

## 9 Substitution over Time in Consumption and Work\*

Robert E. Hall

STANFORD UNIVERSITY AND NATIONAL BUREAU  
OF ECONOMIC RESEARCH

Fifty years after the publication of *Value and Capital*, the view of the role of the interest rate in macroeconomic fluctuations expounded there has never been more influential. According to that view, a higher interest rate brings a deferral of spending and an acceleration of production; a temporarily high real wage causes a burst of work effort. A huge body of research in the past decade has sought to clarify the details of equilibration through the interest rate and to measure the strength of intertemporal substitution effects. My purpose here is to comment selectively on the new developments in intertemporal substitution research as it bears on macroeconomic fluctuations. All the research fits nicely into the intertemporal framework of *Value and Capital*.

### 1 INTERTEMPORAL SUBSTITUTION IN *VALUE AND CAPITAL*

The feature of *Value and Capital* that makes it good reading on intertemporal issues fifty years later is its full exploitation of the principle that intertemporal general equilibrium is an application of standard general equilibrium with the goods identified by date. The later chapters of the book apply the principle systematically, first with respect to intertemporal production and then with respect to consumption.

In Chapter XV, 'The Planning of Production', intertemporal equilibrium conditions for the firm appear. The key condition is, 'The marginal rate of substitution [in production] between outputs of any two dates must equal the ratio of their discounted prices' (p. 197).<sup>1</sup>

When a future discounted price rises, intertemporal substitution can take a number of forms:

This substitution may take place at the expense of output earlier than the critical date (reduction of output from now on, in order to accumulate stocks which can be sold at the critical date), or at the expense of later output (acceleration of production, using up of the stock of goods in process, in order to have as much as possible ready at the critical date), or perhaps of both. How far these methods are available depends upon the technical character of the product and the technical character of the initial equipment: the durability of the product, the durability of the unfinished goods which go to make it, the quantity of such unfinished goods available in the initial equipment, and so on (p. 208).

The tenor of the discussion is that intertemporal substitution often – but not invariably – makes short-run product supply fairly elastic. The largest substitution effects occur when the price change is temporary, but fully anticipated, for in that case the accumulation effect operates as well as the drawing against future production. When the price change is unexpected, only the latter effect operates.

Chapter XVII, ‘Interest and the Production Plan’, rightly insists on the discipline of discounted prices (now often called ‘Arrow–Debreu prices’). Within that framework, changes in interest rates have effects that can be restated as changes in sets of prices. An increase in a single one-period interest rate lowers all subsequent discounted prices by the same proportion relative to earlier discounted prices (p. 214). Equal changes in all future interest rates ‘affect the “tilt” or *crescendo* of the production plan’ (p. 224, emphasis in original).

Chapter XVIII on the household, ‘Spending and Lending’, takes the view that ‘the dynamic problem of the private individual ought to be conceived as the choice of a most preferred collection of streams of commodities, out of the various collections of streams which the individual could expect to be able to purchase out of a given expected stream of receipts’ (p. 227). Modigliani (with Brumberg, 1954, and Ando, 1963) and Milton Friedman (1957) pursued this view in the life-cycle and permanent income models of the consumer. The view continues to dominate thinking about consumer behaviour. This view yields a critically important insight whose importance for fluctuations theory has only recently been appreciated:

If the price of X rises, and the rise is not expected to be permanent, the income effect will usually be very slight or indeed quite negli-

gible. The substitution effect, however, may well be much more considerable than in the preceding case [of a permanent price increase]. For substitution may now proceed, not only in favour of other commodities, but also in favour of future purchases of X itself. The main effect of such a temporary rise may well consist in the postponement of expenditure (p. 232).

From this point, it is an easy step to the conclusion that an increase in the current short-term interest rate, with future short-term interest rates held constant, will cause households to decrease current spending and to make plans to spend more in the future instead.

Chapter XX, 'The Temporary Equilibrium of the Whole System, I: Its Imperfect Stability', notes that an economy with significant intertemporal substitution is highly stable with respect to temporary disturbances:

So long as all changes in current prices are regarded as being temporary changes, any changes in current prices will induce very large substitution effects in a large number of markets. A rise in price will make people postpone expenditure, entrepreneurs postpone input and accelerate output; a fall in price will work the opposite way. This substitution over time will be strongly stabilizing; small rises in price will produce large excesses of supply over demand; indeed the forces making for stability are likely to be so potent that it will take a very violent disturbance of data to have any considerable effect on the price system at all (pp. 250–1).

## 2 INTERTEMPORAL SUBSTITUTION IN SUBSEQUENT MACROECONOMIC THOUGHT

Hicks's themes in *Value and Capital* saw some important developments in the macroeconomics of the 1950s and 1960s. As I noted above, the intertemporal analysis of the household flourished in Modigliani's and Friedman's work, though the emphasis on substitution over time was slight. Interest in intertemporal substitution arose primarily in connection with investment theory – acquisition of equipment, structures, housing, and consumer durables could be readily deferred or accelerated, depending on the interest rate. A major turning point in this area was Jorgenson's (1963) derivation of an investment function following the paradigm of *Value and Capital* explicitly. Jorgenson derived a formula expressing the optimal

amount of postponement of acquisition of durable goods given the crescendo of discounted prices.

The next major development in macroeconomic applications of intertemporal substitution was Lucas and Rapping's (1969) paper on labour supply. Lucas and Rapping applied Hicks's insight that short-run supply may be highly elastic because the offsetting income effect is small. Their application was to a supply issue that received little attention in *Value and Capital* – labour supply – but which has received more than its due in business cycle theory in the past two decades. Lucas and Rapping sought to give an equilibrium interpretation to a fact that has occupied the attention of almost every macroeconomist since Keynes – employment has business-cycle fluctuations of about the same amplitude as those of output. In their account of a boom, workers see temporarily high real wages and short-term real interest rates, which makes them substitute toward work and away from leisure. As *Value and Capital* explained, the income effects from these temporary changes is small, so short-run labour supply is quite elastic as a function of these two variables. Lucas and Rapping's claim that the intertemporal substitution view could explain the ups and downs of the US economy, including the Great Depression, was vigorously disputed by Rees (1970) and others, but the idea has remained centrally important in business cycle theory.

Lucas's (1972) classic paper on monetary non-neutrality gives full theoretical expression to intertemporal substitution in labour supply in general equilibrium. He examines a model economy subject to a temporary real shock and a permanent monetary shock. Analytical convenience requires him to assume that people live only two periods. As a result, the minimal income effect stressed by Hicks does not apply to Lucas's model. Instead, Lucas makes an explicit assumption that the substitution effect dominates the income effect. The model has a business cycle because some years appear to be better than others for work and production.

Recent work by Kydland and Prescott (1982) and Prescott (1986) has seen the full application of the intertemporal substitution paradigm from *Value and Capital*. Households choose the scheduling of consumption by equating the marginal rate of substitution between current and future consumption to the discounted price ratio, suitably adjusted for stochastic variations. Households also choose the scheduling of work with reference to the expected ratio of the current real wage to the discounted future real wage. Businesses choose the

timing of investment to maximise business value, in the sense appropriate to the stochastic setting. One of the most novel features of this line of thought is its assumption that the major driving force in macroeconomic fluctuations is stochastic shifts in the production function. When conditions are favourable, the real wage is high and workers have an incentive to work harder than usual – a boom occurs. Recessions are periods when productivity and real wages are low. Almost all the features of this type of model are controversial – see Summers (1986) for a critique – but the line of thought has become well established.

Barro (1981) has developed a related model in which the driving force is government purchases of goods and services, rather than productivity. The government sets off a boom by purchasing more output. The real interest rate rises temporarily, which stimulates added work effort. In addition, the higher rate should cause businesses to defer investment and consumers to defer consumption. The first prediction is fulfilled in data for most countries, but not the second. Consumption and government purchases have a zero or slightly positive correlation, not a negative correlation. To fit the data, the model would have to invoke a high intertemporal elasticity of substitution in labour supply and a low elasticity in consumption. To some extent, the evidence supports this combination of elasticities.

### 3 EVIDENCE ON INTERTEMPORAL SUBSTITUTION IN SUPPLY

The hypothesis of high intertemporal substitution in labour finds little apparent support in the very large body of research on labour supply at the level of the individual. This research has consistently failed to find substitution elasticities in labour supply much above 0.1 or 0.2 for the dominant demographic groups in the labour force (Pencavel, 1986). However, the source of wage variation exploited in panel data is almost entirely life cycle in nature, not the year-to-year temporary fluctuations that would reveal the short-run elasticity of substitution. Existing studies of panel data do not test the hypothesis that modest temporary added incentives to work will call forth significant additional work effort. Consider a worker earning \$10 per hour and working a standard 40-hour, five-day work-week. Would he add 8 hours of work on Saturdays in response to an opportunity to earn \$12

per hour for a few months? Studies of panel data are unable to isolate instances of this type. It is common to find episodes when work effort rises 20 per cent above average. It is not possible to isolate the wage changes that might have induced the increase in work effort, in the face of large amounts of random noise in measured wages for individuals. In principle, an exogenous variable that signalled a true increase in the wage could solve this problem, but the available instrumental variables do not seem to be correlated with these temporary fluctuations in employment opportunities.

Aggregate labour supply schedules are generally found to be quite inelastic as well. Again, the findings do not settle the issue of the short-run elasticity of labour supply. When that elasticity is high, correctly measured real wages will be close to constant. If there is any noise in the actual measure of the real wage, the correlation of work effort and the real wage will be close to zero. In addition to standard measurement problems, wage measurement may be complicated by the tendency for actual wages to fluctuate less than the underlying shadow value of time (see Hall, 1980).

My own view is that the facts about the labour market actually point in the direction of a combination of reasonably high intertemporal substitution from one year to another (though not from one decade to another) and fairly strong fluctuations in the shadow value of work. The evidence is necessarily indirect. First, substantial annual fluctuations in hours of work are the rule, not the exception. Volatility is significant among older workers with long-term jobs in stable sectors, as well as in markets well known for employment variations, such as construction. Institutions in the labour market accommodate, rather than resist, variations in annual weeks of work. Perhaps the most convincing evidence is Ramey's (1988) finding that firms schedule output as if they faced flat or declining marginal cost. If the short-run labour supply schedule were inelastic, firms would accumulate inventories during periods of low demand and decumulate in high demand, so as to smooth employment and limit the disamenity from variable weeks of work. If anything, according to Ramey, firms do the opposite. Seasonal movements of employment are substantial, which points in the same direction as Ramey's findings.

### 3.1 Characterisation of Preferences about Work over Time

*Value and Capital* raised the possibility that short-run supply may be more elastic than long-run supply, because of the absence of income

effects. Recent work on labour supply has pushed the idea a little further by recognising a difference between short- and medium-run intertemporal substitution. A convenient family of intertemporal preferences for this purpose follows suggestions of Sargent (1979, p. 371) and Kydland and Prescott (1982). Let  $z_t$  be the accumulated stock of current and past work effort, with persistence factor  $\omega$ :

$$z_t = (1-\omega) \sum_{s=0}^t \omega^s L_{t-s} \quad (1)$$

Because variations in hours of work per week are small, the measure of work effort,  $L_t$ , is weeks of work in period  $t$ . The parameter  $\omega$  controls the memory of past work and leisure. If  $\omega$  is 0, there is no memory; only current work effort matters. If  $\omega$  is large (close to its upper limit of 1), then  $z_t$  depends on a long distributed lag of past work effort. The worker orders work schedules with a (dis)utility function that is separable over time in the  $z_t$ s:

$$\sum_{t=0}^T \frac{(\gamma - z_t)^{1/\sigma-1}}{1/\sigma-1} \quad (2)$$

Define effective leisure as  $\gamma - z_t$  and actual leisure as  $\gamma - L_t$ . The parameter  $\sigma$  is both the intertemporal elasticity of substitution in effective leisure and the long-run elasticity of substitution in actual leisure. In the short run, the elasticity of substitution in actual leisure is greater than  $\sigma$  by an amount that is controlled by the memory parameter,  $\omega$ . The parameter  $\gamma$  can be interpreted as the number of weeks physically available for work.

A worker with a high  $\sigma$  will suffer little from a work schedule involving many weeks of work per year in one decade and few weeks per year in another decade, in comparison to putting in the same number of lifetime weeks with no variation from decade to decade. In a situation with free choice of weeks, such a worker will concentrate weeks disproportionately during the years of highest wages.

On the other hand, a worker with low intertemporal substitution,  $\sigma$ , but high memory persistence,  $\omega$  (that is, close to one), will tolerate short-term fluctuations in weeks of work but resist decade-to-decade movements. This feature makes the preferences attractive within a theory in which the deadweight burden from employment fluctuations is small or absent. Kydland and Prescott (1982) rely on a slightly more general class of preferences in their model that attempts to rationalise employment fluctuations as fully efficient responses to

vibrations in technology. In their specification, current work can have a role in the utility function beyond the role implicit in the variable  $z_t$ .

To illustrate the difference between the short-run and the medium-run responses of labour supply to wage changes, consider the following question: let  $2N$  be the number of periods considered to define the medium run, say 24 months. Suppose a worker increases his weeks of work by 1 per cent in periods  $t-N, \dots, t, \dots, t+N$ . By what per cent does his supply price of a week of work in period  $t$  increase? The elasticity of labour supply over the  $2N+1$  period run is the ratio of the two numbers.

It is convenient to use the  $\lambda$ -constant or Frisch labour supply schedule to answer this question. Let  $\lambda$  be the Lagrangian multiplier associated with the worker's intertemporal budget constraint. The first-order condition associated with labour supply is

$$\frac{\partial U(L_1, \dots, L_t, \dots, L_T)}{\partial L_t} = \lambda w_t \quad (3)$$

Here  $w_t$  is the real wage in period  $t$  stated in period 0 prices. The Frisch inverse labour supply function is simply the marginal disamenity of work stated in wage units:

$$\frac{1}{\lambda} \frac{\partial U(L_1, \dots, L_t, \dots, L_T)}{\partial L_t} \quad (4)$$

When  $U$  is additively separable in labour, this can be solved to give current labour supply as a function of the current wage. Absent separability, it states the supply price of work in one period as a function of the level of work in that and other periods. Keeping  $\lambda$  constant has two interpretations. First, equation (4) gives the supply price of labour at different points in time along the same labour-supply trajectory. Under this interpretation, statements about the response of the supply price to different levels of work are comparisons of the supply price at different points in time; the change in the level of work is fully anticipated. Second, the supply price conditional on  $\lambda$  has a comparative statics interpretation when the change has little or no effect on  $\lambda$ . Under this interpretation, equation (4) is very similar to (but not quite the same as) the compensated labour supply schedule.

The Frisch labour supply function associated with the preferences considered here is



$$w_t = \frac{1}{\lambda} (1-\omega) \sum_{s=t}^T \omega^{s-t} (\gamma - z_s)^{-1/\sigma} \quad (5)$$

Let  $x$  be the common increment to  $L_{t-N}, \dots, L_t, \dots, L_{t+N}$ . For simplicity, assume that the horizon,  $T$ , is infinite and that  $L_t$  and  $z_t$  have the common value  $L$  in all periods. Then some manipulations show that the slope of the inverse labour supply schedule is

$$\frac{dw}{dx} = \frac{1}{\lambda\sigma} (\gamma - L)^{-1/\sigma-1} \left( 1 - \frac{2\omega^{N+1}}{1+\omega} \right) \quad (6)$$

The elasticity,  $\epsilon(N)$ , of the labour supply schedule is

$$\epsilon(N) = \sigma \frac{\gamma - L}{L} \left[ 1 - \frac{2\omega^{N+1}}{1+\omega} \right]^{-1} \quad (7)$$

If there is no memory of past work ( $\omega = 0$ ) or if the displacement of work is lengthy ( $N$  is large), then the elasticity is just the intertemporal elasticity of substitution in leisure,  $\sigma$ , multiplied by the ratio of non-work time to work time:

$$\epsilon(\infty) = \sigma \frac{\gamma - L}{L} \quad (8)$$

The elasticity  $\epsilon(\infty)$  controls labour supply over the life cycle. A worker with an  $\epsilon(\infty)$  of 1 with a wage at age 40 double his wage at age 20 would work twice as many hours at age 40. Because life-cycle variations in weeks of work are not nearly this large, there is a presumption against high values of  $\epsilon(\infty)$ ; rather, the evidence appears to favour values of 0.1 to 0.2. If  $(\gamma - L)/L$  is  $5/47 = 0.11$  and the medium-run elasticity of labour supply is 0.15, then  $\sigma$  is  $0.15/0.11 = 1.4$ . I assume that the medium run is short enough for income effects not to offset substitution effects.

By contrast, the elasticity of labour supply in the context of a one-period displacement ( $N = 0$ ) is

$$\epsilon(0) = \sigma \frac{\gamma - L}{L} \frac{1+\omega}{1-\omega} \quad (9)$$

If memory decays at a rate of 20 per cent per period, as might be

appropriate in a quarterly model, the very short-run elasticity is  $1.8/0.2 = 9$  times as large as the medium-run elasticity. The specification is successful in delivering a high short-run elasticity of labour supply without relying on significant decade-to-decade elasticity of labour supply.

### 3.2 The Burden of a Fluctuating Work Schedule

Except in the polar case of perfect intertemporal substitution, workers resist variable work schedules. That is, they would choose jobs with constant weeks over jobs with variable weeks but the same total hours and total compensation.<sup>2</sup> With the intertemporal utility function, it is possible to compute the work equivalent of variability in a work schedule. The work equivalent is the added amount of work that a worker would be willing to do in exchange for a completely constant work schedule.

Most variability of work schedules is idiosyncratic to the worker. Even with a generous allowance for measurement error in annual hours of work, it appears that the standard deviation of hours around the labour supply schedule is at least 13 weeks per year, or about 25 per cent of the normal level of work.<sup>3</sup> Variations occur within jobs – temporary lay-offs and deferred vacations – and between jobs, as spells of unemployment. Among groups who normally work close to 50 weeks per year, there is an inevitable asymmetry in the variability – weeks can drop by a large amount but cannot rise above 52.

The central question of variability, from the point of view of the worker, is whether the occasional interruptions in work create valueless idle periods (inelastic labour supply) or whether workers can make valuable use of stretches of non-work (elastic labour supply). The preferences described in this section can give different answers depending on the duration of time out of work. The high- $\sigma$  worker can make good use of time off in both the short and medium runs. The worker has numerous opportunities to benefit from the use of time other than in work and does not exhaust the stock of those opportunities even over several years. The low- $\sigma$ , high- $\omega$  worker has good alternatives to work in the short run, but exhausts them after a few months or a year and finds himself unproductively idle thereafter. The low- $\sigma$ , low- $\omega$  worker cannot make good use of time away from work even in the short run.

### 3.3 Life-cycle Wage and Work Profiles

Because the typical life-cycle wage pattern is smooth, memory as measured by the parameter  $\omega$  is unimportant in the life-cycle relation between wages and work effort. To see this, consider the simple case of an individual with exponential decline in leisure hours over the life cycle:

$$\gamma - z_t = \gamma_0 e^{-\theta t} \quad (10)$$

The inverse labour supply function from equation (4) is

$$\begin{aligned} w_t &= \frac{1}{\lambda} (1 - \omega) \sum_{s=t}^T \omega^{s-t} e^{\frac{\theta}{\sigma} s} \\ &= \frac{1}{\lambda} (1 - \omega) \gamma_0 \left[ 1 - \omega e^{\theta/\sigma} \right]^{-1} e^{\frac{\theta}{\sigma} t} \end{aligned} \quad (11)$$

The rate of growth of the wage is the rate of growth of leisure,  $\theta$ , divided by the elasticity of substitution,  $\sigma$ , without any adjustment for the persistence of memory of past work,  $\omega$ . To put it differently, studies of labour supply using panel data on individuals, with instruments such as age and education that are associated with gradual life cycle changes in wages, will reveal the low elasticity,  $\sigma \frac{\gamma - L}{L}$ , not the high elasticity,  $\sigma \frac{\gamma - L}{L} \frac{1 + \omega}{1 - \omega}$ . The key parameter  $\omega$  is identified by the high frequency movements of the wage, such as the coming and going of special opportunities, not the low frequency movements over the life cycle.

### 3.4 Panel Studies of Labour Supply

Pencavel (1986) and Heckman and Killingsworth (1986) survey the direct evidence on labour supply for panel data on individuals. MaCurdy (1981) is one of the leading studies they consider. The basic approach is the following: the elasticity of labour supply is the ratio of the change in work effort to the change in wage that occurs as the result of a change in one or another instrumental variable. For example, as a worker moves from age 29 to age 30, his wage typically rises because this is a steep part of the age-wage-rate profile. If hours

of work also rise from age 29 to age 30, there is a positive wage-elasticity of labour supply. Given that age effects are largely anticipated, the response is a pure substitution effect. Pencavel's Table 1.22, p. 85 suggests that the elasticity estimated on this basis is somewhere between 0 and 0.45 for men.

In view of the fact that researchers have not been able to find instruments that identify temporary changes in wages (ones lasting two years or less), it seems appropriate to view the estimates from panel data as measuring the medium-run elasticity of labour supply,  $\epsilon(\infty)$  in the notation of section 1. The estimates are thus consistent with short-run elasticities of 1 or 2 or even more, if amplification occurs through memory of recent work effort.

Pencavel devotes considerable attention to the one group of short-run exogenous events whose labour supply effects are well documented – the negative income tax experiments. Experimental subjects experienced a three-year reduction of 30, 50, or 70 per cent in effective wages. Pencavel's survey reaches the conclusion that the elasticity of the response of labour supply to these wage reductions was in the fairly narrow range from 0.06 to 0.19 (Table 1.21, p. 80). This finding is the single most telling criticism of the view that intertemporal substitution is high in the short run.

The negative income tax findings are less than definitive for the following reason. For good reasons relating to asymmetric information, workers delegate to their employers the determination of weeks of work. Workers shop among employers with different policies for setting weeks, but once the worker accepts a job, the weeks he puts in at that job are largely out of his control. In particular, if an event occurs that is personal to the worker, but not within the class of events (such as disability) contemplated by the employment arrangement, it is unlikely that the employer will agree to a reduction in weeks *ad hoc*. Employment arrangements with given, understood rules are the solution to the problem of opportunistic behaviour by both employers and workers. Of course, workers can always take extra weeks off by quitting one job and delaying taking another job, but that step involves the sacrifice of the advantages of seniority. The finding of small reductions in weeks of work in the negative income tax experiments is not inconsistent with the hypothesis that much larger reductions can occur when the marginal revenue product of labour declines in a downturn. One is unprecedented and unfamiliar, completely new to the environment under which employment ar-

rangements have evolved; the other is exactly within the historical experience that shaped those arrangements.

Another reason that the panel studies, both survey and experimental, are not good evidence against elastic short-run supply is the amount of variability they reveal in annual work effort. According to MaCurdy (1981), the standard deviation of annual hours of work around the predictions of his labour supply function is several hundred hours, a significant fraction of the normal level of around 2000 hours. The supply function has a different intercept for each worker, so all this noise represents variability over time in relation to the worker's own normal level. If the intertemporal elasticity of labour supply is as low as the numbers in Pencavel's survey, with respect to substitution between one year and the next, then the deadweight burden of the unexplained variability of work is extremely high. A more reasonable conclusion is that the low elasticities apply to life-cycle influences but that much higher elasticities operate at year-to-year frequencies.

### 3.5 Abowd and Ashenfelter's Study of Wage Premiums for Variable Work Schedules

Abowd and Ashenfelter (1981) investigate the trade-off between the level of earnings and the variability of work effort by asking what wage premium is required to attract workers to variable employment. They consider the case of a worker whose alternative employment is at a constant wage and where the worker can choose the number of hours to work, given that wage. The employer at hand also pays a constant wage, but it must embody a premium because the employer has the right to make random interruptions in the worker's schedule. Each year, hours are reduced by a random variable with mean  $\mu$  and standard deviation  $\delta$ . In slightly simplified form, the theoretical expression for the wage premium they derive is

$$-\gamma \frac{\mu}{h} + \frac{1}{2\epsilon} \left[ \frac{\mu}{h} \right]^2 + \frac{1}{2} r \left[ \frac{\delta}{h} \right]^2 \quad (12)$$

Here  $\gamma$  is the fraction of lost earnings replaced by unemployment insurance (UI),  $\epsilon$  is the elasticity of the compensated labour supply schedule,  $r$  is the coefficient of relative risk aversion, and  $h$  is the unconstrained level of hours. The first term measures the desirability of employment fluctuations when there is unemployment

compensation. For small amounts of unemployment, workers prefer variability because they value leisure at the market wage, but receive unemployment compensation as well during periods of extra leisure. The second term measures the area above the labour supply schedule associated with the average depression,  $\mu$ , of work below the worker's optimum. The third term measures the burden of stochastic variability.

Abowd and Ashenfelter develop their theory within a one-period stochastic setting. An alternative would be to consider the variability of hours of work over time instead of over states of the world. In that case, with time-separable preferences, the coefficient of relative risk aversion,  $r$ , would be the reciprocal of the labour supply elasticity,  $\epsilon$ . The second two terms would combine into a single term involving expected squared unemployment,  $\mu^2 + \delta^2$ .

The empirical work carried out by Abowd and Ashenfelter examined the relation in the Panel Study of Income Dynamics of the mean and variance of unemployment to the actual wage and observed determinants of the alternative wage. A preliminary study of annual hours of unemployment yielded the estimated mean and variance of unemployment for each worker. Then a regression of the log of the actual wage on mean unemployment divided by hours, the mean squared divided by hours squared, and the variance divided by hours squared yields estimates of the key parameters.

Although the theory invoked by Abowd and Ashenfelter deals with a wage premium for employment variability, the data reveal that wages for workers facing more unemployment are *lower* than are wages for otherwise identical workers in jobs with little unemployment. In order to fit this feature of the data, the estimation process has to come up with a very high estimate of the UI replacement rate,  $\gamma$ . The estimate claims that the worker receives \$2.49 in unemployment benefits for each dollar of lost earnings. In such a world, employers with frequent temporary lay-offs could attract workers with lower base wages than those offered by other employers, because each layoff is a paid vacation at over double normal earnings.

Abowd and Ashenfelter are aware that their estimated replacement rate is high. They offer some calculations to try to convince the reader that the estimates are only a little too high, and that an alternative estimate of around one is completely reasonable. I believe that these calculations are quite erroneous. First, they contradict evidence presented in Table 4.A.1 of the paper which show that industry average replacement rates are in the range of 10 to 25 per

cent. These replacement rates seem quite reasonable in the light of nominal replacement rates of around 50 per cent subject to limitations in the first week of unemployment and exhaustion of benefits. Second, the calculations themselves seem to contain an outright error. Rather than simply compare the replacement rates themselves, Abowd and Ashenfelter look at average UI benefits as a fraction of total earnings. This is just the first term in equation (12). To convert the actual data on benefits/lost earnings to the form of benefits/total income, one should multiply by lost earnings/total income, which is the unemployment rate of hours. According to Table 4.A.1, the average worker lost 27 hours to unemployment in 1973. The average amount of work and unemployment in the same year was 2360 hours, so the unemployment rate of hours was 1.2 per cent. Thus the observed average UI benefits in 1973 were 0.130 (the 1973 replacement rate) times 1.2 per cent, or 0.15 per cent of earnings. By contrast, the 'observed average' presented in Table 4.5 is 1.10 per cent, almost an order of magnitude higher. The explanatory notes for Table 4.5 indicate the 'observed average' is calculated as the product of the replacement rate (13 per cent) and the fraction of the sample who had positive unemployed hours (10.8 per cent). Apparently, Abowd and Ashenfelter confused the fraction with some unemployment with the unemployment rate of hours. The two differ by almost an order of magnitude (10.8 per cent against 1.2 per cent).

The fact of the matter is that the estimated replacement rate of 2.5 is more than an order of magnitude too high. Even in the alternative estimates of Table 4.4, the replacement rate of 1.2 is about 6 times too high. What accounts for the gross overstatement? I think that the answer is probably the problem that dogs all research on wages – there are important unmeasured individual wage determinants that are correlated with the right-hand variable. In this instance, workers who are in jobs with unusually large amounts of unemployment tend to be workers with adverse unmeasured earning potential. It appears to be a hopeless task to tease out the small effects of employment variability from data contaminated by unmeasured wage determinants.

### 3.6 Ramey's Evidence on Product Supply

Valerie Ramey (1988) has assembled evidence on intertemporal substitution in production. All the production substitution issues discussed in *Value and Capital* are considered in her work. She

examines the joint behaviour of output and finished goods inventories in industries that produce to stock rather than to order. The basic idea is that a convex technology with rising marginal cost and limited intertemporal substitution creates an incentive to produce to inventory in anticipation of increases in demand. The incentive to use inventories to avoid fluctuations in output is even stronger when the firm internalises the short-run labour supply schedules of its workers. Earlier work by West (1986) and others showed that inventory investment does not smooth production. Ramey's work interprets that fact as revealing that marginal cost is flat or even downward-sloping. Her results, taken at face value, require a combination of a high elasticity of intertemporal substitution and large movements of the shadow wage. That is, a boom is either a time when workers work much harder in response to a modest increase in the incentive to work, or a time when economies of scale permit a substantially higher shadow real wage.

The following simplification of Ramey's approach shows how inventory behaviour reveals the extent of intertemporal substitution opportunities. Within a broader optimisation problem through which the firm determines its sales, there is a sub-problem of minimising the cost of those sales. Suppose the expected cost of producing to meet given sales is proportional to

$$\frac{1}{2} E_t \sum_{\tau=t}^T \left[ \gamma y_{\tau}^2 + (x_{\tau-1} - \alpha s_{\tau})^2 \right] \quad (13)$$

Here  $E_t$  is the expectation conditional on information at time  $t$ ,  $y$  is output,  $x$  is the end-of-period stock of finished-goods inventories, and  $s$  is the level of sales. The parameter  $\gamma$  controls intertemporal substitution; if the firm perceives an upward-sloping short-run labour supply schedule for its workers or if the technology itself is convex,  $\gamma$  will be positive. The parameter  $\alpha$  controls the inventory/sales ratio. An identity links the variables:

$$x_t = x_{t-1} + y_t - s_t \quad (14)$$

A first-order condition necessary for the optimal scheduling of production is

$$E_t [\gamma(y_t - y_{t+1}) + x_t - \alpha s_{t+1}] = 0 \quad (15)$$

This condition characterises the cost-minimising policy, for negative as well as positive values of  $\gamma$  (see Ramey, 1988).



Now let

$$\begin{aligned} h_t &= x_t - \alpha s_{t+1} \\ &= x_t - x_{t-1} - (\alpha s_{t+1} - x_{t-1}) \end{aligned} \quad (16)$$

The variable  $h_t$  is inventory investment in excess of the amount needed to maintain the level of inventories at its usual relation to sales;  $h_t$  measures inventory investment undertaken to smooth production plus a purely random element related to surprises in sales. The first-order condition in terms of  $h_t$  is

$$E_t h_t = -\gamma(y_t - E_t y_{t+1}) \quad (17)$$

Alternatively,

$$h_t = -\gamma(y_t - y_{t+1}) + \epsilon_t \quad (18)$$

Here  $\epsilon$  is an expectation error satisfying  $E_t \epsilon_t = 0$ .

Equation (18) strips the first-order condition to its bare essentials. A firm with low intertemporal substitution ( $\gamma \gg 0$ ) will deplete its inventories by setting  $h_t < 0$  when it is producing more this period than it plans to produce next period ( $y_t - y_{t+1} > 0$ ). Note that the inventory draw-down affects the magnitude of  $y_t - y_{t+1}$ ;  $h_t$ ,  $y_t$ , and  $y_{t+1}$  are all variables controlled directly by the firm. When the optimal output plan calls for lower output this period than next period, the firm with rising marginal cost will accumulate inventories in excess of the level required by maintenance of the inventory/sales ratio.

Imperfect intertemporal substitution has a sharp and robust implication. When an outside event stimulates product demand temporarily, it should also cause an inventory draw-down, in the sense of a negative value of  $h_t$ . To put it differently, an instrumental variable positively correlated with  $y_t - y_{t+1}$  should be negatively correlated with  $h_t$ . The negative of the ratio of the covariances is the instrumental variable estimator of  $\gamma$ . By contrast, the firm with perfect intertemporal substitution (highly elastic labour supply and no convexity of technology) is indifferent to the scheduling of production. Its only objective is to maintain its inventory/sales ratio at the prescribed level  $\alpha$ , so it always plans for  $h_t = 0$ . An instrumental variable positively correlated with  $y_t - y_{t+1}$  will have zero correlation with excess inventory accumulation. The instrumental variable estimator for  $\gamma$  will be zero.

For a firm with decreasing marginal cost ( $\gamma < 0$ ), it is efficient to bunch production. The firm would produce its output for all periods in the first period but for the cost of departing from the normal inventory/sales ratio,  $\alpha$ . Even in the face of that cost, the firm amplifies fluctuations in output so as to obtain the economies of bunching output. Equation (18) shows that the firm builds extra inventories in the same periods when current output exceeds expected future output. An instrument positively correlated with  $y_t - y_{t+1}$  will also be positively correlated with  $h_t$ , and the IV estimate of  $\gamma$  will be negative.

Ramey's model as estimated is considerably more elaborate than the one just discussed. The cost function is cubic in output and the linear term depends on the wage, the price of materials, and the price of energy. There is a time trend in the quadratic term. There are costs of adjustment of the level of output in the form of a term involving the square of  $y_t - y_{t-1}$ . Finally, there are random shifts in the technology itself, in the adjustment cost term, and in the target inventory/sales ratio. In principle, the use of a wage in the model means that the estimate of the slope of marginal cost does not measure the slope of the labour-supply schedule. However, there is very little variation apart from trend in the real wage. Hence the results are almost exactly the same if the wage is omitted and the slope parameter is interpreted to include the slope of labour supply. To put it another way, the stability of the real wage means either that labour supply is highly elastic or that the wage does not reflect the true shadow price of labour over the cycle.

In Ramey's work, the exogenous variables that shift product demand and do not shift product supply are three measures of federal military spending, the relative price of oil, a dummy variable for the political party of the president, and population. The seven industries Ramey studies are food, tobacco, apparel, chemicals, petroleum, and autos. Except for the auto industry, a dummy for auto strikes also serves as an instrument. For the tobacco industry, a set of variables characterising federal regulation is included. In all seven industries, the estimate of  $\gamma$ , the slope of the marginal cost schedule, is negative. In four of the seven industries, the point estimate of  $\gamma$  is more than two standard errors below zero, so the evidence against rising marginal cost is statistically unambiguous. All seven of the industries tend to bunch production during times of high sales. They typically accumulate inventories beyond the amount needed to maintain the normal inventory/sales ratio at the same time that output is strong

because sales are high. Firms with rising marginal cost would behave in the opposite way, building inventory stocks in times of weak sales and drawing them down when sales are high. Ramey's strong statistical evidence in favour of production bunching is inconsistent with rising marginal cost.

The bunching of production revealed in Ramey's study means that work effort is equally bunched. Nothing in the results rules out the possibility that management is setting a work schedule that is painful to workers because of its irregularity, but that some kind of coordination failure leaves management with no incentive to smooth the work schedule. The opportunity to accumulate and decumulate inventories means that management could smooth work schedules at relatively low cost. But the most reasonable conclusion from these observations, in my opinion, is that the disamenity of irregular work schedules is sufficiently low to be offset by fluctuations in the shadow value of work. Workers would gain relatively little by trading the current practice of taking extra time off when demand is low for taking the same time off every August.

### **3.7 Bils's Evidence on Overtime**

Mark Bils (1987) presents evidence that appears to be unfavourable to the hypothesis of high intertemporal substitution in labour supply. Bils argues that the effective marginal cost of labour input to the firm rises sharply when output and labour input rise in an expansion. He reaches this conclusion through a detailed analysis of data on the cost of overtime hours and on the way that firms divide an increase in labour input into increases in weekly hours and in the number of workers at work each week. He makes a convincing case that a larger fraction of incremental hours of work have to be paid at the 150 per cent overtime rate when hours are close to the 40-hour limit on straight-time hours mandated by the Fair Labor Standards Act than when average hours are below that limit. If firms perceive overtime premiums as an extra cost associated with the use of labour at the time the premiums are paid, then Bils's conclusion about fluctuations in the effective marginal cost of labour are inescapable. However, as Bils points out, it is far from an obvious hypothesis that overtime premiums are a cost. Firms may be committed to paying premiums of a certain amount on the average, as part of their employment agreements. Then paying a premium during one episode relieves them of a future compensation obligation. Bils attempts to show that

overtime premiums are treated as true costs by the firm. Rather than pay extensive overtime premiums when demand is strong, he argues, firms incur substantial adjustment costs to add more workers. Because the firm should always equate the marginal costs of labour input from adding hours and adding workers, the cyclical pattern of marginal adjustment costs also reveals the cyclical pattern of the cost of a marginal hour. That pattern is quite similar to the one inferred directly from overtime premium costs.

Bils's position on marginal labour costs is fundamentally at odds with high intertemporal substitution in labour supply. One of the ways that firms can adjust labour input is to vary the annual weeks of work of their long-term workers. The firm should equate the marginal cost of labour from this type of adjustment to the marginal costs of adding weekly hours or adding more workers. If the labour-supply schedule is flat, and adjustment costs for variations in weeks are minor, then cyclical fluctuations in marginal labour cost should be small. Bils's claim to finding large fluctuations contradicts the flatness of the supply of weeks, even though he gives no explicit consideration to the weeks dimension. There is no reason why firms should be running up substantial overtime costs when they could increase weeks without raising the marginal cost of labour. Bils's finding needs to be reconciled with Ramey's evidence of flat or declining marginal cost. If the bunching of output raises average cost because of overtime premiums, then firms should smooth output. Their failure to smooth suggests that overtime payments are not a cost or that the cost reductions associated with the bunching are large enough to offset the extra overtime expense.

Bils's evidence is completely convincing that compensation paid for the marginal hour rises sharply when employment is high. Data on total and overtime hours show that an increase in total hours from 40 to 41 increases overtime by 0.39 hours whereas an increase from 41 to 42 increases overtime by 0.50 hours (Bils, 1987, p. 845). The marginal wage is 4.6 per cent higher at 41 hours than at 40 hours. Average weekly hours in manufacturing decline by a little over an hour in the typical recession. Although these calculations depend on the assumption that the 50 per cent overtime premium dictated by the Fair Labor Standards Act is actually paid, Bils cites evidence that compliance is virtually universal.

Bils explains in detail why the evidence on overtime payments is inconclusive:

The wage a firm pays takes a very large jump of 50 per cent at 40 hours per week due to the overtime premium. Workers' disutility of working is presumably smoothly increasing in hours. This implies workers would strictly prefer working some overtime hours to working 40 hours per week (in fact, overtime hours are rationed in many instances). By offering workers overtime hours, therefore, a firm may incur some goodwill, which allows it to lower compensation in another form, if not then, at some other time. The implication is that the effective cost premium of an overtime hour may be less than the 50 per cent explicit payment (p. 843).

Bils's response is to try to infer the slope of the marginal cost of labour from variations in the cost of adding workers. In equilibrium, the firm will equate the marginal costs of labour along both the hours and workers dimensions. Bils hypothesises adjustment costs for adding workers. His formula for the marginal cost of an hour of work obtained along the worker dimension is

$$(1 + f) \frac{W(H)}{w} + q \Delta^2 \log N_{t+1} \quad (19)$$

Here  $f$  is the amount of fringe benefits not related to hours of work, expressed as a fraction of total earnings (reported by the Chamber of Commerce),  $H$  is weekly hours,  $W(H)$  is the average hourly wage at  $H$  hours,  $w$  is the straight-time hourly wage,  $q$  is the adjustment cost parameter, and  $N_t$  is the number of workers. Note that marginal cost is measured in units of the straight-time wage. The marginal cost of an added hour in the weekly hours dimension is  $\bar{W}(H)$ , which is specified as a cubic in  $H$  multiplied by the straight-time wage.

Bils's estimating equation is

$$\Delta^2 \log N_{t+1} = \phi (1 + f) \frac{W(H)}{w} - \phi \frac{\bar{W}(H)}{w} + \eta \quad (20)$$

Here  $\phi$  is the reciprocal of the adjustment cost parameter and  $\eta$  is a random disturbance. All of the variables in  $\bar{W}(H)/w$  have unknown parameters, so identification hinges on variation in  $W(H)/w$  that is independent of variation in  $\bar{W}(H)/w$ . At the most general level, it is clear that identification from this source of variation must fail;  $\bar{W}$  is just the marginal to  $W$ , so there can be no independent variation. Bils's equation does not suffer from perfect collinearity for three

reasons. First, there is some variation over time in  $f$ , the element of compensation that is independent of weekly hours. Second, although  $W$  and  $\bar{W}$  have an exact functional relation, they are not precisely linearly dependent, so the estimation does not break down numerically. Third, Bils uses actual data on average hourly compensation to calculate  $W/w$ , rather than imposing the relation between  $W$  and  $\bar{W}$  dictated by his theory. Because the Bureau of Labor Statistics does not use exactly Bils's formula to derive straight-time earnings  $w$  from average hourly earnings  $W$ , the dependence between  $W$  and  $\bar{W}$  is broken.

The basic train of logic implicit in Bils's work is the following: we observe the ratio between incremental hours and incremental workers when total labour input rises. This ratio identifies the ratio of the slopes of the two implicit supply schedules for hours and workers. If we can determine the slope of the implicit supply schedule for workers, then we can infer the slope of the supply of hours from our knowledge of the ratio. The slope of the marginal cost schedule for workers can be inferred from the way that the number of workers changes when there is a change in the cost of workers without any change in the cost of hours. In the data, such independent changes occur because the BLS method for determining straight-time hourly earnings has some noise in it and because of variations in the fixed element of compensation. Absent this variation, only the ratio of the slopes would be identified.

One of the implicit identifying assumption in Bils's work is that the noise in the calculation of straight-time wages is a genuine measure of the cost of adding a worker to the pay-roll, independent of variations in the cost of adding an hour to a worker's weekly schedule. There is no basis for this assumption. If the BLS approach to straight-time wage measurement is really better than the one based on Bils's formula, then fluctuations in  $W/w$  ought to be incorporated in  $\bar{W}/w$  as well.

Quite apart from the objection that the basic estimating equation is identified only by the modest variability of the fixed component of compensation, there is an even more fundamental objection to Bils's procedure for measuring the slope of the implicit supply of hours. Measurement of the slope of the marginal cost schedule for workers rests on knowledge of the allocational price of added workers. Without comment, Bils assumes that the out-of-pocket current pay-roll cost,  $W$ , is the true effective cost of obtaining an added hour of work through a new hire. But  $W$  includes current overtime and any

other element of compensation that may not be part of the effective cost of labour. There is no more basis for the use of average hourly earnings,  $W$ , as the measure of the effective cost of an hour in the worker dimension than there is for the use of its marginal,  $\bar{W}$ , as the effective cost of an hour in the hour's dimension. The criticisms quoted above from Bils apply with equal force to his measure of the effective cost of an hour obtained by hiring.

Bils is unsuccessful, in my opinion, in his attempt to corroborate his finding of highly cyclical marginal cost of labour by measuring it without assuming that overtime premiums are a true cost. The reliable finding of his work is that marginal cost is cyclical *assuming* that the out-of-pocket costs of overtime are the true costs. Ramey's evidence that firms behave as if marginal cost was flat or downward sloping suggests that the assumption does not hold or that the economies of bunching output are extremely strong. If overtime costs are not true costs, firms are happy to bunch output even though it appears to add to compensation costs through overtime premiums, because the overtime payments go into a bank, from which they can withdraw in the future.

### 3.8 Seasonal Fluctuations in Production and Work Effort

Closely related to the finding that firms behave over the cycle as if they had near-perfect intertemporal substitution in production is the considerable amplitude of seasonal movements in labour input and in production. If labour supply is not itself seasonal, and if the marginal disamenity of work is a rising function of the amount of work, then work schedules should not be seasonal. Because seasons are completely predictable, the test of this proposition is simple. Barsky and Miron (1987) find the following seasonal averages in per cent growth at quarterly rates for the private non-agricultural sector of the US economy:

Table 9.1

	Quarter			
	1	2	3	4
Number of workers employed	-1.65	1.09	0.84	-0.27
Weekly hours	-1.08	0.74	0.95	-0.61
Total	-2.73	1.83	1.79	-0.88

The obvious candidate for a seasonal shift in supply would be the third quarter, when most families take their vacations. But both employment and hours reach their peak levels in the third quarter (July, August, and September). Labour input reaches its trough in the winter. Shifts in demand, related to weather, appear to be the seasonal driving force. If workers had a strong aversion to uneven work schedules, institutions would develop to smooth employment over the seasons. For example, construction workers, who currently work very hard during the spring, summer and autumn, and take the winter off, would develop ways to apply their skills in the winter and would work normal schedules the rest of the year.

#### 4 EVIDENCE ON INTERTEMPORAL SUBSTITUTION IN CONSUMPTION

Although research on consumption in an explicit intertemporal setting has flourished since the publication of *Value and Capital* (see Hall, forthcoming, for a survey), measurement of the elasticity of intertemporal substitution has not been a major focus. My own work (Hall, 1988) will be the primary subject of my discussion. The work is nothing more than a systematic investigation of the proposition developed in Chapter XVIII of *Value and Capital* that a higher real interest rate will cause households to defer consumption through an intertemporal substitution effect. The research measures that deferral in a framework in which income effects are cleanly controlled. Earlier research on the responses of consumption or saving to interest rates failed to control for income effects. *Value and Capital*, through its rigorous application of standard consumer theory to the intertemporal problem, insisted on the importance of distinguishing income and substitution effects in the intertemporal problem.

The essential idea of the research is that consumers plan to change their consumption from one year to the next by an amount that depends on their expectations of real interest rates. Actual movements of consumption differ from planned movements by a completely unpredictable random variable that indexes all the information available next year that was not incorporated in the planning process the year before. If expectations of real interest rates shift, then there should be a corresponding shift in the rate of change of consumption. The magnitude of the response of consumption to a change in real



interest expectations measures the intertemporal elasticity of substitution.

The model of the joint behaviour of consumption and asset returns employed in this work was developed by Breeden (1977, 1979) and Hansen and Singleton (1983). The joint distribution of the log of consumption in period  $t$ ,  $c_t$ , and the return earned by the consumer from period  $t-1$  to period  $t$ ,  $r_{t-1}$ , is normal with a covariance matrix that is unchanging over time. The means of the two variables obey the linear relation,

$$\bar{c}_t = \sigma \bar{r}_{t-1} + c_{t-1} + k \quad (21)$$

That is, the expected change in the log of consumption is the intertemporal elasticity of substitution,  $\sigma$ , times the expected real return plus a constant.

The interpretation of equation (21) has been assisted by recent developments in the theory of intertemporal choice under uncertainty. Epstein and Zin (1987a, 1987b) and others have applied a general framework due to Kreps and Porteus (1978) in order to understand how equation (21) can be derived from explicit preferences and probability distributions and to interpret the coefficient  $\sigma$ . The central issue is the relation between risk aversion and intertemporal substitution. If the consumer orders uncertain future streams of consumption by computing the expected value of an intertemporally separable utility function, then the coefficient  $\sigma$  is both the intertemporal elasticity of substitution and the reciprocal of the coefficient of relative risk aversion. Both substitution and risk aversion are controlled by the curvature of the one-period utility function, which is controlled in turn by  $\sigma$ . In order to make a distinction between substitution and risk aversion, an assumption must be relaxed. One possibility is to drop intertemporal separability. Another is to drop expected utility. In both cases, equation (21) results from utility maximisation and the coefficient  $\sigma$  is unambiguously the intertemporal elasticity of substitution. Risk aversion affects the constant,  $k$ , but not the slope,  $\sigma$ .

If data on the expected real interest rate,  $\bar{r}_t$ , are available directly (from a market for indexed bonds or by subtracting a measure of expected inflation from a nominal interest rate), then the intertemporal elasticity of substitution,  $\sigma$ , is the slope coefficient in the regression,

$$\Delta c_t = \sigma \bar{r}_{t-1} + k + \epsilon_t \quad (22)$$

Estimation is nothing more than checking whether consumption has tended to grow more in periods of high expected returns. On the other hand, if data are available only for the realised real interest rate,  $r_t$ , but there are instrumental variables,  $x_t$ , so that

$$r_t = x_t \beta + \eta_t \quad (23)$$

then an estimate of  $\sigma$  is available by applying instrumental variables estimation to

$$\Delta c_t = \sigma r_{t-1} + k + \epsilon_t + \eta_{t-1} \quad (24)$$

using  $x_{t-1}$  as instruments. In effect, the second procedure calculates a proxy for the expected real interest rate as the fitted values from a regression of the realised real interest rate on the instruments, and then estimates the intertemporal elasticity of substitution by regressing the rate of change of consumption on the proxy.

The first approach yields estimates of the intertemporal elasticity of substitution,  $\sigma$ , using post-war data on consumption changes and on expected inflation. The latter come from Joseph Livingston's survey of inflation forecasters conducted every November. I have obtained separate estimates of  $\sigma$  for the returns on each of three assets: treasury bills, savings accounts, and a diversified portfolio of common stocks (the S&P 500 portfolio). For the first two, the point estimates of the elasticity are about 0.3, but with a fairly large standard error (also about 0.3) because there has been relatively little variability in expected real returns. For common stocks, the estimate is 0.07 with a standard error of 0.05, so the finding of a  $\sigma$  close to zero is quite strong.

The second approach can be applied to a longer period of data because it does not require survey evidence on expected inflation. With annual data on consumption changes and real returns on Treasury bills over the period from 1924 through 1940 and 1950 through 1983, the estimate of  $\sigma$  is  $-0.40$  with a standard error of 0.20 (lagged values of the endogenous variables served as instruments). Evidence against intertemporal substitution is statistically unambiguous. A negative value of  $\sigma$  is inconsistent with standard theory, of course, but the data do not show the tendency for consumers to postpone consumption in times of high real returns that would occur with significant intertemporal substitution.

The time-series evidence on intertemporal substitution in consumption is far from definitive. Similar procedures applied to data on hours of work suggest that intertemporal substitution in labour supply is low, contrary to the conclusion reached in the previous section – see Mankiw, Rotemberg, and Summers (1985). Some of the same objections can be raised in the case of consumption as in labour supply. For example, the expected real interest rate used in my work may be no better a measure of the true allocational intertemporal price ratio than is the measured real wage a measure of the true price ratio of labour to goods. Moreover, panel data on the consumption of individual families shows a large transitory component (see Hall and Mishkin, 1982). Just as high variability in work schedules is evidence in favour of substantial intertemporal substitution in labour supply, high variability in consumption may be evidence in favour of intertemporal substitution in consumption.

## 5 CONCLUSIONS

Both as a matter of theory and empirical findings, the following view of economic fluctuations seems attractive. Intertemporal substitution in labour supply is reasonably high and in product supply is extremely high. The movements in employment and output that occur over the business cycle carry with them relatively small changes in the real interest rate and the real wage, though there may be a good deal of concealed variation in the shadow value of work. On the other hand, intertemporal substitution in consumption is very low; consumption hardly changes over the cycle. When an increase in government purchases of goods and services or other increase in product demand occurs, current production rises. Investment falls, but consumption remains more or less unchanged. Although the simple Keynesian expenditure model describes the process roughly, the richer model of *Value and Capital*, with its emphasis on intertemporal equilibrium, gives a much more satisfactory account.

### Notes

\* This research was supported by the National Science Foundation and is part of the Economic Fluctuations Program of the National Bureau of Economic Research. I am grateful to James Conklin for assistance.

1. References are to the second edition, 1946.
2. On the other hand, in a world of perfect foresight, workers would prefer

to work in a labour market with variable wages and consequently variable weeks, as against a market with a constant wage equal to the mean of the wage in the other market. The workers in the variable market would earn and consume more than those in the stable market, because they would work harder when wages were high.

3. See MaCurdy (1981) and related papers.

### References

- Abowd, John M. and Ashenfelter, Orley (1981) 'Anticipated Unemployment, Temporary Lay-offs, and Compensating Wage Differentials', in Sherwin Rosen (ed.) *Studies in Labor Markets*, (Chicago: University of Chicago Press), pp. 141–70.
- Ando, Albert and Modigliani, Franco (1963) 'The "Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests', *American Economic Review*, vol. 53 (March) pp. 55–84.
- Barro, Robert J. (1981) 'Output Effects of Government Purchases', *Journal of Political Economy*, vol. 89 (December) pp. 1086–121.
- Barsky, Robert B. and Miron, Jeffrey A. (1987) 'The Seasonal Cycle and the Business Cycle', University of Michigan, April.
- Bils, Mark (1987) 'The Cyclical Behavior of Marginal Cost and Price', *American Economic Review*, vol. 77 (December) pp. 838–55.
- Breeden, Douglas T. (1977) 'Changes in Consumption and Investment Opportunities and the Valuation of Securities', Ph.D. dissertation, Stanford University.
- Breeden, Douglas T. (1979) 'An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities', *Journal of Financial Economics*, vol. 7 (September) pp. 265–96.
- Epstein, Larry G., and Zin, Stanley E. (1987a) 'Substitution, Risk Aversion, and the Temporal Behaviour of Consumption and Asset Returns. I: A Theoretical Framework', Working Paper 8715 (Toronto: University of Toronto).
- Epstein, Larry G., and Zin, Stanley E. (1987b) 'Substitution, Risk Aversion and the Temporal Behaviour of Consumption and Asset Returns. II: An Empirical Analysis', Discussion Paper no 698 (Kingston, Ontario: Queens University).
- Friedman, Milton (1957) *A Theory of the Consumption Function* (Princeton: Princeton University Press).
- Hall, Robert E. (1980) 'Employment Fluctuations and Wage Rigidity', *Brooking Papers on Economic Activity*, 1, pp. 91–141.
- Hall, Robert E. (1988) 'Intertemporal Substitution in Consumption', *Journal of Political Economy*, vol. 96 (April) pp. 339–57.
- Hall, Robert E. (forthcoming) 'Consumption', in R. J. Barro (ed.) *Handbook of Modern Business Cycle Theory* (Harvard: Harvard University Press).
- Hall, Robert E., and Mishkin, Frederic S. (1982) 'The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households', *Econometrica*, vol. 50 (March) pp. 461–81.
- Hansen, Lars Peter, and Singleton, Kenneth J. (1983) 'Stochastic Consump-

- tion, Risk Aversion, and the Temporal Behavior of Asset Returns', *Journal of Political Economy*, vol. 91 (April) pp. 249–65.
- Heckman, James J. and Killingsworth, Mark R. (1986) 'Female Labor Supply: A Survey', in Orley Ashenfelter and Richard Layard (eds) *Handbook of Labor Economics*, (Amsterdam: North Holland).
- Hicks, J. R. (1946) *Value and Capital* (Oxford: Oxford University Press) 2nd edn.
- Jorgenson Dale W. (1963) 'Capital Theory and Investment Behavior', *American Economic Review*, vol. 53 (May) pp. 247–59.
- Kreps, David M. and Porteus, Evan L. (1978) 'Temporal Resolution of Uncertainty and Dynamic Choice Theory', *Econometrica*, vol. 46 (January) pp. 185–200.
- Kydland, Finn E. and Prescott, Edward C. (1982) 'Time to Build and Aggregate Fluctuations', *Econometrica*, vol. 50 (November) pp. 1345–70.
- Lucas, Robert J. (1972) 'Expectations and the Neutrality of Money', *Journal of Economic Theory*, vol. 4, pp. 103–24.
- Lucas, Robert J. and Rapping, Leonard A. (1969) 'Real Wages, Unemployment and Inflation', *Journal of Political Economy*, vol. 77 (September/October) pp. 721–54.
- MaCurdy, Thomas (1981) 'An Empirical Model of Labor Supply in a Life Cycle Setting', *Journal of Political Economy*, vol. 89 (December) pp. 1059–85.
- Mankiw, N. Gregory, Rotemberg, Julio J., and Summers, Lawrence H. (1985) 'Intertemporal Substitution in Macroeconomics', *Quarterly Journal of Economics*, vol. 100 (February) pp. 225–56.
- Modigliani, Franco, and Brumberg, R. E. (1954) 'Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data', in K. K. Kurihara (ed.) *Post-Keynesian Economics* (New Brunswick).
- Pencavel, John. (1986) 'Labor Supply of Men: A Survey', in Orley Ashenfelter and Richard Layard (eds) *Handbook of Labor Economics* (Amsterdam: North Holland).
- Prescott, Edward C. (1986) 'Theory ahead of Business Cycle Measurement', in Karl Brunner and Allan Meltzer (eds) *Real Business Cycles, Real Exchange Rates and Actual Policies*, vol. 25 in the *Carnegie-Rochester Conference Series on Public Policy*, (Amsterdam: North Holland) pp. 11–44.
- Ramey, Valerie A. (1988) 'Non-convex Costs and the Behavior of Inventories', Department of Economics, University of California, San Diego, June.
- Rees, Albert O. (1970) 'On Equilibrium in Labor Markets', *Journal of Political Economy*, vol. 78 (March) pp. 306–10.
- Sargent, Thomas J. (1979) *Macroeconomic Theory* (New York: Academic Press).
- Summers, Lawrence H. (1986) 'Some Skeptical Observations on Real Business Cycle Theory', *Quarterly Review*, Federal Reserve Bank of Minneapolis (Fall) pp. 23–7.
- West, Kenneth D. (1986) 'A Variance Bounds Test of the Linear Quadratic Inventory Model', *Journal of Political Economy*, vol. 94 (April) pp. 374–401.