectation errors in macroeconomics. Much criticism of the intertemporal substitution model has been directed at Lucas and Rapping's formulation of it and, specifically, at the large role it assigns to errors in perceiving the current real wage.[2] Many economists have received the impression that intertemporal substitution takes place *only* because of workers' misunderstanding of movements in relative prices in the intertemporal substitution model. However, in fact, the issues of substitution and expectation or perception errors can be completely divorced. For now, I want to concentrate on substitution that is induced by genuine and appropriate changes in relative prices.

Another source of misunderstanding of the intertemporal substitution model is the claim that it makes all movements in employment and unemployment voluntary. A recurrent theme in the discussion of the model is the criticism that it makes employment fall by raising quits, when, in fact, quits are low during a downturn.[3] This criticism has not been taken very seriously by the proponents of the theory of intertemporal substitution. An attempt is made in this paper to reconcile the apparent unilateral determination of employment by employers with the hypothesis that the consequent changes in employment represent *intertemporal substitution of work at one time for work at another in* line with changes in relative wages.

Probably the most serious misunderstanding of the substitution model is the claim that it rests on movements of real wages that do not, in fact, take place.[4] This paper shows that the role of real wages in intertemporal substitution depends on the source of the disturbance. Fluctuations of real interest rates are equally important—the most general version of the intertemporal substitution model rests on variations in the relative prices of *current* labor against *future* labor. This point was clearly spelled out in Lucas and Rapping's original paper, but was not fully incorporated in their empirical results. Recently, Barro (1978) has developed a more complete theoretical model of intertemporal substitution in general equilibrium.

My purpose here is to set forth a simple empirical version of the intertemporal substitution model, to estimate *its key parameter, and then to inquire,* at length, whether the story told by the model about aggregate fluctuations makes sense in view of other evidence about intertemporal substitution. In particular, I will draw on recent microeconomic evidence on labor supply based on panel data and the negative income tax experiments.

Within a formal model, the common-sense view; i.e., that people work more when there is more work to do, requires that aggregate supply be reason-

ably elastic with respect to the relative wage and that aggregate demand be reasonably inelastic. Then, a temporary exogenous burst of demand, say, from the government's decision to pursue a costly war, brings about an increase in the current wage compared to the discounted future wage, so people work harder and labor input rises. The evidence presented here suggests that an increase of one percent in the relative price raises labor supply by about 0.5 percent. The amount of intertemporal substitution of labor implied by these results is roughly consistent with the findings of microeconomic research.

Viewed on its own grounds, the intertemporal substitution model has to be judged a success. Its two central implications; i.e., that exogenous demand drives up employment and drives up the current real wage compared to the discounted future real wage, are clearly confirmed by the data. The apparent elasticity of labor supply is not implausibly high. However, this evidence does not confirm the model in any strong sense or distinguish it from competing hypotheses. As a later section of the paper will demonstrate, the intertemporal substitution model in its pure form seems quite unambiguously refuted by other data. The next step in the exposition is to set forth a simple version of the pure form of the model.

Microeconomic Theory of Intertemporal Labor Supply

It is useful to review the standard theory as developed in Lucas and Rapping (1969) and in several other places. Individuals have preferences over present and future work and present and future consumption of goods. They have endowments of time available in the present and future, valued by wage rates that, for now, will be considered known with certainty. They may also hold some wealth. Consumption goods are available at known prices, and consumers may borrow or lend at a given real interest rate to move resources from present to future or the reverse. Formally, with

c_1, c_2 : present and future consumption of goods

L_1, L_2 : present and future work

r : real interest rate

w_1, w_2 : present and future real wages

A : real assets other than the value of time,

there are goods demand and labor supply equations for year 1:

$$c_1 = d_1(r, w_1, w_2, A)$$

$$L_1 = s_1(r, w_1, w_2, A)$$

and a similar pair for year 2. The substitution effects of changes in the relative prices are as follows:

Effect of higher real interest rates on present consumption: negative

Effect of higher current real wage on present consumption: presumably positive

Effect of higher future real wage on present consumption: ambiguous

Effect of higher real interest rate on present labor supply: presumably positive

Effect of higher present real wage on present labor supply: positive

Effect of higher future real wage on present labor supply: presumably negative

The basic logic of the intertemporal substitution model holds that the observed responses to cyclical movements in relative prices should be essentially pure substitution effects, as cyclical variations have little effect on lifetime well-being, which is the relevant concept of income. Thus, the central question treated in this section is whether the set of relative price changes that accompany a boom or a recession is compatible with the increase or decrease in labor supply that takes place. Possible changes that might explain, say, the rise in employment in a boom are:

1. The present real wage rises,
2. The future real wage falls,
3. The real interest rate rises.

Of these, the second seems unlikely and has not received any attention. Most discussion of cyclical labor supply has focused on the role of temporary changes in the real wage. But the third suggestion, a rise in the real interest rate, is clearly pointed out by Lucas and Rapping and deserves an examination as well.

The transplantation of the two-period analysis to the real world where people work and consume for many years requires some additional comment. Of course, the complete intertemporal model has the relative prices for all future years as arguments of the demand and supply functions for this year. Condensation to the two-period version requires the hypothesis that all future years can be combined into a Hicksian composite good. If any departure of the real wage and the real interest rate this year from their normal values is expected to be corrected by next year and all succeeding years, then the composite-good theorem applies, and the demand and supply functions can be written just as

functions of the current real wage, real interest rates, and real wealth. More generally, if the current values of the real wage and interest rate provide all the information that is needed to predict the path by which they will return to normal over the future, then it will still be true that the demand functions can be written as functions of contemporaneous variables alone. The latter explanation of the role of the variables will be relied upon in the empirical work presented here, as it turns out that movements in the variables persist for well over a year.

Under a general specification of preferences, changes in real wages and in real interest rates have distinct substitution effects and, so, should enter separately in econometric work. This observation led Lucas and Rapping (1969) to include both variables in their regressions, though the interest rate was eventually omitted from their final equations. I am not aware of any successful attempts to measure the independent contribution of interest rates in intertemporal labor supply; certainly, microeconomic research has nothing to say on this point, because there is no usable cross-sectional variation in interest rates. An additional assumption about preferences implies that the effects of real wages and interest rates combine in a certain way: the intertemporal utility function is assumed to be separable in goods and work. The marginal rate of substitution between present work and future consumption then depends on the amount of work today (but not on the amount of consumption today) and on the level of future consumption, which is independent of r and w_1, owing to the assumption that transitory fluctuations have no wealth effects. The return to labor in the current period relative to the price of future consumption is $(1 + r)w_1$. Thus, a composite substitution variable with the same form,

$$s_t = (1 + r_t)w_t,$$

indexes the effects of both variables in time series work under the assumptions set forth earlier about the information content of current wages and interest rates and the separability of preferences. Reasonably general specifications of preferences, such as the constant elasticity of substitution (CES), obey separability.

Econometric Evidence on the Joint Movements of the Relevant Variables

One potential empirical weakness of the intertemporal substitution model is the high elasticity of labor supply it seems to require: aggregate fluctuations involve *large movements in employment and small movements in the substitution variable*. If the changes in employment are movements along a labor supply curve, the curve must be quite elastic, or so the argument goes.

The first step in examining this line of attack on the intertemporal substitution model is to find out how elastic the labor supply curve implicit in the aggregate data actually is. Now, because any model has a right to some residuals, just to look at the plot of employment against the substitution variable is not adequate. Nothing less than a full econometric untangling of the supply function from the demand function will answer the question. Of course, the elasticity of the supply function is not the only issue. If it turns out that the supply function is plausibly inelastic, but that almost all fluctuations in employment are attributed to unexplained shifts of the supply curve rather than to movements along the curve, then the intertemporal substitution model has little that is interesting or new about it.

Separation of demand curves from supply curves is a well-understood econometric problem. If the model hypothesizes random shifts of both curves, then identification requires that there be some exogenous variable that has predictable effects on one curve and not the other. The obvious candidate in aggregate supply and demand is government purchases of goods and services (I will call this government expenditures for short, but it should be understood to exclude transfers). Periodically, the government absorbs an abnormal fraction of employment and output, mainly to pursue wars. If these contributions to demand are unrelated to the random shifts in the schedules (in other words, intertemporal substitution is not a major factor in the determination of government expenditure), then the relation between government expenditure, on the one hand, and employment and the substitution variable, on the other, provides information about the slope of the supply curve. From the econometric point of view, all that is involved is the use of government or military expenditures as an instrumental variable in the estimation of the aggregate labor supply schedule. However, the issue of how to measure the supply schedule is much more than just a technical econometric one. It may be helpful at this stage to sketch a general equilibrium analysis of the effects of military expenditures on the economy as a whole.[5]

Consider an economy with one produced good, called output, and labor. Suppose, for simplicity, that the capital stock is given, so that production takes place with labor alone, subject to diminishing returns. Suppose further that changes in military expenditures have negligible effects on consumers' wealth, either because the changes are brief or because the government cuts the permanent level of other expenditures to finance temporary military activities.[6] Suppose finally that the current values of the real interest rate and the real wage provide complete information about the future paths of relative prices, as proposed earlier. Then, it is reasonable to assume that the demand for output depends only on the real interest rate (of course, it also depends on real wealth,

but that is taken as given). *The supply of output depends only on the price of output relative to labor*, namely, the real wage, as in the standard theory of the firm. In the labor market, the demand for labor is just a rewriting of the supply of output and also depends just on the real wage. Finally, labor supply is a function of the substitution variable, which depends on both the real interest rate and the real wage.

The government enters this simple economy in two ways. It purchases *g* units of output, and it hires *n* workers directly. It will turn out that these two demands have opposite effects on the real wage: purchases of goods depress the real wage because of the diminishing marginal product of labor, while direct government employment raises the real wage. Both actions raise the real interest rate.

For a given real wage, *w*, and level of government expenditures, *g*, there is a real interest rate, *r*, which clears the market for output. The market-clearing real interest rate is an increasing function of the real wage (a higher real wage depresses supply and leaves demand unchanged) and is an increasing function of government demand for goods. Similarly, for a given real interest rate and level of direct government employment, *n*, there is a real wage which clears the labor market. The market-clearing real wage is a decreasing function of the real interest rate (*a higher real interest rate stimulates labor supply but leaves labor demand unaffected*) and is an increasing function of direct government employment. All this leads to general equilibrium, as depicted in Figure 1. The *QQ* curve describes clearing of the output market, and the *LL* curve describes clearing of the labor market. An increase in government purchases of output shifts the *QQ* curve upward and leaves the *LL* curve unaffected (see Figure 2). Output and employment rise. The real wage *falls*, but labor supply rises, because the rise in the real interest rate more than offsets the fall in the real wage; the substitution variable rises. When the government stimulates employment by purchasing output, more than all of the intertemporal substitution effect operates through the real interest rate. In a conventional labor market diagram with employment in the horizontal axis and the real wage on the vertical axis, the interest rate effect would show up as a rightward shift of the labor supply function. The effect from the real wage is in the opposite direction.

An increase in direct government employment shifts the *LL* curve to the right and leaves the *QQ* curve unaffected (see Figure 3). Again, both output and employment rise. In this case, however, both real interest rates and real wages contribute to the intertemporal substitution toward higher current labor supply.

Since the typical military buildup involves a combination of purchases

of output and direct employment, the model does not make an unambiguous prediction about the relation between real wages and military expenditures. It is clear, however, that military expenditures will cause workers to substitute toward higher current labor supply and that the inducement will be evident as a higher value of the composite substitution variable. This general equilibrium analysis establishes the relevance of military expenditures as instrumental variables in the estimation of the aggregate labor supply schedule. It points out the difference between the effects of purchases of output and of direct employment. Accordingly, both variables will be used as instruments.

The value of military expenditures as econometric instruments depends on the transitory nature of wars. A permanent change in government expenditures should simply displace an equal amount of private consumption, for the following reason: sooner or later, an increase in the government's diversion of resources must be financed with higher taxes. Consumers should make permanent adjustments in consumption levels in response to any information they receive about present or future taxes. A permanent increase in government expenditures should bring about an immediate permanent decline in consumption of about the same magnitude. A transitory increase in government expenditure—say, for a war—should bring about a much smaller, but still permanent, decrease in consumption. Consequently, the overall response of employment or output to a measure of exogenous demand should be stronger for those measures with large transitory elements--the military component, for example-- than for measures that are close to random walks. It is also important that the measure of exogenous demand not be one that simply displaces private demand. Government expenditures on standard consumption goods distributed to the populace would not be a satisfactory variable for the purposes of this study. On the other hand, diversion of resources to an expensive war in another part of the world is very close to the ideal for these purposes.

The labor supply equation to be estimated within this framework is

$$L_t = a + \beta s_t + u_t,$$

where L_t is total labor supply (including government employment), s_t is the substitution variable, approximated henceforth as $r_t + \log w_t$, and u_t is the random unexplained component of labor supply. Because the labor supply equation has a central role in the general equilibrium of the economy, there is every reason to think that u_t feeds back into the determination of s_t, so they are statistically correlated. For this reason, a two-stage least squares estimator will be used with military expenditures and military employment as instrumental variables.
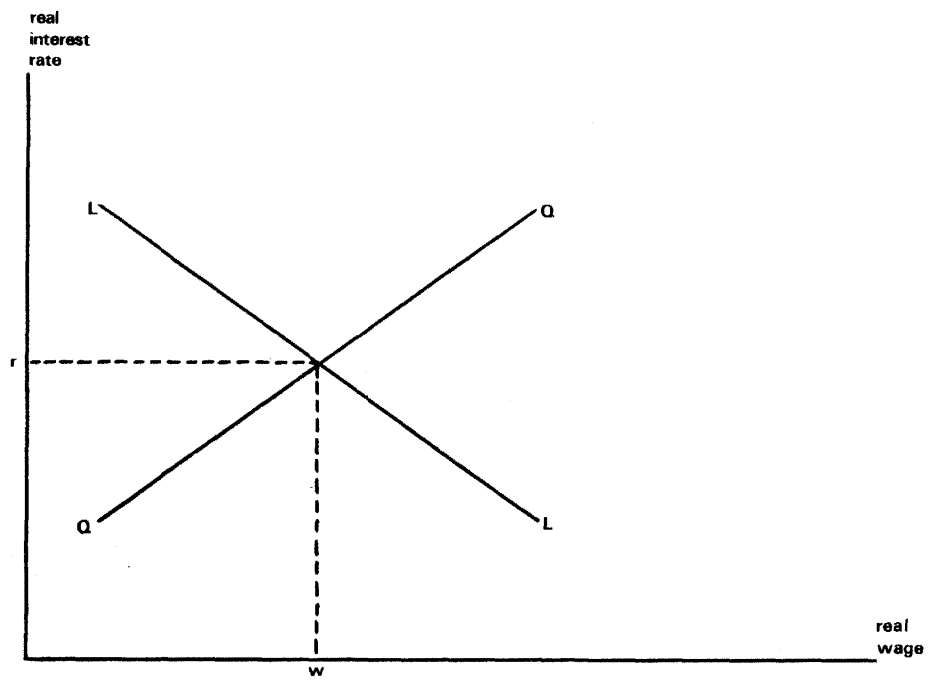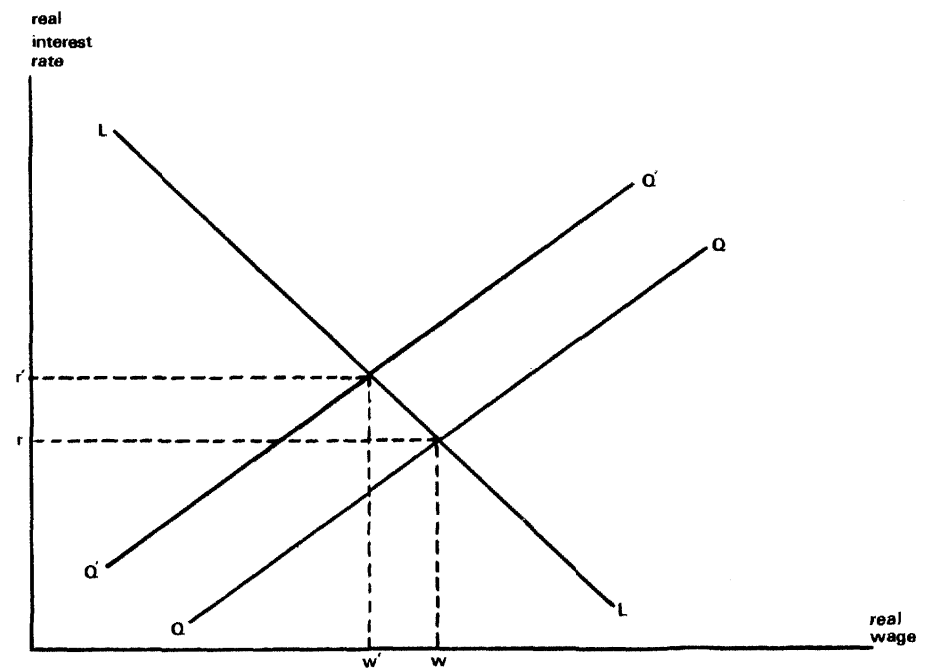
real
interest
rate

L

Q

r - - - - - - - -

Q

L

real
wage

w

Figure 1

real
interest
rate

L

Q'

Q

r' - - - - - - - -

r - - - - - - - -

Q'

Q

L

real
wage

w' w

Figure 2

real
interest
rate

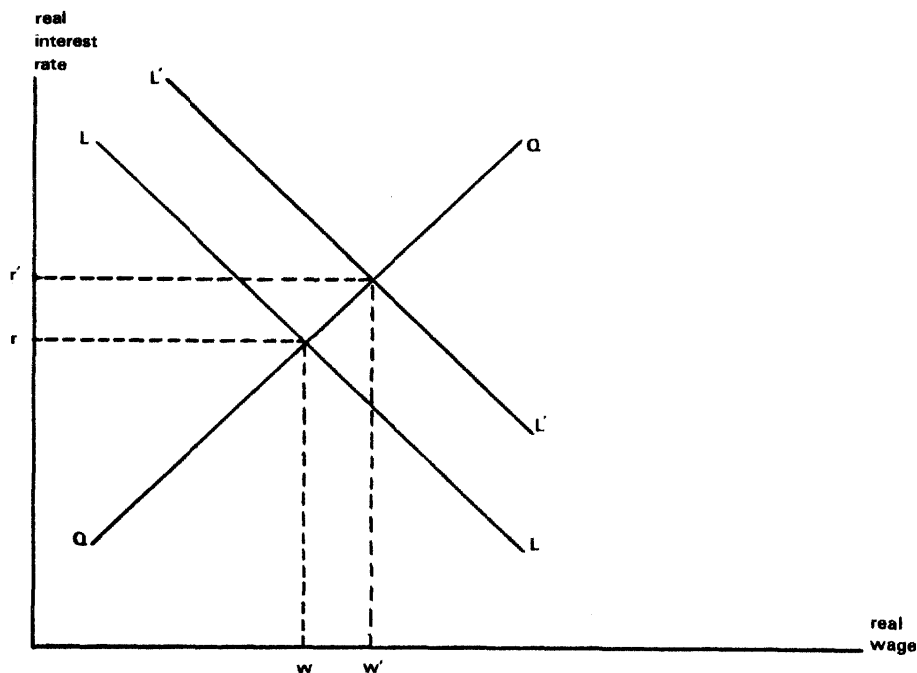
Figure 3

Estimation of the labor supply equation in this form is not directly feasible, because the real interest rate, one of the components of s_t, is not observed. Instead, the real rate must be inferred from the nominal rate and the rate of inflation. Restating the substitution variable in nominal terms gives

$$s_t = r_t^N + \log p_t - \log p_{t+d} + \log w_t - \log p_t,$$

where r_t^N is the nominal interest rate, and d is the number of time periods over which the interest rate applies. Actual behavior at time t responds not to s_t, but to $s_t + \epsilon_t$, where ϵ_t is the error made in time t in forecasting the future price, $\log p_{t+d}$. In terms of observed variables, the labor supply function is

$$L_t = a + \beta s_t + u_t + \beta \epsilon_t.$$

Under the assumption that expectations about future prices made in period t make use of information about military expenditures in period t, the expectation error, ϵ_t, is uncorrelated with military expenditures, and they remain eligible as instrumental variables when s_t is measured from nominal quantities.

In the results presented here, L_t is measured as employee-hours in the total U.S. economy with an exponential trend removed by regressing the log on a linear time trend. The price in the substitution variable is measured by the GNP deflator, the wage by hourly compensation in the private nonfarm sector, and the nominal interest rate by the rate on one-year commercial paper. The future price in s_t is, thus, four quarters ahead. The instruments are military purchases of goods and services, again with an exponential trend removed in the same way, and detrended total military employment. The latter is an annual variable and is given its annual value in all four quarters of each year; this does not affect its eligibility as an instrument. The other data are quarterly and span 1948 through 1978, third quarter. Employee-hours are measured in billions per year (a representative value is 142 billion in 1972, first quarter). The substitution variable is measured in percentage points (137 in 1972:1). As a correction for serial correlation in the residuals from the labor supply equation, all data, including the instruments, were transformed autoregressively with $\rho = 0.86$.

The two-stage least squares estimate of the labor supply equation is

$$L_t = 7.0 + 0.66 s_t \qquad \hat{\sigma} = 1.59.$$
$$\quad (5.4) \quad (0.28)$$

(standard errors are in parentheses)

18

19

According to the point estimate of the slope, an increase of one percentage point in the substitution variable raises labor supply by 0.66 billion hours per year. The corresponding elasticity of the labor supply function with respect to, say, the real wage, is 0.46 (computed as 100 x 0.66/142). But the slope parameter and elasticity are estimated with a certain amount of sampling error. Elasticities much above one are pretty clearly ruled out, but elasticities as low as 0.3 are apparently compatible with the data.

The two-stage least squares estimator is not invariant to normalization; that is, different choices about which jointly dependent variable is on the left-hand side yield different estimates of the parameters. Running the labor supply equation in the opposite direction gives

$$s_t = -10.5 + 1.51 L_t \qquad \hat{\sigma} = 2.40.$$
$$\quad\;\; (12.7) \quad (0.65)$$

The reciprocal of 1.51 is 0.63, very close to the earlier estimate of the slope, 0.66. The low estimate of the slope of the labor supply schedule is not just an artifact of the normalization.

To summarize, real military expenditures can reasonably be taken as one of the major exogenous influences on the demand for output in the U.S. economy and, hence, on the derived demand for labor. Statistical evidence shows that military expenditures push the economy along a well-defined upward sloping labor supply schedule whose elasticity is around a half. There have been enough large movements in real military expenditures to make the sampling variation in the estimates of the slope of the labor supply function adequately small. The next step is to decide if an elasticity of around a half is reasonable in the light of other evidence about labor supply.

Microeconomic Evidence on Intertemporal Labor Supply

The elasticity of labor supply with respect to transitory variations in effective wages can be measured econometrically using data collected from individuals over time in panel studies. The variation found in micro data is less balanced than that found in the aggregate data examined in the previous section: there is little observable variation in interest rates, but a great deal in real wage rates. This part of the paper will look at evidence from surveys of individual workers and from various negative income tax experiments.

Thomas MaCurdy (1978) has recently completed a study of the evidence in the Panel Study of Income Dynamics for adult men (a longitudinal survey conducted by the Survey Research Center, University of Michigan). Everyone in his sample works at least part of each year, so the relevant margin

of substitution is the number of hours of work, not the discrete decision about whether to work at all. MaCurdy assumes that workers' perceptions of lifetime well-being do not change from year to year, but rather can be treated as unobserved permanent individual characteristics. Then, the observed response of hours to changes in wages is interpreted as a pure substitution effect. He estimates the elasticity directly by regressing the log of hours of work on the log of the wage. The resulting elasticities range from 0.09 to 0.23, depending on the details of the specification. This research tends to confirm the general view that adult males do not vary their hours of work in a way that is sensitive to wages.

A similar study has been carried out by James Heckman and MaCurdy (1977) for adult women. They find, again in accord with earlier research, that women are much more sensitive to changes in wages. For women whose wages are beyond the reservation point that draws them into the labor market, the estimated elasticity of hours of work with respect to the wage is 6.6.

As far as I know, there are no comparable studies of intertemporal labor supply among teenagers, who are probably similar to adult women.

As it stands, this body of evidence suggests an aggregate wage elasticity of labor supply of around 2, but virtually all of that comes from women and teenagers. Observed cyclical fluctuations in employment are disproportionately concentrated among women and teenagers, but are noticeable among men as well.

Another highly relevant body of evidence has been collected by the various negative income tax experiments. I cannot do justice to all of the findings of the experiments, but will offer the following broad summary: the typical experiment offered a three-year period of increased income to its subjects, along with a tax of about 50 percent on earnings. The typical adult male subject reduced his labor supply by three hours per week, and the typical adult female subject, by about two hours per week. However, only about half of the subjects were actually influenced by the tax and transfer provisions of the experiment; the rest had total incomes in excess of the critical point where payments reached zero under the benefit formula. Putting all this together gives a rough estimate of the wage elasticities of 0.26 for men and 0.66 for women. This is somewhat higher than MaCurdy's estimates for men and very much lower than Heckman and MaCurdy's estimates for women. Part of the difference may be attributable to the rather different compositions of the two samples. For aggregate labor supply, the evidence from the negative income tax experiments suggests an overall wage elasticity of about 0.40.

Except possibly for Heckman and MaCurdy's rather extreme estimate for adult women, the microeconomic evidence on the labor supply response to temporary movements of wages seems quite consistent with aggregate evidence.

The long-standing criticism of the intertemporal substitution model accusing it of resting on implausibly high elasticities of labor supply is not sustained by the evidence. On its own grounds, the model seems to work out quite well.

III. The Irrelevance of Money in the Pure Intertemporal Substitution Model

Money has no explicit role in a simple version of the intertemporal substitution model of labor supply. Whatever influence money has on labor supply ought to operate through the substitution variable. There is no mechanism within the model by which the money stock can shift the supply curve.[7] Nothing in the pure theory of intertemporal substitution suggests that labor supply is the source of the apparent influence of the nominal money stock on real output and employment. Of course, several authors, including Lucas and Rapping at the outset, have coupled the intertemporal substitution model to models of short-run limitations in the diffusion of information. In these hybrids, monetary neutrality fails in the short run. An unexpected monetary expansion can create the impression of high real interest rates (or higher real wages) because of limited information among employers and workers. The effect can last only as long as the lag in the arrival of information. As Lucas (1975) has argued, the transitory real effect of a monetary disturbance could set in motion a longer response within the real economy. For example, costs of adjusting the level of output will make inventories fall in response to a monetary expansion, and then, gradually, output will rise to restore the stock of inventories.

As a simple test of the irrelevance of money, I reestimated the earlier equation with the addition of the log of the money supply (M1 definition) lagged one quarter:

$$L_t = -2.0 + 0.70\,s_t + 0.25\,m_{t-1} \qquad \hat{\sigma} = 1.40.$$
$$ (6.7) \quad (0.28) \quad (0.06)$$

Here, lagged money is treated as exogenous. Contemporaneous money is excluded because of the possibility of feedback from the current state of the economy to monetary policy. The t-statistic for the exclusion of money from this equation is in excess of four, which is statistically overwhelming evidence that the money stock shifts the labor supply function. A neutral response of labor supply to the money stock is not a plausible hypothesis with respect to the postwar American economy. This finding invites a closer look at theories of labor supply which attempt to explain the vulnerability of the real economy to nominal shocks.

IV. Theories of Labor Supply under Limited Information and Uncertainty

Robert Lucas (1972) presents a completely worked-out theory of the non-neutrality of money based upon an explicit intertemporal theory of labor supply. Barro and Fischer (1976) give a similar but more general and less rigorous development of the same ideas. In essence, this line of thought relies on the strength of intertemporal substitution in labor supply, but suggests that workers sometimes make adjustments that they would find inappropriate in general equilibrium, where they and all the other agents in the economy would have better information about the current state of the economy. In the simplest models, the economy is subject to two kinds of shocks, one real and one nominal. In general equilibrium with full information, labor supply would respond only to the real shock. The nominal shock would simply change all prices and wages in equal proportion. However, workers have access to only a single indicator of the current state of the economy, say, the wage level (or, in Lucas's model, the local price level). Then, as Lucas and various other authors have shown, labor supply over-reacts to the nominal disturbance and under-reacts to the real disturbance.

One line of attack on this theory, as applied to the contemporary U.S. economy, claims that the elasticity of labor supply is too small to explain the observed movements in employment. To put it another way, workers never see, or think they see, large enough movements in real wages for them to make sufficiently large adjustments in hours of work, whether the adjustments are appropriate or not. It seems to me that this criticism is untenable in view of the evidence presented earlier in this paper. Real wages and interest rates *do* move enough to accord with observed movements in employment, given microeconomic evidence about labor supply.

Another line of attack questions the relevance of the critical assumption about limitations in information for labor supply in the modern U.S. economy.[8] Within the general class of models pioneered by Lucas, monetary neutrality fails only when workers do not observe enough different aggregate variables to figure out the nature of the aggregate shock (Barro and Fischer, 1976, give a particularly clear exposition of this point). A large amount of information about prices, wages, employment, unemployment, and other aggregate variables is available virtually instantaneously and essentially for free. It is hard to see how the hypotheses of the theory of monetary non-neutrality on account of limited information can apply when such a wealth of information is available. It is clear, for example, that permitting the participants in Lucas's 1972 model to know the average price level in the economy would restore complete monetary neutrality in the short run. The criticism is not that workers

actually do know what is happening in the U.S. economy, but rather that we must question the success of this type of model in capturing the important source of non-neutrality. At this stage, there seems good reason to doubt the relevance of the existing models of limited information.

A second body of thought has developed with objectives similar to those of the theory of limited information. Its point of departure is the observation that many workers have formal contracts which predetermine wages and let employers choose employment so as to maximize profit once they observe demand. It is clear that this kind of contract will generate monetary non-neutrality (Fischer, 1977; Gray, 1976; and Phelps and Taylor, 1977). It is much less clear why such contracts would ever come into being in the first place. Barro (1977) has pointed out the inefficiency of contracts in which nominal shocks have real consequences. Imposition of the hypothesis of efficiency (marginal product of labor equal to marginal value of workers' time) within labor contracts has the simple implication that the level of employment in a contract economy is essentially the same as in an economy with an open labor market. Contract economies can have monetary non-neutrality but for the same reasons discussed earlier in this section if limitations on the availability of current information make it impossible to distinguish real from monetary shocks, then contracts that are contingent on whatever information is available will be vulnerable to purely nominal shocks. But this is just a restatement of the conclusions of Lucas's theory of incomplete information. Contract theory has not so far made any independent contribution to the explanation of the large shifts in employment and output in response to movements of the money supply that are revealed in aggregate U.S. data. Contract theory has contributed a good deal to the understanding of a variety of phenomena in the labor market, most notably temporary layoffs, but has not so far fulfilled its original promise of explaining the cyclical behavior of employment.

V. Wage Rigidity and the Denial of the Labor Supply Function

Economists have not yet produced a theory of the transmission of monetary shocks to employment and output which is based on realistic hypotheses of the optimal behavior of individuals. The result is a dramatic split between those who conclude that we know almost nothing about business cycles and what to do about them (Barro, 1979; and Lucas and Sargent, 1978) and the proponents of the view that nominal wages are rigid and employment is frequently off the labor supply functions of workers (Gordon, 1976). The debate is all the more intense because wage rigidity leads directly to activist policy—if the labor market won't clear by itself, it can be kept in balance by adroit manipulation of the money supply.

It is unnecessary to elaborate upon the point that wage rigidity and the denial of labor supply are sufficient to explain the phenomena that concern this paper. The textbook IS-LM model, together with any sensible empirical Phillips curve, easily rationalizes all of the empirical findings reported here: exogenous demand drives up output and interest rates, because the LM curve slopes upward and the IS curve shifts upward. Money has a strong influence on real output and employment, because it shifts the LM curve horizontally. The great majority of practical economists will continue to use this model to understand the operation of the economy. No econometric test has so far shown them to be wrong.

What is troublesome about this second position is the utter irrationality it seems to attribute to employment arrangements. It amounts to nothing less than a revocation of the law of supply and demand. In the short run, preferences about work and time away from work are ignored in the determination of employment—only the concerns of employers matter. In a recession or a boom, workers are pushed far off their labor supply functions. Labor supply influences the economy only in the rather long run established by the very flat empirical Phillips curve. The central problem is the explanation *not* of wage rigidity per se, but of the unilateral determination of employment by employers who are free to maximize profit subject to the given nominal wage. The rational worker or group of workers who granted an employer the right to unilateral employment determination would be careful to set up an agreement that respected its labor supply function (Hall and Lilien, 1979). Such an arrangement should not be vulnerable in any serious way to nominal shocks. Some further aspects of this issue are treated in the next section.

Arthur Okun (1975) has written one of the most thoughtful discussions of wage and price rigidity. It gives a convincing rationale for long-term relations between buyers and sellers in a variety of markets. In these markets, the alternative of repetitive auctions or spot negotiation of the terms of sales or employment would be more costly than long-term arrangements. But Okun and the authors who have argued in this vein are less convincing as to why the upshot of these arrangements is to predetermine wages or prices and let the buyer determine quantity unilaterally. I am aware of no important commercial contracts with these provisions. Rather, the typical formal contract predetermines *both* price and quantity. Such contracts cannot explain the mystery of large fluctuations of output in response to nominal influences. In more informal arrangements where prices are apparently stabilized (as, for example, in the copper market), sellers are actively involved in determining quantity, and the

arrangements stabilize output as well as price. In the labor market, long-term relations tend to convert labor into a fixed factor, as suggested by Oi (1962). Again, employment as much as wages is stabilized by this influence. The theory of long-term relations or implicit contracts has some important implications for interpreting aggregate data, but does not seem to provide a coherent explanation of the hypothesis that the level of employment frequently departs from the labor supply function because of the rigidity of wages.

VI. Unilateral Determination of Employment by Employers

Though the simple hypothesis of unilateral employment determination by employers subject to a rigid nominal wage has escaped successful rationalization, it does seem to be a fact that employers have more to say in the short run about employment than do workers. Instead of consulting the relevant prices, wages, and interest rates to decide how much time to spend in the labor market, workers simply show up for their jobs every day and do whatever work is available. When business is strong, employers ask for and receive extra effort from their workers; in some cases they pay for it through overtime hours, and sometimes it is just part of the job. In times of slack, employers unilaterally depress hours of work through temporary layoffs, reduced work weeks, and so on. Even outside the blue-collar industrial sector where provisions for varying annual hours of work are highly formalized, it is clear that employers expect more from workers in times of brisk demand than in times of slack. By and large, it is employers, not workers, who actually make decisions about the volume of work.

Under an employment contract, a worker may grant to the employer the right to determine the level of employment, subject to limitations spelled out in the contract. Herbert Simon (1957) was one of the first to investigate such contracts. A voluminous recent literature examines contracts where employers bear the risk of temporary fluctuations in demand; under these contracts, efficiency generally requires that employers vary the level of employment even though compensation is certain or nearly certain (see Calvo and Phelps, 1977). Part of the original motivation for this line of research was the belief that contracts could provide an explanation for wage rigidity. To put it most precisely, the researchers believed that contracts provided an explanation for the apparent ability of purely nominal shocks to affect the level of employment and output. Though there is still a serious question whether this explanation is successful, the idea of contracts by which workers grant the power of employment determination to employers is an important one and is not at all linked to wage-rigidity theory.

Within theories based on intertemporal substitution, employment contracts could arise because employers have better information about the value of work at different points in the business cycle than do workers. Though markets with sensitively varying wages and interest rates could induce the efficient pattern of work, it is better, from the point of view of managing information for employers, to determine the pattern of work unilaterally. When there is more work to do, employers will ask their workers to work harder and longer, and the workers will cooperate even though it costs them foregone time at home because their long-term contract is attractive. If employers act properly, they will simulate the market by equating the marginal product of labor to the marginal value of time. Of course, all the standard problems in the theory of labor contracts interfere with this process—employers can cheat workers by asking them to work long hours too frequently, workers can default by quitting, and so on.

This line of thought may also help explain cyclical fluctuations in labor productivity. It is a notorious contradiction of the theory of production functions that the same factors produce more output when demand is strong than when it is weak.

Unilateral determination of employment by employers and the consequent "involuntary" nature of employment reductions do not by themselves tell us that only demand matters in determining employment. One of the important contributions of contract theory is to help us understand how unilateral actions by employers can respect the labor supply functions of workers.

VII. Unemployment

Many critics of theories in which employment is always on the labor supply function find them at odds with the facts about unemployment. If a recession is just a period when people recognize that lower levels of work are appropriate, why is it accompanied by a bulge in the number of people who are looking for work?

Part of the answer comes immediately from what is known about temporary layoffs. About a third of the increase in the unemployed during a recession have not lost their jobs, but are on furlough as part of the process of diminished total work effort. A reduction in effort has three components: less intensive work each hour; short hours of work each week; and fewer weeks each year. Those on layoff as part of the last component are counted as unemployed even though they still have jobs, a fact pointed out by Martin Feldstein (1975). However, the remaining two-thirds of the increase during a recession are truly jobless. The great majority of them have lost jobs through permanent layoff; the

rest of the increase consists of people who have just entered the labor force and are taking longer than usual to find work. None of the increase in the unemployed comes from job-quitters. I am not suggesting there is any fundamental economic distinction between quits and layoffs. Rather, the point is that employers more than workers are responsible for the immediate decision to end the employment relationship.

Labor contracts can explain the narrow issue of the active role of employers and the passive role of workers in the process of reducing total labor supply, it seems to me. A contract that provides for unilateral action by employers in terminating the employment relationship when the marginal value of workers' time threatens to exceed the marginal product of labor makes good sense. We are not surprised that a shift in the composition of demand away from one sector causes employers in that sector to lay workers off, and the same principle ought to apply in the aggregate.

The harder issue is the explanation of the strong tendency for the population to spend a larger fraction of its time looking for work in times of slack. Why does the duration of the typical spell of unemployment lengthen from six weeks in normal times to eight weeks in a recession? This question may have an answer somewhere in the microeconomic theory of job search. Indeed, I have made one attempt in this direction by looking at the widening of wage differentials that occurs during a recession (Hall, 1975). Still, the lack of a convincing, complete account of the cyclical behavior of unemployment is a serious weakness of existing theories which respect labor supply.

VIII. Concluding Remarks

We have made important bits of progress in understanding the effects of economic fluctuations within the labor market, without really solving the central problem. The pure intertemporal substitution theory establishes the principle that employment can fluctuate even in an ideal economy; i.e., there is not necessarily a "natural" level of employment. An attempt to understand the movements of employment in the postwar U.S. economy in terms of the pure model yields sensible results. However, very strong results contradicting monetary neutrality reveal the inadequacy of the pure substitution model. A consistent theory of monetary non-neutrality has been advocated by Robert Lucas and others, but there are questions about the informational assumptions that underlie it. The long-standing theory of wage rigidity and determination of employment by demand alone is fully capable of explaining observed movements of the economy. But it amounts to a denial of labor supply in the short

run. No good rationalization for this gross departure from standard economic postulates has yet been offered.

In view of our lack of a complete account of fluctuations in the labor market, it is no surprise that major disagreements persist within the profession. The intellectual case for wage rigidity and the irrelevance of labor supply is an uneasy one, but most practical economists are more persuaded by the evidence of the disproportionate role of demand in short-run fluctuations than they are by existing theories that respect the labor supply function. They will continue to believe in the efficacy, and perhaps in the desirability, of active countercyclical policy until a much more convincing theory based on labor supply is developed and tested.

Notes

1. "Real Wages, Employment, and Inflation" (1969), with Leonard Rapping.

2. For example, in Tobin (1972).

3. See, for example, Tobin (1972) and Modigliani (1977).

4. Again, see Tobin (1972) and Modigliani (1977).

5. The discussion owes a great deal to suggestions by Stanley Fischer.

6. More accurately, it is wealth measured in units of output of next period that should be independent of transitory expenditure policy. Movements of the current real interest rate should be accompanied by corresponding movements of wealth measured in units of current output.

7. The supply function could be augmented by real balance effect, in which case changes in the real stock of money might have a small influence.

8. Tobin (1972), for example.

References

Barro, R.
 (1977) Long-term Contracting, Sticky Prices, and Monetary Policy. *Journal of Monetary Economics*, 3: 305-16.

———— (1978) A Capital Market in an Equilibrium Business Cycle Model. (Unpublished, University of Rochester).

———— (1979) Second Thoughts on Keynesian Economics. *American Economic Review Papers and Proceedings*, 69: 54-59.

Barro, R. and Fischer, S.
 (1976) Recent Developments in Monetary Theory. *Journal of Monetary Economics*, 2: 133-67.

Calvo, G.A. and Phelps, E.S.
 (1977) Employment Contingent Wage Contracts. *Stabilization of the Domestic and International Economy.* Carnegie-Rochester Conference Series on Public Policy, 5. K. Brunner and A.H. Meltzer, eds. Amsterdam: North-Holland.

Feldstein, M.S.
 (1975) *The Importance of Temporary Layoffs: An Empirical Analysis.* Brookings Papers on Economic Activity, 3. Washington, D.C.: The Brookings Institution.

Fischer, S.
 (1977) Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule. *Journal of Political Economy*, 85: 191-205.

Gordon, R.J.
 (1976) Recent Developments in the Theory of Inflation and Unemployment. *Journal of Monetary Economics*, 2: 185-219.

Gray, J.
  (1976)   Wage Indexation: A Macroeconomic Approach. *Journal of Monetary Economics*, 2: 221-36.

Hall, R.E.
  (1975)   *The Rigidity of Wages and the Persistence of Unemployment.* Brookings Papers on Economic Activity, 2. Washington, D.C.: The Brookings Institution.

Hall, R.E. and Lilien, D.
  (1979)   Efficient Wage Bargains under Uncertain Supply and Demand. *American Economic Review.* (Forthcoming; NBER Working Paper, no. 251).

Lucas, R.
  (1972)   Expectations and the Neutrality of Money. *Journal of Economic Theory*, 4: 103-24.

_____
  (1975)   An Equilibrium Model of the Business Cycle. *Journal of Political Economy*, 83: 1113-44.

Lucas, R. and Rapping, L.
  (1969)   Real Wages, Employment, and Inflation. *Journal of Political Economy*, 77: 721-54.

Lucas, R. and Sargent, T.
  (1978)   After Keynesian Macroeconomics. (Unpublished, University of Chicago).

MaCurdy, T.
  (1978)   An Econometric Model of Labor Supply in a Life Cycle Setting. (Unpublished, Stanford University).

Modigliani, F.
  (1977)   The Monetarist Controversy, or, Should We Forsake Stabilization Policies? *American Economic Review*, 67: 1-19.

Oi, W.
  (1962)   Labor as a Quasi-fixed Factor. *Journal of Political Economy*, 70: 538-55.

Okun, A.
  (1975)   *Inflation: Its Mechanics and Welfare Costs.* Brookings Papers on Economic Activity, 2. Washington, D.C.: The Brookings Institution.

Phelps, E.S. and Taylor, J.B.
  (1977)   Stabilizing Powers of Monetary Policy under Rational Expectations. *Journal of Political Economy*, 85: 163-90.

Simon, H.
  (1957)   *Models of Man.* New York: Wiley.

Tobin, J.
  (1972)   Inflation and Unemployment. *American Economic Review*, 62: 1-18.