The Tax-Transfer System and Labour Supply

Keane, Michael

2010

Chapter Seven

The Tax–Transfer System and Labour Supply

Michael P Keane*

7.0 Introduction

This chapter provides a survey of the male labour supply literature, while also asking what that literature implies for the design of the tax system. Much of the labour supply literature is concerned with how peoples' decisions about whether and how much to work are influenced by taxes on labour income and government transfers. To begin, it is important to have a clear understanding of *why* these labour supply decisions matter for the design of the tax system. So, by way of motivation, I'll start with a brief and very informal summary of the 'optimal taxation' literature, pioneered by Mirrlees (1971). Loosely speaking, this literature implies that welfare costs of taxation are smaller, and the optimal tax rate on labour income higher, if labour supply is relatively unresponsive to tax rates.

I will then give an overview of the male labour supply literature, which, according to conventional wisdom, generally concludes that labour supply is fairly insensitive to tax rates. This, in turn, implies that the welfare losses from taxation are in fact small. However, I will argue: (i) that the literature is not really so uniform as the conventional wisdom suggests (i.e. quite a few well-executed papers do find that labour supply is responsive to wages/taxes and that welfare costs of taxation are high), and (ii) much of the literature that does find labour supply is unresponsive to after-tax wages is not actually relevant for the setting of tax policy. This is because much of this literature has ignored human capital. I will argue that once one accounts for the effects of income taxation on the incentive to accumulate human capital, one finds evidence that labour supply is much more sensitive to income taxation than previously thought—implying that optimal tax rates are correspondingly lower.

7.1 The Literature on 'Optimal Taxation': Basic Ideas

The optimal tax literature starts with two key problems:

1. The government needs to raise a certain amount of revenue to pay for public goods (such as education, health care and defence forces), unemployment insurance, income support for the poor, and other programs.
2. The use of income taxation to raise this revenue causes people to work less. This leads to a decline in overall economic output (and generates what economists call an efficiency or welfare loss).

There is clearly a tradeoff between the desirable aspects of taxation listed in point one—that is, taxes provide more funds to pay for desirable programs—versus the undesirable effect listed in point two, which is the decline in overall economic activity. A familiar metaphor to describe the problem is the economic 'pie'. We can view government programs such as education, health care and income support as providing people with more of a 'fair go', leading to a more equal division of the economic pie. But as we attempt to split the pie more evenly it tends to shrink. That is, as we raise income taxes, people know their share of the pie is less tied to how much they work, and hence their incentive to work is reduced. So we face a tradeoff between achieving a more even division of the pie versus achieving a larger pie.

The optimal tax literature, pioneered by Mirrlees (1971), develops mathematical models of this tradeoff, and uses them to derive optimal levels of taxation and government spending. The basic conclusion of this literature is that the optimal tax rate depends on the severity of the shrinking pie problem. That, in turn, depends on how much people reduce their labour supply, or work effort, if you tax them. This is what economists call the 'labour supply elasticity'.

Economists define the 'labour supply elasticity' as the percentage reduction in a person's labour supply (i.e. hours of work or effort) if their after-tax wage is reduced by 1 per cent[1]. Labour supply is 'inelastic' if this labour supply elasticity is small. In this case, people won't work very much less if the income tax rate is increased, so the shrinking pie problem is not very serious. Thus, the basic solution of the optimal tax literature is that government should tax people more if their labour supply is 'inelastic'—this is equivalent to the simple statement that it is optimal to tax people more if they won't reduce their work effort much when you do so.

A concrete example will demonstrate what this means. Suppose we want to choose the marginal tax rate for the top income bracket. To simplify matters, let's assume that this bracket is sufficiently high that government (or society) places no value whatsoever on an extra dollar of income for people in this bracket. The government's only goal is to raise as much revenue from people in the highest bracket as possible. In this case, Saez, Slemrod and Giertz (2009) give the following simple formula for the revenue maximising top bracket tax rate:

$$\tau = \frac{1}{1 + a \cdot e} \quad (1)$$

* This research has been supported by Australian Research Council grant FF0561843 and by the AFTS Secretariat of the Australian Treasury. The views expressed in this chapter are entirely my own.

1 Actually, there are several labour supply elasticities, depending on whether one is talking about the response to a change in the after-tax wage that is expected to be only temporary or, alternatively, long-lived, and depending on whether the change is or is not accompanied by changes in non-labour income (i.e. government transfers). These issues will be central to the later discussion.

Here, τ is the tax rate applied to the top bracket. The parameter e is the labour supply elasticity (i.e. the per cent increase in work for a 1 per cent increase in after-tax wage $w(1 - \tau)$, where w is the pre-tax wage). Finally, a is called the 'Pareto' parameter. It is an (inverse) measure of the amount of income dispersion within the top bracket. I'll say more about a below. For now, it is sufficient to know that estimates of a are generally in the range of 1.6 to 2.0 (for a wide range of countries), and there is not much controversy about the value of this parameter.

Brewer, Saez and Shepard (2008) report a value of 1.67 for the United Kingdom, so I'll use that value to illustrate the influence of the labour supply elasticity e on the optimal tax rate:

(2) $$\tau = \frac{1}{1 + (1.67) \cdot e}$$

Now, let's look at what this implies for the optimal tax rate for different values of the labour supply elasticity e:

Table 7.1 Optimal Top Bracket Tax Rates for Different Labour Supply Elasticities

Elasticity (e)	Tax rate (τ) (%)
2.0	23
1.0	37
0.5	54
0.2	75
0.1	86
0.0	100

Note: These rates assume the government places essentially no value on giving extra income to the top earners.

Table 7.1 reveals quite strikingly how sensitive the optimal top bracket tax rate is to the labour supply elasticity. For instance, if the elasticity is only 0.2, which means a 1 per cent reduction in the after-tax wage rate would only reduce labour supply by 0.2 per cent, then the optimal top rate is a very high 75 per cent. In contrast, if the elasticity is 2.0, which means a 1 per cent reduction in the after-tax wage rate would reduce labour supply by a substantial 2 per cent, the optimal top rate is only 23 per cent.

What may appear puzzling for the non-economist about Table 7.1 is why, given the assumptions I've made, the optimal tax rate is not simply 100 per cent? As I indicated earlier, I am assuming that the government (or society) places no value on additional income for people in the top bracket, and its only goal is to raise as much revenue from people in the top bracket as possible. So why not tax them at 100 per cent?

The answer illustrates the shrinking pie problem quite clearly: if income in excess of the level at which the top bracket begins is taxed at 100 per cent then no-one would have any incentive to earn income above that level. As a result, revenue collection on the 100 per cent tax would (in theory) be zero. So even if revenue collection is the only goal, the optimal tax is less than 100 per cent. The one exception, as we see in Table 7.1, is if the labour supply elasticity is zero (i.e. labour supply is totally inelastic). This would arise if, for reasons that are unrelated to income itself, high wage people still choose to earn in excess of the top bracket threshold because, for example, they enjoy the work, or they gauge success by earnings relative to peers (even if it doesn't translate into extra take-home pay).[2]

Now, let's discuss the 'Pareto' parameter a in more detail. The definition of this parameter is $a = z_m/(z_m - z)$, where z is the level of income where the top bracket starts, and z_m is the average income of people in the top bracket. For example, if the top bracket starts at \$500,000, and the average income of people in that bracket is \$1,000,000, then a = 1,000,000/(1,000,000 – 500,000) = 2. In contrast, if average income in the top bracket were \$2,000,000 (implying more dispersion or less equality) we would have a = 2/(2 – 0.5) = 1.33. Thus, we see how a decreases as the degree of dispersion (or inequality) in income increases.[3] Note from equation (1) that as a decreases the optimal tax rate increases (because a decrease in a makes the denominator smaller).

Notice that for a flat rate tax system without brackets, i.e. a single flat rate tax on all income starting at \$0, we would have $z = 0$. Then we would just have $a = z_m/z_m = 1$. If the government's goal is purely revenue maximisation, then equation (1) becomes simply[4]:

(3) $$\tau = \frac{1}{1 + e}$$

Given this formula, Table 7.2 reports the optimal flat rate tax rates for different values of the labour supply elasticity e. Increasing tax rates to levels above those listed in Table 7.2 would actually reduce government revenue, because the shrinking pie problem becomes so severe.

As in Table 7.1, we see that the optimal tax rate increases sharply as e decreases (i.e. as labour supply becomes less responsive to after-tax wages). For instance, if the elasticity is only 0.5, which means a 1 per cent reduction in the after-tax wage rate would only reduce labour supply by 0.5 per cent, then the optimal tax rate is a very high 67 per cent. But if the elasticity is 2.0, which means a 1 per cent reduction in the after-tax wage rate would reduce labour supply by a substantial 2 per cent, the optimal tax rate is only 33 per cent.

Notice that, because a is now smaller (a = 1.0 vs 1.67), the tax rates in Table 7.2 are generally higher than those in Table 7.1. This may seem surprising, given that we are now talking about a flat rate tax, as opposed to a top bracket

2 We can also consider the case where the government (or society) does place some value on extra income for people in the top bracket. Suppose this value is g dollars for each extra dollar of income. g is less than 1 if the society has egalitarian preferences. In that case, and assuming for simplicity that all government revenue is used for redistribution (i.e. there is no minimum tax level needed to provide essential services), Brewer, Saez and Shepard (2008) show that (1) becomes $T = (1 - g)/(1 - g + a \cdot e)$. Thus, we see that for $g > 0$ the tax rates in Table 7.1 would be reduced. Table 7.1, of course, corresponds to $g = 0$.

3 In other words, the thicker the right tail of the income distribution, the smaller is a.

4 It is also easy to derive (3) directly. Just assume that $\ln(h) = e \cdot \ln(w(1 - \tau))$, so e is the labour supply elasticity. Then we have that $h = [w(1 - \tau))]^e$. Let R denote tax revenue. We have $R = (wh)\tau = w[w(1 - \tau))]^e \cdot \tau$. It is instructive to look at the derivative of R with respect to τ, which is $dR/d\tau = w[w(1 - \tau))]^e - ew^2[w(1 - \tau))]^{e-1} \cdot \tau$. This first term, which is *positive*, is the mechanical effect of the tax increase holding labour supply fixed. The second term, which is *negative*, is the loss in revenue due to reduced labour supply. Setting this derivative equal to zero and solving for the revenue maximising τ gives equation (3).

Table 7.2 Revenue Maximising Flat Rate Income Tax Rates for Different Labour Supply Elasticities

Elasticity (e)	Tax rate (τ) (%)
2.0	33
1.0	50
0.5	67
0.2	83
0.1	91
0.0	100

tax. It should be recalled, however, that the models in the optimal tax literature assume that taxes are used largely to finance income inequality-reducing transfers. Under the flat rate scheme in Table 7.2, low-to-middle income taxpayers pay higher taxes, but also receive larger transfers.

It also is worth emphasising that the tax rates in Table 7.2 are revenue maximising rates, not welfare maximising rates. That is, they are only optimal under the extreme assumption that the government places no value on an extra dollar of private income, and seeks only to maximise revenue. This assumption is presumably a better approximation to reality with regard to the top bracket rate (Table 7.1) than in the case of a flat rate (Table 7.2). Thus, the figures in Table 7.2 should not be viewed as plausible estimates of optimal flat rate tax rates given different labour supply elasticities. But they are indicative of the rapid rate of growth of optimal tax rates as the labour supply elasticity falls.

Both Tables 7.1 and 7.2 illustrate the key role of labour supply elasticities in the optimal tax literature, with smaller elasticities implying higher optimal tax rates. As I noted earlier, this is to be expected since smaller elasticties imply that raising taxes is less 'costly', in the sense that it leads to less reduction in work effort and less shrinking of the economic pie. With this background in mind, we will look at what the labour supply literature implies about labour supply elasticities and the welfare costs of taxation.

Section 7.2 describes the standard models of labour supply used by economists. I'll show how these models lead to several alternative definitions of the elasticity of hours of work with respect to the after-tax wage, so that in fact it is not correct to talk about *the* labour supply elasticity as if there were only one. Then, section 7.3 provides the survey of the male labour supply literature. It discusses estimates of the various labour supply elasticities, and what they imply about the costs of taxation.

Section 7.3 is divided into four parts. Section 7.3.1 discusses the main econometric problems that arise in attempting to estimate labour supply models. The next three sections cover results from three main classes of labour supply model. Section 7.3.2 covers 'static' models that consider only the choice of work hours but take assets and human capital as given. Section 7.3.3 covers 'life-cycle' models that incorporate decisions about saving. Section 7.3.4 covers life-cycle models that also account for how wages depend on work experience (i.e. human capital).

Sections 7.2 and 7.3 are at times somewhat technical (as they present mathematical models), so it is worth summarising in advance what is discussed. Essentially, it is fair to say that, regardless of which of the various definitions of the labour supply elasticity you use, the consensus of the economics profession—whether accurate or not—has been that labour supply elasticities are quite small (i.e. less than 0.50). This implies, for instance, that the optimal top-bracket tax rate is towards the high end of the figures given in Table 7.1.

The consensus is summed up nicely in a recent survey by Saez, Slemrod and Giertz (2009), who state:

> …optimal progressivity of the tax–transfer system, as well as the optimal size of the public sector, depend (inversely) on the…elasticity of labour supply… With some exceptions, the profession has settled on a value for this elasticity close to zero… In models with only a labour-leisure choice, this implies that the efficiency cost of taxing labour income…is bound to be low as well.

However, I believe that section 7.3 presents evidence that challenges this consensus. First, I show that many well-executed papers in the literature have produced reasonably large estimates of labour supply elasticities, as well as the welfare costs of taxation. The extent of agreement among existing studies is not nearly so great as the conventional wisdom would suggest. Second, and perhaps more importantly, I'll argue that a serious problem with the existing labour supply literature is that it is based almost entirely on models that ignore human capital. Section 7.3.4 shows how, in a model with human capital, conventional econometric methods tend to seriously understate labour supply elasticities. Hence, inclusion of human capital into standard labour supply models leads to a conclusion that labour supply elasticities may well be higher than the conventional wisdom would indicate.

7.2 Basic Models of Labour Supply

Before discussing the empirical literature on labour supply, it is necessary to lay out the theoretical framework on which it is based. Labour supply models can be broadly classified into two main types, static and dynamic. There are many variations within each type, but for our purposes this simple division will prove useful.

7.2.1 The Basic Static Labour Supply Model

In the basic *static* labour supply model, a person's utility in period t depends positively on consumption and negatively on the hours of work needed to attain that consumption.[5] One commonly used utility function has the form:

$$(1) \quad U_t = \frac{C_t^{1+\eta}}{1+\eta} - \beta_t \frac{h_t^{1+\gamma}}{1+\gamma} \qquad \eta \le 0,\ \gamma \ge 0$$

Here U_t is utility in period t. It depends on consumption C_t and hours of labour supplied h_t. To keeps things simple, I assume that consumption is simply equal to labour earnings, so that $C_t = w_t(1-\tau)h_t$ where w_t is the pre-tax wage rate and τ is the tax rate. η and γ are parameters that describe

5 The definition of a 'period' in labour supply models is somewhat arbitrary. In empirical work it is often chosen to be a year, although shorter periods are sometimes examined.

preferences. As $\eta < 0$ consumption is raised to a power less than one, so we have diminishing marginal utility of consumption.[6] And $\gamma > 0$ means hours are raised to a power greater than one, so that people find an additional hour of work more painful as the level of hours increases. Both are very standard economic assumptions. The parameter β_t captures the person's tastes for leisure versus consumption, and this may change over time.

The static model has two key features that distinguish it from dynamic models. First, it assumes that workers do not borrow or save, so that current period consumption is simply equal to current after-tax income. Second, it ignores human capital accumulation. This means that workers decide how much labour to supply today based only on today's wage rate. They do not consider the possibility that working more today may have the effect of raising future wages (because by working more today one acquires more work experience).

To solve this model for optimal hours of work, use the budget constraint $C_t = w_t(1 - \tau)h_t$ to substitute for consumption C_t in equation (1), obtaining:

$$(2) \qquad U_t = \frac{[w_t(1-\tau)h_t]^{1+\eta}}{1+\eta} - \beta_t \frac{h_t^{1+\gamma}}{1+\gamma} \qquad \eta \le 0,\ \gamma \ge 0$$

We have now expressed utility as a function of hours, and we can solve for the level of hours that maximises utility. To do this, we simply differentiate (2) with respect to h_t, set the derivative to zero to maximise U_t, and then solve for the optimal h_t. Doing this we obtain:

$$(3) \qquad \frac{dU_t}{dh_t} = [w_t(1-\tau)h_t]^{\eta}\, w_t(1-\tau) - \beta_t h_t^{\gamma} = 0$$

This can be reorganised into the more familiar form:

$$(4) \qquad \frac{MUL(h)}{MUC(h)} = \frac{\beta_t h_t^{\gamma}}{[w_t(1-\tau)h_t]^{\eta}} = w_t(1-\tau)$$

This is one of the most basic equations in economics. The left-hand side is the ratio of the marginal utility of leisure, $\beta_t h_t^{\gamma}$, (which is simply the negative of the marginal disutility of work hours) to the marginal utility of consumption, $[w_t(1 - \tau)h_t]^{\eta}$. Utility is maximised by choosing hours of work so as to set this ratio equal to the after-tax wage rate, $w_t(1 - \tau)$.

Notice that as hours increase, income increases, and hence consumption increases. Thus, the marginal utility of consumption ($MUC(h)$) falls, given the assumption of diminishing marginal utility of consumption ($\eta < 0$). And as hours increase, the marginal utility of leisure $MUL(h)$ increases. Thus, as hours increase, the ratio on the left side of (4) gets smaller.[7] Hours increase up to that point where the left and right sides of (4) are equalised.

Solving for h_t we obtain:

$$(5) \qquad h_t^{\gamma-\eta} = \frac{[w_t(1-\tau)]^{1+\eta}}{\beta_t}$$

This equation is easier to work with if we take logs, giving:

$$(6) \qquad \ln h_t = \frac{1+\eta}{\gamma-\eta}\ln[w_t(1-\tau)] - \frac{1}{\gamma-\eta}\ln\beta_t$$

As I indicated earlier, the labour supply elasticity is simply the percentage reduction in labour supply (i.e. hours of work) with respect to a 1 per cent change in the after-tax wage. Formally, this is defined as:

$$(7) \qquad e = \frac{w(1-\tau)}{h}\frac{\partial h}{\partial w(1-\tau)} = \frac{\partial \ln h}{\partial \ln w(1-\tau)}$$

Thus, the labour supply elasticity e is obtained by taking the derivative of the log of hours with respect to the log of the after-tax wage. Given the form of equation (6), this derivative is simple to calculate:

$$(8) \qquad e = \frac{\partial \ln h_t}{\partial \ln w_t(1-\tau)} = \frac{1+\eta}{\gamma-\eta}$$

This quantity is called the 'Marshallian' labour supply elasticity (after the great economist Alfred Marshall), and is sometimes also called the 'uncompensated' or 'total' elasticity. It is certainly the simplest labour supply elasticity concept. Recall that standard economic assumptions of diminishing marginal utility of consumption and leisure imply that $\eta < 0$ and $\gamma > 0$. Thus, we know that the denominator in (8), which is $(\gamma - \eta)$, is positive.

But apart from this result, economic theory tells us little. Obviously, the magnitude of the Marshallian elasticity depends on the utility function parameters γ and η. I'll discuss plausible values for these parameters in the literature review. For now let us just note that it is possible for the numerator $1 + \eta$ to be negative if $\eta < -1$. In that case, an increase in the wage would actually reduce hours of work. Several of the empirical studies that I review below do find this. But most studies find that $1 + \eta > 0$. In that case the Marshallian elasticity $e = (1 + \eta)/(\gamma - \eta)$ is positive, meaning that an increase in the after-tax wage increases hours of work.

Conceptually, an increase in the wage rate can be thought of as having *two* effects. First, given a higher wage, a person can now earn more income just by maintaining his/her original level of hours. Given diminishing marginal utility of consumption, this creates an incentive to reduce hours of work and take more leisure time. This negative effect of a wage increase on desired hours of work is called the 'income effect'.

Second, given a higher wage, the rate at which a person can increase his/her income by working more hours increases. This gives the person an incentive to work more hours, or, in economists' terminology, to 'substitute' work for leisure. This positive effect of a wage increase on desired hours of work is called the 'substitution' effect.

Knowledge of both income and substitution effects is important for understanding the impact of changes in tax

6 That is, the utility to a person of the first dollar of consumption is less than that of the millionth dollar.

7 The ratio $MUL(h)/MUC(h)$ is known as the marginal rate of substitution (MRS) between consumption and leisure. Evaluating this at $h = 0$ one obtains the 'reservation wage', the minimum wage at which a person is willing to supply positive hours. Notice that MRS = $(\beta_t/[wt(1-\tau)]^{\eta})h_t^{\gamma-n}$. Since $(\gamma - \eta)$ is positive, we see that the MRS equals zero if $h = 0$. Thus, the reservation wage is zero and people in this model will work positive hours for any positive wage. The model can be easily modified to account for people who choose not to work by including some non-labour income (e.g. government transfers) so that consumption does not fall to zero when hours equal zero. But this extension is not critical for the points I wish to make in this section.

and transfer policy. For example, suppose we have a flat rate tax system. Further suppose that we decide to increase the flat rate tax rate and use the revenue to finance grants to every member of the population (perhaps with the goal of guaranteeing a minimum income). Economists refer to such grants that do not depend on income or hours of work as 'lump sum payments'. This policy discourages work in two ways. The tax increase itself reduces the reward from work, but the lump sum payments, which have the effect of increasing the income attained by working any given level of hours, also discourage work via the income effect.

In contrast, suppose the revenue from the increased income tax is used to finance public goods (e.g. schools, public transport, carbon capture). In that case, the negative effect on labour supply will be less because the income effect that comes from transferring the tax revenue directly back to the population in the form of grants is avoided.

Eugene Slutsky developed a method for decomposing the Marshallian labour elasticity into the separate substitution and income effects. This is known as the 'Slutsky equation':

$$\text{(9)} \quad \frac{\partial h}{\partial w} = \left.\frac{\partial h}{\partial w}\right|_u + h\frac{\partial h}{\partial N}$$

Here N represents non-labour income. In the previous example, non-labour income comes from the grants or lump sum payments that the government makes to members of the population.

In equation (9) the first term on the right-hand side is the substitution effect, while the second term is the income effect. The second term can be understood as follows. First, suppose a person is working h hours, and their wage increases by a dollar. If they do not change their hours, then their income will go up by h dollars. The idea behind the Slutsky equation is that this is like giving the person a grant (or lump sump payment) of h dollars. $\partial h/\partial N$ stands for the effect on hours of an extra dollar of grant income. Recall that this must be negative given diminishing marginal utility of consumption. Thus, the second term, $h \cdot \partial h/\partial N$ tells us the overall reduction in hours that occurs because the person has, in effect, been given h extra dollars of grant income.

The first term on the right, the substitution effect, is more subtle. The idea here is roughly the following: we can think of giving a person a wage increase and simultaneously taking away the same h dollars that we gave them above—perhaps through a poll or head tax. This means that if the person sticks with their original hours level, their net income won't change. Obviously their leisure is unchanged as well, so their overall utility level is unchanged. The person will have to increase hours in order to take advantage of the higher wage rate and raise consumption. Thus, we see that this 'compensated substitution effect' of a wage increase—that is, raising the wage while simultaneously 'compensating' by taking away enough income (through a poll or head tax) so that the person can't be better off by 'standing pat'—must be positive. The notation $\left.\frac{\partial h}{\partial w}\right|_u$ stands for this operation: it is the effect on hours of raising the wage by one unit while taking away h units of non-labour income so as to 'compensate' for the wage increase and hold utility fixed.

Another way to think about (9) is that we hypothetically give a person a wage increase in two steps. First, we give them the wage increase but simultaneously apply a poll or head tax to counteract it, so the person is not made better off. At their original hours level the person's net income and consumption will be unchanged, but their marginal wage rate is higher. Hence, according to theory, the person must choose to increase hours. In the second stage we remove the head tax. This increases the person's income level at any given level of hours, so, according to theory, the person should reduce hours.

It is convenient to write the Slutsky equation in elasticity form, so that the Marshallian elasticity appears on the left-hand side. To do this we just pre-multiply equation (9) by w/h, and multiply and divide the income effect term by N, to obtain:

$$\text{(10)} \quad \frac{w}{h}\frac{\partial h}{\partial w} = \frac{w}{h}\left.\frac{\partial h}{\partial w}\right|_u + \frac{wh}{N}\left[\frac{N}{h}\frac{\partial h}{\partial N}\right]$$

The first term on the right is called the 'compensated' or 'Hicks' labour supply elasticity (after the famous economist John Hicks). The second term is the income effect, which includes the elasticity of hours of work with respect to non-labour income, $\frac{N}{h}\frac{\partial h}{\partial N}$.

Now we see why the Marshallian elasticity is sometimes called the 'total' elasticity, as it is the sum of the Hicks elasticity and the income effect. We also see why the Marshallian elasticity is sometimes called the 'uncompensated' elasticity; in contrast to the Hicks elasticity, it is simply the total effect of a wage increase, without any compensating head tax.

It should now be obvious why the Hicks elasticity is of practical importance for tax policy. For example, an after-tax wage increase induced by an income tax cut may in some cases be financed via reduced transfer payments. Depending on the size of the tax cut versus the cut in transfers, the Hicks elasticity may well be the relevant one for predicting the overall effect of the policy change on labour supply.

Another point is that given a progressive tax system (i.e. a system with brackets such that marginal tax rates increase with income) it can be shown that the effect of a change in upper bracket tax rates on the labour supply of upper income workers depends mostly on the Hicks elasticity, not the Marshallian elasticity. Thus, the extent to which a highly progressive tax system generates a welfare cost by shrinking the economic pie is largely a function of the Hicks elasticity. I'll discuss this key point in more detail later.

In most empirical applications, the Hicks, elasticity is 'backed out' by estimating the Marshallian elasticity and income elasticity and applying equation (10). But some applications estimate the parameters of preferences (γ and η) directly, and then construct the elasticities using theoretical formulas. To obtain the Hicks and income elasticities we need to modify the budget constraint of our static model to include non-labour income, giving $C_t = w_t(1-\tau)h_t + N_t$. Equation (2) then becomes:

$$U_t = \frac{[w_t(1-\tau)h_t + N_t]^{1+\eta}}{1+\eta} - \beta_t\frac{h_t^{1+\gamma}}{1+\gamma} \qquad \eta \le 0, \ \gamma \ge 0$$

The mathematics is a bit cumbersome, but it can be shown that, in this simple model, the income elasticity of labour supply, evaluated at small values of N_t, is approximately:

(11) $$e_I = \frac{N_t}{h_t}\frac{dh_t}{dN_t} = \frac{N_t}{w_t h_t(1-\tau)}\frac{\eta}{\gamma-\eta}$$

which means the income effect in equation (10), which I'll denote '*ie*', is just:

(12) $$ie = \frac{w_t h_t(1-\tau)}{N_t} e_I = \frac{\eta}{\gamma-\eta} < 0$$

Of course, the income effect is negative because $\eta < 0$ and $\gamma > 0$, which are conditions required for diminishing marginal utility of consumption and leisure. Intuitively, the magnitude of the negative income effect is increasing in the magnitude of the parameter η. If η is a larger negative number it implies that the incremental utility from extra consumption diminishes more quickly as consumption increases. Thus, the tendency to reduce labour supply in response to an increase in non-labour income is greater.

It is instructive to note that the income effect *ie* in (10) can be written as:

(13) $$ie = \frac{wh}{N}\left[\frac{N}{h}\frac{\partial h}{\partial N}\right] = w\frac{\partial h}{\partial N} = \frac{\partial (wh)}{\partial N}$$

Thus, the income effect is also the effect of an increase of non-labour income on labour income (i.e. given an extra dollar of non-labour income, how much does a worker reduce his/her earnings?). As Pencavel (1986) notes, if both leisure and the composite consumption good (C_t) are normal goods, then *ie* must be between zero and –1. Indeed, we can see from (12) that as η runs off towards negative infinity, *ie* runs off towards –1. But Pencavel (1986) argues that values of *ie* near –1 would be quite implausible. Simple introspection suggests that people would be unlikely to react to an increase in non-labour income by reducing hours so sharply that total consumption $C_t = w_t h_t + N_t$ does not increase.[8]

Now, using the Slutsky equation we can obtain the Hicks elasticity as the difference between the Marshallian elasticity and the income effect:

(14) $$e_H = e - ie = \frac{1}{\gamma-\eta}$$

Notice that because $\eta < 0$, the Hicks elasticity in (14) must be greater than the Marshallian elasticity in (8). The two approach each other as $\eta \to 0$, in which case there are no income effects. Much of the literature on optimal taxation makes the assumption of no income effects in order to simplify the analysis (e.g. see Diamond 1998). However, in my view the assumption that income effects can be ignored is questionable, for reasons I discuss later.

7.2.2 The Basic Dynamic Model with Savings

Consider next the basic *dynamic* labour supply model, also known as the 'life-cycle' model. The pioneering work by MaCurdy (1981) introduced dynamics in empirical labour supply models by introducing savings. In his model, workers are free to borrow and lend across periods (rather than being constrained to consume their earnings in each period).

MaCurdy (1981) considered a multi-period model, but in order to emphasise the key points it is sufficient to have two periods in the working life.[9] As before, the per-period utility function is given by:

(15) $$U_t = \frac{C_t^{1+\eta}}{1+\eta} - \beta_t \frac{h_t^{1+\gamma}}{1+\gamma} \qquad t = 1,2 \qquad \eta \le 0,\ \gamma \ge 0$$

where C_t is consumption in period t and h_t is hours of labour supplied in period t.

The key change in the dynamic model is that now we have $C_1 = w_1(1-\tau_1)h_1 + b$, where b is the net borrowing in period 1, while $C_2 = w_2(1-\tau_2)h_2 - b(1+r)$, where $b(1+r)$ is the net repayment of the loan in period 2. The amount that must be repaid is $b(1+r)$ where r is the interest rate. Of course, b can be negative, meaning the person saves in period 1. Note that w_1 and w_2 are wage rates in periods 1 and 2, while τ_1 and τ_2 are tax rates on labour earnings in periods 1 and 2, respectively.[10]

In the dynamic model, a person makes decisions so as to maximise his/her lifetime utility over the two periods. The present value of lifetime utility is given by:

(16) $$V = U_1 + \rho U_2$$

where the parameter ρ is the discount factor. Substituting the values of period $t = 1$ and $t = 2$ utility into (16) we obtain:

(17) $$V = \frac{[w_1 h_1(1-\tau_1)+b]^{1+\eta}}{1+\eta} - \beta_1\frac{h_1^{1+\gamma}}{1+\gamma} + \rho\left\{\frac{[w_2 h_2(1-\tau_2) - b(1+r)]^{1+\eta}}{1+\eta} - \beta_2\frac{h_2^{1+\gamma}}{1+\gamma}\right\}$$

In the standard life-cycle model, there is no human capital accumulation via returns to work experience. That is, hours of work in period 1 do not affect the wage rate in period 2. Thus, the worker treats the wage path $\{w_1, w_2\}$ as exogenously given (i.e. it is unaffected by the worker's own decisions).

In the life-cycle model, a new labour supply elasticity concept is introduced. This is the response of a worker to a temporary change in the after-tax wage rate. For instance, this could be induced by a temporary tax cut in period 1 that is rescinded in period 2. Since the worker can now save, the response to such a tax change may be to work more in period 1, save part of the extra earnings, and then work less in period 2. Economists call such a reaction (i.e. shifting one's labour supply toward periods where wages are relatively high) 'inter-temporal substitution'. The magnitude of this response is called the 'inter-temporal elasticity of substitution'. It is also sometimes called the 'Frisch' elasticity, after the economist Ragnar Frisch.

The first order conditions for the worker's optimisation problem are simply:

8 And, as I have already noted, even $\eta < -1$ implies that income effects dominate substitution effects, so that an increase in the wage reduces labour supply.

9 He also considered that the change in a person's wage rate from one period to the next might be in part unexpected, but to keep things simple I put aside uncertainty about future wages for now.

10 As in the static model I assume there is no non-labour income. This simplifies the analysis while not changing any key results.

(18) $$\frac{\partial V}{\partial h_1} = \left[w_1 h_1(1-\tau_1)+b\right]^{\eta} w_1(1-\tau_1) - \beta_1 h_1^{\gamma} = 0$$

(19) $$\frac{\partial V}{\partial h_2} = \left[w_2 h_2(1-\tau_2)-b(1+r)\right]^{\eta} w_2(1-\tau_2) - \beta_2 h_2^{\gamma} = 0$$

(20) $$\frac{\partial V}{\partial b} = \left[w_1 h_1(1-\tau_1)+b\right]^{\eta} - \rho\left[w_2 h_2(1-\tau_2) - b(1+r)\right]^{\eta}(1+r) = 0$$

Equation (20) can be simplified to read $[C_1]^{\eta}/[C_2]^{\eta} = \rho(1+r)$, which is the classic inter-temporal optimality condition that requires one to set the borrowing level b so as to equate the ratio of the marginal utilities of consumption in the two periods to $\rho(1+r)$.

An important special case is when $\rho = 1/(1+r)$, so that people discount the future using the real interest rate. In that case, we have $\rho(1+r)=1$, so that $[C_1]^{\eta}/[C_2]^{\eta} = 1$ and hence $C_1 = C_2$. So, we have complete consumption smoothing, that is, the consumer desires to have equal consumption in both periods.

Utilising the inter-temporal condition, we can divide (19) by (18) to obtain:

(21) $$\left(\frac{h_2}{h_1}\right)^{\gamma} = \frac{w_2(1-\tau_2)}{w_1(1-\tau_1)}\frac{1}{\rho(1+r)}\frac{\beta_1}{\beta_2}$$

And taking logs we obtain:

(22) $$\ln\left(\frac{h_2}{h_1}\right) = \frac{1}{\gamma}\left\{\ln\frac{w_2}{w_1} + \ln\frac{(1-\tau_2)}{(1-\tau_1)} - \ln\rho(1+r) - \ln\frac{\beta_2}{\beta_1}\right\}$$

From (22) we obtain:

(23) $$\frac{\partial\ln(h_2/h_1)}{\partial\ln(w_2/w_1)} = \frac{1}{\gamma}$$

Thus, the Frisch elasticity of substitution, the rate at which a worker shifts hours of work from period 1 to period 2 as the relative wage increases in period 2, is simply $1/\gamma$. The elasticity with respect to a change in the tax ratio $(1-\tau_2)/(1-\tau_1)$ is identical.

There is an important relationship between the Frisch, Hicks and Marshallian elasticities:

(24) $$\frac{1}{\gamma} > \frac{1}{\gamma-\eta} > \frac{1+\eta}{\gamma-\eta}$$

That is, the Frisch elasticity is larger than the Hicks, which is larger than the Marshallian. This follows directly from $\eta < 0$ (i.e. diminishing marginal utility of consumption). This implies that if we can obtain an estimate of the Frisch elasticity we have an upper bound on how large the Hicks and Marshallian elasticities might be. With these concepts in hand, we are in a position to talk about estimation of labour supply elasticities.

7.3 A Survey of Labour Supply Elasticity Estimates

There have been many surveys of the labour supply literature and of labour supply elasticity estimates. These include Hausman (1985b), Pencavel (1986), Killingsworth and Heckman (1986), Blundell and MaCurdy (1999) and Meghir and Phillips (2008). Here I will start by summarising the main econometric problems this literature faces, and then move on to describe the main empirical results on male labour supply elasticities.

7.3.1 Econometric Issues

Broadly speaking, there are two main approaches to estimating labour supply elasticties in the literature. One is simply to run a regression of hours of work on the wage rate and non-labour income. An alternative is to specify and estimate a structural model of labour supply behaviour, which would include specifying utility and wage functions. I'll begin by discussing a regression approach.

Various functional forms could be chosen for an hours regression but, as a starting point, let's consider a logarithmic specification of the form:

(25) $$\ln h_{it} = \beta + e\ln w_{it}(1-\tau) + \beta_I N_{it}(1-\tau) + \varepsilon_{it}$$

where I now include person subscripts i to indicate that we have data on a sample of people. Thus h_{it} is hours of work for person i in period t. Similarly w_{it} is the wage rate faced by person i at time t, and N_{it} is their non-labour income.

It is important that equation (25) controls for non-labour income, N_{it}. As a result, the coefficient on the log after-tax wage rate (e) is the effect of a wage change holding non-labour income fixed. Thus it is interpretable as the Marshallian elasticity (i.e. when the wage changes there is no compensating change in non-labour income).[11] The coefficient on the non-labour income variable ($\beta_I = \partial h_{it}/\partial N_{it}$) can be multiplied by the after-tax wage rate to obtain the income effect $ie = w_{it}(1-\tau)\beta_I$. Of course, given estimates of (25), the Hicks elasticity can be backed out using the Slutsky equation as $e_H = e - w_{it}(1-\tau)\beta_I$.

In section 7.2, I considered models of the labour supply of a single individual, so it was not necessary to consider heterogeneity in tastes for work. In (25), the error term ε_{it} captures the notion that different people may have different tastes for work. That is, facing the same wage and non-labour income, some people may choose to work more than others.

It is also important to note that equation (25) does not follow directly from the utility function specification I gave in (1). I adopt the functional form in (25) because it is simple to interpret. One should be aware that many alternative specifications for the labour supply function have been estimated in the literature, and there is no consensus on the 'right' functional form. But (25) will suffice for explaining the main issues/problems that arise in attempting to estimate labour supply elasticities.

Indeed, there are a multitude of econometric problems that arise in attempting to estimate labour supply elasticities, so I will just highlight some of the most important.

Problem One

The first main problem is that there is no reason to think that the tastes for work captured by ε_{it} would be uncorrelated with either the wage rate w_{it} or the level of non-labour income N_{it}. For example, people who are

11 Blundell and MaCurdy (1999) provide an extensive discussion about how different sets of controls lead to different interpretations of the wage coefficient.

relatively hard working (or, in other words, have a relatively low taste for leisure) might also work harder and be more productive when they do work. Thus, ε_{it} could be positively correlated with the wage rate. Furthermore, those who are relatively hard working might also tend to save more, leading to relatively high asset income. This would create a positive correlation between ε_{it} and non-labour income. Either of these problems would violate standard 'exogeneity' assumptions on the error term used to justify OLS regression. Econometricians refer to such problems as 'endogeneity' problems.

These problems are not merely academic. Pencavel (1986, p. 23) reports a simple OLS regression of annual male hours of work on wage rates, various types of non-labour income, and a long list of demographic controls (e.g. education, age, marital status, children, race, health, region) using data from the 1980 US census. He finds that the coefficient on asset income is actually positive, implying that \$10,000 in additional non-labour income would *increase* annual hours by 46 hours. This contradicts the assumption that income effects should be negative.[12] He also finds that the coefficient on the wage rate is negative, implying that a dollar per hour wage increase would reduce annual hours by 14. As noted earlier, a negative Marshallian elasticity is theoretically possible (i.e. 'backward bending labour supply'), but only given a strong negative income effect. Thus, taken at face value, the sign pattern found here would seem to completely contradict economic theory. But it is quite likely the result of endogeneity (or other econometric problems I'll list later).

One approach to deal with such endogeneity problems is to adopt a fixed effects specification, where the error term is decomposed as:

(26) $$\varepsilon_{it} = \mu_i + \eta_{it}$$

Here μ_i is the individual fixed effect, which captures person i's taste for work (assumed to be time invariant), while η_{it} is a purely idiosyncratic shock to tastes for work (e.g. person i may have been sick in a particular period). In the fixed effects approach, it is assumed that the fixed effect μ_i may be correlated with wages and non-labour income, but that the idiosyncratic shocks η_{it} are not. Methods such as first differencing or de-meaning the data can be used to eliminate μ_i from the error term. Then, the η_{it} that remain are assumed to satisfy the conditions required for OLS regression.[13] In addition, labour supply studies typically also include various observable control variables that might shift tastes for work, such as age, number and ages of children, marital status, and so on.

A second approach is to use an instrument variables approach. An 'instrument' is a variable that is correlated with the variable of interest—in this case wages and non-labour income—but that is uncorrelated with the regression error term ε_{it}. For example, changes in the price of iron ore or bauxite might shift wage rates in Australia, but changes in these prices are presumably uncorrelated with changes in tastes for work. Thus, mineral prices would be sensible instruments to use for wage rates.

In an instrumental variable (IV) regression, one exploits only the variation in the variable of interest induced by the instrument to calculate the effect of that variable on the dependent variable. For instance, one might use only variation in wages induced by changes in mineral prices to calculate the effect of wage rates on hours of work. In most contexts, the choice of whether instruments are valid is quite controversial. We'll see some examples of this in the discussion of particular papers below.

Problem Two

The second main problem involved in estimation of (25) is that real world tax schedules are typically not the sort of flat rate schedules I assumed in the theoretical discussion of section 7.2. The typical schedule in OECD countries involves transfers to low income individuals, a rate at which these transfers are taxed away as income increases, and then a set of brackets, with progressively higher rates in higher income brackets. We can summarise this by saying the tax rate τ_i that a person faces, as well as their non-labour income N_{it}, are actually functions of their wage rate and hours of work. I'll denote these functions as $\tau_i(w_{it}, h_{it})$ and $N_{it}(w_{it}, h_{it})$. Then (25) becomes:

(27) $$\ln h_{it} = \beta + e \ln w_{it}(1 - \tau_i(w_{it}, h_{it})) + \beta_I N_{it}(w_{it}, h_{it}) + \varepsilon_{it}$$

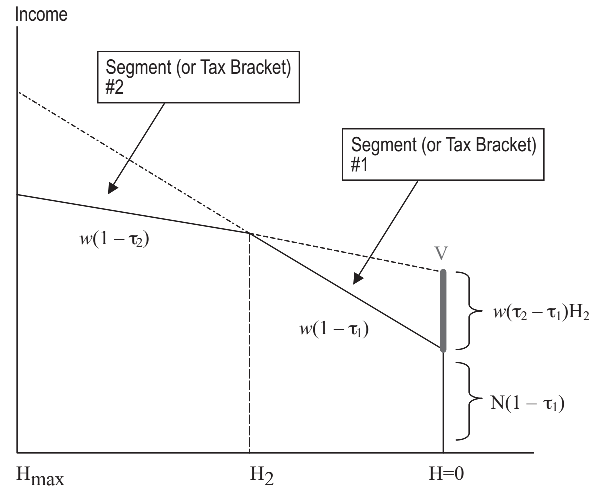
This creates a blatant endogeneity problem, as the after-tax wage rate and non-labour income depend directly on hours, which is the dependent variable. For example, as noted earlier, a person who is a hard worker—that is, has a high value of ε_{it}—will work more hours for a given wage and non-labour income. With a progressive tax system, this may drive such a person into a higher bracket and/or lower their level of transfers. Hence, the progressivity of the tax system creates a negative correlation between the error term ε_{it} and both the after-tax wage and non-labour income. Again, OLS assumptions are violated.

An additional problem created by transfers and progressive taxation is that tax rates and transfer amounts do not usually vary smoothly with income. Rather, they tend to take discrete jumps at certain income levels. An example is given in Figure 7.1, which shows the sort of budget constraint created by simple tax system with two brackets. In bracket #1, the tax rate is τ_1, while in bracket #2 the tax rate jumps to τ_2. The person represented by the graph moves into the upper bracket if he/she works more than H_2 hours, at which point his/her income exceeds the cut-off level which is assumed to be $wH_2 + N$. Notice that at this income level the slope of the budget constraint suddenly

12 A positive income effect for hours, implying a negative income effect for leisure, would mean that leisure is not a normal good. That is, it is not a good that people demand more of as they become wealthier. While not theoretically impossible this seems highly unintuitive.

13 A limitation of the fixed effects approach, which is rather technical, is that the η_{it} must be 'strictly exogenous' as opposed to merely exogenous. This means the η_{it} must be uncorrelated with all leads and lags of wages and non-labour income, not just the contemporary values. Strict exogeneity is actually a much stronger assumption. It implies, for example, that an adverse health shock that lowers one's taste for work today cannot affect one's wage in the next period. Yet, one could easily imagine that it would (e.g. if working less in the current period causes one's human capital to depreciate). Keane and Runkle (1992) provide an extensive discussion of this issue.

Figure 7.1 The Piecewise Linear Budget Constraint Created by Progressive Taxation



drops from $w(1-\tau_1)$ to $w(1-\tau_2)$. This is what is known as 'kink' in the budget constraint. At that point the constraint does not have a well-defined slope. Note that whole labour supply theory discussed in section 7.2 was based on the idea that hours are determined by setting the marginal rate of substitution between consumption and leisure equal to the after-tax wage rate, which is the slope of the budget line. This approach breaks down if the budget constraint contains kinks.

There have been three main approaches to these problems in the literature. The 'structural approach' in which one models in detail how people make labour supply decisions when facing a non-linear tax schedule, is described in the pioneering papers by Burtless and Hausman (1978), Wales and Woodland (1979), Hausman (1980, 1981), Blomquist (1983) and Moffitt (1983). A second idea is to approximate the non-linear budget constraint by a smooth (i.e. kink free or differentiable) polynomial function, as suggested by MaCurdy, Green and Paarsch (1990). Suppose that tax rate is a differentiable function of earnings, which I'll denote by $\tau(w_t h_t)$. Then equations (2)–(4) just become:

(2') $$U_t = \frac{[w_t h_t - \tau(w_t h_t)]^{1+\eta}}{1+\eta} - \beta_t \frac{h_t^{1+\gamma}}{1+\gamma} \qquad \eta \le 0,\ \gamma \ge 0$$

(3') $$\frac{dU_t}{dh_t} = [w_t h_t - \tau(w_t h_t)]^{\eta}[w_t\left(1-\tau'(w_t h_t)\right)] - \beta_t h_t^{\gamma} = 0$$

(4') $$\frac{MUL(h)}{MUC(h)} = \frac{\beta_t h_t^{\gamma}}{[w_t(1-\tau)h_t]^{\eta}} = w_t\left(1-\tau'(w_t h_t)\right)$$

Comparing (4) and (4'), we see that the constant tax rate τ in (4) is simply replaced by $\tau'(w_t h_t)$, the derivative of the tax function evaluated at earnings level $w_t h_t$ (or, in other words, the tax on a marginal dollar of earnings).

An older approach, dating back to Hall (1973), is to model each person as if they choose labour supply subject to a simple *hypothetical* linear budget constraint created by taking the segment (or bracket) on which they are observed to locate, and extending this segment from $h = 0$ to $h = \mathrm{H}_{max}$. In Figure 7.1, these extensions of segments 1 and 2 are indicated by the dotted lines. As noted by Hall (1973), as long as preferences are strictly convex (which is implied by diminishing marginal returns to consumption and leisure) a person facing such a hypothetical budget constraint would make the same choice as a person facing the actual budget constraint.[14] It is common in applying this method to instrument for wages and non-labour income to deal with measurement error.

Problem Three

The third main problem, which was emphasised by Pencavel (1986, p. 59), is that in estimating an equation like (25) we can't be sure if we are estimating a labour supply curve or a labour demand curve, or just some combination of the two. The key question here is *why* wages and non-labour income vary across people. (Note that this general issue can be taken as subsuming the more specific issues raised under Problems One and Two above). For clarity, let me focus on the problem of wages (assuming for now that non-labour income can be treated as exogenous). A common (although not universal) perspective on the issue is that wages represent a payment for skill. Each person has a skill level determined by their skill endowment, education, experience, and so on, and the economy as a whole determines an equilibrium rental price on skill. Thus, we have that the wage rate is given by:

(28) $$w_{it} = p_t S_{it}(X_{it})$$

Here p_t is the skill rental price at time t, and S_{it} is the person i's level of skill. It is determined by a set of variables X_{it} that would typically include things like education and experience.

Now let's consider explicitly modifying (25) to include a set of observables Z_{it} that shift tastes for work:

(29) $$\ln h_{it} = \beta + e \ln w_{it}(1-\tau_t) + \beta_I N_{it} + \beta_T Z_{it} + \varepsilon_{it}$$

One approach to identification of the supply curve in (29) is that there exist some variables in (28) that can be plausibly excluded from (29). Unfortunately, it is far easier to think of variables that fail to satisfy this requirement than to think of ones that do.

For example, some authors have assumed that education enters X_{it} in (28) but not Z_{it} in (29). Yet it is perfectly plausible that education is related to tastes for work (e.g. people who are relatively hard working may also tend to get more education), and hence that education belongs in Z_{it} as well. Indeed, the profession has had difficulty agreeing on any particular variable or set of variables that could be included in X_{it} and excluded from Z_{it}.

14 It should be noted, however, that this approach does not deal with the endogenous choice of segment. If tastes for work are stochastic, as in (27), then which segment one locates on is not determined solely by one's wage rate and non-labour income, but also by the value of the taste shock ε_{it}. If we take the segment on which a person chooses to locate as a *given* we are in effect truncating the range of the taste shock (e.g. people who locate on a high hours segment will tend to be people with high tastes for work). As I noted earlier, this induces a negative correlation between the after-tax wage and tastes for work, which will tend to bias the Marshallian and Hicks elasticities in a negative direction. The approach of Burtless and Hausman (1978) accounts for the taste shock, which makes the segment a person chooses *probabilistic*. Thus, when estimating the labour supply elasticities, their method accounts for the correlation between taste shocks and segment location.

Another approach becomes apparent if we assume that $X_{it} = Z_{it}$, but then substitute (28) into (29) to obtain what economists call a 'reduced form' equation:

$$\text{(30)}\quad \begin{aligned} \ln h_{it} &= \beta + e\{\ln p_t + \ln(1-\tau_t) + \ln S_{it}(X_{it})\} + \beta_I N_{it} \\ &\qquad + \beta_T X_{it} + \varepsilon_{it} \\ &= \beta + e\{\ln p_t + \ln(1-\tau_t)\} + \beta_I N_{it} + \beta_T^* X_{it} + \varepsilon_{it} \end{aligned}$$

Here I have written $\beta^* X_{it} = e\ln S_{it}(X_{it}) + \beta_T(X_{it})$ to subsume all of the common skill and taste shifting variables into one term. We see from (30) that one approach to identify the Marshallian elasticity e in the supply equation is to exploit exogenous variation in the skill rental price p_t and/or in tax rates τ_t.

As I already alluded to under Problem One, prices of raw materials such as oil, iron ore or bauxite could plausibly serve as 'demand side instruments' that shift the rental price of skill but are unrelated to tastes for work. Also, as I discussed under Problem Two, it may well be inappropriate to treat the actual marginal tax rates that people face as exogenous (as these are determined by labour supply decisions which alter tax brackets). But the tax *rules* that people face may (perhaps) be plausibly be treated as exogenous. Thus, one might think about estimating an equation like (29) using raw material prices and/or tax rules as instruments for after-tax wages.

All of the issues I have discussed here potentially apply to non-labour income as well. As with wages, one possible approach is to instrument for non-labour income using the rules that determine transfer benefits. This approach is taken in Bernal and Keane (2009).

Problem Four

The fourth main problem involved in estimation of (25) is that wages are not observed for people who choose not to work. This tends to be more of a problem when studying labour supply of married women (who have a fairly high rate of non-participation) versus other groups like men or single women. The reason non-participation creates a problem can be explained as follows. Assume that, other things being equal, the probability of working increases as the wage rate increases. Then, the people we see working despite relatively low wages will be those with relatively high tastes for work (i.e. large values of the error term ε_{it}). Suppose we try to estimate (25) using only the population of workers with observed wages—the negative correlation between wage rates and tastes for work amongst the population of workers will cause us to underestimate the positive impact of wages on labour supply.

In econometrics this is known as the 'selection bias' problem, as we must estimate (25) using only the people who select to be employed, not the whole population. Pioneering work by Heckman (1974) began a large literature on methods to deal with the selection problem. Unfortunately, there is no solution that does not involve making strong assumptions about how people select into employment. This means that empirical results based on these methods are necessarily subject to some controversy.

In the literature on male labour supply it has been common to ignore the selection problem on the grounds that a very large majority of adult non-retired men do participate in the labour market, so the selection problem can safely be ignored. Whether this is actually true is unclear, but this is the approach of almost every paper I will review.

Problem Five

The fifth main problem concerns the interpretation of the non-labour income variable. In the static labour supply model, one's current non-labour income is treated as a measure of one's wealth. But much of non-labour income is asset income, which is the consequence of a person's decisions about consumption and savings over the life-cycle. We expect assets to follow an inverted U-shaped path over the life-cycle: low when people are young and have low incomes (need to borrow to buy houses, etc.), high in the middle of the life-cycle as people build up assets for retirement, and then declining in retirement. This means that a person's asset level at a particular point in time is not a good indicator of their actual wealth. For example, a 35 year old with a high level of skills who has just gone rather heavily in debt in order to buy a house may in reality be wealthier (in a life-cycle sense) than a 60 year old who has positive savings but at a level that is inadequate to fund retirement. The income effect creates a greater inducement to supply labour for the latter than the former, despite the fact that the latter person has a higher level of current assets.

This brings us back to consideration of the dynamic (or life-cycle) model. Let's return to equation (22) and write it in a slightly modified form:

$$\text{(22')}\quad \ln h_2 - \ln h_1 = \frac{1}{\gamma}\left\{\ln w_2(1-\tau_2) - \ln w_1(1-\tau_1)\right\} - \frac{1}{\gamma}\ln\rho - \frac{1}{\gamma}\ln(1+r) - \frac{1}{\gamma}\ln\frac{\beta_2}{\beta_1}$$

We see that to obtain an estimate of the Frisch elasticity ($1/\gamma$) we essentially need to regress changes in log hours on changes in log after-tax wages, while also including controls for interest rates and the discount rate. To put this theoretical equation into a form that is amenable for econometric estimation we'll need to make several changes. Obviously, we need person (i) and time (t) subscripts on the hours, wage and interest rate variables. And we will again need to account for taste shocks as in (26). This can be done by letting the taste shift variable β_{it} be given by $\beta_{it} = \exp(X_{it}\alpha + \varepsilon_{it})$, where the X is observed taste shifters and the ε is unobserved taste shifters.

Furthermore, (22′) assumes that the change in the after-tax wage from period $t = 1$ to $t = 2$ is fully anticipated by the worker. In fact, there may well be a surprise component to the wage change from $t - 1$ to t. In the life-cycle model a surprise wage change has a different effect than an expected wage change. A surprise wage increase would make a person feel wealthier, and thus it has a negative income effect. I'll denote this surprise wealth effect by ζ_{it}. An expected wage change does not make the person feel wealthier (after all, it is what he/she expected already) and so it has no income effect, only a substitution effect.

Given these changes, we can rewrite (22′) as:

$$\text{(31)}\quad \ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}\left\{\ln w_{it}(1-\tau_t) - \ln w_{i,t-1}(1-\tau_{t-1})\right\} - \frac{1}{\gamma}\ln\rho(1+r_{it}) - \frac{\alpha}{\gamma}\left\{X_{i,t+1} - X_{it}\right\} + \zeta_{it} + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

Many of the papers on life-cycle labour supply that I will discuss estimate versions of (31).

Earlier I noted that tastes for work may plausibly include an individual fixed effect that is constant over time (i.e. some people are just more hard working than others). So in equation (26) we wrote $\varepsilon_{it} = \mu_i + \eta_{it}$, where μ_i is this individual effect. One useful aspect of differencing the data as in (31) is that it causes μ_i to drop out, avoiding the endogeneity problems that its presence would otherwise cause. Thus, the change in ε_{it} from $t-1$ to t may be interpreted as capturing only 'idiosyncratic' shocks to tastes for work (e.g. person i may have been sick in a particular period). The X_{it} in (31) is typically specified to include various observable control variables that might shift tastes for work, such as age, number and ages of children, marital status, and so on.

A few important econometric issues arise in the estimation of (31). Most importantly, the change in the (log of) the after-tax wage from $t-1$ to t is correlated with the error component ζ_{it}. This is essentially by construction: ζ_{it} arises due to the surprise part of the change in the wage, and that must be correlated with the wage change itself. Typically, an instrumental variable procedure is used to deal with this problem (see MaCurdy 1981). In the life-cycle model with saving (but no human capital accumulation) valid instruments for estimation of (31) would be variables that people use to predict wage growth. As long as a variable X_{t-1} is used to predict the growth in the after-tax wage rate from $t-1$ to t it should be uncorrelated with ζ_{it}, as the latter derives from errors in forecasting wage growth.[15]

Good examples of variables that predict wage growth are age and education. This is because wages over the life-cycle follow a well-defined hump shape—tending to grow quickly when people are young, levelling off in middle age and actually declining in real terms for older workers. Also, the shape of the hump varies with education; the peak of the 'hump' comes at a later age for the more educated. To capture these patterns, one might use age, age squared, education and an interaction between age and education as instruments for (i.e. predictors of) wage growth.

To gain intuition for how such a procedure works, it is useful to note that an instrumental variables estimator is typically implemented in a two-stage procedure known as two-stage least squares (2SLS). In the first stage, one regresses the endogenous variable (in this case, wage growth) on the instruments (in this case, the functions of age and education). From this regression one obtains a predicted path of wage growth based on age and education. In the second stage one regresses hours growth on predicted wage growth. Thus, the estimated coefficient on predicted wage growth captures how hours respond to predictable variations of wages in the life-cycle—that is, the extent to which people substitute their time inter-temporally and allocate more work hours to those periods when wages are relatively high. This is exactly the Frisch elasticity concept.

Unfortunately, as we will see below, the typical instruments used to predict wage growth (i.e. age and education) actually predict it quite poorly. As a result, the Frisch elasticity has proven difficult to estimate with any precision. Furthermore, I will show in section 7.3.4 that standard IV procedures generate seriously downward biased estimates of the Frisch elasticity if wages rise with work experience.

Problem Six

A sixth major problem in estimation of labour supply elasticities is measurement error in wages and non-labour income. There is a broad consensus that wages are measured with considerable error in available micro data sets. As is well known, classical measurement error will cause OLS estimates of the coefficient on the wage variable to be biased towards zero, thus leading to underestimates of labour supply elasticities. Furthermore, the measurement error may not be classical. In many data sets, such as the US census, wage rates are constructed by dividing annual earnings by annual hours. Suppose that hours are measured with error—we then have an equation with the error-ridden hours variable as the dependent variable and a constructed wage measure, with the error-ridden hours variable in the denominator, as the independent variable. This creates what is known as 'denominator bias'; the measurement error induces negative correlation between the hours measure and the constructed wage measure. Then, not only will the wage coefficient be biased towards zero, it may be biased in a negative direction. This may in part account for the negative wage coefficient found by Pencavel (1986, p. 23).

Measurement error creates more severe bias when estimating an equation in differences, such as (31), than when estimating equations in levels, such as (25) or (27). This is because if a variable is measured with error then taking the change in the variable over time compounds the error. Again, there are two basic approaches to this problem. One is a 'structural' approach where one models the measurement error process (see Keane & Wolpin 2001; Imai & Keane 2004). The second is to instrument for the change in after-tax wages using variables that are correlated with the true wage change but presumably uncorrelated with the measurement error. Notice that in discussing estimation of (25), (27) and (31) I have already indicated that instrumental variables procedures may be necessary to deal with endogeneity problems. Thus, use of instrumental variables may serve the dual role of dealing with endogeneity and measurement error.

It is likely that error in measuring non-labour income is even more severe than that in measuring wages. As we'll see below, popular econometric methods to model labour supply in the presence of taxes require modelling the details of workers' budget constraints. Yet knowing the actual budget constraint that workers face given modern tax systems is quite difficult. One of the most difficult problems arises because taxes apply to taxable income, and the typical tax system offers an array of deductions. In commonly used data sets it is difficult, if not impossible, to know which

15 This idea of using variables that economic agents use to make forecasts as instruments in dynamic models originated in work by McCallum (1976) and Sargent (1978).

deductions a worker is eligible for and/or actually takes, so deductions must often be imputed. A related problem is the difficulty in measuring fixed costs of work, which are especially important for modelling participation decisions.

7.3.2 Summary of Estimation Results—Static Labour Supply Models

As should be clear from the previous section, there are many econometric problems one must face when estimating labour supply elasticities. And there are many alternative approaches to dealing with these problems. Unfortunately, no consensus has emerged in the economics profession on a 'correct' approach. Indeed, the controversy between advocates of alternative approaches has often been rather intense. The major surveys of the labour supply literature that I cited earlier have tended to break down results both by demographic group and/or by the econometric methods/models employed. In this chapter I'll focus on labour supply of men, and consider in turn the results from static models, life-cycle models with savings, and life-cycle models with both savings and human capital

Pencavel's (1986) classic survey of male labour supply emphasises that the income effect given in equation (13), which I repeat here for convenience:

$$(13)\quad ie = \frac{wh}{N}\left[\frac{N}{h}\frac{\partial h}{\partial N}\right] = w\frac{\partial h}{\partial N} = \frac{\partial (wh)}{\partial N}$$

could also be called the 'marginal propensity to earn' or *mpe*. This is because it indicates how a dollar increase in non-labour income N would shift earnings wh. He notes that, in the static model, this quantity could also be calculated from consumption data. In fact Deaton (1982) did this, using the British Family Expenditure Survey of 1973, and obtained an estimate of ie near zero. (This means consumption increases nearly one-for-one with an increase in non-labour income, while wh hardly declines at all.) Based on this result, Pencavel (1986) concludes that estimates of the income effect that differ much from zero are suspect. He goes on to largely discount the results of several studies that obtain fairly large estimates of the income effect, such as Wales and Woodland (1979) and Hausman (1981).

While Pencavel's survey is generally excellent, I think this conclusion goes too far. The Deaton (1982) result is hard to interpret as a causal effect of non-labour income on consumption, given that non-labour income is likely to be endogenous in a consumption equation. And in a life-cycle model, a high level of non-labour income may simply indicate a high level of permanent income, causing it to be highly positively correlated with consumption.[16] Furthermore, there is substantial evidence that people mostly save the proceeds from temporary tax rebates. As I indicated earlier, introspection may suggest that very large effects of N on wh (that is, values of ie very near –1) are implausible, but I would not conclude based on Deaton (1982) that only effects near zero are plausible.

Pencavel (1986) also largely discounts studies that use estimation methods that impose restrictions from economic theory *a priori*, such as the restriction that the Hicks elasticity be positive or the income effect negative. I would again disagree on this point. Any attempt to estimate labour supply elasticities necessarily involves a long list of assumptions, many of which I discussed in section 7.2.1. These include: exogeneity assumptions (or exclusion restrictions), functional form assumptions, issues of how variables are measured and what is assumed about measurement error, how missing wages of non-workers are handled, assumptions about expectations (i.e. are people forward looking or not?), assumptions about how wages are determined, and so on. It is not clear to me why it is more or less defensible to assume the restrictions that derive from the basic economic theory of consumer behaviour than it is to make these other types of assumptions.[17]

An important point stressed by Pencavel (1986) is that, in work that takes the approach of first specifying a utility function (as opposed to first specifying a labour supply function), one should be aware of what restrictions the utility function imposes on elasticities. For example, utility functions in the constant elasticity of substitution (CES) family have often been used:

$$(32)\quad U_{it} = \left((1-B)C_t^{\rho} + B(L-h_{it})^{\rho}\right)^{1/\rho} \qquad \rho < 1$$

Here L is the maximum hours of work in a period and $(L - h_{it})$ is leisure time. The parameter ρ governs the elasticity of substitution between consumption and leisure, which is $1/(1-\rho)$. Note that as $\rho \uparrow 1$ the elasticity of substitution approaches infinity (perfect substitutes). B is just the CES share parameter. Now, given this utility function, it can be shown that if $N \approx 0$ we have:

$$(33)\quad e = \frac{\rho}{1-\rho}\frac{L-h_{it}}{L} \qquad e_H = \frac{1}{1-\rho}\frac{L-h_{it}}{L} \qquad ie = -\frac{L-h_{it}}{L}$$

Notice that the single parameter ρ governs the Marshallian and Hicks elasticities and the income effect (or *mpe*). To see clearly how this is restrictive, let's think of a period as a day, and assume L = 16. Let's consider a person working 8 hours per day. For such a person, ie must equal –0.5. The model has no flexibility to make it more or less. This, in turn, means that if I told you e, you could back out both e_H and ie. Imposing a particular value on the income effect (as opposed to simply imposing the theoretical restriction that it be negative) does appear to be an unwise modelling choice.

Contrast this situation to that for the functional form I gave in equation (2), which leads to:

16 Note that a one-for-one increase in consumption, if interpreted causally, is wildly at variance with the life-cycle model. In a dynamic model, only unanticipated changes in non-labour income would alter consumption, as an anticipated change would not make a person feel wealthier. Furthermore, even an unanticipated change would be smoothed out over the whole life-cycle, and therefore would have little effect in any one period. Only an unanticipated change in non-labour income that is also expected to be highly persistent should have much impact on current consumption.

17 I believe Pencavel's point is that the restrictions of economic theory should be tested rather than imposed. But, given that the theory cannot be tested without a wide range of auxiliary assumptions, I don't feel this position is completely tenable. Furthermore, assumptions about exogeneity are theoretical restrictions as well.

(34) $e = \frac{1+\eta}{\gamma-\eta}$ $e_H = \frac{1}{\gamma-\eta}$ $ie = \frac{\eta}{\gamma-\eta}$

This is clearly more flexible, as it allows these three quantities to be governed by two parameters (γ and η). The income effect ie can take on a range of values, and knowledge of e alone would not pin down e_H.

Pencavel (1986) notes that the first labour supply function estimation using individual (as opposed to aggregate) level data was by Kosters (1969). He looked at employed married men aged 50–64 in the 1960 US census. Estimating an equation with log hours as the dependent variable, and logs of wages and non-labour income as independent variables (along with various control variables) he obtained an estimate of the Marshallian elasticity of –0.09 (i.e. backward bending labour supply) and a small (negative) income effect (–0.14). However, this early study ignored endogeneity, taxes, and essentially all the key problems listed in section 7.2.1. As Pencavel (1986) discusses, a number of subsequent studies attempted to instrument for the wage to deal with measurement error. But these studies generally continued to obtain small negative Marshallian elasticities. For instance, Ashenfelter and Heckman (1973) obtained e = –0.156 and ie = –0.27. This study continued to ignore taxes.

The first studies to consider effects of after-tax wages and non-labour income on labour supply were Boskin (1973) and Hall (1973). But these studies did not deal with the endogeneity of after-tax wages created by the progressive tax system, which I discussed in section 7.2.1. Also, they did not model the tax system exactly, but instead treated people as if they were choosing labour supply subject to a linear approximation to the piecewise linear budget line created by progressive taxes. Boskin (1973) estimated a Marshallian elasticity of –0.29, an income effect of –0.41, and a Hicks elasticity of 0.12. As I'll discuss in more detail below, Hall (1973) presents his results in a rather complicated form. But my interpretation is that they imply backward bending labour supply but a Hicks elasticity of at least 0.40.

The first study to model the full complexity of the budget constraint created by progressive taxation, and model men as choosing labour supply subject to this constraint, was Wales and Woodland (1979). To achieve this, however, they assume that wages and non-labour income are measured without error. Their estimates, obtained using married men from the PSID, were quite different from the earlier literature. They estimated a Marshallian elasticity of 0.14 (finally positive!), a large income effect of –0.70, and a Hicks elasticity of 0.84. Adopting a similar approach, Hausman (1981) also used married men in PSID and obtained a Marshallian elasticity of close to zero and an income effect of –0.77.

An important point, stressed by Hausman (1981), is that, even with a small (or zero) Marshallian elasticity, large Hicks elasticities of the type estimated by Wales and Woodland (1979), Hall (1973) and Hausman (1981) imply large negative labour supply effects of progressive taxation for people in the upper brackets, as well as large welfare losses.[18] To understand why it is the Hicks elasticity that matters, consider Figure 7.1, which shows a progressive tax system with just two brackets. A person in bracket #1 has an after-tax wage rate of $w(1-\tau_1)$, which of course is also the slope of segment #1. At $H=0$ segment #1 has a height of $N(1-\tau_1)$, which is the person's after-tax non-labour income. Now, suppose the person increases his/her hours above level H_2, so that he/she earns enough to be in tax bracket #2—suddenly the person has a flatter budget constraint with a slope of only $w(1-\tau_2)$. Notably, if we project this budget line all the way over to $H=0$, we arrive at point V. Point V plays an important role in the subsequent analysis which is explained below.

Consider the hypothetical linear budget constraint that starts from V and has slope $w(1-\tau_2)$. Provided preferences have the standard concave shape (as is guaranteed by diminishing marginal utility of consumption and leisure), then any person who would choose to locate on segment #2 given the actual non-linear budget constraint in Figure 7.1 will make the same choice if he/she were presented with the hypothetical linear constraint that originates at point V. Thus, the quantity V is known as 'virtual' non-labour income for a person on segment #2, because such a person acts *as if* V were his/her level of (after-tax) non-labour income.

Now consider what happens when a person moves from segment #1 up to segment #2. Not only does his/her after-tax wage rate fall from $w(1-\tau_1)$ to $w(1-\tau_2)$, but the person also shifts to a linear budget constraint with income V at $H=0$. How does V compare to the actual level of after-tax income $N(1-\tau_1)$ that is obtained if $H=0$? Some simple geometry shows that it exceeds this level by the amount $wH_2(\tau_2-\tau_1)$. That is, segment #1 and segment #2 have the same height at $h = H_2$. But their slopes differ by $w(1-\tau_2)-w(1-\tau_1) = w(\tau_2-\tau_1)$. As we run from $h = H_2$ to $h = 0$, the height of the two segments must diverge by the run multiplied by the difference in the slopes, or $H_2 \cdot w(\tau_2-\tau_1)$. Thus, we see that $V = wH_2(\tau_2-\tau_1) + N(1-\tau_2)$.

Thus, moving from segment #1 to segment #2 has a 'double whammy' effect on labour supply. Not only does a worker face a lower marginal wage rate but, in addition, the amount of the 'virtual' level of non-labour income that is relevant for his/her decision-making has increased by $wH_2(\tau_2-\tau_1)$. This is precisely Hausman's point: even if the Marshallian elasticity is close to zero, there can be a large negative effect of the progressive tax on labour supply if the income effect is large. Of course, since we are talking about a case where the Marshallian elasticity is small, this is equivalent to saying there can be a large negative effect if the Hicks elasticity is large.

Following MaCurdy (1992), we can formalise this as follows. First, suppose the tax rate on segment #2 is increased from τ_2 to $(\tau_2+\Delta)$. This causes the after-tax wage rate to fall by Δw and virtual non-labour income to increase by ΔwH_2. Now, to keep things simple, let's assume a simple

18 In fact, Hausman (1981) found that the welfare loss from progressive taxation was 22 per cent of tax revenues. He found that a shift to a flat rate tax would reduce this to 7 per cent.

linear labour supply function (which is in fact one of the most common specifications in this literature):

(35) $h = \beta + \beta_w w(1-\tau_2) + \beta_I V_2 + \varepsilon$

where the notation τ_2 and V_2 denote the tax rate and virtual income on segment #2, respectively. Plugging in the new values for the tax rate and virtual income we get:

(35′) $h' = \beta + \beta_w w(1-\tau_2-\Delta) + \beta_I (V_2 + \Delta w H_2) + \varepsilon$

Thus, we have that $h' - h = -\beta_w w\Delta + \beta_I w\Delta H_2 = -(\Delta w)(\beta_w - H_2\beta_I)$. The first term here is the change in the after-tax wage, and the second term is precisely the definition of the substitution effect from equation (9), evaluated at the hours level H_2.[19] Thus, we see that, to a good approximation, it is the Hicks elasticity that determines the labour supply response of taxpayers in the higher brackets. Given their findings of substantial Hicks elasticities, Hall and Hausman became strong advocates for a flat rate tax.

In the previous sections I have discussed only literature based on US data. As Pencavel (1986) notes, the literature based on British data took a somewhat different tack for two reasons. First, it has always focused on the effect of taxation, so that wages and non-labour income are always treated as after-tax. Second, it has been largely based on the Family Expenditure Survey, which contains both labour supply and consumption data. Thus, it has generally estimated equations for labour supply and consumption jointly. This is not surprising as once one specifies a utility function defined over both leisure (or hours) and consumption, as in (1) or (32), along with a budget constraint, it is, of course, possible to derive not only labour supply functions but also consumer demand functions. The results from the eight British studies Pencavel cites all find small negative Marshallian elasticitics (with a mean of –0.16), income effects in the range of –0.04 to –0.50 (with a mean of –0.29), and Hicks elasticities ranging from 0.30 to slightly less than zero, with an average of 0.13.

A good deal of work on labour supply was stimulated by the negative income tax (NIT) experiments that were conducted in several US cities beginning in 1968. The NIT experiments were intended to have treatment and control groups. Members of the treatment groups received a grant level G which was taxed away, at a fairly high rate, as they earned income. Thus, G would serve as the guaranteed minimum income for a person with no labour earnings or non-labour income. At a certain level of income a person reaches the 'break-even point' where G has been totally taxed away. Beyond that, they revert to the conventional income tax rate, which is typically less than the benefit tax rate. This creates a *non-convex* budget constraint, because tax rates *fall* as income rises.

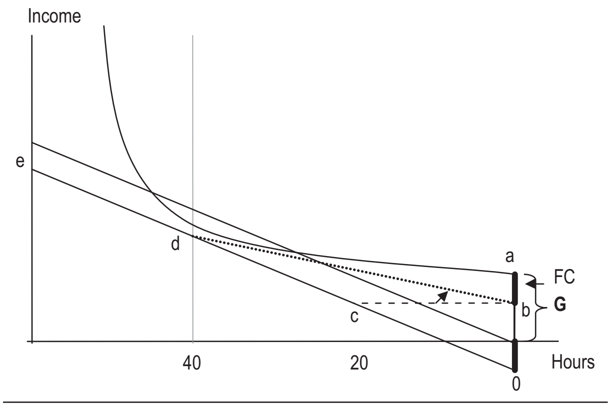
Figure 7.2 illustrates the shape of a typical non-convex budget constraint created by an NIT or other types of welfare programs. The budget constraints connects points *a*, *b*, *c*, and *e*. The figure has been drawn so a person who works zero hours receives G. If they begin to work their income actually drops (from *a* to *b*), due to fixed costs of working, represented by FC. I have drawn an example where, as the person works more hours, his or her grant money is taxed away at a 100 per cent rate as earnings increase. This is represented by the flat dashed line from point *b* to point *c*. The tax rate in the NIT program was only 40 per cent or 60 per cent, but it has not been uncommon for other types of welfare programs that generate non-convex budget constraints to generate tax rates as high as 100 per cent. A good example is the Aid to Families with Dependent Children (AFDC) program in the United States. Finally, point *c* is the break-even point. Above that the person is off the program and faces the regular income tax schedule.

Unfortunately, people in the NIT experiments were not actually assigned randomly to the 'treatment' and 'control' groups, and there is a substantial literature on why this was the case. Nevertheless, the NIT experiments generated useful variation in budget constraints across workers that can be used to help estimate labour supply elasticities.

A very well-known analysis of the NIT experiments by Burtless and Hausman (1978), takes an approach similar to the Wales and Woodland (1979) and Hausman (1981) studies mentioned earlier. That is, the authors model the complexity of the non-linear budget constraint created by the NIT program, and model males as choosing labour supply subject to this constraint (including the choice of which segment to locate on). The difference here is that, while the previously mentioned studies dealt with the *convex* budget constraint created by progressive taxation (i.e. taxes *rising* as income increases), the present study was the first to deal with the *non-convex* budget constraint created by the typical transfer program (i.e. tax rates *falling* as income rises).

As Burtless and Hausman (1978) discuss, given a non-convex budget constraint, if one wants to model the point on the budget constraint that a person chooses, it is necessary to specify the person's utility function. This is because each segment or point (like the non-working point

Figure 7.2 The Non-Convex Budget Constraint Created by NIT or AFDC Types of Programs



Notes: The budget constraint created by the program goes through points *a*, *b*, *c*, *e*. It is generated by the program grant level (G), the fixed cost of working (FC) and the program tax rate, which render the constraint non-convex. The line straight through the origin is the after-tax wage line that would be the budget constraint in the hypothetical situation of a flat rate tax. The dotted line shows the shift in the budget constraint when the program tax rate on earnings is reduced to 50 per cent.

19 Note that in the linear specification $\beta_w = \partial h/\partial w$ is the uncompensated wage effect and $h \cdot \beta_I = h \cdot \partial h/\partial N$ is the income effect.

a in Figure 7.2) implies a different after-tax wage rate, a different level of non-labour income, and a different optimal hours choice. Thus, no single function exists that maps 'the wage' and 'the level of non-labour income' into optimal hours, as wages and non-labour income are themselves functions of the hours choice.

Still, Burtless and Hausman (1978) argued that, since we are more used to specifying hours equations directly than specifying utility functions, it is more intuitive to specify a familiar hours equation and work back (using Roy's identity) to the implied utility function. As an aside, I suspect that many economists today would be more accustomed to specifying utility functions than hours equations. Regardless, Burtless and Hausman choose to use a double log specification:

(36) $$\ln h_{it} = \beta + e \ln w_{it}(1-\tau_i(w_{it}, h_{it})) + e_I N_{it}(w_{it}, h_{it}) + \varepsilon_{it}$$

The parameter e in this equation would be the Marshallian elasticity in the hypothetical case that the person faced a linear budget constraint, but he does not. Thus the estimate of e will not tell us how the person would respond to a change in wage or tax rates. In a model of this type, that would require simulating the person's optimal behaviour under the new regime.

The implications of this point are far reaching. In particular, given piecewise linear budget constraints, utility function parameters are no longer tightly linked with any particular elasticity concept. Thus, labour supply could appear to be 'elastic' or 'inelastic', depending on the type of budget constraint shift one considers.[20] This point is illustrated in Figure 7.2. As I described above, the budget constraint in the figure goes through points *a*, *b*, *c*, *e*. Now consider the indifference curve, which is drawn in such a way that utility is maximised at point *a*, where $h = 0$. I have drawn the shape of the indifference curve so that the Marshallian elasticity given a linear budget constraint would be rather small. That is, the person would choose to work close to 40 hours per week for a wide range of wage rates. But, this elasticity tells us nothing about how the person would respond to changes in the program tax rate on earnings (sometimes called the benefit reduction rate). The dotted line in the figure represents how the budget constraint shifts if the tax rate on earnings is reduced from 100 per cent to 50 per cent. As we see, this has no effect whatsoever on hours of work (i.e. the worker stays at zero).

In contrast, the figure is also drawn so that a small increase in the worker's actual market wage rate would cause him/her to jump from zero to 40 hours of work per week (by slightly raising point *d*). This is true whether the program tax rate is 100 per cent or 50 per cent. Similarly, reductions in the grant level or in the fixed costs of working would have large effects.

Thus, given data that contained wide historical variation in program tax rates (say between 50 per cent and 100 per cent), a researcher studying a program like that described in Figure 7.2 might well conclude labour supply elasticities are small, so that it would be very difficult to induce members of the target population to work. As a brief diversion into the literature on labour supply of lone mothers, let me note that historically this is roughly what happened with the AFDC program in the United States. Years of tinkering with the AFDC tax rate in an attempt to create work incentives had little effect, leading to a conventional wisdom that labour supply was 'inelastic' for single mothers.

Thus, most of the economics profession was taken completely by surprise when a change in policy in the mid-1990s, towards wage subsidies (EITC) and child-care subsidies (CCDF), as well as a strong macroeconomy that raised wage rates, led in a short period of time to very dramatic labour supply increases for this group (see Fang & Keane (2005) for a more detailed discussion). Notably, however, in my work with Moffitt (Keane & Moffitt 1998) and an earlier study (Keane 1995), we modelled the budget constraint created by AFDC in great detail (along with the Foodstamp program and fixed costs of work), and suggested that, while substantial AFDC tax rate reductions would have little effect, labour supply of single mothers would be quite sensitive to wage subsidies, EITC and fixed cost of work subsidies (or work bonuses). This illustrates the value of a structural approach.[21]

Still, the labour supply literature has had a strong tendency to report parameters like e in (36) as 'the' Marshallian elasticity obtained by the study in question. I will generally follow this ingrained tradition, but the reader should always keep this strong caveat in mind: when one sees a typical labour survey that contains a list of Marshallian and Hicks elasticities, one should recall that in many cases these are statements about the shape of workers' utility functions, not about how they would respond to particular tax changes.

That being said, I'll note that Burtless and Hausman obtained a 'Marshallian elasticity' of $e \approx 0$ and an elasticity of hours with respect to non-labour income of $e_I = -0.048$. As we see from (13), to obtain the income effect from the income elasticity we need to multiply by wh/N. Given the population under study, reasonable values (on a weekly basis) would appear to be roughly w = \$3.00, h = 35, N = \$70 so that wh/N = 105/70 = 1.5, giving a typical value of $ie \approx -0.072$.[22] Burtless' and Hausman's overall conclusion was that the income guarantee in the NIT experiments led

20 This point was emphasised by all the authors who pioneered this literature. For instance, Blomquist (1983) states: 'A change in the gross wage rate, non-labour income, or parameters of the tax system changes the whole form of the budget set … the elasticities presented above should therefore not be used to calculate [their] effects …'.

21 As noted by Hausman (1980): 'Structural econometric models which make labour force participation a function of…wages, income transfer levels and the tax system can attempt to answer questions such as the effect of lowering the marginal tax rates on labour force participation. The more traditional reduced form models which do not explicitly parameterise the tax system will be unable to answer such questions'.

22 The discussion in Burtless and Hausman (1978) does not go into much detail about characteristics of the sample. I choose $h = 35$ because they indicate this was the mean of hours, and I choose N = \$70 because their examples imply that that G was approximately \$3,500 per year. w = \$3.00 seems plausible given the time and sample, which was very low income. Alternatively we could, for example, evaluate wh/N at the first kink point in the budget constraint for control subjects, reported in the first row of Table 2. This gives (1.67)(43.16)/(27.8) = 2.6. Then we obtain a higher ie of (–0.048)(2.6) = –0.125.

to only modest reductions in labour supply (i.e. an hours reduction of about 7.5 per cent).[23]

Pencavel (1986) summarises results from eight other studies that also examined the NIT experiments. Again, the estimates of the Marshallian elasticity are all small, but at least here the mean is positive (0.03). Income effects range from about 0.02 to –0.29 (mean –0.10). The Hicks elasticity estimates are bunched fairly tightly around the mean of 0.13.

Next, I turn to Blomquist (1983), who used the piecewise-linear method to study labour supply behaviour in Sweden in 1973. The country had a highly progressive tax structure at that time. Blomquist studied married men who were of prime working age (i.e. 25–55 years old). His estimates implied a Marshallian elasticity of 0.08 and an income effect of ie = –0.03 at mean values in the data. The implied Hicks elasticity is 0.11.

Blomquist (1983) stressed the key point that in non-linear budget constraint models labour supply elasticities cannot tell us how people will respond to changes in the budget constraint. Hence, he went on to use his estimated model to simulate the consequence of Sweden switching from the highly progressive tax regime in place in 1973 to a flat rate tax, a lump sum tax, and a no-tax regime. Under the progressive income tax, the model predicts average annual hours of work of 2,143 hours (close to the sample average). The model predicts that complete elimination of taxes would increase annual hours of work from 2,143 to 2,443.[24] This is a 14 per cent increase. Blomquist also calculates that a 34 per cent flat rate tax would raise the same revenue as the progressive tax. Given a flat rate tax, average annual hours would be 2,297 hours (a 7.2 per cent increase).

Comparing the proportional and no-tax worlds, we see that a 34 per cent tax increase (wage reduction) leads to a 6 per cent reduction in hours. The implied Marshallian elasticity is therefore roughly 6/34 = 0.18. This is quite a bit larger than the Marshallian elasticity of 0.08 implied by the estimates at the mean values of after-tax wages and hours in the data.[25] This illustrates how elasticities calculated assuming linear budget constraints can be quite misleading in a piecewise-linear context. It may also indicate that mean values of elasticities can be quite misleading with regard to population responses in models with heterogeneous workers.[26]

The compensating variation is the lump sum payment that would be needed to make a person in the progressive or flat rate tax worlds as equally well off as a person in the no-tax world. For the progressive tax, this is SEK16,417 while for the flat tax it is SEK18,059. This compares to SEK16,103 in revenue per person raised (under either tax). One method for calculating deadweight loss from the tax is to take the amount by which the compensating variation exceeds the tax revenue, and divide by the tax revenue. This gives (18,059 – 16,103)/16,103 = 12 per cent of revenue for the progressive tax and 2 per cent of revenue for the flat rate tax. Thus, at least for the progressive tax system, the implied welfare losses are rather large. This is despite the quite modest estimates of the Marshallian and Hicks elasticities at the mean of the data (0.08 and 0.11 respectively).

At this point it is worth taking stock of the state of the literature on male labour supply up until the early to mid-1980s. I have discussed four papers that used sophisticated econometric methods to model the structure of progressive tax systems and the choice of hours of work subject to the full complexity of those systems (thus dealing with the endogeneity of wages created by progressive taxation). These were: (i) Wales and Woodland (1979), who obtained a Marshallian elasticity of 0.14, a large income effect of –0.70, and a Hicks elasticity of 0.84, (ii) Hausman (1981) who obtained a Marshallian elasticity of zero and an income effect of –0.77, and so a Hicks elasticity of 0.77, (iii) Burtless and Hausman (1978), who obtained a Marshallian elasticity of zero and an income effect of about –0.07, and (iv) Blomquist (1983), who obtained Marshallian and Hicks elasticities of roughly 0.08 and 0.11, respectively. Notably, all these studies obtained positive (although typically quite small) values for the Marshallian elasticities and, even more importantly, obtained Hicks elasticities that were positive and sometimes quite large. In general, the work of these authors was taken as evidence supporting the idea of a flat rate tax.

However, this conclusion, and the whole approach to estimating models with piecewise-linear budget constraints originated by Burtless and Hausman (1978) and Wales and Woodland (1979), became the subject of considerable controversy. This controversy is often referred to as the 'Hausman-MaCurdy controversy'. In a very influential paper, MaCurdy, Green and Paarsch (1990) argued that the Hausman approach to handling piecewise-linear tax models was, in effect, biased towards finding large Hicks elasticities. To see why, let's use a linear specification as in (35). For a person on segment #1 in Figure 7.1, the labour supply equation is:

(37a) $h = \beta + \beta_w w(1-\tau_1) + \beta_I N(1-\tau_1) + \varepsilon$

while, for a person located on segment #2, the labour supply equation is:

(37b) $h = \beta + \beta_w w(1-\tau_2) + \beta_I [N(1-\tau_1) + w(\tau_2-\tau_1)H_2] + \varepsilon$

23 The Burtless and Hausman (1978) study has been criticised because the authors let the income elasticity e_I be heterogeneous in the population, and a large fraction of the estimates were bunched up near zero (see Heckman & MaCurdy 1981). The implication is that much of the mass would have been on positive values for the income elasticity if this had been allowed in the estimation. But even so, it seems the main conclusion of small income effects would not be altered.

24 It is important to note that this is a partial equilibrium analysis. Such a massive increase in labour supply would presumably lead to a reduction of wages in equilibrium.

25 Of course, for such a large change, the direction in which we do the calculation matters. Going from the proportional tax world to the no-tax world, hours increase 6.4 per cent while wages increase 52 per cent, so the implied elasticity is 6.4/52 = 0.12. This is still 50 per cent greater than Blomquist's calculation at mean values.

26 It is also interesting to compare a no-tax world to lump sum tax world. Blomquist simulates that a SEK16,103 lump sum tax would increase hours from 2443 to 2506, an increase of 63 hours or 2.6 per cent. His estimated non-labour income coefficient of –0.0042 (per thousand) implies an increase in hours of (0.0042)(16,103) = 0.068 thousand hours = 68 hours, which is quite close.

Note that in (37b) I have simply replaced V_2 in (35) by its value $V_2 = N(1-\tau_1) + w(\tau_2-\tau_1)H_2$ which one can derive from Figure 7.1. Now, the key point of this whole approach is that the taste shock ε has to fall in a certain range in order for a person to locate on one of the segments. The ε has to be above a threshold such that desired hours are at least H_2 in order for the person to locate on segment #2, and ε has to be below some threshold in order for the person to choose to locate on segment #1.[27]

Furthermore, there is an intermediate range of ε such that a person will choose to locate precisely at the kink point H_2. This occurs if:

(38a) $\beta + \beta_w w(1-\tau_2) + \beta_I[N(1-\tau_1) + w(\tau_2-\tau_1)H_2] + \varepsilon < H_2$

(38b) $\beta + \beta_w w(1-\tau_1) + \beta_I N(1-\tau_1) + \varepsilon > H_2$

The first equation says that, given the hypothetical budget line that extends segment #2 past $h = H_2$ all the way down to $h = 0$, the person would choose hours *less* than H_2. The second equation says that, given the hypothetical budget line that extends segment #1 past $h = H_2$ all the way up $h = H_{max}$, the person would choose hours *greater* than H_2. Given the actual two-segment constraint, this person's best choice is to locate precisely at the kink point H_2.[28]

Now, rearranging (38) to express it as a range on ε, we obtain:

(38') $$\varepsilon < H_2 - \beta - \beta_w w(1-\tau_2) - \beta_I[N(1-\tau_1) + w(\tau_2-\tau_1)H_2] \equiv U(\varepsilon)$$

$$\varepsilon > H_2 - \beta - \beta_w w(1-\tau_1) - \beta_I N(1-\tau_1) \equiv L(\varepsilon)$$

I have adopted the notation $U(\varepsilon)$ and $L(\varepsilon)$ to denote the upper and lower bounds on ε such that the person would want to locate at the kink point. Now, obviously we must have $U(\varepsilon) > L(\varepsilon)$ in order for the probability of locating at the kink point to be positive. Indeed, the opposite of would imply the logical impossibility that the probability is negative, implying an internal inconsistency within the model. The condition that $U(\varepsilon) > L(\varepsilon)$ can be written as:

$$-\beta_w w(1-\tau_2) - \beta_I[N(1-\tau_1) + w(\tau_2-\tau_1)H_2] > -\beta_w w(1-\tau_1) - \beta_I N(1-\tau_1)$$

which can be further simplified to:

(39) $\beta_w[w(1-\tau_1) - w(1-\tau_2)] - \beta_I w(\tau_2-\tau_1)H_2 > 0$

or simply:

(40) $\beta_w - \beta_I H_2 > 0$

The left-hand side is simply the definition of the Hicks compensated substitution effect from equation (9). Thus, MaCurdy, Green and Paarsch (1990) argued that the Hausman approach to handling non-linear tax models requires compensated substitution effects, and hence the Hicks elasticity, to be positive in order to avoid generating negative probabilities.

Notice that if $\beta_I > 0$ (i.e. the income effect has the 'wrong' sign, implying that leisure is not a normal good) then (39) will have to turn negative for large enough values of H_2. Thus, if confronted with a tax system with kinks at high levels of income, this approach also requires for all practical purposes that $\beta_I < 0$.[29] Indeed, papers such as Burtless and Hausman (1978), Hausman (1981) and Blomquist (1983) restrict $\beta_I < 0$ in estimation.[30]

To get an intuition for why (39) is necessary to induce people to locate at kink points, suppose that $\beta_I > 0$. Then, for a person located at H_2, the increase in virtual non-labour income that would occur should he/she increase hours above H_2 would actually be an inducement to increase hours, not a deterrent. Thus, the only thing that can keep the person from increasing hours beyond H_2 is if the uncompensated wage effect is strong enough to outweigh the perversely signed income effect (as the wage will drop if the person moves above H_2). But if the uncompensated wage effect is strong enough to outweigh the (perverse) income effect it means by definition that the Hicks elasticity is positive.

Referring to the surveys of Pencavel (1986) and Hausman (1985), MaCurdy, Green and Paarsch (1990) note that papers that used 'simple' empirical methods that did not attempt to model the full complexity of the budget constraint tended to obtain small Hicks elasticities, including even perverse negative values. In contrast, the papers that used the piecewise-linear budget constraint approach advocated by Hausman tended to get large values for the Hicks elasticity. MaCurdy, Green and Paarsch (1990) argued that the difference in results did not arise because the Hausman type models did a better job of incorporating taxes. Instead, they argued the difference arose simply because the piecewise-linear budget constraint approach imposed the restriction in (40) that the Hicks elasticity be positive.[31] This criticism was highly influential, leading many

27 This dependence of the range of the errors on the observed segment is precisely why the errors do not satisfy standard OLS assumptions in models with progressive taxation.

28 In other words, if a person with ε in the range given by (38) faced a flat tax at rate τ_1 (i.e. if the tax rate didn't increase from τ_1 to τ_2 at H_2), then he/she would want to work more hours than H_2. However, given the reality that the tax rate does jump at H_2, this person does not want to move up into segment #2, and is content to locate precisely at H_2.

29 Equation (39) says that the uncompensated wage effect (β_w), times the drop in the wage in going from segment #1 to segment #2, must exceed the income effect (β_I) times the increase in virtual non-labour income in moving from segment #1 to segment #2. One would normally expect $\beta_I < 0$, so that the second term in (39) is positive, and the equation simply constrains how negative β_w, the sign of which is theoretically ambiguous, can be. But if β_I has the 'wrong' sign (i.e. $\beta_I > 0$) then the second term is negative and increasing in H_2. Then, it becomes very difficult to satisfy (39) for large values of H_2.

30 These papers all adopt specifications where the income effect is randomly distributed in the population but the distribution is truncated at zero.

31 To quote MaCurdy, Green and Paarsch (1990): 'As documented in the surveys of Pencavel (1986) and Hausman (1985), empirical studies of men's labour supply based on econometric approaches incorporating piecewise-linear constraints produce results that…imply larger estimates of compensated substitution responses that have the sign predicted by economic models of consumer choice, which is in contrast to much of the other empirical work on labour supply. This finding of greater consistency with economic theory has been interpreted…as evidence confirming the merits of accounting for taxes using the piecewise-linear approach. Contrary to this interpretation, this paper shows that the divergence in the estimates…follows directly from features of the econometric models that implicitly restrict parameters… The simple estimation approaches impose no restrictions, but maximum likelihood techniques incorporating piecewise-linear budget constraints require…the Slutsky condition to hold at various points in estimation'.

to discount the large Hicks elasticities obtained in many of the studies cited by Hausman (1985), and contributing to the consensus that the Hicks elasticity is small.

While it is undeniable that the piecewise-linear budget constraint approach requires the Hicks elasticity to be positive in order to generate a sensible econometric model (in the sense that probabilities are guaranteed to be positive), it is not so obvious that this can explain the difference in results between the piecewise-linear budget constraint studies and those that use simpler linear regression methods. There are two reasons for saying this. First, a number of studies that use a piecewise-linear budget constraint approach do nevertheless find Hicks elasticities and income effects that are close to zero. And, conversely, some papers using simpler econometric approaches to handle taxes have found large Hicks elasticites and/or large income effects.

To begin, consider what happened when MaCurdy, Green and Paarsch (1990) applied the same approach as Hausman to a sample of 1,017 prime age men from the 1975 PSID. Like Hausman (1981), they assume a linear hours equation as in (35) with a random coefficient on non-labour income. Strikingly, using the same econometric approach as Hausman, MaCurdy, Green and Paarsch obtained a wage coefficient of essentially zero and a (mean) income coefficient of –0.0071 (see their Table 2, first column). The latter implies an income effect of roughly $w \cdot (\partial h / \partial N) = (4.4)(-0.0071) = -0.031$ and hence a Hicks elasticity of roughly 0.031 at the mean of the data. Thus we have an example where the Hausman approach does yield a very small Hicks elasticity. And there have been other applications of the piecewise-linear budget constraint approach that also obtain small Hicks elasticities and small income effects. A good example is Triest (1990) who applies methods very similar to Hausman (1981) to study 978 married men aged 25–55 in the 1983 PSID. He obtains an income elasticity of essentially zero and Marshallian and Hicks elasticities of roughly 0.05. And recall that the Blomquist (1983) study that I discussed earlier obtained a Hicks elasticity of roughly 0.11 and an income effect of –0.03, which can hardly be called large.

Turning to the simpler approach of assuming a smooth approximation to the kinked budget constraint, MaCurdy, Green and Paarsch (1990) note that this approach also constrains the Hicks elasticity, except now the constraint is a bit weaker: instead of requiring it to be positive, it requires that it can't be 'too negative'. But I don't see this situation as fundamentally different. As the smooth approximation to the budget constraint is made more accurate, the bound on the Hicks elasticity gets tighter, converging to a lower bound of zero as the approximation approaches the true constraint. When MaCurdy, Green and Paarsch (1990, p. 458) apply this approach, they conclude that 'there is no perceptible difference in the estimates obtained assuming differentiable and piecewise-linear tax functions'.

As for papers that use simpler methods but still obtain a large Hicks elasticity and/or a large income effect, a prime example is, in fact, the classic paper by Hall (1973) that initiated this line of research. He used the simple method of linearising the budget constraint around the observed wage/hours combination (as in Figure 7.1), but he did not model the choice of segment. But, like Hausman, he obtained large income effects and fairly large estimates of the Hicks elasticity. Pencavel (1986) excluded Hall's paper from his summary because 'many different estimates are presented and I gave up the attempt to summarise them adequately with a few numbers'. However, Hall's Figures 3.5 and 3.6 appear to provide a concise summary of the results. Hall's sample consisted of all men and women from the 1967 US Survey of Economic Opportunity (SEO), which is an augmented version of the CPS, to include better wage and hours measures and an over-sample of the low income population. As I understand it, Figures 3.5 and 3.6 present labour supply curves averaged across the various demographic groups. Figure 3.6 shows backward bending labour supply above an after-tax wage rate of about \$2.00 per hour. But Figure 3.5 shows a Hicks elasticity evaluated at 2,000 hours of approximately 0.45.[32]

Given these results, I don't think that the use of piece-wise linear budget constraint methods versus simpler methods can explain the large divergence in results across the studies I've discussed. It is particularly puzzling that Wales and Woodland (1979), Hausman (1981), MaCurdy, Green and Paarsch (1990) and Triest (1990) all applied Hausman-like approaches to PSID data on married men from the PSID, using data from nearby (and sometimes identical) waves, and yet the former two studies obtained very large Hicks elasticities and income effects while the latter two studies obtained negligible values for each. Indeed, the latter two papers clearly make note of the fact that this is puzzling.

The excellent replication study by Eklöf and Sacklén (2000) sheds a great deal of light on the reasons for the divergence in results between Hausman (1981) and MaCurdy, Green and Paarsch (1990). Both papers study married men aged 25–55 in the 1976 wave of the PSID. The MaCurdy, Green and Paarsch sample size is a bit smaller (1,018 vs 1,084), because they apply slightly more stringent selection criteria[33], but Eklöf and Sacklén (2000) show this is not a main reason for differences in results. Rather, the difference appears to arise because the two studies adopt very different definitions of the wage and non-labour income variables.

A key point about the PSID is that it contains questions both about the interview week (e.g. what is your current wage rate?) and about the prior year (e.g. what were your annual earnings and annual hours during the past year?). This is a common feature of panel data sets. Hausman (1981) uses the current wage question as his measure of the wage rate, while MaCurdy, Green and Paarsch (1990)

32 I'm able to estimate this figure because the graph that Hall (1973) presents of the compensated labour supply function in Figure 3.5 is rather flat over a very wide range. This is not true of the uncompensated graph.

33 The main difference arises because Hausman (1981) requires that workers not be self-employed at the time of the 1976 interview, while MaCurdy, Green and Paarsch (1990) require they not be self-employed in *both* 1975 and 1976. This costs 55 people.

use the ratio of annual earnings to annual hours. Both of these wage measures have problems.

Hausman's current wage measure is missing for 87 workers and for 4 workers who were not employed in the survey week, and it is top coded at $9.99 per hour for 149 workers. Hausman imputes these missing wage observations for 240/1084 = 22 per cent of the sample using a regression method. In addition, even an accurately measured current wage is presumably a noisy measure of the wage rate that is relevant for the prior year.

MaCurdy, Green and Paarsch's ratio wage measure suffers from the denominator bias problem discussed in section 7.3.1. That is, if observed hours are equal to $h^* = h + \varepsilon$, where h is true hours and ε is measurement error, and we construct the wage as $w^* = E/(h + \varepsilon)$, where E^* is measured earnings, then the measurement error in hours tends to induce a negative covariance between h^* and w^*.[34] As discussed earlier, this denominator bias has the potential to drive the wage coefficient negative.

In addition, Hausman (1981) and MaCurdy, Green and Paarsch (1990) take radically different approaches to measuring non-labour income. Hausman simply imputes an 8 per cent return to equity in owner-occupied housing (this is the only financial asset measured in the PSID). In contrast to this very narrow measure, MaCurdy, Green and Paarsch (1990) construct a very broad measure by taking total household income minus total labour earnings of the husband. The broad measure has the problem that it includes the wife's income, which may be endogenous. That is, the husband's decision on how much to work may affect the wife's labour supply. In contrast, Hausman's narrow measure simply leaves out many types of non-labour income. Not surprisingly, the sample mean of MaCurdy, Green and Paarsch's non-labour income measure is roughly three times greater than that of Hausman's. Neither measure includes imputed flows of services from durables.

Finally, Hausman (1981), and MaCurdy, Green and Paarsch (1990), use different hours measures. The latter study uses the answer to a direct question about hours of work in 1975. Hausman (1981) uses questions about usual hours per week and number of weeks worked in 1975. The mean of the MaCurdy, Green and Paarsch hours measure is 2,236 while that of Hausman's hours measure is 2,123.

Using the same data as MaCurdy, Green and Paarsch, Eklöf and Sacklén (2000) are able to replicate their results almost exactly. That is, the wage coefficient bumps up against the non-negativity constraint and has to be pegged at zero. And the mass of the random non-labour income coefficient also piles up near zero. Then Eklöf and Sacklén (2000) report results of an experiment where, either one by one or in combination, they shift to Hausman's wage measure, non-labour income measure, sample selection criteria and/or hours measure. A subset of the results is reproduced in Table 7.3.

The first row of Table 7.3 presents the authors' replication of MaCurdy, Green and Paarsch (1990). The only difference is a slight change in the computation procedure that leads to a small increase in the estimated income effect (from about –0.037 to –0.068).[35] The second row shows the effect of adopting Hausman's sample selection criteria. This leads to a doubling of the income effect to –0.136. But the wage coefficient remains pegged at zero.

In the third row, the authors switch to Hausman's narrower definition of non-labour income. This has a dramatic effect on the results, with the income effect jumping to –0.488. This result is actually rather disconcerting. Given that each paper's definition of non-labour income is quite debatable, and that, as noted in section 7.3.1, it is not at all obvious how one should define non-labour income in a static model (given that in the real world non-labour income evolves over the life-cycle as a result of savings decisions), it seems unfortunate that results are so sensitive to how non-labour income is defined.[36]

The fourth row shows the results using Hausman's wage measure. Strikingly, the wage coefficient now converges to a positive value, implying a small but positive

Table 7.3 Eklöf and Sacklén (2000) Analysis of Hausman vs MaCurdy-Green-Paarsch (M-G-P)

Wage measure	Non-labour income measure	Sample selection criteria	Hours measure	Coefficient on: Wage	Coefficient on: Non-labour income	Marshall elasticity	Income effect	Hicks elasticity
M-G-P	M-G-P	M-G-P	M-G-P	0.0	–0.011	0.000	–0.068	0.068
M-G-P	M-G-P	Hausman	M-G-P	0.0	–0.022	0.000	–0.136	0.136
M-G-P	Hausman	M-G-P	M-G-P	0.0	–0.079	0.000	–0.488	0.488
Hausman	M-G-P	M-G-P	M-G-P	10.3	–0.004	0.030	–0.025	0.055
Hausman	Hausman	M-G-P	M-G-P	26.5	n.a.	0.078	n.a.	n.a.
Hausman	Hausman	Hausman	M-G-P	26.9	–0.036	0.078	–0.222	0.300
Hasuman	Hausman	Hausman	Hausman	16.4	–0.036	0.048	–0.222	0.270
Hausman's Reported Results			0.2	0.2	–0.120	0.000	–0.740	0.740

Notes: For the sake of comparability all elasticities and income effects are calculated using the mean wage of $6.18 and the mean hours of 2123 from Hausman (1981). In the authors' attempt to replicate Hausman's data set the corresponding figures are 6.21 and 2148. The mean values of both hours and wages are a bit higher in the MaCurdy, Green and Paarsch data set, but this makes little difference for the calculations. For the random non-labour income coefficient, the table reports the median. n.a. denotes not available.

34 Of course, if $E^* = (w + \upsilon)(h + \varepsilon)/(h + \varepsilon) = (w + \upsilon)$, where υ is a stochastic term independent of h^*, then the denominator bias problem does not arise. But this is a highly implausible special case.

35 MaCurdy, Green and Paarsch (1990) reported it was necessary to constrain the variance of the random income effect to obtain sensible estimates, but Eklöf and Sacklén (2000) did not have this problem in the replication.

36 I was also puzzled as to why the authors maintained the peg of the wage coefficient at zero in this model. With an income effect as large as –0.488, there is plenty of leeway for the Marshallian elasticity to go negative while maintaining a positive Hicks elasticity.

Marshallian elasticity of about 0.03. But the income effect remains very small at –0.25, implying a Hicks elasticity of only 0.055.

The fifth row shows the effect of simultaneously adopting Hausman's wage and non-labour income measures. This causes the Marshallian elasticity to jump further to 0.078, but unfortunately the authors do not report the income coefficient for this case. The sixth row shows the effect of simultaneously adopting Hausman's wage and non-labour income measures, and his sample selection criteria. The Marshallian elasticity remains at 0.078 and now we see the income effect is –0.222, giving a Hicks elasticity of 0.300.

The sixth row also adopts Hausman's hours measure. Having adopted all of Hausman's variable definitions and sample selection criteria, this is, in fact, the author's attempt to replicate Hausman (1981). The results have a similar flavour to Hausman's: the Marshallian elasticity is a modest 0.048 but the income effect is a solid –0.220, giving a fairly large Hicks elasticity of 0.270.

Based on these results, the authors conclude it is not the piecewise-linear budget constraint approach itself that explains why Hausman (1981) obtained much larger values for the Hicks elasticity and the income effect than did other authors who adopted 'simpler' econometric approaches. Instead, Eklöf and Sacklén (2000) argue that the key differences were Hausman's use of a direct wage measure and his narrow definition of non-labour income. In particular, the evidence suggests that measuring the wage as annual earnings divided by annual hours does lead to a severe denominator bias that tends to drive the wage coefficient negative.

As further evidence of this assertion, they point to the special issue on labour supply of the *Journal of Human Resources* (1990). They note that in three studies where the wage measure is the ratio of earnings to hours (Triest 1990; MaCurdy, Green & Paarsch 1990; Colombino & del Boca 1990) the estimated Hicks elasticity is either negative or runs up against the non-negativity constraint, while in the two studies where a direct wage measure is used (Blomquist & Hansson-Busewitz 1990; van Soest, Woittiez & Kapteyn 1990), as well as in the authors' own version of MaCurdy, Green and Paarsch (1990), the Hicks elasticity is positive.

Finally, the last two rows of Table 7.3 compare the authors' replication of Hausman (1981) with the results that Hausman actually reports. As is clear, the authors are not able to replicate Hausman very precisely. While Hausman obtained a Marshallian elasticity close to zero, the authors obtain 0.048. And while Hausman obtained a very large income effect of –0.740, Eklöf and Sacklén (2000) obtain a perhaps more plausible value of –0.222.[37]

How do we account for these substantial differences? The authors note that they were unable to match Hausman's sample as accurately as they matched MaCurdy, Green and Paarsch's. They also note that the likelihood function was quite flat in the vicinity of the optimum. In particular, they found that a fairly wide range of different values for the mean and variance of the random coefficient on non-labour income produced similar likelihood values. Given this, they speculate that fairly minor changes in the data set could have produced a fairly large change in the estimates.

Recall that Hausman (1981) calculated that the progressivity of the tax system led to a welfare loss equal to 22 per cent of tax revenues. As we have seen, this value is driven largely by his large estimate of the Hicks elasticity. Given that Eklöf and Sacklén (2000) obtain a mean Hicks elasticity about a third as large as Hausman's, one is tempted to conclude that the implied welfare loss is about a third as large as well. However, as these models assume a distribution of income effects, and as Eklöf and Sacklén (2000) obtain not only a lower mean but also a higher variance, it is not at all clear what a simulation of their model would imply about welfare effects. (It is unfortunate that such a simulation is not available.)

At this point, it is worth taking a closer look at the work by Blomquist and Hansson-Busewitz (1990). They study the labour supply of married men in Sweden using 1980 data from the Level of Living Survey. They restrict attention to those aged 25–55 and have a sample size of 602. One innovation in this study is the use of an hours equation that includes a quadratic in wages. They find that this provides a significantly better fit than a linear specification, although the difference has little impact on the main results. The authors use a direct wage measure (as does Hausman 1981) and a broad measure of non-labour income (the same as MaCurdy 1981). Based on the results in Eklöf and Sacklén (2000) we would predict this combination to lead to a modest positive Marshallian elasticity and a small income effect. It is somewhat comforting that this is roughly what happens. Blomquist and Hansson-Busewitz (1990) obtain Marshallian elasticities of 0.12 to 0.13 in their preferred models, and income effects of only about –0.005.

One nice feature of the Blomquist and Hansson-Busewitz (1990) paper is that they plot both the 'basic' or 'structural' labour supply equation—that is, the equation that would apply if people maximised utility subject to a linear budget constraint (and the structural parameters of the equation could be used to infer the underlying utility function)—and the 'mongrel' or 'reduced form' wage equation that gives desired hours as a function of wages, non-labour income and the *existing tax structure*. This reduced form hours equation will vary as the tax system varies. Strikingly, even though the true labour supply curve is linear with a positive Marshallian elasticity throughout, the reduced form labour supply curve becomes backward bending for wage rates above about SEK26 per hours. This compares to an average gross wage rate of SEK41.75 and an average marginal after-tax rate of only SEK14.83. Thus, a reduced form analysis that fails to account for progressive taxation could easily conclude that labour supply is backward bending when, in fact, this is a feature induced by the tax system, not by underlying preferences.

Finally, when Blomquist and Hansson-Busewitz (1990) simulate the consequence of shifting to a flat rate tax

37 Recall that the income effect $w{\cdot}\partial h/\partial N$ in equation (10) can also be thought of as the derivative of earnings with respect to non-labour income $\partial(wh)/\partial N$—see equation (13)—and that Pencavel (1986) argued that the value obtained by Hausman (1981) was implausibly large based on the consumption literature.

(which needs to be 37 per cent to generate equivalent revenue) they find that the welfare loss from taxation falls from 16 per cent to 5 per cent of revenue collected, while annual hours of work increase from 2,099 to 2,238 (or 6.7 per cent). They also simulate a cut in the national tax rate in the top several brackets by 5 percentage points, from a range of 44–58 per cent to a range of 39–53 per cent. They simulate that this would increase labour supply by 0.4 per cent while actually increasing tax revenue by 0.6 per cent. This implies that the upper bracket tax rates in Sweden in 1980 actually exceeded the revenue maximising rates (see section 7.1, equation (3)).[38]

The paper by van Soest, Woittiez and Kapteyn (1990) uses data from the Dutch Organization of Strategic Labour Market Research (OSA) 1985 survey. This survey contains a direct question about wages on a weekly or monthly basis (which in the latter case is converted to weekly). Consistent with the above conjectures, the authors obtain a Marshallian elasticity of 0.19 and an income effect of –0.09 at the mean of the data, and so have no problems with the non-negativity constraint on the Hicks elasticity (0.28).[39] In my view, the more important aspect of this paper is that, as far as I can discern, it was the first to use simulated data from the model to actually examine model fit. A rather striking failure of the labour supply literature (which it shares with many other literatures in economics) is the lack of effort to examine model fit. The authors find, perhaps not surprisingly, that the simple linear labour supply function (like equation (35)), combined with a piecewise-linear budget constraint, does a very poor job of fitting the observed distribution of hours. In particular, it is completely unable to generate the substantial bunching of male hours at exactly 40 hours per week (see their Figure 1).

The authors attempt to rectify this problem by introducing a demand side constraint on possible hours choices. Each worker is assumed to draw a set of hours points at which he may locate, and the probability of each point is estimated. Of course, offers of 40 hours are estimated to be much more likely than offers of lower hours levels. So this model does fit the spike in hours at 40 (as well as the distribution over other points) quite well. What seems unsatisfactory about this procedure is that the model contains no rationale for why offers of lower levels of hours are uncommon. One explanation would be start-up costs at work, so that productivity rises with hours but starts to decline somewhere after 40. An alternative supply side story for why low levels of hours are uncommon would be fixed costs of work.

Returning to our main theme, Eklöf and Sacklén (2000) found that the major differences in results between Hausman (1981) and MaCurdy, Green and Paarsch (1990), as well as between several other studies in the *Journal of Human Resources* (1990) special issue on labour supply[40], could be explained by differences in the definitions of the wage rate and non-labour income. Specifically, the use of a ratio wage measure (i.e. annual earnings over annual hours), rather than an hourly or weekly wage measure, led to much smaller estimates of wage elasticities, presumably due to denominator bias. And the use of more narrow definitions of non-labour income lead to larger estimates of the income effect. Given the problem of denominator bias, it seems fairly clear that the use of ratio wage measures should be avoided in favour of hourly measures.[41] But the best way to measure non-labour income is not at all clear.

In general, non-labour income may include many components, such as interest income from assets, the service flow from durables, government transfer payments, transfers from relatives, and, in a household context, spouse's income (or some share thereof). Determining the 'right' measure of non-labour income in a static labour supply model is difficult in part because the static model does not provide a framework to even think about asset income. Indeed, in a static model assets should not even exist, as there is no motive for saving. This leads us to an examination of life-cycle labour supply models with savings.

7.3.3 Life-Cycle Labour Supply Models with Savings

In dynamic models, workers make labour supply decisions jointly with decisions about consumption/savings, and the evolution of non-labour income becomes part of the model. But as I'll discuss below, estimation of such dynamic models is difficult. Thus, some authors have sought to develop an alternative approach that maintains the simplicity of static models while producing estimates that are still consistent with life-cycle behaviour.

In an important paper, MaCurdy (1983) developed a scheme for estimating the parameters of a life-cycle labour supply model using techniques no more complicated than instrumental variables estimation. To see how his method works, it is useful to return to the simple two-period model of section 7.2.2. Begin by modifying (17) to include an exogenous source of non-labour income N_t whose level is independent of the person's labour supply decisions (e.g. N_t might represent a lump sum government transfer and/or transfers from relatives):

$$(17') \quad V = \frac{[w_1 h_1(1-\tau_1) + N_1 + b]^{1+\eta}}{1+\eta} - \beta_1 \frac{h_1^{1+\gamma}}{1+\gamma} + \rho\left\{\frac{[w_2 h_2(1-\tau_2) + N_2 - b(1+r)]^{1+\eta}}{1+\eta} - \beta_2 \frac{h_2^{1+\gamma}}{1+\gamma}\right\}$$

Now we can modify equation (18), the first order condition for optimal choice of hours in period $t = 1$, to obtain:

38 Note that Sweden had an array of payroll, value-added and local taxes that brought the overall rates to well above the 58 per cent top bracket national rate. In 1980, the upper limit for the sum of national and local taxes was set at 85 per cent.

39 The paper does not give information on the construction of the non-labour income variable, but in private correspondence the authors have told me that they used a fairly narrow measure that consists only of child benefits (which do not depend on income) and capital income (which few households have).

40 Specifically, these are Triest (1990), Colombino and del Boca (1990), Blomquist and Hansson-Busewitz (1990) and van Soest, Woittiez and Kapteyn (1990).

41 This is not to say that an hourly wage measure is ideal. Its drawback is that we are typically modelling labour supply over a longer period, such as a year. Indeed, this is presumably the reason that many studies chose to use annual wage measures (to better match the time period of the wage with that of the observed labour supply behaviour).

(18′) $\frac{\partial V}{\partial h_1} = \left[w_1 h_1(1-\tau_1)+N+b\right]^{\eta} w_1(1-\tau_1) - \beta_1 h_1^{\gamma} = 0$

Recall that b is net borrowing (or net dissaving) in period 1, which can be positive or negative. Thus, $C_1 = w_1 h_1(1-\tau_1) + N_1 + b$ is consumption in period 1. We can rewrite (18′) as:

(41) $$\frac{\beta_1 h_1^{\gamma}}{[w_1(1-\tau_1)h_1 + N_1 + b]^{\eta}} = \frac{\beta_1 h_1^{\gamma}}{[C_1]^{\eta}} = w_1(1-\tau_1)$$

Recognising that in a T period model an analogous optimality condition will hold in every time period $t = 1,\ldots,T$, we have:

(42) $$\frac{\beta_t h_t^{\gamma}}{\left[w_t(1-\tau_t)h_t + N_t + b_t\right]^{\eta}} = \frac{\beta_t h_t^{\gamma}}{\left[C_t\right]^{\eta}} = w_t(1-\tau_t)$$

It is also important to note that, while (17) and (17′) assumed a flat rate tax, an optimality condition analogous to (42) will also hold in a world with progressive taxation. Then, τ_t is the marginal tax rate the person faces at time t, for the tax bracket in which he/she sits at that time.[42] Also, the equation for consumption must be modified. This is illustrated in Figure 7.3 for a system with two brackets.

Consider a person who chooses to locate on segment #2, which means he/she chooses a level of hours $h_t > H_2$, where H_2 is the hours level that renders the person's earnings high enough that he/she enters tax bracket #2. The consumption level for this person is:

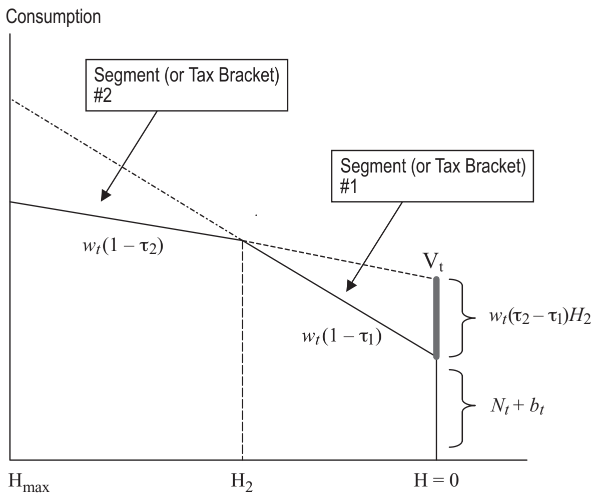
(43a) $C_t = w_t(1-\tau_2)h_t + V_t$

where 'virtual' non-labour income V_t is given by:

(43b) $V_t = w_t(\tau_2 - \tau_1)H_2 + N_t + b_t$

MaCurdy (1983) noted these points, and also noted that the optimality condition (42) contains only variables dated at time t. Hence, despite the fact that we have a dynamic model with saving, the parameters γ and η that describe preferences can be estimated from a *single period of data*, provided we utilise not only data on hours and wages but also data on consumption. MaCurdy proposed two methods for doing this:

Figure 7.3 The Budget Constraint Created by Progressive Taxation in the Presence of Saving



Method 1: Estimate (42) using two-stage least squares. As MaCurdy notes, (42) must hold at a person's optimal hours choice, regardless of whether there is a progressive income tax or a flat tax (provided the person is not at a kink point). To put (42) in a form that can be estimated we need to introduce a source of stochastic variation in hours and consumption choices. Let the parameter β which shifts the marginal rate of substitution (MRS) between consumption and leisure, be given by:

(44) $\beta_{it} = \exp(X_{it}\alpha + \varepsilon_{it})$

Here X_{it} represents *observed* characteristics of person i that shift his/her tastes for consumption versus leisure, and ε_{it} represents *unobserved* taste shifters. Now, taking logs of (42), and putting i subscripts on all variables to indicate person-specific values, we get:

(45) $\ln w_{it}(1-\tau_{it}) = \gamma \ln h_{it} - \eta \ln C_{it} + X_{it}\alpha + \varepsilon_{it}$

Note that (45) is not a typical labour supply equation (which would have hours as the dependent variable). Rather, it is simply a relationship among three endogenous variables—the after-tax wage, hours and consumption—that must hold if person i is making work/consumption choices as suggested by economic theory. All three variables are endogenous because they are correlated with the taste shocks ε_{it}. This occurs for reasons we have already discussed.[43] As a result, it is really just a matter of convenience which endogenous variable we call the 'dependent' variable.

Given the endogeneity of hours and consumption, we should estimate (45) using instrumental variables. The instruments must be correlated with wage, hours and consumption but not correlated with the unobserved tastes for work ε_{it}. Naturally, the choice of instruments tends to be controversial in any such approach. Estimation of (45) gives us the values of the structural parameters of preferences γ and η, from which we could construct the Marshall, Hicks and Frisch elasticities, as in equation (24).

Method 2: Estimate a labour supply function that is consistent with the life-cycle framework. The idea here is to extend the Hall (1973) approach to the dynamic case simply by redefining virtual non-labour income for period t to include b_t. The approach is illustrated in Figure 7.3, for the case of a two-bracket tax system. Note that if the person locates on segment #1 then their after-tax wage is $w_t(1-\tau_1)$ and their virtual non-labour income is $V_t = N_t + b_t$. If the person locates on segment #2 then their after-tax wage is $w_t(1-\tau_2)$ and their virtual non-labour income is $V_t = w_t(\tau_2 - \tau_1)H_2 + N_t + b_t$. Notice that, regardless of segment, virtual non-labour income is given by:

42 Condition (42) would fail to hold for a person who locates at a kink point. Thus, MaCurdy assumes the tax system is approximated by a smooth function, ruling out kink points.

43 Obviously, hours are endogenous because a person who is hard working (i.e. has a low value of ε_{it}) will tend to work more hours, other things being equal. The after-tax wage is endogenous because a person who is hard working will: (i) tend to have a high pre-tax wage because he/she puts in greater effort, and (ii) tend to face a higher tax rate because he/she works enough hours to be pushed into a high bracket. And consumption is likely to be endogenous because it is a function of the endogenous w and h.

(46) $V_t = C_t - w_t(1-\tau_t)h_t$

where τ_t denotes the tax rate for the segment on which the person locates at time t. Thus, MaCurdy suggests estimating labour supply equations of the form:

(47) $h_{it} = h(w_{it}(1-\tau_{it}),\ V_{it},\ X_{it})$

To implement this procedure one must pick a particular functional form for the labour supply function in (47). For example, one might choose the linear specification we saw in (37a), or the double log specification we saw in (36). Also, because both the after-tax wage rate and virtual non-labour income are endogenous, we must instrument for them, analogous to the approach in Method 1.

MaCurdy (1983) implements both Method 1 and Method 2 using a sample of 121 married males who were part of the control group in the Denver Income Maintenance Experiment (a negative income tax experiment) in 1972 to 1975. To implement Method 1—equation (45)—MaCurdy includes as observed taste shifters (X_{it}) the number of children and race indicators. His main instruments are quadratics in age and education, as well as interactions between the two. This makes sense given the strong correlation between education and lifetime earnings, and the fact that both wages and hours follow hump shapes over the life-cycle. The interactions capture the fact that the peaks of these humps tend to come at later ages for those with more education. But the use of these instruments does require the strong assumption that age and education are not correlated with tastes for work ε_{it}.

MaCurdy's estimates indicate that $\gamma = 0.16$ and $\eta = -0.66$. To compare these figures to prior literature, MaCurdy calculates what they would imply about labour supply elasticities given a linear budget constraint. It turns out that the estimates imply highly elastic labour supply behaviour. Using our formulas for a person with no non-labour income—equation (24)—the implied Marshallian elasticity is $(1+\eta)/(\gamma-\eta) = 0.42$, the Hicks elasticity is $1/(\gamma-\eta) = 1.22$, the income effect is $\eta/(\gamma-\eta) = -0.80$, and the Frisch elasticity is $1/(0.16) = 6.25$. MaCurdy calculates elasticities at the mean of the data, and obtains a Marshallian elasticity of 0.70, a Hicks elasticity of 1.47, and an income effect of $w\partial h/\partial N = -0.77$.

Turning to Method 2, MaCurdy considers both linear and double log specifications, using the same control variables and instruments as in Method 1. For the double log specification he obtains:

$$\underset{}{\ln h_{it}} = \underset{(0.53)}{0.69}\ln w_{it}(1-\tau_{it}) - \underset{(0.0010)}{0.0016}\,V_{it} + X_{it}\alpha + \varepsilon_{it}$$

and for the linear specification he obtains:

$$h_{it} = \underset{(13.8)}{19.4}\ w_{it}(1-\tau_{it}) - \underset{(0.07)}{0.16}\,V_{it} + X_{it}\alpha + \varepsilon_{it}$$

where the figures in parentheses are standard errors. The log specification gives a Marshallian elasticity of 0.69, almost identical to that obtained via Method 1 when evaluated at the mean of the data.

MaCurdy (1983) evaluates the other elasticities at the mean of the data. To do this we need to know the mean of $V_{it} = C_{it} - w_{it}(1-\tau_{it})h_{it}$ = \$133 per month, the mean of the after-tax wage is \$2.75 per hour and the mean of hours is 170 per month. In the double log model the income effect is then $(wh)(1/h)\partial h/\partial V = (468)(-0.0016) = -0.75$. This is again almost identical to the value obtained using Method 1. The Hicks elasticity is thus 0.69 + 0.75 = 1.44.

For the linear specification, the Marshallian elasticity is (2.75/170)(19.4) = 0.31, the income effect is (2.75)(–0.16) = –0.44 and the Hicks elasticity is 0.75. Thus, the linear specification produces more modest elasticity estimates. Nevertheless, as MaCurdy (1983) notes, all three approaches (Method 1 and Method 2 with a double log or linear specification) produced estimates of labour supply elasticities that are quite large relative to most of the prior literature. MaCurdy notes that this may indicate that prior estimates were misleading because: 'Existing studies of male labour supply rarely treat measures of wages and income as endogenous variables… Many of these studies ignore taxes or fail to account properly for the endogeneity of marginal tax rates, and none of them recognises that a household may save or dissave during a period'. But MaCurdy also notes that other factors, such as possibly invalid instruments or the small and unrepresentative nature of the Denver Income Maintenance Experiment sample, could have led to upward biased estimates of labour supply elasticities. It is also the case that the parameter estimates are rather imprecise (see above).

Altonji (1986) noted that one could rewrite (45) as:

(48) $$\ln h_{it} = \frac{1}{\gamma}\ln w_{it}(1-\tau_{it}) + \frac{\eta}{\gamma}\ln C_{it} - X_{it}\frac{\alpha}{\gamma} + \frac{\varepsilon_{it}}{\gamma}$$

By estimating this equation using instrumental variables we would uncover the Frisch elasticity ($1/\gamma$) directly. Recall that the Frisch elasticity is defined as the effect of a change in the wage holding lifetime wealth fixed. In (48) consumption serves as a summary statistic for lifetime wealth. If the wage changes but consumption stays fixed it means that perceived wealth remained fixed. This means either (i) that the person expected the wage change, so it does not affect his/her perception of lifetime wealth, or (ii) that the person expects the wage change to be very short-lived, so that it has a negligible effect on lifetime wealth. Estimation of (48) also enables us to back out the preference parameter η as the ratio of the consumption coefficient to the wage coefficient.

Altonji (1986) estimates (48) using data on married men, aged 25–60, taken from the 1968–1981 waves of the PSID. Two key differences with MaCurdy (1983) are that Altonji does not use after-tax wage rates, and the PSID measure of consumption includes only food consumption. Altonji also includes a more extensive set of observed taste shifters in X than does MaCurdy (i.e. in addition to children and race he also includes age, health, region and year dummies). Recall that we must instrument for wages and consumption both because they are measured with error and because they are presumably correlated with the unobserved taste shifter ε_{it}. A novel feature of Altonji's paper is that he uses a ratio wage measure (annual earnings over hours) as the independent variable in (48), and then uses as an instrument a direct question about the hourly wage. As long as the measurement error in these two measures is uncorrelated (as seems plausible), the latter is a valid instrument.[44] As an additional instrument Altonji uses a measure of the 'permanent wage', constructed by regressing the observed wage on individual fixed effects, education, a quadratic in age, an interaction between age and education, year dummies, health and region.

Altonji (1986) estimates that $(1/\gamma)$ = 0.172 (standard error = 0.119) and that (η/γ) = –0.534 (standard error = 0.386). The implied values of γ and η are 5.81 and –3.10. These, in turn, imply Frisch, Hicks and Marshall elasticities of 0.17, 0.11 and –0.24, respectively, and an income effect of –0.35. It is interesting to compare these values with those obtained by MaCurdy (1983) of 6.25, 1.22, 0.42 and –0.80. Reminiscent of the 'Hausman-MaCurdy' debate discussed earlier, we again find ourselves in the situation—now in the context of life-cycle models—of authors finding very different estimates of labour supply elasticities for reasons that are not evident. Does MaCurdy get much higher elasticities because he accounts for taxes and/or has a more complete measure of consumption? Or, because he uses different instruments? Or are his results unreliable due to the small and unrepresentative nature of the Denver Income Maintenance Experiment sample? Does rearranging (45) to obtain (48) actually matter? Unfortunately, there is no replication study that attempts to reconcile the Altonji (1986) and MaCurdy (1983) results so we don't know the answer to these questions.

Blundell and Walker (1986) published a closely related paper. Like MaCurdy (1983), these authors also develop a scheme for estimating the parameters of a life-cycle labour supply model using techniques no more complicated than instrumental variables estimation of a static model, and again the method involves a redefinition of the virtual non-labour income variable. Blundell and Walker adopt the approach of 'two-stage budgeting'. In the first stage, the worker/consumer decides how to allocate his/her 'full income' across all periods of his/her life. Full income is defined as the wage rate times the total hours in a period, plus any exogenous non-labour income, plus net dissaving. Within each period, full income is allocated between consumption and leisure. Thus we have the within-period budget constraint:

(49) $$F_t = w_t(1-\tau_t)T + N_t + b_t = w_t(1-\tau_t)(T-h_t) + C_t$$

where F_t is full income, T is total time in a period and $T - h_t$ is leisure.[45] Similar to MaCurdy's Method 2, where one estimates labour supply equations that condition on the virtual income variable $V_t = C_t - w_t(1-\tau_t)h_t$ (see equation (47)), here one conditions on the full income allocated to period t [46]:

(50) $$h_{it} = h(w_{it}(1-\tau_{it}),\ F_{it},\ X_{it})$$

Notice that full income allocated to period t plays a role analogous to that of consumption in MaCurdy's Method 1 or virtual income in his Method 2. That is, if the wage increases but the full income allocated to the period is held fixed, it means that the wage increase has not made the person feel wealthier (i.e. it has not relaxed his lifetime budget constraint).

Blundell and Walker (1986, p. 545) argue that there is no need to instrument for F_{it}, even though it is a choice variable, because it reasonable to assume that taste shifters which affect that allocation of resources over the life-cycle are independent of those that affect choices within a period. But this seems like an odd argument. For instance, one would plan to allocate more resources to periods when tastes for consumption and/or leisure are high than towards other periods.

In contrast to the direct utility function, which expresses utility as a function of the goods a person consumes (i.e. consumption and leisure), the indirect utility function expresses the maximum utility a person can attain as a function of his/her budget constraint variables, in this case full income and the after-tax wage rate (under the assumption that he/ she will make consumption and leisure choices optimally given F_t and $w_t(1-\tau_t)$). Blundell and Walker (1986) consider a case where the indirect utility function has the form:

(51) $$U_t = G\left[\frac{F_t - a(w_t(1-\tau_t))}{b(w_t(1-\tau_t)}\right]$$

Actually, as we will see below, Blundell and Walker consider a more complex model of joint labour supply of couples, where the price of consumption goods varies over time in addition to the wage. But I will omit those complications for now in order to focus on how the two-stage budgeting idea is implemented. Obviously, we have that:

(52) $$\frac{\partial U_t}{\partial w_t(1-\tau_t)} = h_t \cdot \frac{\partial U_t}{\partial F_t}$$

This equality holds because if the after-tax wage increases by one unit then the person has h_t extra units of income to spend on consumption.[47] But if full income increases by one unit the person has only one extra unit of income to spend on consumption. Thus the derivative on the left of (52) must be h_t times greater than that on the right. Applying (52) to (51) we can obtain the labour supply equation:

(53) $$h_t = \frac{G'(\cdot)\left\{\frac{-b'(\cdot)}{b^2(\cdot)}[F_t - a(\cdot)]\right\} - \frac{a'(\cdot)}{b(\cdot)}}{G'(\cdot)\frac{1}{b(\cdot)}}$$
$$= -a'(w_t(1-\tau_t)) - \frac{b'(w_t(1-\tau_t))}{b(w_t(1-\tau_t))}[F_t - a(w_t(1-\tau_t))]$$

44 To be in the estimation sample a person must have both wage measures. There are 4367 men who satisfy the criteria. Note that this tilts the composition of the sample towards hourly workers.

45 Given progressive taxes, N_t could be defined to include the virtual non-labour income for the linearised budget constraint, just as before.

46 Indeed, the methods are not just analogous but identical, as one can always write a one-period budget constraint in terms of after-tax wage and either full income or virtual non-labour income. That is, the expressions (47) and (50) are alternative expressions for the same labour supply function.

47 The assumption that h_t stays fixed when the wage increases is a simple application of the 'envelop theorem'. This says that for very small changes in the wage rate, the consumer can't do better than to spend all the extra income on consumption. Any utility gain that he/she might achieve by reallocating his/her full income between consumption and leisure is trivially small.

As Blundell and Walker (1986) note, the researcher has a great deal of flexibility in choosing the $a(\cdot)$ and $b(\cdot)$ functions. Thus, labour supply can be allowed to depend on the wage and full income in rather complex ways. This has a downside in terms of interpretability, in that, in contrast to the MaCurdy (1983) and Altonji (1986) specifications, elasticities will have to be simulated. As best as I can ascertain, the basic specification that Blundell and Walker (1986) estimate for men looks like this:

$$(54) \quad h_t = (T_m - \gamma_m) - \frac{\beta_m}{w_t^m}[F_t - \gamma_m w_t^m - \gamma_f w_t^f - \gamma_c p_t^c - 2\gamma_{fc}(w_t^f p_t^c)^{1/2}]$$

Here w_t^m is the after-tax wage for the husband, w_t^f is the after-tax wage for the wife, p_t^f is the price of consumption goods and F_t is full income as given by (49).

Blundell and Walker (1986) estimate this model on a sample of families from the 1980 UK Family Expenditure Survey. As the focus in this chapter is on labour supply of men I only discuss those results. Some limitations of the analysis should be noted. First, as already noted, the authors do not instrument for full income or wages. Second, the analysis was limited to families where the household head is a manual worker, a shop assistant or a clerical worker. This gives 1378 households with a female participation rate of 64 per cent. (The authors do not indicate why they chose to restrict the sample in this way.) Third, the consumption measure was limited, including food, clothing, services and energy but excluding housing, transport, alcohol and other important categories.

Averaging over the whole sample, the authors simulate a Frisch elasticity for men of only 0.026 and a Hicks elasticity of only 0.024. These values are small for all demographic subgroups examined. The authors report an elasticity of male hours with respect to full income of –0.287. Based on figures reported in the paper, I calculate that full income is £267 per week on average, and that male after-tax earnings are (2.08)(39.8) = £82.78 per week on average. These figures imply an income effect of $(wh/F)(F/h)(\partial h/\partial F) = w(\partial h/\partial F) =$ –0.089[48] and a Marshallian elasticity of –0.065. In this regard, it is notable that the authors use a ratio wage measure (i.e. usual earnings over usual hours) to construct wage rates. As we discussed earlier, this may lead to downward bias in elasticity estimates due to denominator bias, particularly when no instrument is used to correct for measurement error. This may account in part for these low elasticity estimates.

Beginning with MaCurdy (1981), a number of studies have attempted to use equations similar to (31) to estimate the Frisch elasticity directly. Here I repeat (31) for convenience:

$$(31) \quad \ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}\left\{\ln w_{it}(1-\tau_t) - \ln w_{i,t-1}(1-\tau_{t-1})\right\} - \frac{1}{\gamma}\ln\rho(1+r_{it}) - \frac{\alpha}{\gamma}\left\{X_{it} - X_{i,t-1}\right\} + \zeta_{it} + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

MaCurdy (1981) estimates equation (31) using annual data on 513 married men observed from 1967–1976 in the PSID. To be included in the sample the men must have been 25–46 years of age in 1967 and have been continuously married to the same spouse during the sample period. MaCurdy uses a complete set of time dummies to pick up the log interest rate terms in (31), rather than using a particular interest rate variable. No observed taste shifter variables X are included. It is also notable that MaCurdy does not adjust wages for taxes.[49] Hence, the specification is simply a regression of annual log hours changes on log wage changes, along with a set of time dummies.

MaCurdy presents his analysis in a setting where workers have perfect foresight about their future wages. But, as he notes, his results do not hinge on this assumption provided he uses as instruments for wages variables that were known to a worker at time t or before, so that the worker could have used these variables to forecast wage growth from time $t-1$ to t. Provided the worker forecasts wage growth rationally, such instruments will be uncorrelated with the error term ξ_{it} which arises because of errors in forecasting wage growth. Such forecast errors should only be correlated with variables that are revealed during the time $t-1$ to t interval (e.g. an unexpected recession, illness or plant closure). The instruments that MaCurdy uses to predict wage growth are by now familiar: quadratics in age and education as well as age/education interactions, parental education and year dummies.

Using this approach, MaCurdy (1981) obtained a Frisch elasticity of only 0.15 (standard error = 0.98). It is striking to compare this to the Frisch elasticity of 6.25 that MaCurdy (1983) obtained using the Denver data, where he adopted the alternative approach of using consumption to proxy for the marginal utility of wealth. But, given that Altonji (1986) obtained 0.172 using a closely related consumption-based approach, and Blundell and Walker (1986) obtained 0.026 using the two-stage budgeting approach, the high Frisch elasticity figure in MaCurdy (1983) starts to look like a striking outlier. Furthermore, as the Frisch elasticity is in theory an upper bound on the Hicks and Marshallian, this would lead one to a conclusion that labour supply elasticities are small for men in general.

But before reaching this conclusion, it is important to keep two points in mind. First, the large standard error (0.98) on MaCurdy's estimate, which suggests that the instruments are doing a very poor job of predicting wage changes. Second, as noted earlier, wages themselves are measured with error, and taking the change in wages as in (31) greatly exacerbates the problem. This would substantially bias the coefficient on wage changes towards zero. The two issues are related, as one needs good predictors of true wage changes in order to correct the measurement error problem.

Altonji (1986) tried to address this problem by using a better instrument for wage changes. As I noted earlier, he

48 Consistent with calculation, when the authors simulate a £50 reduction in non-labour income it leads to an average 2.5 hour increase in weekly hours for males, implying a derivative of about –2.5/50 = –0.05. Multiplying this by the mean male after-tax male wage rate of 2.08 gives an income effect of about –0.10.

49 It may be argued that taxes will largely drop out of (31) if the marginal tax rate a person faces does not change too much from year to year. Altonji (1986) makes this argument explicitly.

uses two wage measures from the PSID, one serving as the wage measure in the labour supply equation, the other serving as an instrument. Using a PSID sample very similar to MaCurdy's, and using similar predictors of wage changes (quadratic in age and education, etc.) he gets an R-squared in the first stage prediction equation of only 0.008. Then, in estimating the main labour supply equation, he gets a Frisch elasticity of 0.31, again with a large standard error of 0.65. However, when Altonji uses the alternative wage change measure as an additional instrument, he gets a much better R-squared of 0.031 in the prediction equation.[50] Then, in estimating the labour supply equation, he gets Frisch elasticity of 0.043, with a standard error of only 0.079. Thus, we seem to have a rather tight estimate of a small Frisch elasticity. The problem here, of course, is that the use of an alternative wage change measure as an instrument is only valid under the strong assumption that workers do have perfect foresight about wage changes. Otherwise, any wage change measure will be correlated with ξ_{it}.[51]

Angrist (1991) proposes dealing with the measurement error problem by using grouped data estimation. That is, he works with the equation:

$$(55)\quad \overline{\ln h_{it}} - \overline{\ln h_{i,t-1}} = \frac{1}{\gamma}\left\{\overline{\ln w_{it}(1-\tau_t)} - \overline{\ln w_{i,t-1}(1-\tau_{t-1})}\right\} + f(t) + \overline{\zeta_{it}} + \overline{(\varepsilon_{it} - \varepsilon_{i,t-1})}$$

Here, $\overline{\ln h_{it}}$ denotes the sample mean of variable $\ln h_{it}$ over all people i observed in year t. The idea is that, while the individual log hours and log wage variables may be measured with error, this measurement error will cancel out when we average over people.[52]

Notice that I have substituted a function of time $f(t)$ for the interest rate variable that appears in (31). MaCurdy (1981) and Altonji (1986) both used a complete set of year dummies to pick up the interest rate variable in estimating versions of (31). But that will not work here because a complete set of year dummies would enable one to fit changes in average hours perfectly and the Frisch elasticity $(1/\gamma)$ would not be identified. Identification requires that $f(t)$ be specified as a low order polynomial in time.

Also notable is that (55) includes the mean of the surprise variable $\overline{\zeta}_{it}$. We would expect this to be negatively correlated with the mean wage change. That is, an unexpected aggregate productivity shock that increases the average wage rate would tend to make people feel wealthier, inducing a negative income effect. For estimation of (55) to identify the Frisch elasticity, $f(t)$ must capture such unexpected aggregate shocks, so that $\overline{\zeta}_{it}$ drops out.

At this point, it is important to consider whether estimation of (55) will actually uncover labour supply parameters, or some mongrel of supply and demand factors. For estimation of (55) to identify the Frisch elasticity, it is necessary that the variation in average wages be induced by *anticipated* shifts in labour demand (e.g. anticipated productivity growth). For this to be true we must not only rule out aggregate unexpected productivity shocks, but aggregate shocks to tastes for work as well. Or, if these are present, we must assume they are all captured by the time polynomial $f(t)$.

Note that $\overline{(\varepsilon_{it} - \varepsilon_{i,t-1})}$ is the average change in tastes for work in the sample. If there are aggregate shocks to tastes for work, we would expect to have a negative correlation between $\overline{(\varepsilon_{it} - \varepsilon_{i,t-1})}$ and the change in average wages (as an increased supply of labour would drive down wages in equilibrium). In that case, we should estimate (55) using demand side variables known at $t-1$ as instruments. But in the absence of aggregate shocks to tastes for work $\overline{(\varepsilon_{it} - \varepsilon_{i,t-1})}$ is simply noise, and (55) does represent a supply equation.

With these caveats in mind, let's consider Angrist's results. He uses the PSID data from 1969–1979, and takes a sample of 1,437 male household heads aged 21–64 with positive hours and earnings in each year. He then constructs average hours and earnings for the sample members in each year, and uses these to estimate (55), with $f(t)$ either left out or set to be a linear or quadratic time trend. When no trend is included the estimate of the Frisch elasticity is –0.132 (standard error = 0.042), which violates economic theory. However, this obviously occurs only because during the 1969–1979 period there was a secular downtrend in average hours and a secular upward trend in the wage rate. When a linear trend is included in the model it picks this up, and Angrist obtains a Frisch elasticity of 0.556 (standard error = 0.124).[53] Using a quadratic trend he obtains 0.634 (standard error = 0.205). Specification tests do not reject the model with a linear trend, although it is likely that the test has little power given the small sample size.

Regardless, these estimates provide some evidence for higher values of the Frisch elasticity than results from most of the prior literature would suggest. However, it is unclear whether Angrist obtains the higher value because of a superior method of handling measurement error or because the results are contaminated by unanticipated labour demand shocks that induce a positive correlation between wages and hours.

A related paper by Browning, Deaton and Irish (1985) shows how one can estimate the Frisch elasticity using repeated cross-section data instead of true panel data. First, they show how to derive a version of the Frisch labour supply function that has the wage change in levels (not logs) as the dependent variable, giving an equation of the form:

$$(56)\quad h_{it} - h_{i,t-1} = \beta\left\{\ln w_{it} - \ln w_{i,t-1}\right\} - \beta \ln \rho(1+r_{it}) - \alpha\left\{X_{it} - X_{i,t-1}\right\} + \zeta_{it} + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

50 This may still seem small, but is actually not bad given the large sample size of roughly 4000 observations, as indicated by the highly significant F statistic of 129.

51 Given this problem, Altonji tried using the lagged wage change as an instrument, as the lag would have been known at time t. But it is a poor predictor, and the standard error jumps to 0.45.

52 Of course, this requires that the measurement error be additive.

53 It is interesting that when Angrist simply estimates (55) on the micro data, using a linear trend, he obtains a Frisch elasticity of –0.267 with a standard error of 0.008. But when he estimates the hours equation in levels he obtains –0.063 (standard error = 0.005). This illustrates how first differencing exacerbates the downward bias in the wage coefficient.

To estimate this equation, the authors use data on married men from the UK Family Expenditure Survey (FES) for the seven years from 1970–1976. The FES does not track individual people through time. Rather, it takes a random sample of the population in each year. Thus, it is not possible to take first differences like $h_{it} - h_{i,t-1}$ for individual people i. Instead Browning, Deaton and Irish construct eight cohorts from the data: men who were 18–23 in 1970, men who were 24–28 in 1970, up to men who were 54–58 in 1970. (Note that members of the first cohort are 24–29 in 1976 when the data ends, while members of the last cohort are 60–64. Thus, the data cover all ages from 18 to 64.) The authors then take the cohort-specific means of each variable in (56) for each year of the data. This gives:

$$h_{ct} - h_{c,t-1} = \beta\left\{\ln w_{ct} - \ln w_{c,t-1}\right\} - \beta \ln \rho(1+r_t) - \alpha\left\{X_{ct} - X_{c,t-1}\right\} + \zeta_{ct} + (\varepsilon_{ct} - \varepsilon_{c,t-1}) \quad (57)$$

Here, for instance ln w_{ct} is the mean of the log wage for people in cohort c, $c = 1,\ldots,8$ in year t, $t = 1970,....,76$. Notice that ζ_{ct} is the mean of the surprise shock to wealth for members of cohort c in year t. It is important to note that this may differ among cohorts because different cohorts are affected differently by aggregate shocks in period t. For example, an unexpected recession in year t may lead to larger unexpected wage reductions for younger workers.

Similarly, ε_{ct} is the mean of the taste shock for cohort c in year t. As we discussed earlier, writing the labour supply equation in terms of aggregate or cohort means highlights the potential existence of aggregate taste shocks. If aggregate taste shocks exist, they will alter equilibrium wages, and (57) will no longer represent a labour supply relationship. To deal with this problem we would need to find instruments that generate exogenous variation in wages (i.e. variation that is not induced by supply shocks). And, given the existence of aggregate surprise changes to lifetime wealth (captured by the ζ_{ct}) it is necessary that any instruments we use to predict wage growth from $t - 1$ to t be known at time $t - 1$.

Browning, Deaton and Irish (1985) use time dummies to pick up the aggregate shock, an approach that is feasible for them (in contrast to Angrist) because they observe multiple cohorts at each point in time. However, it should be noted that this does not address the possibility that aggregate shocks may differ by cohort, a point I'll return to below.

To estimate (57) Browning, Deaton and Irish (1985) use as instruments a quadratic in age along with lagged wages. They include number of children as observed taste shifters in X_{ct}. The wage measure is 'normal' weekly salary divided by normal weekly hours, and taxes are not accounted for. The main results, which they report in their Table 4/row 4.6, indicate that the Frisch elasticity is very small. The estimate of β in (57) is 0.13 (standard error = 0.27) and given this functional form the Frisch elasticity is roughly β/h which is 3.77/43 = 0.09 at the mean of the data, implying very little inter-temporal substitution in labour supply. Indeed, only the time dummies (and, marginally, children) are significant in the equation.

Based on this result, the authors argue that, 'there is a marked synchronisation over the life-cycle between hours worked and…wage rates...' but 'the characteristic hump-shaped patterns of…hours…though explicable in terms of life-cycle wage variation…can be explained as well as or better…as the response of credit-constrained consumers to the variation in needs accompanying the birth, growth and departure of children'. This quote from Browning, Deaton and Irish (1985) illustrates one of two possible reactions to a finding that the inter-temporal elasticity of substitution is very small.

One possibility is to maintain that the life-cycle model is valid but that preferences are such that people are not very willing to inter-temporally substitute hours (i.e. that $\gamma >> 0$). In this case, since the Frisch elasticity is an upper bound on the Hicks and Marshallian, we must conclude the other elasticities are small as well.

Alternatively, one could conclude, as do Browning, Deaton and Irish (1985), that consumers are credit constrained. In this case, the life-cycle model is invalid, and the static model of labour supply is in fact appropriate. Under these circumstances the Frisch elasticity is meaningless, and the estimates that imply it is small tell us nothing about possible values of the Hicks and Marshallian elasticities. Of course, if we abandon the life-cycle model we need some alternative explanation for the variation in assets over the life-cycle.[54]

Now let's consider further the issue of aggregate shocks. It is important to note that the presence of aggregate surprise variables such as $\bar{\zeta}_{it}$ or ζ_{ct} is not an issue only in studies like Angrist (1991) and Browning, Deaton and Irish (1985) which work with sample or cohort means. Taking means just makes the issue more salient. In fact, the same issue is implicitly present in the studies by MaCurdy (1981) and Altonji (1986) which used micro panel data to estimate versions of (31). The potential problems created by aggregate shocks for the estimation of (31) or (57) were stressed by Altug and Miller (1990). In particular, they argue that use of time dummies to soak up the mean of the aggregate shock in each period may not solve the problem.

Specifically, let $(\xi_{it} - \bar{\zeta}_{it})$ be the idiosyncratic surprise for household i at time t. Having included time dummies D_t in equation (31), it now takes on the form:

$$\ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}\left\{\ln w_{it}(1-\tau_t) - \ln w_{i,t-1}(1-\tau_{t-1})\right\} + D_t - \frac{\alpha}{\gamma}\left\{X_{it} - X_{i,t-1}\right\} + (\zeta_{it} - \bar{\zeta}_{it}) + (\varepsilon_{it} - \varepsilon_{i,t-1}) \quad (31')$$

where now the error term includes only the idiosyncratic surprise terms $(\xi_{it} - \bar{\zeta}_{it})$ along with the unobserved taste shifters. Despite the fact that $(\xi_{it} - \bar{\zeta}_{it})$ is mean zero by construction, it may be systematically related to instruments like age and education that are typically used to predict wage growth in this literature.

54 There is, of course, a huge parallel literature testing the life-cycle model by looking for evidence of liquidity constraints that prevent people from using assets to smooth consumption over the life-cycle (e.g. Keane & Runkle 1992). It is beyond the scope of this survey to discuss that literature, except to mention that whether liquidity constraints are important determinants of savings behaviour remains controversial.

For example, suppose the sample period contains an adverse productivity shock in year t, but that low education workers were much more adversely affected. This would cause low education workers to have relatively large values for ζ_{it} (recall that a positive value for ζ_{it} represents a surprise negative shock to lifetime wealth). Thus, letting S_i denote education, we have that $\text{Cov}[S_i\ ,\ (\zeta_{it} - \overline{\zeta}_{it})] < 0$. Now, this would not invalidate education as an instrument if the sample contained some other years where shocks to lifetime wealth tended to favour low education workers. However, the key point is that we would want our sample to consist of a fairly large number of years before we could be confident that such favourable and unfavourable shocks cancelled out.

Altug and Miller (1990) adopt a rather radical approach to this problem, which is to adopt assumptions that make it vanish. Specifically, they assume that workers have complete insurance against idiosyncratic shocks, so that the idiosyncratic shock terms $(\zeta_{it} - \overline{\zeta}_{it})$ vanish. Of course, all economic models are abstractions, so we should not dismiss a model simply because it contains some implausible assumptions. And, as Altug and Miller argue, given the existence of unemployment insurance, family transfers and so on, the existence of complete insurance, while obviously false, might not be such a terrible assumption. The real questions are: What does the assumption buy you? Does its falsity severely bias our estimates of parameters of interest?

Now, let's see what Altug and Miller (1990) gain from the complete insurance assumption. Let's return to equations (18)–(20), the first order conditions for the worker's optimisation problem, and extend them to a many-period setting. Rewrite (18)–(19) as:

(58) $$\beta_{it} h_{it}^{\gamma} = w_{it}\left[C_{it}\right]^{\eta} \qquad \eta \le 0$$

and rewrite (20) as:

(59) $$\left[C_{it}\right]^{\eta} = \rho(1+r_{t+1})[C_{i,t+1}]^{\eta} \qquad \eta \le 0$$

Recall that $[C_{it}]^{\eta}$ is the marginal utility of consumption at time t. Equation (59) describes how the marginal utility of consumption evolves over the life-cycle, given that the worker makes optimal consumption/savings decisions. For example, if $\rho(1 + r_t) = 1$, the worker will want to equate the marginal utility of consumption across all periods. Given the particular utility function in (1), this means equating consumption itself across all periods.[55] But if $\rho(1 + r_t) > 1$, meaning the rate of return on assets exceeds the discount factor, the consumer will choose to have $[C_{it}]^{\eta}$ fall over time—that is, he/she will tend to save early in order to have higher consumption later in life.

It is important to note that, given optimal behaviour, the marginal utility of consumption in period t is equivalent to what economists call the 'marginal utility of wealth' at time t. Let's call this λ_{it}. This is the increment in lifetime utility that the consumer can achieve if we give him/her an extra unit of assets (or wealth) at the start of period t. The equivalence $\lambda_{it} = [C_{it}]^{\eta}$ arises because, for a very small increment in wealth at time t, the consumer can't do significantly better than to simply spend it all at once.[56]

Now, (59) describes a situation of perfect foresight, where a consumer knows how λ_{it} will evolve over time.[57] That is a very strong assumption, as, in order to know how λ_{it} will evolve, the consumer must know how wages, interest rates and his/her own preference shocks will evolve over time. Taking logs of (59), and grouping terms, we have:

(60) $$\ln[C_{i,t+1}]^{\eta} - \ln\left[C_{it}\right]^{\eta} = -\ln \rho(1+r_{t+1})$$

Thus, with perfect foresight, the marginal utility of consumption at time t evolves in a known way with the interest rate and the discount factor. As MaCurdy (1981) pointed out, this makes estimation of equation (58) possible without consumption data. Taking logs of (58) we have:

(61) $$\ln h_{it} = \frac{1}{\gamma}\ln w_{it} - \frac{1}{\gamma}\ln \beta_{it} + \frac{1}{\gamma}\ln\left[C_{it}\right]^{\eta}$$

Now, by first differencing (61) and using (60) we can make the marginal utility of consumption terms vanish:

$$\begin{aligned}\ln h_{it} - \ln h_{i,t-1} &= \frac{1}{\gamma}(\ln w_{it} - \ln w_{i,t-1}) - \frac{1}{\gamma}(\ln \beta_{it} - \ln \beta_{i,t-1}) \\ &\quad + \frac{1}{\gamma}\Big[\ln\left[C_{it}\right]^{\eta} - \ln\left[C_{i,t-1}\right]^{\eta}\Big] \\ &= \frac{1}{\gamma}(\ln w_{it} - \ln w_{i,t-1}) - \frac{1}{\gamma}(\ln \beta_{it} - \ln \beta_{i,t-1}) \\ &\quad - \frac{1}{\gamma}\left[\ln \rho(1+r_t)\right]\end{aligned}$$

And, if we assume that the taste shifters β_{it} have a stochastic component, as in $\beta_{it} = \exp(X_{it}\alpha + \varepsilon_{it})$ we obtain the estimable equation:

55 For more general utility functions this consumption smoothing result does not follow. For example, suppose that we generalise (1) to have:

$$U_t = G\left[\frac{C_t^{1+\eta}}{1+\eta} - \beta_t \frac{h_t^{1+\gamma}}{1+\gamma}\right]$$

where $G(\ )$ is a concave function. Then the marginal utility of consumption is, $G_t'(\cdot)C_t^{\eta}$, and we have that $\lambda_t = G_t'(\cdot)C_t^{\eta}$. Now if $\rho(1 + r_t) = 1$ it is that $G_t'(\cdot)C_t^{\eta}$ the consumer seeks to equate across periods. Notice that, for given C_t, the derivative $G_t'(\cdot)$ is increasing in h_t, so that, if consumption were equalised across periods the marginal utility of consumption would be higher in periods when hours of work are higher. This will cause the consumer to allocate more consumption to periods when h_t is relatively high. I discuss this issue more below.

56 This is a simple application of the 'envelop theorem'. If we give the consumer a very small increment of assets at the start of period t, then he/she can't do significantly better than to consume it all at once. Any incremental gain in lifetime utility that he/she might achieve by optimally allocating tiny increases in consumption over all remaining periods of the life, so as to satisfy (59), would be trivially small.

57 Notice that (59) implies:

$$[C_{t+1}]^{\eta} = \frac{1}{\rho(1+r_t)}\left[C_t\right]^{\eta} = \left[\frac{1}{\rho(1+r_t)}\right]\left[\frac{1}{\rho(1+r_{t-1})}\right]\left[C_{t-1}\right]^{\eta} = \cdots = \frac{1}{\rho^t}\left[\prod_{k=1}^{t}\frac{1}{(1+r_{t-k})}\right]\left[C_1\right]^{\eta}$$ or, in logs: $$\ln\left[C_{t+1}\right]^{\eta} = -t\ln\rho + \sum_{k=1}^{t}\ln(1+r_{t-k}) + \ln\left[C_1\right]^{\eta}$$

Thus, the marginal utility of consumption at any time t is simply a function only of the marginal utility of consumption in the first period, along with the discount rate and interest rates in subsequent periods, all of which the consumer is assumed to know at $t = 1$.

(62) $$\ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}(\ln w_{it} - \ln w_{i,t-1}) - \frac{\alpha}{\gamma}(X_{it} - X_{i,t-1}) - \frac{1}{\gamma}\left[\ln \rho(1+r_t)\right] + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

Given the perfect foresight assumption, this equation can actually be consistently estimated using OLS, provided the unobserved taste shocks are uncorrelated with wages. However, as MaCurdy (1981) pointed out, as a practical matter we would still want to use instrumental variables to deal with measurement error in wages.

Now, consider a situation where, rather than having perfect foresight, the worker is uncertain about future wage realisations. It is simple to introduce such uncertainty into the life-cycle model of (58)–(59) just by modifying (59) to read:

(63) $$[C_{it}]^{\eta} = E_t \rho(1+r_{t+1})[C_{it+1}]^{\eta} \qquad \eta \leq 0$$

Here E_t denotes the worker's expectation of his/her future state, given all the information he/she has available at time t. Specifically, the worker is forecasting what his/her consumption will be at time $t+1$, and this in turn depends on what his/her wage rate realisation will be.

Now, we can rewrite (63) as:

(64) $$\rho(1+r_{t+1})[C_{i,t+1}]^{\eta} = [C_{it}]^{\eta}(1+\xi_{i,t+1})$$

where $\xi_{i,t+1}$ is a forecast error that is independent of information known at time t. Taking logs of (63) we obtain:

(65) $$\ln[C_{t+1}]^{\eta} - \ln[C_t]^{\eta} = -\ln \rho(1+r_{t+1}) + \ln(1+\xi_{i,t+1}) \approx -\ln \rho(1+r_{t+1}) + \xi_{i,t+1}$$

In the second line I've made use of the approximation that $\ln(1+\xi_{it}) \approx \xi_{it}$ as long as ξ_{it} is not too large.[58]

At this point it appears as if we can introduce uncertainty into the model with little added complication. If we proceed as before, first differencing (61) but simply using (65) in place of (60), we obtain:

(66) $$\ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}(\ln w_{it} - \ln w_{i,t-1}) - \frac{\alpha}{\gamma}(X_{it} - X_{i,t-1}) - \frac{1}{\gamma}\left[\ln \rho(1+r_t)\right] + \frac{1}{\gamma}\xi_{it} + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

This is precisely our familiar equation (31) that we derived informally earlier, except that in (31) we defined $\zeta_{it} = \xi_{it}/\gamma$ and we introduced after-tax wages. Now, given (65), we can see quite clearly what we previously stated informally: ξ_{it} represents a surprise increase in the marginal utility of consumption from $t-1$ to t, or, equivalently in this case, a surprise decrease in consumption. This would be induced by a surprise wage reduction, which makes the person feel less wealthy than expected.

Now, the point made by MaCurdy (1981) and Altonji (1986) is that estimation of (66), the labour supply model that accounts for uncertainty, is hardly any more difficult than estimation of (62), the model that assumes perfect foresight. Estimation of (62) already made use of instrumental variables to deal with the measurement error in wages. The only complication apparent in estimating (66) is that we must also make sure the instruments are uncorrelated with the forecast errors ζ_{it}.

But as Altug and Miller (1990) argue, things are not quite so simple. If aggregate shocks are present, so that the ζ_{it} do not have mean zero within each period, this violates the assumptions that allow us to estimate the Frisch elasticity ($1/\gamma$) by applying instrumental variables to (66). And the obvious solution of using time dummies to 'sop up' the period-specific means does not necessarily work if there is idiosyncratic uncertainty.

Altug and Miller (1990) deal with the problem by assuming away idiosyncratic risk. Specifically, they write that:

(67) $$\lambda_{it} = \eta_i \lambda_t$$

A person i with a low η_i has a relatively low marginal utility of wealth, meaning he/she is relatively rich. But a person's position in the wealth distribution stays constant over time. Aside from interest rates and discounting, the only source of variation (and hence uncertainty) in the marginal utility of wealth over time are aggregate shocks that cause movements in λ_t.

Given this assumption, Altug and Miller can rewrite (63) as:

(68) $$\lambda_{it} = E_t \rho(1+r_{t+1})\lambda_{i,t+1} \Rightarrow \eta_i \lambda_t = \rho \eta_i E_t(1+r_{t+1})\lambda_{t+1} \Rightarrow \lambda_t = \rho E_t(1+r_{t+1})\lambda_{t+1}$$

Then (64) and (65) become:

(69) $$\rho(1+r_{t+1})\lambda_{t+1} = \lambda_t(1+\xi_{t+1}) \Rightarrow \ln \lambda_{t+1} - \ln \lambda_t \approx -\ln \rho(1+r_{t+1}) + \xi_{t+1}$$

Notice also that, using (67), equation (61) becomes:

(70) $$\ln h_{it} = \frac{1}{\gamma}\ln w_{it} - \frac{1}{\gamma}\ln \beta_{it} + \frac{1}{\gamma}\ln \eta_i + \frac{1}{\gamma}\ln \lambda_t$$

So first differencing (61) and using (69) in place of (65), we obtain:

(71) $$\ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}(\ln w_{it} - \ln w_{i,t-1}) - \frac{\alpha}{\gamma}(X_{it} - X_{i,t-1}) - \frac{1}{\gamma}\left[\ln \rho(1+r_t)\right] + \frac{1}{\gamma}\xi_t + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

Compared to (66), this equation has the almost imperceptible difference that the surprise term $\zeta_t = \xi_t/\gamma$ no longer has an i subscript, so it really is just an aggregate shock, and it can be appropriately captured with time dummies.

But Altug and Miller (1990) do not make this point simply as a critique of other work (or at least its interpretation). They note that if we adopt the assumption (67) then we can first difference (70) to obtain:

(72) $$\ln h_{it} - \ln h_{i,t-1} = \frac{1}{\gamma}(\ln w_{it} - \ln w_{i,t-1}) - \frac{\alpha}{\gamma}(X_{it} - X_{i,t-1}) - \frac{1}{\gamma}\left[\ln \lambda_t - \ln \lambda_{t-1}\right] + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

That means we can actually estimate the changes in $\ln \lambda_t$ as the dummy coefficients in estimating equation (72). That in turn means that, given data on interest rates r_t we can actually estimate the asset pricing equation (69), with the only unknown parameter being ρ.

58 A more formal derivation would define the mean zero error term $\zeta_{it} = \ln(1+\xi_{it}) - E_t\ln(1+\xi_{it})$. The derivation in the text would then go through except that the mean $E_t\ln(1+\xi_{it})$ would show up as part of the intercept of hours equation.

So the main point of the Altug and Miller (1990) paper is to use data on hours, wages, consumption and rates of return to *jointly* estimate (i) a within-period optimality condition like (48), (ii) a first difference hours equation like (72), and (iii) an asset price equation like (69), using the cross equation restrictions among the equations (e.g. γ appears in multiple places) to get a more efficient estimate of the Frisch elasticity.[59] They estimate their model on a sample of married men from the PSID. To be in the sample the men had to be continuously married from 1967–1980, and be no older than 46 in 1967.

A complication is that Altug and Miller (1990) do not use the simple utility function (1) that was used by MaCurdy (1981) and Altonji (1986). They use a more complex function where the wives' leisure is allowed to be non-separable with consumption and husbands' leisure. For instance, the within-period optimality condition gives the following demand for husband leisure equation that is the analogue to (48):

$$(73)\quad \ln l_{it}^h = \frac{1}{\tilde{\gamma}}\ln w_{it} + \frac{\eta}{\tilde{\gamma}}\ln C_{it} - \frac{\alpha}{\tilde{\gamma}}X_{it} + \frac{\pi}{\tilde{\gamma}}\ln l_{it}^s + \frac{\varepsilon_{it}}{\tilde{\gamma}}$$

where now $(1/\tilde{\gamma})$ is the inter-temporal elasticity of substitution in leisure and l_{it}^s is the wives' lesiure. Estimating (73) jointly with the rest of the system, the authors obtain a Frisch elasticity of leisure with respect to the wage of 0.037 with a standard error of 0.013. This precise estimate contrasts with an estimate of 0.018 with a standard error of 0.087 that they obtain when they do not include the first difference hours equation and the asset equation in the system. Thus we see that their approach does lead to a substantial efficiency gain.

Given that leisure is normalised to a fraction of total time, we have the Frisch elasticity of male labour supply with respect to the wage implied by the authors' estimate as:

$$\frac{\partial \ln h}{\partial \ln w} = \frac{w}{h}\frac{\partial h}{\partial w} = \frac{w}{h}\frac{\partial(1-l)}{\partial w} = -\frac{w}{h}\frac{\partial l}{\partial w} = -\frac{w}{h}\frac{l}{w}\left[\frac{w}{l}\frac{\partial l}{\partial w}\right]$$
$$= -\frac{l}{h}(-0.037) \approx \frac{8760}{2300}(0.037) = 0.14$$

Thus, despite the different methodology, the estimate is similar to the rather small values obtained by MaCurdy (1981) and Altonji (1986). To conclude the discussion of Altug and Miller, it is worth pointing out some limitations of the study. One is that it does not incorporate taxes and another is that it uses a ratio wage measure.

Finally, one odd aspect of the Altug and Miller (1990) results is that the coefficient on consumption in (73) is 0.003, which implies that $\eta = 0.08$ (i.e. the coefficient on consumption in the utility function is $1 + \eta = 1.08$). This violates the theoretical restriction that $\eta < 0$ (i.e. diminishing marginal utility of consumption). However, the coefficient is so imprecisely estimated that one can't reject that utility is linear in consumption ($\eta = 0$). On the other hand, log utility ($\eta = -1$) is rejected. So all that can be discerned is that $-1 < \eta < 0$, which is essentially the entire plausible range for the parameter.[60]

Now let's consider further developments in the line of work that adopts the two-stage budgeting approach. Blomquist (1985) noted that there was problem in applying the two-stage budgeting approach of MaCurdy (1983)—Method 2—and Blundell and Walker (1986) in contexts with progressive taxation. The basic idea is that an increase in hours of work in period t, holding consumption fixed, will cause the person to have more assets at the end of period t. This, in turn, will lead to higher asset income in period $t+1$. And this in turn may increase the person's tax bracket at $t+1$. More generally, a worker's decisions at time t may affect the tax rates that he/she faces at time $t+1$. This means we no longer achieve the simplification that, *conditional* on the full income allocated to time t, we can model the person's time t decisions as if he/she were choosing labour supply subject to a one-period budget constraint.

The paper by Ziliak and Kniesner (1999) attempts to deal with the problem of progressive taxation within the two-stage budgeting framework (specifically, MaCurdy's Method 2). Following Blomquist (1985) they note that the two-stage budgeting approach can be salvaged in a world of progressive taxation by writing labour supply in period t as conditional on assets at both the start and end of the period. The idea is that, by holding end-of-period assets fixed, you shut down any channel by which increased labour supply in period t might affect the budget constraint in period $t+1$ (or later). Thus, they estimate a labour supply equation of the form:

$$(74)\quad h_{it} = \beta + \beta_w w_{it}(1-\tau_{it}) + \beta_{A1}A_{i,t-1}^{*} + \beta_{A2}A_{i,t} + X_{it}\alpha + \mu_i + \varepsilon_{it}$$

In this equation A_{t-1}^{*} is 'virtual wealth', which plays a role analogous to virtual non-labour income in static piecewise-linear budget constraint models (see Figure 7.1). It is defined as:

$$(75)\quad A_{t-1}^{*} = A_{t-1} + \frac{(\tau_{it} - \tau_{it}^{A})I_{it}}{r_t}$$

where τ_{it}^{A} is the average tax rate paid by person i in period t, and r_t is the risk-free rate of interest.

Ziliak and Kniesner (1999) estimate equation (74) using data on 532 married men from the PSID who are 22–51 years old in 1978 and who are observed to work in every year from 1978–1987. The asset measure is home equity plus the capitalised value of rent, interest and dividend income. In constructing the wage measure, Ziliak and Kniesner seek to avoid the denominator bias problem by using hourly wage rates for hourly workers. For workers paid weekly, they divided weekly earnings by 40 hours rather than actually observed hours (and so on for workers paid over other time periods). This procedure avoids denominator bias, at the cost

59 For good measure they throw in a wage equation as well. There are no cross equation restrictions between this and the other three equations, but allowing for the error covariance increases efficiency.

60 The greater imprecision of the η estimate compared to prior studies may stem from attempting to estimate the extent of non-separability between female non-market time and both consumption and male labour supply, which adds female non-market time as an additional regression in the labour supply equations. Altug and Miller (1990) reject the joint hypothesis that female non-market time is separable from both consumption and male labour supply, but their estimates are too imprecise to determine from which quantity it is non-separable.

of introducing a different type of measurement error in hours. In forming taxable income and marginal tax rates, the authors use information in the PSID to estimate standard and itemised deductions.[61] Observed taste shifters included in X_{it} are age, health and number of children.

Equation (74) contains three endogenous variables: the after-tax wage, end-of-period assets, and start-of-period virtual assets. All three variables may be correlated with the individual fixed effect μ_i (i.e. a person with high tastes for work will tend to have both a high wage and high asset levels). Thus, as a first step towards estimating (74), the authors take first differences to eliminate the individual fixed effect μ_i:

$$(76)\quad h_{it}-h_{i,t-1}=\beta_w[w_{it}(1-\tau_{it})-w_{i,t-1}(1-\tau_{i,t-1})] + \beta_{A1}[A^*_{i,t-1}-A^*_{i,t-2}]+\beta_{A2}[A_{it}-A_{i,t-1}] + [X_{it}-X_{i,t-1}]\alpha+(\varepsilon_{it}-\varepsilon_{i,t-1})$$

Now, $w_{it}(1-\tau_{it})$ and A_{it} are presumably correlated with ε_{it} as a high taste for work in period t will tend to both (i) shift a person into a higher tax bracket and (ii) lead to higher assets at the end of the period. Furthermore, even start-of-period virtual wealth A^*_{t-1} is presumably correlated with ε_{it}. If a high ε_{it} tends to shift a worker into a higher tax bracket at time t, then it affects A^*_{t-1} directly as shown in (75). Now, valid instruments for estimation of (76) must be uncorrelated with both ε_{it} and ε_{it-1}. For this reason, Ziliak and Kniesner (1999) argue that one must lag the wage and asset variables by two periods (i.e. $w_{it-2}(1-\tau_{it-2})$, A_{it-2} and A^*_{t-3}) so as to obtain valid instruments that are uncorrelated with ε_{it-1}. They also include a quadratic in age, age interacted with education, and home ownership as additional instruments.

The main estimation results imply a Marshallian elasticity evaluated at the mean of the data of $(w/h)\partial h/\partial w =$ (10.19/2179)(24.66) = 0.1153, and a very small income effect of $w_t\partial h_t/\partial A_{t-1}$ = (10.19)(–0.00162) = –0.0165. Thus, the Hicks elasticity is 0.1318. In a second stage, which I will not describe in detail, they estimate the Frisch elasticity as 0.163.[62]

Ziliak and Kniesner (1999) go on to use their model to simulate the impact of various tax reform experiments. The average marginal tax rate in their data was 29 per cent. One experiment simulates an across the board 10 per cent rate cut by the United States in 1987. In the authors' simulation, this would only increase average annual hours for prime age married men by 13 hours (or 0.6 per cent).

This small effect on hours is not too surprising given the Marshallian elasticity of 0.12. The authors also simulate the effect of the 1986 tax reform that substantially reduced the progressivity of the tax system. As we've seen, it is the Hicks elasticity, which they estimate to be 0.13, that is relevant for determining the welfare effects of changing progressivity. They simulate only a 2 per cent hours increase, but a substantial welfare gain from these changes.

It is interesting to compare this result to those in the papers by Blomquist (1983) and Blomquist and Hansson-Busewitz (1990) which I discussed earlier. These papers found Hicks elasticities of 0.11 and 0.13 respectively, but they both found large welfare gains from switching to a flat rate tax. So all three of these papers are similar in finding that a fairly modest value of the Hicks elasticity can imply substantial welfare gains from reducing progressivity.

At this point I'd like to discuss the issue of non-separablility between leisure and consumption, which up until now I have largely ignored. To further explore the implications of non-separability, suppose we modify the utility function in equation (1) to read:

$$(77)\quad U_t = G\left[\frac{C_t^{1+\eta}}{1+\eta}-\beta_t\frac{h_t^{1+\gamma}}{1+\gamma}\right] \qquad \eta\le 0,\ \gamma\ge 0$$

Let's assume that $G[\cdot]$ is a concave function, such as $G[X] = \log(X)$ or $G[X] = (1+\sigma)^{-1}X^{1+\sigma}$ for $\sigma \le 0$. Notice that now the marginal utility of consumption is given by:

$$(78)\quad \frac{\partial U_t}{\partial C_t} = G'_t(X_t)\cdot C_t^{\eta} = \lambda_t \qquad \text{where} \quad X_t \equiv \frac{C_t^{1+\eta}}{1+\eta}-\beta_t\frac{h_t^{1+\gamma}}{1+\gamma}$$

Thus, unlike in (1), the marginal utility of consumption is no longer a function only of consumption itself. It also depends on X_t which is a composite of consumption and hours of work. Notice that, for a given level of consumption, X_t is a decreasing function of h_t. Given our assumption that G is concave, this means $G'_t(X_t)$ is increasing in h_t. Thus, if the consumer were (naively) to equate consumption across periods, he/she would have a higher marginal utility of consumption in periods when hours of work are higher. Thus, *ceteris paribus*, the consumer would like to allocate *more* consumption to periods when hours are higher. Thus, a concave G function generates a situation where hours and consumption are complements.

The consumer still seeks to satisfy an inter-temporal optimality condition like (63) but, with our new expression for the marginal utility of consumption, we must revise (63) to obtain:

$$(79)\quad G'_t(X_{it})[C_{it}]^{\eta} = E_t\rho(1+r_{t+1})\left\{G'_t(X_{i,t+1})[C_{i,t+1}]^{\eta}\right\} \quad \eta\le 0$$

Notice that now, even if $\rho(1+r_t)=1$, the consumer will not seek to equalise consumption across periods. As indicated above, with a concave G function he/she will seek to make consumption higher when hours are higher. But of course, in the life-cycle model, hours are high when wages are high. Hence, the consumer will seek to make consumption high when earnings are high. Thus, if G is sufficiently concave, the life-cycle model can generate consumption and earnings paths that look very much like liquidity constrained behaviour!

61 Basically, they use IRS data to calculate the average level of itemised deductions for a person's income level. Beginning in 1984, the PSID asks whether or not a person itemised, so the authors can assign them either the standard deduction of the itemised deduction. Prior to 1984, the authors assign either the standard or (estimated) itemised deduction, whichever is larger.

62 The authors demonstrate that if they use a ratio wage measure (annual earnings over annual hours) and apply exactly the same estimation procedure, they obtain a Marshallian elasticity of –0.083 and a Hicks elasticity of –0.072. This highlights the severe bias created by use of ratio wage measures which I discussed earlier.

Now, consider what the MRS condition in (58) will look like, given the new utility function in (77):

$$(80)\quad \frac{\partial V}{\partial h_t} = \left\{G'(X_t)C_t^{\eta}\right\} w_t(1-\tau_t) - \left\{G'(X_t)\beta_t h_t^{\gamma}\right\} = 0$$
$$\Rightarrow \quad C_t^{\eta} w_t(1-\tau_t) = \beta_t h_t^{\gamma}$$

That is, it doesn't change at all. The factor $G_t'(X_t)$ appears in both the marginal utility of consumption and the marginal utility of leisure, and so it cancels out. This point was made by MaCurdy (1983): the G function does not affect within-period decisions about work and consumption, and so estimation of the MRS condition does not tell us anything about the form of G.

This is why MaCurdy (1983), in his Method 1, proposed estimating the form of G in a second stage. The first stage (discussed earlier) uses the MRS condition to obtain estimates of the parameters of the X_t function, which are γ and η in our example (77). One can then use these estimates, along with a person's actual hours and consumption data, to construct estimates of the X_t. One then treats these estimates of the X_t as data. In the second stage, one uses data on X_t, C_t and r_t from multiple periods to estimate the unknown parameters of (79). These include the discount rate ρ and the parameters of G. For instance, if we assume that $G[X] = (1+\sigma)^{-1}X^{1+\sigma}$ then the only parameter of G is σ. In his study, MaCurdy actually estimated $\sigma = -0.14$ but with a standard error of 0.23. Thus, he couldn't reject the simple linear G case ($\sigma = 0$).

Now, let's return to the first order conditions for consumption and hours in (18)–(19), which will now take the form:

$$(81)\quad \frac{\partial V}{\partial h_t} = G'(X_t)\left[C_t\right]^{\eta} w_t(1-\tau_t) - G'(X_t)\beta_t h_t^{\gamma} = 0$$
$$\Rightarrow \quad \lambda_t w_t(1-\tau_t) = G'(X_t)\beta_t h_t^{\gamma}$$

Thus, we now have that:

$$(82)\quad \ln h_t = \frac{1}{\gamma}\left\{\ln w_t + \ln(1-\tau_t) + \ln\lambda_t - \ln G'(X_t) - \ln\beta_t\right\}$$

So the Frisch elasticity, which is the effect of a change in the wage holding the marginal utility of consumption λ_t fixed, is no longer simply $(1/\gamma)$, because in general a change in w_t will affect $G'(X_t)$.

To explore further how a concave G affects willingness to substitute labour across periods, let's assume $G[X] = (1+\sigma)^{-1}X^{1+\sigma}$ so that (82) becomes:

$$(82')\quad \ln h_t = \frac{1}{\gamma}\left\{\ln w_t + \ln(1-\tau_t) + \ln\lambda_t - \sigma\ln X_t - \ln\beta_t\right\}$$

Clearly, the elasticity of hours with respect to the wage rate holding λ_t fixed—the Frisch elasticity—is not simply $(1/\gamma)$ in this case, because we have to worry about the $\ln X_t$ term, and X_t contains C_t and h_t. The exception, of course, is if $\sigma = 0$ so the $\ln X_t$ term drops out.

To determine what (82′) implies about the Frisch elasticity, we can use the within-period MRS condition in (80) to substitute out for consumption in X_t (see equation (78)) obtaining an expression for X_t solely in terms of hours. Then (82′) becomes an implicit equation that relates hours and the wage, holding λ_t fixed. Implicitly differentiating this equation one obtains:

$$(83)\quad e_F = \left.\frac{\partial \ln h_t}{\partial \ln w_t}\right|_{\lambda_t\ fixed} = \frac{1}{\gamma}\left\{\frac{X_t + (\sigma/\eta)C_t^{1+\eta}}{X_t + (\sigma/\eta)C_t^{1+\eta} - \sigma(\beta/\gamma)h_t^{1+\gamma}}\right\}$$

First note that if $\sigma = 0$ then the (83) reduces to just $(1/\gamma)$ as we would expect. However, as $\sigma \to -\infty$ the fraction in curly brackets becomes less than one. We can see this because the extra term $-\sigma(\beta/\gamma)h_t^{1+\gamma}$ in the denominator is positive. Thus, greater complementarity between hours and consumption reduces the Frisch elasticity to less than $(1/\gamma)$.

Numerical simulations of the simple two-period model in (77)–(82) reveal a lot about how σ influences behaviour, and give a clear intuition for why the Frisch elasticity falls as $\sigma \to -\infty$. I start from a base case where the wage is 100 in both periods, hours are 100 in both periods, the tax rate is 40 per cent and consumption is 6,000 in both periods (as I set $\rho(1+r_1) = 1$). In a two-period model where each period corresponds to roughly twenty years of a working life, a plausible value for $1+r$ is about $(1+0.03)^{20} \approx 1.806$, or $\rho = (1+r)^{-1} \approx 0.554$. The utility function parameters are set to $\gamma = 0.5$ and $\eta = -0.5$

Then, from (24), the Marshallian elasticity, which indicates how hours respond to a permanent (i.e. two-period) wage change, is 0.5, while the Hicks elasticity which also indicates how hours respond to a permanent (i.e. two-period) wage change, but in this case compensated by a lump sum transfer that keeps utility fixed, is 1.0. Importantly, these elasticities are invariant to σ. The Frisch elasticity (i.e. how hours respond to an anticipated one-period wage change) is equal to $(1/\gamma) = 2.0$ if $\sigma = 0$. But this elasticity varies with σ.

In Table 7.4, I simulate the effect of a 1 per cent after-tax wage increase in period 1 (from 60 to 60.6) induced by cutting the tax rate from 0.40 to 0.394. Results are shown for values of σ ranging from zero to –40. Notice that when $\sigma = 0$ the worker increases work hours in period 1 by 1.03 per cent and reduces work hours in period 2 by 0.96 per cent, leading to the expected 2 per cent increase in labour supply in period 1 relative to period 2 implied by the Frisch

Table 7.4 How Frisch Elasticity Varies with Willingness to Substitute Utility over Time

		Changes in hours (%)		Changes in consumption (%)		Changes in utility (%)	
σ	Frisch elasticity	Hours(1)	Hours(2)	C(1)	C(2)	G(X(1))	G(X(2))
0.0	2.00	+1.03	–0.96	+0.97	+0.97	–0.05	+1.44
–0.5	1.40	+0.82	–0.58	+1.18	+0.58	+0.27	+0.87
–1.0	1.25	+0.76	–0.48	+1.24	+0.48	+0.38	+0.72
–2.0	1.14	+0.73	–0.41	+1.27	+0.42	+0.41	+0.62
–5.0	1.06	+0.70	–0.36	+1.30	+0.36	+0.45	+0.54
–10.0	1.03	+0.69	–0.34	+1.31	+0.34	+0.46	+0.51
–40.0	1.01	+0.68	–0.33	+1.32	+0.33	+0.48	+0.49

elasticity of 2.0. Note also that the consumer continues to smooth his/her consumption across the two periods, as consumption is increased by 0.97 per cent in each period.[63] The consequence of this is that utility actually falls slightly in period 1 while rising by 1.44 per cent in period 2 (when the consumer gets to consume more and work less than under the baseline).

As σ increases, the consumer is less willing to sacrifice utility in period 1 in order to achieve higher utility in period 2. By the time we get to σ = –40 the consumer is almost completely unwilling to substitute utility across periods. Notice he/she allocates consumption and hours so that utility increases by 0.48 per cent in period 1 and 0.49 per cent period 2. To achieve this, the worker shifts consumption into period 1 to compensate him or herself for having to work more hours in that period (i.e. consumption increases by 1.32 per cent in period 1 versus only 0.33 per cent in period 2). Simultaneously, the worker shifts less labour supply towards period 1 than in the $\sigma = 0$ case. Now, hours only increase by 0.68 per cent in period 1 and only fall by 0.33 per cent in period 2. This implies a Frisch elasticity of 1.01. If we send σ all the way to $-\infty$ we end up with a Leontieff utility function where the consumer only cares about maximising the minimum utility in any period. As $\sigma \to -\infty$ the Frisch elasticity approaches 1.0. Interestingly, this is exactly the Hicks elasticity for a permanent (two-period) wage change.[64]

In summary, in the linear case of $G(X) = X$, combined with a within-period utility function that is additive between consumption and hours, there is a separation of the labour supply and consumption problems. The worker chooses savings to smooth consumption across periods. The worker also chooses to work more hours in those periods when wages are higher. This means sacrificing utility in the high wage periods.

But if G is concave, the worker/consumer tries to equalise utility across periods. This tends to reduce inter-temporal substitution in labour supply. But inter-temporal substitution still takes place. However, the worker tries to smooth utility by shifting consumption into the high wage/high hours periods. As $\sigma \to -\infty$ the consumer insists on equal utility in all periods. Thus, any increase in hours worked during a high wage period must be fully compensated by an increase in consumption. Then, the consumer's willingness to substitute consumption across periods puts a damper on his/her willingness to substitute hours. As a result, the Frisch elasticity is less than $(1/\gamma)$ and in the limit it equals the Hicks elasticity.

It is worth recalling that the within-period conditions (45) or (48) still hold regardless of the form of G. Thus, one might estimate such an equation and uncover $(1/\gamma)$, but fail to realise that this is not the Frisch elasticity. However, one can still use such equations to obtain the Hicks and Marshallian elasticities.

Returning to the empirical literature, Ziliak and Kniesner (2005) also allow for non-separablity between leisure and consumption, and adopt MaCurdy's estimation Method 1. However, they introduce non-separablity not only via the G function but also by allowing for an interaction between leisure and consumption in the within-period utility function X_t. Specifically, they adopt a translog within-period utility function:

$$\text{(84)} \quad U_t = G\Big[\alpha_1 \ln(\bar{L} - h_t) + \alpha_2 \ln C_t - \alpha_3 \ln(\bar{L} - h_t) \ln C_t - \alpha_4 [\ln(\bar{L} - h_t)]^2 - [\ln C_t]^2\Big]$$

with $G[X] = (1 + \sigma)^{-1} X^{1+\sigma}$. If $\alpha_3 > 0$ then hours and consumption are complements in the within-period utility function. This appears to be the only paper that allows for within-period non-separability while also incorporating taxes in a dynamic framework.

As in most of the US-based work that estimates versions of the within-period MRS condition, Ziliak and Kniesner (2005) use the PSID, which only contains a measure of food consumption.[65] However, they try to improve on this by using a method proposed by Blundell, Pistaferri and Preston (2001) to impute nondurable consumption in the PSID. Essentially, they use the Consumer Expenditure Survey (CES), which has much more complete information on consumption, to develop an equation for predicting total nondurable consumption based on food consumption, food prices and demographics. They also try a method proposed by Skinner (1987) that predicts total consumption based on food consumption, house value and rent.

Ziliak and Kniesner use the 1980–1999 waves of the PSID, which have data on the years 1979–1998. One advantage over prior work is the long sample period, which encompasses five significant tax law changes.[66] This provides

63 Earnings increase by about 2 per cent in period 1 and drop by about 1 per cent in period 2. This causes the present value of lifetime earnings (and hence of consumption) to increase by [1.02 + (0.554)(0.99)]/1.554 ≈ 1.0097 or 0.97 per cent.

64 Notice that if we take the limit of (83) as $\sigma \to -\infty$ we get that:

$$e_F \to \frac{1}{\gamma}\left\{\frac{(\sigma/\eta)C_t^{1+\eta}}{(\sigma/\eta)C_t^{1+\eta} - \sigma(\beta/\gamma)h_t^{1+\gamma}}\right\} = \frac{1}{\gamma}\left\{\frac{C_t^{1+\eta}}{C_t^{1+\eta} - \eta(\beta/\gamma)h_t^{1+\gamma}}\right\}$$

For the particular parameter values in this simulation, the term in curly brackets is equal to 0.5. Notice that if $\eta = 0$, meaning there are no income effects (utility linear in consumption) then the second term in the denominator vanishes, and the term in curly brackets is exactly equal to 1. Thus, with utility linear in consumption, the value of σ has no impact on the Frisch elasticity. What happens in this case is that, for $\sigma < 0$, if the wage increases in period t the consumer fully compensates him or herself by shifting more consumption into period t, holding utility exactly equal across periods. Unlike the case of $\eta < 0$, there is no 'compromise' solution where the consumer both shifts more consumption into period t while also damping the hours increase in period t. Hence, curvature in G does not dampen the Frisch substitution effect in this case. Ironically, a high degree of substitutability in consumption combined with curvature in G makes the consumer behave in a way that looks a lot like liquidity constrained behaviour (i.e. consumption closely tracks income). Finally note that when $\eta = 0$ the Frisch $(1/\gamma)$ and Hicks $1/(\gamma - \eta)$ elasticities are always exactly equal regardless of the value of σ. Again, this is because the consumer is willing to make all utility equalisations across periods via consumption shifting, leaving him/her free to substitute hours of work towards high wage periods as much as desired.

65 Recall that the Denver data used by MaCurdy (1983) had a very comprehensive consumption measure.

66 The Economic Recovery Tax Act 1981, the Tax Reform Act 1986, the Omnibus Reconciliation Tax Act 1990 (and also1993), and the Taxpayer Relief Act 1997.

more variation in the budget constraint to help identify utility function parameters. The sample includes 3402 male household heads who were at least 25 in 1980, no older than 60 in 1999 and who are observed for at least five years. The authors use the hourly wage rate question for workers paid by the hour, and, in an effort to reduce denominator bias, for salaried workers they use the same procedure of hours bracketing as in their 1999 paper discussed earlier.

Of course, a challenge in incorporating taxes is to estimate taxable income. The authors assume that all married men filed joint returns and that all unmarried men filed as heads of households (the latter being the more likely source of error). The income of working wives is included when calculating adjusted gross income (AGI). To estimate deductions, the authors use IRS estimates of the average levels of itemised deductions by AGI. From 1984 onwards the PSID reports whether a person itemised or took the standard deduction. Following MaCurdy, Green and Paarsch (1990), the authors use a smooth approximation to the piecewise-linear tax schedule.

The parameters α_1 and α_2 in equation (84) are allowed to depend on children, race, and age of youngest child, to capture how these demographic variables may shift tastes for work and consumption. Besides these, the instruments used to estimate the MRS equation, which should be correlated with after-tax wages and consumption but uncorrelated with unobserved tastes for work, are age and education, health, home ownership, and industry, occupation and region dummies.

The authors estimate that $\alpha_3 > 0$, which implies that hours and consumption are complements in the within-period utility function. That is, if work hours are higher then, *ceteris paribus*, the marginal utility of consumption is higher. Ziliak and Kniesner (2005) let σ vary with age, and estimate $\sigma = 0.844 - 0.039 \cdot Age$. This means that σ is roughly zero at age 20 and falls to –1.496 at age 60.[67] However, the age effect is imprecisely estimated.

Given the translog within-period utility function in (84) there is no closed form for the Marshallian and Hicks elasticities. At the mean values in the data, the authors calculate a Marshallian elasticity of –0.468 (standard error = 0.098) and a Hicks elasticity of 0.328 (standard error = 0.064). These results imply a very large income effect of –0.796, which is comparable to the large values obtained by Hausman (1981) and Wales and Woodland (1979). In the second stage, incorporating information from the inter-temporal condition (79), they obtain a Frisch elasticity of 0.535. When they restrict $\alpha_3 = 0$ they obtain Marshallian, Hicks and Frisch elasticities of –0.157, 0.652 and 1.004 respectively (note that the implied income effect is similar). Thus, ignoring the complementarity between work hours and consumption appears to cause upward bias in all three labour supply elasticities.[68]

Interestingly, Ziliak and Kniesner (2005) examine how their results are affected by the use of different consumption measures. The comparison is as follows:

Consumption measure	Marshall	Hicks	Income effect	Frisch
Blundell, Pistaferri and Preston (2001)	–0.468	0.328	–0.796	0.535
Skinner (1987)	–0.313	0.220	–0.533	0.246
PSID unadjusted	–0.442	0.094	–0.536	0.148

Obviously the Hicks and Frisch elasticity estimates are very sensitive to the consumption measure used, while the Marshallian elasticity is relatively insensitive to the consumption measure. This is reminiscent of our earlier observation, when discussing the 'Hausman-MaCurdy controversy' that elasticity estimates tend to be quite sensitive to the wage and non-labour income measures used. In the context of life-cycle models, the same appears to be true of the consumption measure.

Finally, the authors use the estimates of within-period preferences (i.e. the Hicks elasticity) to calculate the marginal welfare cost of tax increases that raise all tax rates proportionately. This turns out to be 16 per cent of the revenue raised. However, if they do the same calculation using the estimates obtained using the PSID unadjusted food consumption measure, the welfare loss is only 5 per cent of the revenue raised.

A novel twist in the literature is taken in the paper by Pistaferri (2003), who estimates regressions for the change in hours (i.e. equation (31) or (66)), as does MaCurdy (1981), Browning, Deaton and Irish (1985) and Altonji (1986) in papers discussed earlier. But Pistaferri adopts a very different approach. Recall that the earlier papers treated the expected change in wages from time $t - 1$ to t as unobserved, and hence they used instruments dated at time $t - 1$ to construct predicted wage growth in the first stage of a two-stage least squares procedure. This approach relies on the assumption that the econometrician knows quite a bit about how workers forecast wage growth. Specifically, he/she must be able to pick instruments that (i) are uncorrelated with the workers' forecast errors, and (ii) are good predictors of the wage growth the workers actually expect. But as we saw, these papers all suffered from the problem that coming up with good predictors for actual wage growth is difficult (i.e. first stage R^2s are low). Perhaps workers can make better predictions of their wage growth than we can. Furthermore, as we don't actually know how workers forecast wage growth, we can't be sure

67 It is a bit difficult to conceptualise what it means for σ to vary with age, given that σ governs how willing a person is to substitute utility across periods. Does a 20 year old with $\sigma \approx 0$ solve his/her lifetime planning problem as if he/she is very willing to substitute utility inter-temporally, and then engage in re-planning each year as his/her σ drops? Or does a person plan out his/her life knowing that his/her σ will fall over time? If so, exactly how does one do that? Does a person fully take into account the preferences of his/her future selves? Apparently, one can circumvent such questions when estimating inter-temporal conditions like (79), but these issues would have to be confronted to actually obtain a full solution of a person's lifetime optimisation problem. I discuss models that involve such full solutions in sesction 7.3.4.

68 I believe the intuition for this result can be explained in the following way. If work and consumption are complements within a period, then a wage increase affects hours through three channels. There are the usual substitution and income effects. But in addition, a wage increase will, *ceteris paribus*, increase consumption. This reduces the marginal utility of leisure at the initial hours level, giving an additional reason for hours to increase. As a result, a smaller substitution effect is required to explain any given level of responsiveness of hours to wages.

that all variables dated at time $t-1$ are in fact used to make forecasts, so we can't be sure they are valid instruments. The Pistaferri (2003) innovation is to use actual data on expectations to construct measures of workers' anticipated and unanticipated wage growth.[69]

The data that Pistaferri uses is the Bank of Italy Survey of Households' Income and Wealth (SHIW) from 1989, 1991 and 1993. The survey is conducted every two years, and a fraction of subjects are re-interviewed (creating a panel component). The survey contains questions about expected *earnings* growth, not *wage* growth. I'll discuss the problems this creates below, but first consider how we could use *wage* expectations data if we had it.

Recall that the hours growth equations (31) and (66) contain actual wage growth as a regressor, while the surprise part of wage growth was relegated to the part of the residual denoted by ζ_{it} or ξ_{it}, respectively, which represented how the surprise wage change altered the marginal utility of consumption λ_{it}. The presence of this term in the residual meant that the instruments used to predict wage growth had to be correlated with expected wage growth and uncorrelated with *unexpected* wage growth. Specifically, recall that:

(85) $$\zeta_{it} \equiv \frac{1}{\gamma}\xi_{it} \equiv \frac{1}{\gamma}\left\{\ln\lambda_{it} - E_{t-1}\ln\lambda_{i,t-1}\right\} = \frac{1}{\gamma}\frac{\partial\ln\lambda_{it}}{\partial\psi_{it}}\left\{\Delta\ln w_{it} - E_{t-1}\Delta\ln w_{it}\right\}$$

where I have defined:

$$\Delta Z_{it} \equiv Z_{it} - Z_{i,t-1} \qquad \psi_{it} \equiv \left\{\Delta\ln w_{it} - E_{t-1}\Delta\ln w_{it}\right\}$$

Equation (85) consists mostly of definitions that have been previously stated and which are collected here for convenience. For the sake of brevity, I have introduced the delta notation ΔZ_t to denote the change in a variable Z_t from $t-1$ to t. And I have introduced ψ_t to denote the unexpected wage change from $t-1$ to t. The only statement of substance embodied in (85) is the third equality, which implies that all surprise changes in the marginal utility of consumption are due to surprise changes in wages. The term $d\,1n\lambda_{it}/d\psi_{it}$ captures how surprise wage growth affects the (log) of the marginal utility of consumption. The assumption that only wage surprises move λ_{it} is a strong one, which rules out, for example, unexpected transfers of assets. I believe that this assumption is important for Pistaferri's approach, as I note below.

Now, if expected wage growth could actually be measured, then, using the definitions in (85), we could rewrite (66) as:

(86) $$\Delta\ln h_{it} = \frac{1}{\gamma}(\Delta\ln w_{it}) - \frac{\alpha}{\gamma}\Delta X_{it} - \frac{1}{\gamma}\ln\rho(1+r_t) + \left[\frac{1}{\gamma}\frac{\partial\ln\lambda_{it}}{\partial\psi_{it}}\right]\left\{\Delta\ln w_{it} - E_{t-1}\Delta\ln w_{it}\right\} + \Delta\varepsilon_{it}$$

where ξ_{it} in (66) has been replaced by

$$\xi_{it} = \frac{\partial\ln\lambda_{it}}{\partial\psi_{it}}\left\{\Delta\ln w_{it} - E_{t-1}\Delta\ln w_{it}\right\}$$

Furthermore, if we decompose the first term on the right-hand side of (86)—actual wage growth—into parts that were anticipated versus unanticipated at time $t-1$, we obtain:

(87) $$\Delta\ln h_{it} = \frac{1}{\gamma}(E_{t-1}\Delta\ln w_{it}) - \frac{\alpha}{\gamma}\Delta X_{it} - \frac{1}{\gamma}\ln\rho(1+r_t) + \left[\frac{1}{\gamma} + \frac{1}{\gamma}\frac{\partial\ln\lambda_{it}}{\partial\psi_{it}}\right]\left\{\Delta\ln w_{it} - E_{t-1}\Delta\ln w_{it}\right\} + \Delta\varepsilon_{it}$$

Equation (87) captures how anticipated wage changes $E_{t-1}\Delta\ln w_{it}$ have only a Frisch substitution effect $(1/\gamma) > 0$ on hours. But unanticipated wage changes $\{\Delta\ln w_{it} - E_{t-1}\Delta\ln w_{it}\}$ have both a substitution effect $(1/\gamma)$ and an income effect $(1/\gamma)(d\ln\lambda_{it}/d\psi_{it}) < 0$. Thus, the sign of the effect of unanticipated wage changes is theoretically ambiguous.

Now, a number of authors, including Blundell and MaCurdy (1999) among others, have argued that tax reforms (that alter after-tax wages) are generally unexpected and, to a reasonable approximation, assumed to be permanent by workers.[70] If we grant this, then, as Blundell and MaCurdy (1999) state, the coefficient on unanticipated wage changes, $(1/\gamma) + (1/\gamma)(d\ln\lambda_{it}/d\psi_{it})$, is what we should be concerned with for evaluating the labour supply effects of tax reforms.[71] But a staggering number of issues are being buried under the rug here.

For instance, the coefficient $(d\ln\lambda_{it}/d\psi_{it})$ depends on many things, including: (i) how do consumers forecast future wages? (i.e. to what extent do they expect surprise wage changes to be permanent or transitory?); (ii) how do consumers forecast future tax changes? (i.e. to what extent do they expect tax rule changes to be permanent or transitory?); and (iii) to what extent do the answers to questions (i) and (ii) depend on the source of the wage or tax surprise? (e.g. if a surprise wage change occurs due to an unexpected change in tax law is it expected to be more or less persistent than if it occurs due to an unexpected promotion or layoff?). Many more questions of this type could be asked.

The first fundamental issue that one must deal with is how workers map unanticipated wage changes into expectations of future wages. To do this one must specify a model of the wage process, and make an assumption

69 This idea is of particular interest to me, as my undergraduate dissertation (Keane 1983), proposed using actual data on money supply growth expectations to inform the then active debate on effects of anticipated versus unanticipated money supply growth on real economic activity. Like Pistaferri, I argued that the agents in the market had more information than the econometrician, so that their forecasts might well be better than the hypothetical forecasts we can construct using regressions of outcomes (e.g. either wage growth or money supply growth) on a set of variables that we assume were in the agents' information sets.

70 Note that the two assumptions are really two sides of the same coin: if one always thinks the current tax regime is unlikely to change, then one will always be surprised by changes.

71 See Blundell and MaCurdy (1999, p. 1603): 'As most tax and benefit reforms are probably best described as once-and-for-all unanticipated shifts in net-of-tax real wages today and in the future, the most appropriate elasticity for describing responses to this kind of shift is $\alpha_l + \gamma_0$', where α_l and γ_0 correspond, in their notation, to the two coefficients $\frac{1}{\gamma} + \frac{1}{\gamma}\frac{\partial\ln\lambda_{it}}{\partial\psi_{it}}$ on *unexpected* wage changes in equation (87).

about how consumers forecast future wages. Pistaferri (2003) assumes that log wages follow a random walk process with drift:

(88) $\ln w_{it} = \ln w_{i,t-1} + X_{i,t-1}\theta + \psi_{it} \qquad E_{t-1}\psi_{it} = 0$

where ψ_{it} is the unexpected shock to wage growth. Pistaferri (2003) further assumes that workers know that (88) is the wage process, and that they use (88) to forecast future wages. Ironically, having data on wage expectations does not enable you to get around the need to make assumptions about how expectations are formed![72] The key behavioural assumption implied by (88) is that workers view all wage innovations as permanent: an unexpected wage change ψ_{it} shifts a worker's expectation of all his/her future wages by exactly ψ_{it}.

Next, similar to MaCurdy (1981), Pistaferri (2003) must make an assumption about how current and expected future wages, as well as current assets, map into the marginal utility of consumption. Specifically, he assumes that:

(89) $\ln \lambda_{it} = \Gamma_{at} A_{it} + \Gamma_{0t} \ln w_{it} + \sum_{\tau=t+1}^{T} \Gamma_{\tau-t,t} E_t \ln w_{i,\tau}$

There is a fundamental difference between MaCurdy (1981) and Pistaferri (2003), however, in that MaCurdy approximates the marginal utility of wealth in a model with perfect foresight. This is obviously a function of the whole life-cycle wage path and initial assets, and it varies over time only according to the deterministic relationship $\lambda_{it} = \rho(1 + r_{t+1})\lambda_{i,t+1}$ (see equation (59)). Thus MaCurdy is trying to estimate a single $\ln \lambda_{i0}$, while Pistaferri is trying to estimate the time varying $\ln \lambda_{it}$.[73]

A key element that an approximation to $ln\ \lambda_{it}$ ought to capture is that, as the remaining time horizon grows shorter, the effect of a one-period increase in the wage should have a larger effect on the marginal utility of consumption (i.e. if you are in the terminal period, a surprise wage increase will be devoted entirely to higher consumption in that period. If you are in an earlier period, then a temporary wage increase will have a smaller effect on current consumption because it will be spread out and used to increase consumption in the current and all future periods). Similarly, a permanent wage increase at T – 1 should have about the same effect on consumption in the last two periods as an (equal sized) wage increase at T would have on consumption at T.[74] This is why the $\{\Gamma_{kt}\}$ terms in (89) are allowed to vary over time. Each term has both a subscript $k = 0,\ldots,\mathrm{T}-t$ that indicates the effect of the expected wage at time k on perceived wealth at time t, and a time subscript t that allows these effects to change over time. Of course, if one allowed the $\{\Gamma_{kt}\}$ terms to vary in an unconstrained way over k and t there would be a severe proliferation of parameters. So Pistaferri constrains them to vary linearly.[75]

From (89) we get that the surprise change in the marginal utility of consumption is related to the surprise change in the wage as follows:

(90)
$$\begin{aligned}\ln \lambda_{it} - E_{t-1} \ln \lambda_{it} &= \Gamma_a [A_{it} - E_{t-1} A_{it}] \\ &\quad + \Gamma_0 \{\ln w_{it} - E_{t-1} \ln w_{it}\} \\ &\quad + \sum_{\tau=t+1}^{T} \Gamma_{\tau-t} \{E_t \ln w_{i,\tau} - E_{t-1} \ln w_{i,\tau}\} \\ &= \Gamma_a [A_{it} - E_{t-1} A_{it}] + \Gamma_0 \{\psi_{it}\} + \sum_{\tau=t+1}^{T} \Gamma_{\tau-t} \{\psi_{it}\} \\ &= \Gamma_a \cdot 0 + \Gamma \cdot \psi_{it}\end{aligned}$$

where I have suppressed the time subscripts on the Γ to conserve on notation.

In (90) all of the Γ terms are negative because a surprise increases in assets, a surprise increase in the current wage, or a surprise increase in any future wage all increase the consumer's perception of his/her wealth. This leads to higher current consumption and hence a lower marginal utility of consumption. The second line of the equation utilises the fact that, given the random walk wage process assumed in (88), the changes in *all* future wage expectations $E_t \ln w_{i,\tau} - E_{t-1} \ln w_{i,\tau}$ for $\tau = t+1,\ldots,\mathrm{T}$ are equal to the current wage surprise ψ_{it}. Again, this is because that surprise is expected to persist forever. At the opposite extreme, if we had instead assumed that consumers perceive all wage surprises as purely transitory, then we would have $E_t \ln w_{i,\tau} - E_{t-1} \ln w_{i,\tau} = 0$ for all $\tau = t+1,\ldots,\mathrm{T}$ and the third term in the second line would vanish. Finally, the last line of (90) invokes Pistaferri's assumption of no unexpected asset changes, and defines $\Gamma = \Gamma_0 + \Gamma_1 + \cdots + \Gamma_{\mathrm{T}-t}$.[76]

72 This point has been made in a different context (forecasting future prices of durable goods) by Erdem et al. (2005). They show that when enough periods are available one can use the expectations data to estimate expectations formation process, but one still has to impose some *a priori* structure on the process. Pistaferri cannot pursue this approach because he only has two periods of expectations data.

73 MaCurdy (1981) can back out $\ln \lambda_{i0}$ in a second stage after he has estimated the differenced hours equation (62) in the first stage. This is because estimation of (62) uncovers all the parameters of the hours equation in levels (61), except for $(1/\gamma)\ln \lambda_{i0}$, which serves as the individual specific constant term (or 'fixed effect') in the levels equation. Having estimated these constants, MaCurdy can regress them on the whole set of life-cycle wages. Of course, the difficulty that MaCurdy faces is that he only observes wages for his ten-year sample period, not for the whole life-cycle. Thus, he must fit a life-cycle wage profile for each person using ten years of data. He then regresses $(1/\gamma)\ln \lambda_{i0}$ on the individual specific parameters of this (assumed quadratic) profile. Using the coefficient on the wage equation intercept in this equation, MaCurdy can determine how an upward shift in the intercept of the whole wage profile would affect $(1/\gamma)\ln \lambda_{i0}$, and hence labour supply. MaCurdy estimates that a 10 per cent increase in wages at all ages would increase labour supply by only 0.8 per cent. Of course, the problem with this procedure relative to MaCurdy (1983) Method 1 or Blundell and Walker (1986) is the need to extrapolate out of sample wage information rather than using current consumption or assets as a proxy for lifetime wealth.

74 The same argument holds for an increase in assets. For instance, a 60 year old who wins a million dollars in the lottery should be much more likely to retire than a 30 year old.

75 This is not indicated in notation in his paper (see Pistaferri, equation (8)), but Pistaferri has confirmed this to me in private correspondence.

76 Notice that the wealth effect term Γ gets (mechanically) smaller as t gets larger, simply because there is less of a future horizon over which wages will increase, so fewer Γ_t terms are being added up. Counteracting that, as I argued earlier, is that the wealth effect of each individual (period specific) wage increase should grow larger as one gets closer to the end of the planning horizon.

Now, given (90), we have that $\frac{\partial \ln \lambda_{it}}{\partial \psi_{it}} = \Gamma$ and hence we can rewrite (87) as:

$$(91) \quad \Delta \ln h_{it} = \frac{1}{\gamma}(E_{t-1}\Delta \ln w_{it}) - \frac{\alpha}{\gamma}\Delta X_{it} - \frac{1}{\gamma}\ln \rho(1+r_t) + \left[\frac{1}{\gamma} + \frac{\Gamma}{\gamma}\right]\left\{\Delta \ln w_{it} - E_{t-1}\Delta \ln w_{it}\right\} + \Delta\varepsilon_{it}$$

This gives us Pistaferri's essential idea. We can use the coefficient on expected wage changes to estimate the inter-temporal elasticity of substitution ($1/\gamma$), while using the coefficient on unexpected wage changes to estimate the 'total' effect of a wage change, which includes both the substitution effect and the income effect. Taking the difference between the two coefficients enables us to isolate the income effect of a permanent wage increase (Γ/γ).

Estimation of this equation would have other key advantages over the conventional approach, as pursued in MaCurdy (1981) and Altonji (1986). First, there is no need to instrument for the wage change variables because now the unexpected wage change is controlled for rather than being relegated to the error. This circumvents the problem, noted earlier, that it is hard to come up with good predictors for wage growth. Second, as Pistaferri notes, the best predictors are usually age and education, but it is a strong assumption that these are excluded instruments that do not appear in the hours equation itself, as we would expect these variables to shift tastes for work. Third, the problem of aggregate shocks that I discussed earlier is avoided, as the average forecast error no longer enters the error term.

Unfortunately, there is quite a gap between this excellent idea and its actual empirical implementation. The first problem Pistaferri faces is that the Bank of Italy Survey does not really contain expectations of wage changes, but only of earnings changes. Pistaferri shows how to construct a version of (91) where expected and unexpected earnings replace wages, and the coefficients are suitably modified. However, as Pistaferri notes, this introduces a major problem: unobserved shifts in tastes for work will, of course, alter earnings (since earnings are a function of hours). And, as in all the models we have considered so far, unobserved tastes for work ($\Delta\varepsilon_{it}$) enter the error term in the hours equation. This renders expected and unexpected earnings changes endogenous.

Second, expected earnings changes are presumably measured with error. Given that variables such as hours and earnings are measured with error (as are even simple demographic variables in some cases), it would be highly implausible to assume that a more subtle variable such as the expected change in earnings is not measured with error as well. Furthermore, there may well be systematic errors arising from how respondents interpret the question. The question reads, 'We are interested in knowing your opinion about your labour earnings or pensions twelve months from now'. Do we think respondents would or would not include expected tax changes when answering such a question? And, while the expectations question asked about earnings twelve months hence, the data on wages, hours and earnings were collected in 1989, 1991 and 1993. In order to align the two-year time interval of the earnings data with the one-year forecast horizon, Pistaferri assumes a person would have projected their earnings growth rate forecast to persist for two years, an additional source of measurement error.

Both of these problems suggest that it may be necessary to instrument for expected and unexpected changes in earnings, using variables that help predict these variables but that are uncorrelated with taste shocks and measurement error. In that case, one of the key advantages of Pistaferri's procedure is lost. Pistaferri (2003) does not actually attempt to deal with these problems, and he estimates his version of (91) by least squares.

For estimation, Pistaferri uses data on male household heads who were aged between 26 and 59 in 1989. There are 1,461 person-year observations in the unbalanced panel. As observed taste shifters, he uses age, education, region, family size, whether the spouse or other household members work, and the number of children in various age ranges. He estimates that the Frisch elasticity is 0.704 (standard error = 0.093) and the income effect (Γ/γ) is –0.199 (standard error = 0.091). Thus, the elasticity of labour supply with respect to a surprise permanent upward shift in the wage profile is 0.51. That is, a permanent unexpected 10 per cent wage increase would cause a 5 per cent increase in labour supply. This is a very large uncompensated wage effect, and it implies that permanent tax changes have very large effects on labour supply. The result contrasts sharply with MaCurdy (1981) whose comparable estimate is only a 0.8 per cent increase. We should view both results with some caution, however, given the data limitations noted above.[77]

It is important to note also that Italy had a recession in 1993. Pistaferri (2003) includes a 1993 dummy in (91) and obtains a coefficient of –0.068 with a standard error of 0.023. This is a large value, implying a 6.8 per cent decline in hours not explained by the model. This would appear to suggest that workers in Italy are not always free to adjust hours in the short run, and that there was demand-induced rationing.

Bover (1989) adopts a different approach to estimating responses to both anticipated wage changes and unanticipated permanent wage changes within the life-cycle framework. Her innovation is to use a Stone-Geary utility function in place of the utility function (1) that was adopted by MaCurdy (1981) and Altonji (1986). The virtue of the Stone-Geary is that it can deliver a closed form solution for the marginal utility of consumption λ. The Stone-Geary functional form is:

77 Pistaferri (2003) contains a couple of other elements that I haven't mentioned. His version of (91) includes a measure of the perceived variance of earnings, also obtained from the survey of expectations. But he finds that variance is not significant in the hours equation. He also tests for separability between leisure and consumption but does not find strong evidence for non-separability. Finally, Pistaferri also uses his data to estimate an hours change regression like that estimated in MaCurdy (1981) and Altonji (1986), using a cubic in age and education and interactions between age and education as instruments. This produces a Frisch elasticity of 0.318 with a standard error of 0.319. The R^2 in the first stage regression is only 0.0025 with an F-statistic of 1.73. One of the key advances in econometric practice since the 1980s is the much greater attention that is now paid to the problem of weak instruments in the first stage of 2SLS regressions. A common rule of thumb is that the F-statistic should be at least 5 before results can be trusted.

(92) $U_t = \beta_{it}\ln(H_{\max} - h_{it}) + (1-\beta_{it})\ln(C_{it} - C_{\min})$

where $H_{\max}$ and $C_{\min}$, which denote maximum feasible hours of work and minimum subsistence consumption respectively, are parameters to be estimated. β_{it} is a parameter that captures tastes for leisure relative to consumption. Given this functional form, the marginal utility of consumption and leisure are given by:

(93)
$$\lambda_{it} \equiv \frac{\partial U_t}{\partial C_t} = (1-\beta_{it})\frac{1}{C_{it} - C_{\min}}$$
$$-\frac{\partial U_t}{\partial h_t} = -\beta_{it}\frac{1}{H_{\max} - h_{it}}$$

Thus, the usual within-period MRS conditional gives us:

(94) $$\frac{\partial U_t}{\partial L_t} = w_{it}\frac{\partial U_t}{\partial C_t} \quad \Rightarrow \quad \beta_{it}\frac{1}{H_{\max} - h_{it}} = w_{it}\lambda_{it}$$

From (94) we solve for labour supply as a function of the marginal utility of consumption λ_{it}:

(95) $$h_{it} = H_{\max} - \frac{\beta_{it}}{w_{it}\lambda_{it}} \quad \Rightarrow \quad \beta_{it} = (H_{\max} - h_{it})w_{it}\lambda_{it}$$

where the equation to the right of the arrow is simply a rearrangement of the labour supply equation that will be useful below.[78] Now, using (95) we can obtain the Frisch elasticity as follows:

(96) $$\left.\frac{\partial h_{it}}{\partial w_{it}}\right|_{\lambda_{it}\,fixed} = \frac{\beta_{it}}{(w_{it})^2\lambda_{it}} \Rightarrow e_F = \frac{w_{it}}{h_{it}}\left.\frac{\partial h_{it}}{\partial w_{it}}\right|_{\lambda_{it}\,fixed} = \frac{\beta_{it}}{w_{it}h_{it}\lambda_{it}}$$

This expression is not particularly useful as it still involves λ_{it}, but we can use the right side of (95) to substitute out for β_{it} and obtain:

(97) $$e_F = \frac{(H_{\max} - h_{it})w_{it}\lambda_{it}}{w_{it}h_{it}\lambda_{it}} = \frac{(H_{\max} - h_{it})}{h_{it}}$$

This equation illustrates the restrictiveness of the Stone-Geary utility function for this purpose. The single parameter $H_{\max}$ will determine the Frisch elasticity (at any given level of hours) but, as we see from (95), this parameter also plays a key role in determining the average level of hours. The restrictiveness creates problems in fitting the data, as we will see below.

Now, returning to the issue of solving for λ, we first use (93) to solve for labour supply as a function of the marginal utility of consumption λ_{it}:

(98) $$C_{it} = C_{\min} + (1-\beta_{it})\frac{1}{\lambda_{it}}$$

Now, following Bover (1989), we assume perfect foresight and assume that $\rho(1 + r) = 1$. In this case, λ_{it} is just a person-specific constant λ_i, and we can write the demand functions as:

(99) $$h_{it} = H_{\max} - \frac{\beta_{it}}{w_{it}}\frac{1}{\lambda_i} \qquad C_{it} = C_{\min} + (1-\beta_{it})\frac{1}{\lambda_i}$$

Note that with perfect foresight the lifetime budget constraint would be:

$$\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}C_{it} = A_{i0} + \sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}w_{it}h_{it}$$

Substituting the (99) into the budget constraint, and approximating finite sums by infinite sums (Bover 1989), we obtain, after some simple algebra:

(100) $$\frac{1+r}{r}\frac{1}{\lambda_i} = A_{i0} + \sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}\left[w_{it}H_{\max} - C_{\min}\right] \equiv F_i$$

The right-hand side of (100) is the definition of lifetime 'full income' in the Stone-Geary setup (i.e. initial assets plus the present value of the maximum amount one could possibly earn in excess of the subsistence consumption level $C_{\min}$). Notice that as lifetime wealth (as measured by full income) increases, the marginal utility of consumption λ_i falls. We can now use (100) to substitute for λ_i in the labour supply equation in (99), obtaining:

(101) $$h_{it} = H_{\max} - \frac{\beta_{it}}{w_{it}}\frac{r}{1+r}F_i$$

Now, to see how a permanent increase in wages would alter labour supply it is useful to specify a process for wages, such as the linear time trend $w_{it} = \alpha_{0i} + \alpha_{1i}t$ that Bover (1989) chooses. Substituting this into (100) we obtain:

(102) $$F_i = A_{i0} + \sum_{t=1}^{T}\frac{(\alpha_{0i} + \alpha_{1i}t)H_{\max} - C_{\min}}{(1+r)^{t-1}}$$
$$= A_{i0} + \frac{1+r}{r}(\alpha_{0i}H_{max} - C_{\min}) + \sum_{t=1}^{T}\frac{\alpha_{1i}\cdot t\cdot H_{\max}}{(1+r)^{t-1}}$$

Now we can calculate the elasticity of labour supply with respect to an anticipated permanent wage increase (i.e. an upward shift in the whole wage profile induced by increasing the wage equation intercept α_{0i}). We have:

$$\frac{\alpha_{0i}}{h_{it}}\frac{\partial h_{it}}{\partial \alpha_{0i}} = \frac{\alpha_{0i}\beta_{it}}{w_{it}^2 h_{it}}\frac{r}{1+r}F_i - \frac{\alpha_{0i}\beta_{it}}{w_{it}h_{it}}H_{\max}$$

The first term on the right is the substitution effect that arises from the increase in the current wage, while the second term is the income effect that arises due to the increase in F_i.[79] Bover shows how to extend this analysis to the uncertainty case where wages evolve stochastically, but I will not discuss this in detail as her empirical results for the uncertainty case are nearly identical to those in the perfect foresight model.

Bover (1989) estimates the labour supply model obtained by combining (101) and (102) using PSID data

78 It is worth noting that, in contrast to (61) or (70), the labour supply equation in (95) does not take a form where it is possible to eliminate the unobserved λ_{it} term via a simple transform such as taking logs and differencing. However, note that $L_{it} \equiv H_{\max} - h_{it}$ is interpretable as leisure. If we observe leisure then we could write the demand for leisure as $\ln L_{it} = \ln\beta_{it} - \ln w_{it} - \ln\lambda_{it}$. Thus, the Frisch elasticity of demand for leisure is simply –1. This means that the Frisch elasticity of demand for labour is:

$$e_F = \left.\frac{d\ln h}{d\ln L}\frac{\partial \ln L}{\partial \ln w}\right|_{\lambda\ fixed} = \frac{h - H_{\max}}{h}(-1) = \frac{H_{\max} - h}{h} = \frac{L}{h}$$

Thus, if leisure were observed the Frisch elasticity would simply be the ratio of leisure hours to labour hours, and it would not depend on any model parameters. This illustrates the lack of flexibility of the Stone-Geary functional form for this purpose.

79 The second term is equivalent to the expression in Bover (1989), equation (11). That expression gives the wealth effect of shifting up the wage profile, but it does not include the current wage effect.

from 1968–1976. She uses data on 785 white men aged 20–50 in 1968, requiring that they have positive annual hours and wages for all periods. In a first stage, she uses the ten years of wage observations for each person to estimate the person-specific parameters in the wage equation $w_{it} = \alpha_{0i} + \alpha_{1i}t$.[80] As usual, randomness is introduced into the labour supply model by letting the taste shift variable β_{it} in (101) be stochastic. It is also allowed to vary with age and number of children. The wage is treated as endogenous, both due to measurement error and because workers with a high unobserved taste for work may work harder and thus achieve a higher wage rate. The instruments are the typical age and education variables, along with the state unemployment rate (interpreted as a demand shifter) and time dummies.

Turning to the estimates, the value of H_{max} is 2,353 hours (standard error = 43). This result seems rather implausible, given that many people do in fact work more than 2,353 hours. Indeed, Bover reports that observed hours exceed $H_{\max}$ for 65 per cent of observations. As I discussed earlier, the Frisch elasticity in the Stone-Geary model is simply $(H_{\max} - h_{it})/h_{it}$, and since h_{it} exceeds 2,000 for most working men, the low value of $H_{\max}$ guarantees that the Frisch elasticity will be quite small. At the mean of the data Bover calculates that it is 0.08. Bover also finds very small income effects. Thus, she reports that the response of hours to a shift in the entire life-cycle wage profile is trivially small. But the main point of these results seems to be to cast doubt on the ability of a model with Stone-Geary preferences to fit observed hours data—which is unfortunate given that the Stone-Geary delivers a simple form for λ.

7.3.4 Incorporating Human Capital into the Life-Cycle Model

A fundamental problem with all of the labour supply models that I discussed in sections 7.3.2 and 7.3.3 is that they treat wages as exogenous. That is, they ignore the fact that work experience may lead to increased wages. If that is the case, it has rather striking implications for all of the proposed estimation methods discussed in those earlier sections.

To see this, let's return to the simple two-period model of (17) but assume that the wage in period 2, rather than being exogenously fixed, is an increasing function of hours of work in period 1. Specifically, I will assume that:

(103) $$w_2 = w_1(1+\alpha h_1)$$

where α is the percentage growth in the wage per unit of work. Given a two-period model with each period corresponding to twenty years, it is plausible in light of existing estimates that αh_1, the percentage growth in the wage rate over twenty years, is around one-third to a half.[81]

Once we introduce human capital accumulation via work experience as in (103), equation (17) is replaced by:

(104) $$V = \frac{[w_1 h_1(1-\tau_1)+b]^{1+\eta}}{1+\eta} - \beta_1 \frac{h_1^{1+\gamma}}{1+\gamma} + \rho\left\{\frac{[w_1(1+\alpha h_1)h_2(1-\tau_2)-b(1+r)]^{1+\eta}}{1+\eta} - \beta_2\frac{h_2^{1+\gamma}}{1+\gamma}\right\}$$

and the first order conditions for the problem are now:

(105) $$\frac{\partial V}{\partial h_1} = C_1^{\eta} w_1(1-\tau_1) - \beta_1 h_1^{\gamma} + \rho C_2^{\eta} w_1 \alpha h_2(1-\tau_2) = 0$$

(106) $$\frac{\partial V}{\partial h_2} = C_2^{\eta} w_1(1+\alpha h_1)(1-\tau_2) - \beta_2 h_2^{\gamma} = 0$$

(107) $$\frac{\partial V}{\partial b} = C_1^{\eta} - \rho(1+r)C_2^{\eta} = 0$$

Comparing (105) to (18) we see that it now includes an extra term $\rho C_2^{\eta} w_1 \alpha h_2(1-\tau_2)$, which captures the effect of an extra hour of work at $t = 1$ on the present value of earnings at $t = 2$. If we perform the usual manipulations on (105) to obtain the within-period MRS condition we now obtain:

$$MRS = \frac{MU_L}{MU_C} = \frac{\beta h_1^{\gamma}}{C_1^{\eta}} = w_1(1-\tau_1) + \frac{\rho C_2^{\eta}}{C_1^{\eta}}\alpha w_1 h_2(1-\tau_2)$$

This can be simplified by using (107) to eliminate $C_2^{\eta}/C_1^{\eta} = [\rho(1+r)]^{-1}$ giving:

(108) $$\frac{\beta h_1^{\gamma}}{C_1^{\eta}} = w_1(1-\tau_1) + \frac{\alpha w_1 h_2(1-\tau_2)}{1+r}$$

It is useful to compare (108) to (42), the MRS condition for the model without human capital. Here the opportunity cost of time is no longer simply the after-tax wage rate. Instead, it is augmented by the term $\alpha w_1 h_2(1-\tau_2)/(1+r)$, which captures the effect of an extra hour of work at $t = 1$ on the present value of earnings at $t = 2$.

Of course, in a multi-period model, this extra term would instead be the effect of an extra hour of work at $t = 1$ on the present value of all future earnings, which depends on hours of work in all future periods. Thus, the essential idea of Method 1 (MaCurdy 1983) and the related two-stage budgeting technique whereby current period consumption can be used as a sufficient statistic for all future period variables no longer holds.

Suppose we ignore this problem and attempt to estimate the parameters of preferences by estimating (42) ignoring the human capital term. Then the resultant bias will depend on the size of the human capital term relative to the after-tax wage. If the human capital term were trivially small then ignoring it might not be a problem. However, a simple back of the envelope calculation in Keane (2009a) suggests that, given plausible values for the return to work experience in the United States,[82] at age 20 the human capital term is roughly the same size as the wage rate, so that for a 20-year-old worker the opportunity

80 As in MaCurdy (1981), a profile is also fit to assets, and the intercept is used to measure initial assets A_{i0}.

81 For instance, using the PSID, Geweke and Keane (2000) estimate that for men with a high school degree, average earnings growth from age 25–45 is 33 per cent. For men with a college degree, they estimate a rate of 52 per cent. Most of this earnings growth is in fact due to wage growth because the growth in hours is modest.

82 Specifically, Keane (2009a) specifies a Mincer-type wage equation with a quadratic in work experience. Parameters are set so a full year of work experience increases the wage rate by 5.7 per cent at age 20, but by only 1.3 per cent at age 40. That is $\ln w_{it} = \text{constant} + 0.057\,x - 0.11\,x^2/100$, where x is years of work experience.

cost of time is roughly double the wage. As a worker ages there is less time to recoup the gains to human capital investment, so the size of the human capital term falls. The same back of the envelope calculation in Keane (2009a) suggests that by age 40 the human capital term is only about 20 per cent as large as the wage.

We can get a more precise idea of how the presence of human capital investment will bias standard methods of estimating the Frisch elasticity based on hours change regressions like (62). If we divide (106) by (105), using (107) to cancel out the consumption terms, we obtain:

$$\left(\frac{h_2}{h_1}\right)^{\gamma} = \frac{\beta_1}{\beta_2}\frac{w_1(1+\alpha h_1)(1-\tau_2)}{\rho(1+r)w_1(1-\tau_1)+\rho\alpha w_1 h_2(1-\tau_2)}$$

We can simplify this to a more intuitive expression if we plug in $w_2 = w_1(1 + \alpha h_1)$ and assume that $\tau_1 = \tau_2 = \tau$. Then we get:

$$\left(\frac{h_2}{h_1}\right)^{\gamma} = \frac{w_2}{w_1}\frac{1}{1+\alpha h_2/(1+r)}\frac{1}{\rho(1+r)}\frac{\beta_1}{\beta_2}$$

Taking logs we obtain:

$$(109)\quad \ln h_2 - \ln h_1 = \frac{1}{\gamma}\left\{\ln\frac{w_2}{w_1} - \ln\left(1+\alpha h_2/(1+r)\right) - \ln\rho(1+r) - \ln\frac{\beta_2}{\beta_1}\right\}$$

This is the same as the first difference log wage equations typically used to estimate the Frisch elasticity (e.g. see equation (22)), *except* now we have the additional term, $-\ln(1 + \alpha h_2/(1 + r))$. Notice that this term is negative: the existence of wage growth with experience ($\alpha > 0$) will, *ceteris paribus*, cause workers to shift hours towards the early part of the life-cycle. As a result, hours grow less over the life-cycle than they would if wage growth were exogenous.

Thus, conventional estimates of $(1/\gamma)$ will be biased downward. Since human capital dampens the association between hours growth and wage growth, a model that ignores human capital will rationalise the observed (smaller) association via a smaller value of $(1/\gamma)$.

How large is the magnitude of this bias likely to be? One way to think about the problem is to note that the correct way to estimate $(1/\gamma)$ would *not* be to regress hours growth on wage growth but to instead to regress hours growth on the growth of the opportunity cost of time, that is, on $\ln w_2/w_1(1 + \alpha h_2/(1 + r))$. Referring to the back of the envelope calculation from Keane (2009a) mentioned earlier, I find that, using conventional estimates of returns to experience, the opportunity cost of time grows roughly six times more slowly than the wage from age 20 to age 40.[83] Given that differential, we would expect conventional estimates of $(1/\gamma)$ that ignore human capital to be biased downward by roughly a factor of 6.

Another way to look at the problem is to simplify (109) by assuming that $\rho(1 + r) = 1$ and that $\beta_1 = \beta_2$. Then, if we solve (109) for $(1/\gamma)$ we obtain:

$$(110)\quad \frac{1}{\gamma} = \ln\left(\frac{h_2}{h_1}\right) \div \left[\ln\left(\frac{w_2}{w_1}\right) - \ln\left(1+\frac{\alpha h_2}{1+r}\right)\right]$$

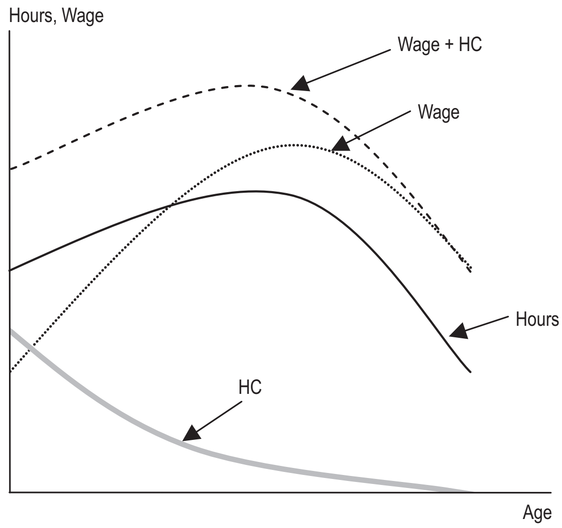
Thus, wage growth from $t = 1$ to $t = 2$ would have to be adjusted downward by roughly $\alpha h_2/(1 + r)$ percentage points in order to correct for the missing human capital term (and obtain a valid estimate of the growth of the opportunity cost of time).

As I noted earlier, a reasonable (and, in fact, conservative) estimate of αh_1 (i.e. wage growth during the first twenty years of the working life) is about 33 per cent. For illustration, let's suppose that h_2 is 20 per cent greater than h_1, so that αh_2 is roughly 40 per cent. As I also noted earlier, in a two-period model a reasonable value for $\rho = 1/(1 + r)$ is 0.554, giving $\alpha h_2/(1 + r) = 22$ per cent. Hence, for these values, we need to subtract roughly 22 percentage points off the rate of wage growth to obtain the growth in the opportunity cost of time (i.e. 33 per cent – 22 per cent = 11 per cent). If we had used observed wage growth to calculate $1/\gamma$, we would obtain $20/33 \approx 0.60$ for the Frisch elasticity. But the correct value is $\ln(1.20)/\ln[1.33/1.22] \approx 2.1$. Thus, for reasonable parameter values, the downward bias in estimates of the Frisch elasticity due to ignoring human capital will tend to be very substantial.

Finally, there is a useful graphical intuition for the same idea. Figure 7.4 presents a stylised plot of male wage rates and hours of work over the life-cycle.[84] The black line represents annual hours of work. It has a hump shape as noted in the descriptive regressions presented by Pencavel (1986), with a peak at roughly age 45 and a fairly rapid decline in the 50s and 60s. The dotted line is the wage rate which also has a characteristic hump shape. As has been noted by many studies, male wages grow very rapidly early in the life-cycle, peak in the 40s, and then decline. (Some studies have also found that wage growth in the early part of the life-cycle is faster for more educated workers.)

Now, as we have seen, the typical study in the male labour supply literature regresses hours (or hours growth)

Figure 7.4 Hours, Wages and Price of Time over the Life-Cycle



Notes: HC denotes the return to an hour of work experience in terms of increased present value of future wages. The opportunity cost of time is wage + HC.

83 This occurs, of course, because the human capital term $\alpha w_t h_{t+1}/(1+r)$ shrinks over time as the wage w_t grows.

84 That is, it does not plot any particular data set, but simply illustrates the typical patterns for male wages and hours observed across a broad range of data sets.

on wages (or wage growth), and, in order to deal with problems of measurement error and endogeneity, it instruments for wages (or wage growth) using primarily polynomials in age and education. These instruments are chosen precisely because they capture the hump shape of the life-cycle wage path depicted in Figure 7.4; of course, the predicted wages based on these instruments will closely track the typical life-cycle wage path depicted in the figure. Thus, when one regresses hours on predicted wages, one will essentially uncover the relative slope of the hours and wage curves in Figure 7.4. Since the wage path is much steeper than the hours path over most of the life-cycle, the estimated elasticity of hours with respect to predicted wage changes will be much less than 1.0.

The thick grey line in Figure 7.4 represents the return to human capital investment. That is, it is the return to an additional hour of work in terms of increased future wages (and hence increased future earnings) captured by the second term in equation (108). As I noted earlier, given reasonable estimates of the return to experience this term will be at least as great as the wage rate itself at the start of a person's working life. But it falls quickly over time and is zero in the terminal period (when investment serves no further purpose). The opportunity cost of time, as indicated in (108), is the sum of the wage rate and this human capital effect. This is represented by the dashed line in Figure 7.4. Note that the dashed line is much less steep than the wage line. Thus, hours appear to be much more responsive to changes in the opportunity cost of time than to changes in the wage rate.

In another paper (Keane 2009b), I pointed out that if returns to work experience are important, there will be important implications for tax policy. Returning to (108) we see that a temporary $t = 1$ tax increase affects only the current wage $w_1(1 - \tau_1)$ but does not alter the return on human capital investment. But a permanent tax increase, which increases both τ_1 and τ_2, will reduce the human capital return $\alpha w_1 h_2(1 - \tau_2)/(1 + r)$ as well. Thus, looking only at the effect of *wages* on hours may have consequences beyond just causing us to understate how responsive workers are to changes in the opportunity cost of time. It may also cause us to understate the responsiveness of workers to permanent tax rate changes.

Put another way, (108) implies that, contrary to conventional wisdom, a permanent tax change *may* have a larger effect on current labour supply than a temporary tax change (as the former alters only the current wage while the latter also alters the return on human capital investment). Keane (2009b) shows that in a model with both human capital and saving it is theoretically ambiguous whether a permanent or transitory tax change has a larger effect on current labour supply. A permanent tax increase has both (i) a larger income effect and (ii) a larger effect on the return to human capital than a temporary tax increase. These factors have opposite effects on current labour supply. Keane (2009b) presents simulations showing that for quite plausible parameter values the human capital effect dominates, so that permanent tax changes have larger effects.

I now turn to the empirical literature on male life-cycle labour supply that includes human capital accumulation. Unfortunately, there are very few papers of this type. As far as I am aware, the first paper to estimate a life-cycle model with human capital was Heckman (1976). The computing technology available at that time did not permit estimation of a model where workers decide jointly on savings and human capital investment, particularly not while also allowing for uncertainty in wages and stochastic taste shocks. Thus, Heckman's model is deterministic and only attempts to fit 'typical' life-cycle paths of wages and hours.

The Heckman (1976) approach is rather different from the 'learning by doing' human capital investment model that I have described here (see equation (103)). Instead of specifying that work experience increases human capital in and of itself, Heckman follows Ben-Porath (1967) and Haley (1973) in using a type of model where a worker may choose to devote some fraction of his/her work time to investment. The worker is paid only for productive time, not time spent learning. But observed labour supply is the sum of all time at work: actual productive time plus investment time. Hence, the observed market wage rate in period t is given by $w_t = w_t^* (1 - S(t))$, where w_t is the worker's actual productivity and $S(t)$ is the fraction of his/her time at work that the worker spends in investing in human capital.

The key similarity between Heckman's model and the learning-by-doing model is that the observed market wage rate w_t is not the true opportunity cost of time. In Heckman's model that is w_t^*, the workers true productivity, as that is what the worker gives up per unit of time spent in leisure or learning. Since $w_t^* = w_t/(1 - S(t))$, the true cost of time exceeds the wage rate by the multiplier $1/(1 - S(t))$, which is an increasing function of the fraction of his/her time at work that the worker spends investing in human capital. So fundamentally his model is quite similar in spirit to the approach taken in equations (109)–(110), in that human capital investment causes the opportunity cost of time to exceed the wage rate.

Heckman (1976) estimates an equation for $S(t)$ jointly with an equation for observed hours and wages (derived from a particular functional form mapping investment time into wages). The model is estimated using data on 23–65-year-old males from the United States in 1970, but as noted, the model is deterministic and it is fit to average wages and hours (by age). Heckman's estimate of the $S(t)$ function implies that 23-year-old male workers spend roughly 35 per cent of their work time in human capital investment activity. Hence, their opportunity cost of time exceeds their observed wage rate by roughly 54 per cent. The fraction of time spent on investment is estimated to drop steadily, becoming near zero per cent at about age 40. Thus, his estimates imply that the opportunity cost of time grows by only 65 per cent as much as the observed wage rate from age 23 to age 40.[85]

Shaw (1989) substantially extended Heckman (1976) by estimating a model where workers make joint decisions about savings and human capital investment, incorporating uncertainty about future wages and hours. Her approach is

85 Note that: $w_{40}^* / w_{23}^* = w_{40}/[w_{23}/(1 - 0.35)] = (0.65)(w_{40}/w_{23})$.

to estimate an equation similar to equation (108), the MRS condition. However, to use the model with data one must first extend it to multiple periods and introduce uncertainty. In that case (108) becomes:

$$(111)\quad \frac{\beta h_t^{\gamma}}{C_t^{\eta}} = w_t(1-\tau_t) + E_t \sum_{\tau=0}^{T-t} \frac{(\alpha w_t) h_{t+1+\tau}(1-\tau_{t+1+\tau})}{(1+r)^{1+\tau}}$$

and the wage equation (103) becomes $w_{t+1} = w_t + (\alpha h_t)w_t$. Note that a one unit increase in hours at time t will raise the wage rate by (αw_t) in all future periods. This induces an increase in earnings of $(\alpha w_t)h_{t+1+\tau}$ for $\tau = t+1,\ldots,T$. We can thus see that the second term in (111) is the expected present value of the increased (after-tax) earnings in all future periods obtained by working an extra unit of time at time t.

The problem with estimating (111) is that it involves hours of work in all future periods, which will not be available in any data set. Shaw (1989) uses the following trick to get around this problem. First, rewrite (111) as:

$$(112)\quad \frac{\beta h_t^{\gamma}}{C_t^{\eta}} = w_t(1-\tau_t) + \frac{(\alpha w_t) h_{t+1}(1-\tau_{t+1})}{(1+r)} + E_t \sum_{\tau=0}^{T-t-1} \frac{(\alpha w_t) h_{t+2+\tau}(1-\tau_{t+2+\tau})}{(1+r)^{2+\tau}}$$

Now take (111) and date it forward one period:

$$(112')\quad \frac{\beta h_{t+1}^{\gamma}}{C_{t+1}^{\eta}} = w_{t+1}(1-\tau_{t+1}) + E_{t+1} \sum_{\tau=0}^{T-t-1} \frac{(\alpha w_t) h_{t+2+\tau}(1-\tau_{t+2+\tau})}{(1+r)^{1+\tau}}$$

Now, notice that the summations in (112) and (112′) are identical, except for a factor of $1/(1+r)$ and the dating of the expectation. So, by pre-multiplying (112′) by $1/(1+r)$ and taking the expectation at time t we obtain:

$$(112'')\quad E_t\left\{\frac{1}{1+r}\left[\frac{\beta h_{t+1}^{\gamma}}{C_{t+1}^{\eta}} - w_{t+1}(1-\tau_{t+1})\right]\right\} = E_t \sum_{\tau=0}^{T-t-1} \frac{(\alpha w_t) h_{t+2+\tau}(1-\tau_{t+2+\tau})}{(1+r)^{2+\tau}}$$

Intuitively, these manipulations are useful because the worker knows (or, rather, we assume he/she knows) that at time $t+1$ he/she will choose hours and consumption to satisfy expression (112″). Thus, we can use the worker's own labour supply and consumption behaviour at $t+1$, described by the simple expression on the left, to infer what he/she believes about the complex expectation term sitting on the right.[86]

So, using (112″) to substitute for the summation term in (112), we obtain:

$$(113)\quad \frac{\beta h_t^{\gamma}}{C_t^{\eta}} = w_t(1-\tau_t) + \frac{(\alpha w_t) h_{t+1}(1-\tau_{t+1})}{(1+r)} + E_t\left\{\frac{1}{1+r}\left[\frac{\beta h_{t+1}^{\gamma}}{C_{t+1}^{\eta}} - w_{t+1}(1-\tau_{t+1})\right]\right\}$$

This is an equation that is feasible to estimate, as it only requires data on hours at t and $t+1$, wages at t and $t+1$, and consumption at time t and $t+1$. The final step is to replace the expectation term with its actual realisation, while appending a forecast error:

$$(114)\quad \frac{\beta h_t^{\gamma}}{C_t^{\eta}} = w_t(1-\tau_t) + \frac{(\alpha w_t) h_{t+1}(1-\tau_{t+1})}{(1+r)} + \frac{1}{1+r}\left[\frac{\beta h_{t+1}^{\gamma}}{C_{t+1}^{\eta}} - w_{t+1}(1-\tau_{t+1})\right] + \xi_{t+1}$$

Equation (114) is the basic type of equation that Shaw (1989) estimates.[87] The estimation is done in two stages. In the first stage, a wage equation is estimated to determine how wages grow with work experience (the parameter α in equation (114)). In the second stage, the wage equation parameters are treated as known and (114) is estimated by instrumental variables. Valid instruments are variables known by workers at time t, so they are uncorrelated with the forecast error ξ_{t+1}.

While (114) is similar to the equation that Shaw (1989) estimates, she does not include taxes. On the other hand, she introduces a number of additional complications. First, rather than that the utility function in (1) she uses a translog utility function as in (84), with $G(X) = X$. As a consequence, the marginal utility of consumption and leisure terms in (114) become more complicated. Second, she lets the taste for work parameter β vary across workers based on schooling level. Third, in the wage equation she allows the rental rate on human capital to vary over time.

It is interesting that Shaw (1989) does not introduce stochastic variation in tastes as in all the previous studies we have examined. The reason why can be seen by looking back at the simple MRS condition for the model without human capital, (58), and following the steps that led to the estimating equation (66), where expectation errors and taste shocks entered as a composite additive error. This meant we could estimate (66) by instrumental variables without having to assume any distribution for the forecast errors and taste shocks. In contrast, looking at (114), we see that if β is allowed to have a stochastic component then that term will enter (114) in a highly nonlinear way. Thus, is not possible to make the taste shock 'pop out' into an additive error that could be combined with the ξ_t. This, in turn, would make the simple application of instrumental variables estimation infeasible.[88]

Turning to the wage equation, Shaw (1989) assumes that a worker's human capital, denoted by K, evolves according to:

$$(115)\quad K_{i,t+1} = \alpha_1 K_{it} + \alpha_2 K_{it}^2 + \alpha_3 K_{it} h_{it} + \alpha_4 h_{it} + \alpha_5 h_{it}^2 + \tau_t + \varepsilon_{it}$$

That is, current human capital is a quadratic function of last year's human capital and last year's hours of work. The $\{\tau_t\}$ are year-specific aggregate shocks to the productivity growth rate, and the $\{\varepsilon_{it}\}$ are person-specific productivity shocks (i.e. illness, involuntary job separations, etc.).

The wage rate is then determined by the aggregate rental price of human capital times the stock of human capital:

86 Interestingly, this is a continuous data analogue of the procedure developed by Hotz and Miller (1993) to infer agents' expectations from their discrete choices in discrete choice dynamic models.

87 Shaw (1989) also substitutes for consumption at $t+1$ using the familiar relationship $C_t^{\eta} = E_t\rho(1+r)C_{t+1}^{\eta}$.

88 Recently, I proposed a method (Keane 2009c) for estimating models where multiple stochastic terms enter the first order conditions nonlinearly.

(116) $w_{it} = R_t K_{it} \quad \Rightarrow \quad K_{it} = w_{it} / R_t$

Shaw (1989) allows the rental prices to vary over time in an unconstrained way. However, as the units of human capital are arbitrary, the rental price must be normalised to one in one year of the data, $R_1 = 1$. An estimable wage equation is obtained by substituting the expression for K_{it} in (116) into (115). The parameters to be estimated are the $\{\alpha\}$, the rental rates, $\{R_t\}_t^T = 2$, and the time dummies, $\{\tau_t\}$.

Shaw (1989) estimates the wage equation using data on white males, aged 18–64, from the 1968–1981 waves of the PSID. The instruments are a polynomial in current (i.e. dated at time *t*) wages and hours, along with schooling, age, the local unemployment rate, a South dummy and year dummies. The assumption is that these variables are uncorrelated with the person-specific productivity shock ε_{it} in (115).

It is worth commenting on the use of current hours h_{it} as an instrument. In general, we would expect the person-specific productivity shock ε_{it} to enter the decision rule for hours of work. For example, if ε_{it} is high, a person realises that his/her human capital is going to rise substantially at time $t+1$, even if he/she has low current hours of work. Thus, assuming there are diminishing returns to human capital, we would expect the person to work less at time *t*. Thus, under this scenario, current hours is not a valid instrument. The key assumption that would validate using hours as an instrument is if the person-specific productivity shock ε_{it} is not revealed until *after* the worker decides on current hours of work.

Another important point is that, unlike conventional studies in the human capital literature, the wage equation estimated here does not include an individual effect to capture a person's unobserved skill endowment. Shaw (1989) makes the point that this is not necessary here, because the lagged level of human capital proxies for unobserved ability.

Given (115), the derivative of human capital with respect to hours of work is:

$$\frac{\partial K_{i,t+1}}{\partial h_{it}} = \alpha_3 K_{it} + \alpha_4 + 2\alpha_5 h_{it}$$

The estimates are α_3 = 0.30, α_4 = –3.55 and α_5 = 0.69. To interpret these figures, let $R = 1$, and note that mean hours in the data is 2160 while the mean wage rate is \$3.91. Then, noting that h_{it} is defined as hours divided by 1000, we have, at the mean of the data:

$$\frac{\partial K_{i,t+1}}{\partial h_{it}} = (0.30)(3.91) - 3.55 + 2(0.69)(2.16) = 0.60$$

This implies, for example, that an increase of 500 in hours of work at time *t* (which is an increase in h_t of 0.5) would increase the wage rate at $t+1$ by 30 cents per hour. In percentage terms, this is a 23 per cent hours increase causing an 8 per cent wage increase. This is a very strong effect of work experience on wages; far stronger that the 'conservative' estimates I used for the back of the envelope calculations of the size of experience effects that I discussed earlier.

Notice that the positive estimate of α_3 implies that hours of work and human capital are complements in the production of additional human capital. That is, wages rise more quickly with work experience for high wage workers than low wage workers.

The estimates also imply that human capital rental rates are quite volatile, although the year-specific rental rates are quite imprecisely estimated. Interestingly, Shaw (1989) reports that the series of annual rental rates for the fourteen years of data has a correlation of –0.815 with an index for the price of fuel. This is consistent with other results (Keane 1993) showing that oil price movements in the 1970s and 1980s had very large effects on real wages in the United States.

Shaw (1989) estimates the first order condition (114) using a subset of the data (ten years) because the PSID did not collect food consumption data in 1967–1968 and 1975. The instruments, which are assumed uncorrelated with the forecast error ξ_{t+1}, include a fully interacted quadratic in the time *t* values of leisure (obtained by taking 8760 minus hours of work), food consumption, and the wage rate (constructed as annual earnings divided by annual hours). Also included are education, age, the local unemployment rate, a South dummy and time dummies.

The parameter estimates are reasonable, implying that the marginal utility of leisure and consumption are both positive, and with diminishing marginal returns. The coefficient on the cross term between consumption and leisure is negative, implying hours of work and consumption are complements. The discount factor is estimated to be 0.958. More interesting, however, are the simulations of the model.

Unfortunately, first order conditions like (114) are generally inadequate to simulate the behaviour of workers in a life-cycle model. The problem is that the first order condition, combined with the law of motion for human capital (equation (115)) and the law of motion for assets ($A_{t+1} = (1+r)(w_t h_t - C_t + A_t)$), only tell us how hours, wages and assets move from one period to the next, *conditional on a particular starting point*. That is, conditional on some levels of assets, the wage rate, hours and consumption at $t=1$, these equations can be used to simulate what levels of hours and consumption the worker would choose at $t=2$, as well as the new levels of assets and wages that these choices would lead to. So, from a particular assumed starting point it is possible to simulate the behaviour of a worker going forward. However, the assumed starting point is arbitrary. The first order conditions cannot be used to determine the *optimal* first period choices for the worker implied by the model. To achieve that we need to obtain what is known as a 'full solution' of the worker's dynamic optimisation problem, which I'll return to shortly.

This criticism is not particular to the model in Shaw (1989). Indeed, it applies to all of the methods based on estimating first order conditions of life-cycle models that I discussed earlier (e.g. MaCurdy 1983, Method 1), as well as to the life-cycle consistent methods of estimating labour supply equations (e.g. MaCurdy 1983, Method 2; Blundell & Walker 1986).[89] Furthermore, the criticism as I have stated

89 MaCurdy (1983) himself emphasised the limitations of all these approaches. As he stated:

Implementing the above procedures yields estimates required to formulate the lifetime preference function, but…this…is not sufficient to determine it omits the further problem that even to use the first order conditions to simulate forward from an arbitrary starting point, one still needs the know the distribution of the stochastic terms (e.g. the forecast error ξ_{t+1} in the case of equation (114)). The instrumental variables estimation techniques that are typically used to estimate first order conditions do not deliver estimates of the distributions of the stochastic terms of the model, making even this limited type of analysis infeasible.[90]

These problems explain why authors who have estimated dynamic models using first order conditions or life-cycle consistent methods have sometimes used the estimated preference parameters to simulate how workers would respond to tax policy changes under the hypothetical situation that they live in a static world (with a static budget constraint). An example of this was MaCurdy (1983). In some cases such simulations are informative. For instance, in the simple life-cycle model of equations (16)–(17) the response of workers to a permanent anticipated tax change is given simply by the Marshallian elasticity of the static model (1), which is given by (8). But only in special cases will such an equivalence hold. It certainly will not hold in a model with human capital because if a tax change alters labour supply at time t it will also alter the pre-tax wage at $t+1$. Thus, the response to tax changes will generally differ by age.[91]

Consistent with the above discussion, Shaw (1989) conducts her simulations by choosing particular arbitrary $t=1$ values for wages, hours and assets, and setting the stochastic terms to zero. Despite these limitations, the simulations are interesting. Take a worker starting out at age 18 with a wage of \$3.30 per hour and working 2200 hours per year. The simulations imply that such a worker's wage would rise to roughly \$3.65 over the first eight years of employment (an 11 per cent increase), but his/her hours are essentially flat (in fact, they decline very slightly). This behaviour illustrates a somewhat extreme version of the phenomenon I described in Figure 7.4. Even though the wage increases by 11 per cent over the first eight years, the opportunity cost of time does not rise because the drop in the human capital return to experience is sufficient to outweigh it. As a result, hours do not rise. Thus, a researcher looking at these simulated data through the lens of a model that ignores human capital would conclude there is no inter-temporal substitution whatsoever in labour supply, yet we know that in the true model that generates the simulated data there is inter-temporal substitution.[92]

In a dynamic life-cycle model, simulating the behaviour of an agent over the whole life (including the initial period) requires not only the first order conditions as in (114), but a 'full solution' of the dynamic optimisation problem. Full solution methods are discussed in detail in a number of references, including Eckstein and Wolpin (1989), Keane and Wolpin (1994) and Geweke and Keane (2001). A full solution requires constructing the value function at every point in the state space, which in the present case means at every possible level of human capital and savings. To see this, let's take the value function for the simple two-period model (104) and extend it to a multi-period setting (with uncertainty):

$$(117)\quad V_t(K_t, A_t, R_t, \beta_t) = \left[\frac{C_t^{1+\eta}}{1+\eta} - \beta_t \frac{h_t^{1+\gamma}}{1+\gamma}\right] + E_t\left\{\sum_{\tau=t+1}^{T} \rho^{\tau-t}\left[\frac{C_\tau^{1+\eta}}{1+\eta} - \beta_\tau \frac{h_\tau^{1+\gamma}}{1+\gamma}\right]\middle| (K_{t+1}, A_{t+1})\right\}$$

The value function now has a t subscript as it is specific to time period t, as opposed to being a lifetime value function. The arguments of the value function, which determine a worker's current state, are human capital, assets, the human capital rental rate and tastes for work.

The first term on the right-hand side of (117) is utility at time t, as opposed to time $t=1$ in equation (104). And the time $t=2$ term on the right-hand side of (104) is replaced by the expected present value of utility in all periods from $t+1$ until the terminal period T. For expositional simplicity I will assume that uncertainty arises from only two sources: the rental rate R_t and tastes for work β_t evolving stochastically over time.[93] As is common in these types of models, I assume the stochastic terms are independent over time.[94]

The notation $E_t\{\cdot \mid (K_{t+1}, A_{t+1})\}$ indicates that the expectation is taken conditional on next period's state variables K_{t+1} and A_{t+1}. How can the worker take an expectation at time t based on variables dated $t+1$? The model is set up so that human capital and assets evolve deterministically. That is, the current human capital and asset

how a consumer will respond to various shifts in budget or asset accumulation constraints, such as those arising from changes in wages or in tax policies…To form predictions for such responses, it is necessary to introduce sufficient assumptions to provide for a complete…formulation of the lifetime optimization problem…which, in addition to a function for preferences, requires a full specification for a consumer's expectations regarding current and future opportunities…Given a particular formulation for the lifetime optimization problem, one…[can conduct]…simulation analysis which involves numerically solving the consumer's optimization problem for the different situations under consideration.

The numerical procedure that MaCurdy describes here is what I refer to as a 'full solution' of the optimisation problem.

90 Keane (2009c) develops an estimation method that involves estimating the distribution of stochastic terms that enter first order conditions.

91 Indeed, Keane (2009b) argues that, in a model with human capital, tax changes cannot be viewed as inducing exogenous changes in after-tax wages because the worker's labour supply response to the tax change affects his/her wage, rendering the wage change endogenous.

92 Shaw (1989) admits that her model actually provides a rather poor fit to the data because hours for youth do in fact exhibit a moderate rise in the first several years after they enter the labour market. She attributes this to factors omitted from the model. Note, however, that it is the very large experience return in her model that drives this result, by causing the opportunity cost of time to greatly exceed the wage at $t=1$.

93 Uncertainty, and hence the need to take an expectation of the time $t+1$ outcome, may arise for a number of other reasons. For instance, wage uncertainty may also arise because there is some stochastic component to how human capital evolves, as in (115). And asset uncertainty may arise because there is some stochastic component to how assets evolve (i.e. interest rates are stochastic). All of these features may be incorporated fairly simply, but they would complicate the exposition.

94 Stochastic terms such as tastes for work are often assumed to consist of a part that is constant over time and a part that is stochastic. The constant part is no different from any other utility function parameter (i.e. η or γ).

level, when combined with today's choices of work hours and consumption, determine the next period's human capital and assets with certainty. Thus, given (K_t, A_t, C_t, h_t), the worker knows the resulting (K_{t+1}, A_{t+1}) with certainty. And these variables, in turn, help to predict future utility flows. For example, higher hours of work today (holding consumption fixed) will tend to generate higher values of human capital and assets at the start of $t+1$, and this will increase the expected present value of utility from $t+1$ to the terminal period.

In contrast to human capital and assets, I have assumed that the stochastic shocks to human capital rental rates and tastes for work are independent over time. As a result, R_t and β_t do not help to predict R_{t+1} and β_{t+1}. Hence, they are excluded from the conditioning set, as they do not help to predict future utility flows. Crucially, however, in forming the expected value of future utilities, the worker must average over all the possible realisations for rental rates and taste shocks.

Obviously, then, the expected present value on the right side of (117) is a very complicated object. One constructs these objects using a 'backsolving' procedure. Note that in the terminal period we have simply:

$$(118)\quad V_T(K_T, A_T, R_T, \beta_T) = \max_{C_T, h_T}\left\{\frac{C_T^{1+\eta}}{1+\eta} - \beta_T \frac{h_T^{1+\gamma}}{1+\gamma}\right\}$$

That is, since there is no future beyond T, we have a simple static problem. Given $w_T = R_T K_T$ and A_T, the consumer chooses consumption and hours of work to maximise utility at time T subject to the static budget constraint $C_T = w_T h_T (1-\tau_T) + A_T$ where A_T are assets at the start of period T.[95] The backsolving procedure starts by calculating $V_T(K_T, A_T, R_T, \beta_T)$ for every possible state (K_T, A_T, R_T, β_T) at which the worker might enter period T.

The solution to such a static problem for any particular state (K_T, A_T, R_T, β_T) is, of course, trivial. It is given by the equation:

$$(119)\quad \frac{\beta_T h_T^{\gamma}}{\left[w_T h_T (1-\tau_T) + A_T\right]^{\eta}} = w_T(1-\tau_T)$$

which can be easily solved for the optimal h_T via an iterative search procedure.[96] Once the optimal h_T has been determined, the optimal C_T is obtained from the budget constraint, and these are both plugged into (118) to obtain $V_T(K_T, A_T, R_T, \beta_T)$ at that particular state point.

A problem arises, however, because the number of possible levels of human capital, assets, the rental rate and tastes for work at the start of period T is extremely large, if not infinite. Thus, it is not computationally feasible to literally solve for $V_T(K_T, A_T, R_T, \beta_T)$ for every possible state (K_T, A_T, R_T, β_T). Keane and Wolpin (1994) develop an approach to this problem that has become quite commonly used in the literature on dynamic models. The idea is to obtain an approximate (rather than exact) solution to the optimisation problem.

The Keane and Wolpin (1994) method is simply to solve for $V_T(K_T, A_T, R_T, \beta_T)$ at a finite (and relatively small) subset of the possible state points. Denote these solutions by $V_T(K_T^d, A_T^d, R_T^d, \beta_T^d)$ for $d = 1,\ldots,D$. One then runs a regression of the $V_T(K_T^d, A_T^d, R_T^d, \beta_T^d)$ on some flexible function of the (K_T^d, A_T^d). Note that R_T^d and β_T^d should *not* be included in this regression, because the worker does not use these variables to forecast $V_T(K_T^d, A_T^d, R_T^d, \beta_T^d)$. The regression is meant to give a prediction of $V_T(K_T^d, A_T^d, R_T^d, \beta_T^d)$ based *only* on (K_T^d, A_T^d).

Once we have fit this regression, we can use it to predict or interpolate the value of $E_{T-1}\{V_T(K_T, A_T, R_T, \beta_T) \mid (K_T, A_T)\}$ at any desired state point (K_T, A_T), including, most particularly, at values of (K_T, A_T) that were not amongst those used to fit the regression. Thus, once we have fit this interpolating regression, we may then proceed as if $E_{T-1}\{V_T(K_T, A_T, R_T, \beta_T) \mid (K_T, A_T)\}$ is known for every possible state (K_T, A_T).

Let's denote the interpolating function that approximates $V_T(K_T, A_T)$ by:

$$\pi_T(K_T, A_T) \approx E_{T-1}\{V_T(K_T, A_T, R_T, \beta_T) \mid (K_T, A_T)\}$$
$$\text{where}\quad \frac{\partial \pi_T}{\partial K_T} > 0,\ \frac{\partial \pi_T}{\partial A_T} > 0$$

We need to assume that π_T is a smooth differentiable function of K_T and A_T (e.g. a polynomial) for the next step. For expositional convenience, let's assume the π_T function is the following simple function of K_T and A_T:

$$(120)\quad \pi_T(K_T, A_T) \approx \pi_{T0} + \pi_{T1}\ln K_T + \pi_{T2}\ln A_T$$

The next step of the backsolving process is to move back to period $T-1$. At that point we have that:

$$V_{T-1}(K_{T-1}, A_{T-1}, R_{T-1}, \beta_{T-1}) = \max_{C_{T-1}, h_{T-1}}\left\{\left[\frac{C_{T-1}^{1+\eta}}{1+\eta} - \beta_{T-1}\frac{h_{T-1}^{1+\gamma}}{1+\gamma}\right] + \rho E_{T-1}\left\{\left[\frac{C_T^{1+\eta}}{1+\eta} - \beta_T \frac{h_T^{1+\gamma}}{1+\gamma}\right] \middle| (K_T, A_T)\right\}\right\}$$

But if we substitute our approximating polynomial for the expectation term on the right we obtain simply:

$$(121)\quad V_{T-1}(K_{T-1}, A_{T-1}, R_{T-1}, \beta_{T-1}) \approx \max_{C_{T-1}, h_{T-1}}\left\{\left[\frac{C_{T-1}^{1+\eta}}{1+\eta} - \beta_{T-1}\frac{h_{T-1}^{1+\gamma}}{1+\gamma}\right] + \rho\pi_T(K_T, A_T)\right\}$$

Now this is actually a simple problem to solve. Substituting in the laws of motion for assets and human capital, which I will assume is simply $K_{t+1} = K_t(1+\alpha h_t)$, we obtain:

95 For exposition simplicity I am assuming that the end of the working life T corresponds to the end of life, and that there are no bequests. Hence, the worker consumes all of his/her remaining assets at time T. In a more general model, the worker might value carrying assets into $T+1$ as savings for retirement and/or to leave bequests. These extensions can be handled by adding to (118) an additional term $f(A_{T+1})$ that represents the value of assets carried into period $T+1$.

96 As an aside, it is notable that the basic idea of the life-cycle model with human capital—that working hard today will improve one's prospects tomorrow—is one which ordinary people would find quite intuitive. Yet one often hears academic economists argue that workers can't possibly behave as if they solve dynamic optimisation problems because the mathamatics involved is too daunting. On the other hand, I can't ever recall hearing an academic economist argue that people can't possibly behave as suggested by a static labour supply model because they can't solve an implicit equation for hours such as (119). I suspect that most people are not very good at solving implicit equations. Perhaps the idea is that most people are sufficiently familiar with Roy's identity that they can choose their indirect utility functions so as to give themselves simple linear labour supply functions.

$$V_{T-1} = \frac{C_{T-1}^{1+\eta}}{1+\eta} - \beta_{T-1}\frac{h_{T-1}^{1+\gamma}}{1+\gamma} + \rho\{\pi_{T0} + \pi_{T1}\ln K_{T-1}(1+\alpha h_{T-1}) + \pi_{T2}\ln(1+r)[w_{T-1}h_{T-1}(1-\tau_{T-1}) - C_{T-1} + A_{T-1}]\}$$

Notice that finding the optimal values of C_{T-1} and h_{T-1} is now just like a static optimisation problem. We have the first order conditions:

(122a) $$\frac{\partial V_{T-1}}{\partial h_{T-1}} = -\beta_{T-1}h_{T-1}^{\gamma} + \rho\pi_{T1}\frac{\alpha}{(1+\alpha h_{T-1})} + \rho\pi_{T2}\frac{w_{T-1}(1-\tau_{T-1})}{w_{T-1}h_{T-1}(1-\tau_{T-1}) - C_{T-1} + A_{T-1}} = 0$$

(122b) $$\frac{\partial V_{T-1}}{\partial C_{T-1}} = C_{T-1}^{\eta} - \rho\pi_{T2}\frac{1}{w_{T-1}h_{T-1}(1-\tau_{T-1}) - C_{T-1} + A_{T-1}} = 0$$

Solving these two equations with two unknowns is numerically not much harder than solving for hours in equation (119).

Thus, given the interpolating function $\pi_T(K_T, A_T)$ we have a simple way to solve for $V_{T-1}(K_{T-1}, A_{T-1}, R_{T-1}, \beta_{T-1})$ at any state point $(K_{T-1}, A_{T-1}, R_{T-1}, \beta_{T-1})$ that might arise at $T-1$. The next step of the approximate solution method proposed by Keane and Wolpin (1994) is to solve for $V_{T-1}(K_{T-1}, A_{T-1}, R_{T-1}, \beta_{T-1})$ at a finite subset of the possible state points. Denote these solutions by $V_{T-1}(K_{T-1}^d, A_{T-1}^d, R_{T-1}^d, \beta_{T-1}^d)$ for $d = 1,\ldots,$D. One then obtains a new interpolating function $\pi_{T-1}(K_{T-1}, A_{T-1})$ by running a regression of the $V_{T-1}(K_{T-1}^d, A_{T-1}^d, R_{T-1}^d, \beta_{T-1}^d)$ on a flexible function of the (K_{T-1}^d, A_{T-1}^d). Using this interpolating function, we can write the (approximate) value functions at time $T-2$ as:

(123) $$V_{T-2}(K_{T-2}, A_{T-2}, R_{T-2}, \beta_{T-2}) \approx \max_{C_{T-2}, h_{T-2}}\left\{\left[\frac{C_{T-2}^{1+\eta}}{1+\eta} - \beta_{T-2}\frac{h_{T-2}^{1+\gamma}}{1+\gamma}\right] + \rho\pi_{T-1}(K_{T-1}, A_{T-1})\right\}$$

Notice that this is exactly like equation (121), the expression for the (approximate) value functions at time $T-1$, except that we have a new polynomial with different coefficients. The first order conditions for C_{T-2} and $h_{T=2}$ will look exactly like (122), except with different π parameters. Thus, we can keep repeating the above steps until we have obtained an approximate solution for every period back to $t = 1$.

The approximate solution consists of the complete set of interpolating functions $\pi_t(K_t, A_t)$ for $t = 2,\ldots,T$. Using these interpolating functions we can solve simple two equation systems such as (122) to find the optimal choice of a worker at any point in the state space. In particular, using $\pi_2(K_2, A_2)$ we can solve for optimal labour supply and consumption in period $t = 1$, the first period of the working life. As I discussed earlier, this is what first order conditions alone do not allow one to do. Furthermore, by drawing values for the taste shocks and rental rates and repeatedly solving equations such as (122) over time, one can simulate entire career paths of workers. This in turn, enables one to simulate how changes in tax rates would affect the entire life-cycle path of labour supply and consumption, as one can re-solve the model and simulate career paths under different settings for the tax parameters.

To my knowledge there are only two papers that have used full solution methods to estimate dynamic life-cycle labour supply models that include *both* human capital investment and savings. These are Keane and Wolpin (2001) and Imai and Keane (2004).

Keane and Wolpin (2001) set up a model where a person, from age 16–65, decides every period whether to work and/or attend school either full-time, part-time or not at all. The choices are not mutually exclusive (e.g. a youth might work part-time while attending college). Somewhat unusually in the literature on life-cycle models, the authors use a model with three decision periods per year (the two school semesters and the summer). They allow for the fact that youth could work summers to finance school. The model is fit to panel data from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY79). This contains people who were 14–21 years old in January 1979. The sample used in estimation consists of 1,051 white males who are followed from when they first reach age 16 until 1992. The maximum age attained in the sample is 30. The NLSY79 collected comprehensive asset data beginning in 1985, making it possible to estimate a model that includes savings. A key feature of the Keane and Wolpin (2001) model is that, while it includes savings, it also allows for liquidity constraints (i.e. an upper bound on uncollateralised borrowing). The model fits data on assets, school attendance and work from age 16–30 quite well.

The focus of the Keane and Wolpin (2001) paper is not on labour supply, it is on school attendance decisions. But the paper is of interest here because it assumes a CRRA utility function in consumption, and so it provides an estimate of the key preference parameter η which governs income effects and inter-temporal substitution in consumption. Keane and Wolpin (2001) obtain $\eta \approx -0.50$, which implies weaker income effects, and less curvature in consumption (i.e. a higher willingness to substitute inter-temporally), than much of the prior literature. Keane and Wolpin (2001, p. 1078) discuss how failure to accommodate liquidity constraints may have led to a downward bias in estimates of η in prior work.[97] Notably, Goeree, Holt and Palfrey (2003) present extensive experimental evidence, as well as evidence from field auction data, in favour of $\eta \approx -0.4$ to -0.5. Bajari and Hortacsu (2005) estimate $\eta \approx -0.75$ from auction data. All of these estimates are closer to the Keane and Wolpin (2001) estimate of –0.5 than to the larger negative values obtained in most prior literature on consumption.

The model in Imai and Keane (2004) is in most respects very similar to the model that I used to exposit full solution methods for life-cycle models. The main difference is that Imai and Keane (2004) use a much richer specification for the human capital production function. Their specification is designed to capture the empirical regularity that wages grow much more quickly with work experience for high wage workers than for low wage workers. Thus, as in Shaw (1989), they specify a function that allows hours of work

97 Specifically, in the absence of constraints on uncollateralised borrowing, one needs a large negative η to rationalise why youth with steep age-earnings profiles don't borrow heavily in anticipation of higher earnings in later life.

and human capital to be complements in the production of additional human capital, and their estimates imply that they are indeed complements. The parameters of the human capital production function are also allowed to differ by education level.

Another difference is that in Imai and Keane (2004) the rental rate on human capital is assumed to be constant. Instead, they allow human capital to evolve stochastically, similar to equation (115). Finally, Imai and Keane (2004) set the terminal period T at age 65, but unlike the simple expositional model above, they include a terminal value function $V_{66}(A_{66})$ which captures the fact that workers value having assets to carry into retirement.

Like Keane and Wolpin (2001), Imai and Keane estimate their model using white males from the NLSY79. They argue that the NLSY79 is preferable to the PSID for this purpose because of its comprehensive asset data. The men in their sample are aged 20–36 and, as the focus of their paper is solely on labour supply, the men included in the sample are required to have finished school. Due to the computational burden of estimation they randomly chose 1000 men from the NLSY79 sample to use in estimation. People are observed for an average of 7.5 years each, and not necessarily starting from age 16.

Imai and Keane (2004) allow for measurement error in observed hours, earnings and assets when constructing the likelihood of the data given their model. They use a ratio wage measure, but account for the resultant denominator bias in an internally consistent way when forming the likelihood. Given that all outcomes are assumed to be measured with error, construction of the likelihood is fairly simple. One can (i) simulate career histories for each worker, and then (ii) form the likelihood of a worker's observed history of hours, earnings and assets as the joint density of the set of measurement errors necessary to reconcile the observed history with the simulated data.[98]

Imai and Keane (2004) estimate that γ = 0.26. In a model without human capital this would imply a Frisch elasticity of $(1/\gamma)$ = 3.8, which implies a much higher willingness to substitute labour inter-temporally than in any estimation we have discussed so far, with the exception of MaCurdy (1983). Imai and Keane explain their high estimate of inter-temporal substitution based on the logic of Figure 7.4, which, as I discussed earlier illustrates why the failure to account for human capital will lead to severely downward biased estimates of $(1/\gamma)$.

Indeed, Imai and Keane show that if they simulate data from their model, and apply instrumental variable methods like those in MaCurdy (1981) and Altonji (1986) to estimate $(1/\gamma)$, they obtain estimates of 0.325 (standard error = 0.256) and 0.476 (standard error = 0.182), respectively. This exercise demonstrates that the model generates life-cycle histories that, when viewed through the lens of models that ignore human capital, imply similarly low inter-temporal elasticities of substitution to those that had been obtained in most prior work. In other words, the model does not generate data that show an oddly high level of positive co-movement between hours and wages compared to the actual data.

As further confirmation of this point, the authors report simple OLS regressions of hours changes on wage changes for both the NLSY79 data and the data simulated from their model. The estimates are –0.231 and –0.293, respectively. This shows two things: (i) the model does do a good job of fitting the raw correlation between hours changes and wage changes observed in the data; and (ii) a negative correlation between hours changes and wage changes in the raw data is perfectly consistent with a high willingness to substitute labour inter-temporally over the life-cycle.

What reconciles these *prima facie* contradictory observations is the divergence between the opportunity cost of time and the wage in a world with returns to work experience. In particular, Imai and Keane (2004) estimate that from age 20–36 the mean of the opportunity cost of time increases by only 13 per cent. In contrast, the mean wage rate increases by 90 per cent in the actual data, and 86 per cent in the simulated data. Thus, the wage increases about 6.5 times faster than the opportunity cost of time. As indicated in the 'back of the envelope' calculations based on equations (109) or (110) these figures imply that conventional methods of calculating $(1/\gamma)$ will understate the opportunity cost of time by a factor of roughly 6.5.[99]

Imai and Keane (2004) also estimate that $\eta = -0.74$. Note that their estimate of η implies a somewhat lower inter-temporal elasticity of substitution in consumption than the Keane and Wolpin (2001) estimate of $\eta \approx -0.5$ (i.e. $(1/\eta) = 1/(-0.74) = -1.35$ versus $1/(-0.5) = -2$).[100] But η is still less negative than in most prior estimates, again implying weaker income effects, and a higher willingness to substitute consumption inter-temporally, than much of the prior literature.

To put the Imai and Keane estimates of γ and η in a familiar context, we can follow MaCurdy's (1983) method and calculate what they imply for the behaviour of a worker with such preferences living in a static world. Intuition suggests that this method may be approximately correct for calculating the effect of a permanent unanticipated tax change on a worker far enough into the life-cycle that the human capital return part of the opportunity cost of time is fairly small relative to the wage rate. The results in Heckman (1976) suggest this occurs in the 40s, and the simulations in Imai and Keane are consistent with this (i.e. at age 45 the return to human capital makes up only 15 per cent of the OCT). In this static context, the implied

98 Keane and Wolpin (2001) first developed this approach to forming the likelihood in dynamic models.

99 It is interesting that French (2005), in a study of retirement behaviour, also obtains a rather large value of $(1/\gamma)$ = 1.33 for the inter-temporal elasticity of substitution for 60-year-old participants in the PSID. As both Shaw (1989) and Imai and Keane (2004) note, human capital investment is not so important for people late in the life-cycle. For them, the wage will be close to the opportunity cost of time, and the bias that results from ignoring human capital will be much less severe.

100 Recall that the inter-temporal elasticity of substitution in consumption will measure the drop in current consumption in response to an increase in the interest rate (i.e. the willingness to sacrifice current consumption for higher future consumption).

Marshallian elasticity with respect to a permanent wage increase would be $(1 + \eta)/(\gamma - \eta) = 0.26/(0.26 + 0.74) = 0.26$ and the Hicks elasticity would be $1/(\gamma - \eta) = 1/(0.26 + 0.74) = 1.0$.

However, simulations of the Imai and Keane model suggest this intuition is not particularly helpful. Specifically, I have used the Imai and Keane model to simulate the impact of a permanent unanticipated 10 percentage point tax on labour earnings for men at age 45, 50, 55 and 65, respectively. Under a scenario where the revenue is simply thrown away, the estimated labour supply effects are –1.1 per cent, –2.3 per cent, –5.3 per cent and –9.5 per cent respectively. Only at age 50 is the impact roughly what the Marshallian elasticity would suggest.

I have also used the Imai and Keane model to simulate the effect of a permanent 10 per cent tax rate increase (starting at age 20 and lasting through to age 65) on labour supply over the entire working life. If the revenue is simply thrown away the model implies that average hours of work from ages 20–65 drops from 1,992 per year to 1,954 per year, a 2 per cent drop. If the revenue is redistributed as a lump sum transfer labour supply drops to 1,861 hours per year, a 6.6 per cent drop. I'll treat this as a reasonable approximation to the compensated elasticity with respect to permanent tax changes implied by the model.

As we would expect, however, the effects are very different at different ages, as Table 7.5 indicates. As we see in the table, tax effects on labour supply are slowly rising from age 20 to about age 40. Starting in the 40s, the effects on labour supply start to grow quite quickly, and by age 60 effects are very substantial. Thus, in response to a permanent tax increase, workers not only reduce labour supply, but also shift their lifetime labour supply out of older ages towards younger ages.

Imai and Keane (2004) also simulate how workers would respond to a 2 per cent temporary and unanticipated wage increase. This generates primarily an inter-temporal substitution effect, as such a short-lived wage increase will have a small effect on lifetime wealth (at least for relatively young workers). For a person at age 20 the increase in the hours is only 0.6 per cent, which, in contrast to the estimate of $(1/\gamma) = 3.8$, would seem to imply rather weak inter-temporal substitution effects. The answer lies in the fact that, according to the Imai and Keane (2004) estimates, at age 20 the wage is less than half of the opportunity cost of time. As we would expect, the strength of the substitution effect rises steadily with age, and at age 60 the increase in hours is nearly 4 per cent, and at age 65 it is about 5.5 per cent. It is important to bear in mind that these are lower bounds on the Frisch elasticities, as, particularly at older ages, the wealth effect of a one-period wage increase may be considerable.

Table 7.5 Effects of a 10 Per Cent Tax on Earnings on Labour Supply at Various Ages

Age	Pure tax (%)	Tax plus lump sum redistribution (%)
20	–0.7	–3.2
30	–0.7	–3.3
40	–0.9	–4.2
45	–1.2	–5.7
50	–2.1	–8.7
60	–9.1	–20.0
20–65 (total hours)	–2.0	–6.6

Finally, in another paper (Keane 2009b), I use the Imai and Keane (2004) and Keane and Wolpin (2001) estimates of the preference parameters γ and η to calibrate the simple two-period model of equation (17), and then use this to provide some simulations of the welfare cost of income taxation. To do this I augment the model to include a public good P that is financed by taxation, as in:

$$(123)\quad V = \lambda f(P) + \frac{[w_1 h_1(1-\tau) + b]^{1+\eta}}{1+\eta} - \beta \frac{h_1^{1+\gamma}}{1+\gamma} + \rho \left\{ \lambda f(P) + \frac{[w_2 h_2(1-\tau) - b(1+r)]^{1+\eta}}{1+\eta} - \beta \frac{h_2^{1+\gamma}}{1+\gamma} \right\}$$

where $\lambda f(P)$ indicates people value the public good. The government provides the same level of the public good P in both periods, and the government budget constraint requires that the present value $P + P/(1 + r)$ equals the present value of tax revenues. The benevolent government sets the tax rate optimally to equate marginal utility of consumption of the public and private goods.[101]

Given that we have a two-period model we can think of each period as twenty years of a forty-year working life (e.g. 25–44 and 45–64). The real annual interest rate is set at 3 per cent, giving a twenty-year interest rate of $r = 0.806$, and the discount factor is set to $\rho = 1/(1 + r) = 0.554$. I set the initial tax rates $\tau_1 = \tau_2 = 0.40$. The wage equation is similar to (103), but augmented to include a quadratic in hours and to accommodate depreciation of skills. The wage equation parameters are calibrated so that the simulations are consistent with roughly 33–50 per cent wage growth for men from age 25–45, which is comparable to what Geweke and Keane (2000) find in the PSID.

Table 7.6 summarises some of the main results from Keane (2009b). The table presents welfare losses from a proportional flat rate income tax, relative to a lump sum tax, expressed as a fraction of consumption, under a number of parameterisations of the simple two-period model. The top panel presents results with the CRRA curvature parameter for consumption (η) set at –0.75, the value estimated by Imai and Keane (2004), while the bottom panel presents results for the value of –0.5 estimated by Keane and Wolpin (2001).[102] Each panel presents results for several values of the CRRA curvature parameter in hours (γ), from a value of 4, which implies little inter-temporal substitution in leisure, up to a value of 0.25, which implies an inter-temporal elasticity of substitution of labour supply of 4, close to the Imai and Keane (2004) estimate.

Under the column labelled 'uncompensated elasticity' the table reports simulated total labour supply elasticities to

101 In the solution, workers ignore the effect of their own actions on P, as each worker makes a trivial contribution to total government revenue. Thus, workers continue to solve equations (105)–(107).

102 As I indicated earlier, these are the only two dynamic life-cycle models for men that include both labour supply and asset accumulation and that are estimated using a full solution method.

Table 7.6 Summary Results for Welfare Losses from Proportional Income Taxes

	γ	Uncompensated elasticity	Compensated elasticity	Welfare loss (C*)		
				f(P) = log(P)	f(P) = $2P^{0.5}$	f(P) = P
η = –0.75	0.25	0.205	0.811	13.35	19.03	35.33
	0.5	0.176	0.698	11.42	15.47	27.57
	1	0.133	0.530	8.92	11.46	19.36
	2	0.088	0.350	6.22	7.62	12.18
	4	0.052	0.206	3.87	4.59	6.98
η = –0.5	0.25	0.532	1.054	12.16	23.08	59.27
	0.5	0.445	0.884	11.38	18.11	41.62
	1	0.318	0.633	9.30	12.83	26.23
	2	0.197	0.392	6.43	8.07	14.90
	4	0.110	0.220	3.88	4.62	7.88

Notes: All results are for α = 0.008. C* = percentage consumption gain needed to compensate for tax distortion (starting from proportional tax world).

permanent tax changes.[103] Note that very high values of the Frisch elasticity $(1/\gamma)$ are consistent with very modest uncompensated elasticities. For example, in the $\eta = -0.75$, $\gamma = 0.25$ case, which corresponds to the Imai and Keane (2004) estimates, the simulated uncompensated elasticity is a modest 0.205. (Interestingly, this is almost identical to the uncompensated elasticity that I obtained when simulating a permanent tax increase in the Imai-Keane multi-period model.)[104] But the welfare cost of proportional income taxation is still substantial (i.e. 13–35 per cent, depending on the measure).

The welfare cost of income taxation is calculated for three cases: (i) where utility is $\log(P)$, where P is the amount of the private good, (ii) where it is $2P^{0.5}$, and (iii) where it is linear in P. This covers a range of degrees of curvature in consumers' utility from the public good, ranging from more than that for the private good to less. The welfare losses in the three cases are equivalent to 13 per cent, 19 per cent and 35 per cent of consumption, respectively.

Even if we reduce $(1/\gamma)$ to the much more modest value of 1, in which case the uncompensated elasticity is only 0.133, the welfare losses in the three cases are 9 per cent, 11 per cent and 19 per cent of consumption, respectively. Thus, it appears that large welfare losses from income taxation are quite consistent with existing (small) estimates of labour supply elasticities.

7.4 Conclusion

The literature on male labour supply is vast, and the number of contentious methodological issues is sizeable. It is therefore impossible to arrive at a simple summary. One very crude way to summarise the literature is to provide a table that reports all the elasticity estimates from all the papers I have discussed. I do this in Table 7.7. In many ways such a table is useless because it makes no attempt to weigh studies based on their relative merits (quality of data, soundness of approach, etc.). Thus, Table 7.7 in effect ignores all the important issues I have been talking about in section 7.3.

On the other hand, Table 7.7 is useful for answering the following question: Is there a clear consensus in the literature on male labour supply that the Hicks elasticity is small? Recall the quote in section 7.1 where Saez, Slemrod and Giertz (2009) indicated that, '…with some exceptions, the profession has settled on a value for [the Hicks] elasticity close to zero'.[105] But, as we see in Table 7.7, the mean value of the Hicks elasticity across the twenty-one studies reviewed here is 0.30. (Note that seven studies do not estimate this parameter.)

As we have seen, a value of 0.30 for the Hicks elasticity is large enough to generate substantial welfare costs of taxation. For instance, Ziliak and Kniesner (2005) obtain a Hicks elasticity of 0.33, and simulations of their model imply substantial welfare costs from progressive taxation. And Blomquist (1983) and Blomquist and Hansson-Busewitz (1990) obtain Hicks elasticities of only 0.11 and 0.13, respectively, yet they also simulate substantial welfare costs from progressive taxation (i.e. 12 per cent and 16 per cent of revenue, respectively, compared to only 2 per cent or 5 per cent under a flat rate tax). Similarly, Ziliak and Kniesner (1999) obtain a Hicks elasticity of 0.13, yet also simulate large welfare losses from taxation. Based on these results, one would have to conclude that a Hicks elasticity of 0.30 is quite large enough to generate substantial welfare losses.

Table 7.7 also shows us that the Hicks elasticity estimates from individual studies range from 0.02 to 1.22, with eight estimates exceeding 0.25. Interestingly, the other thirteen estimates fall in a tight range from 0.02 to 0.13. Thus, the distribution of the estimates across studies has a very odd shape (see Figure 7.5). It is interesting that the distribution exhibits a large gap between 0.13 and the next highest value of 0.27. Regardless, I think it would be difficult to look at Figure 7.5 and conclude there is a broad consensus within the economics profession that the Hicks elasticity is close to zero—unless, that is, one believes that all the studies bunched up in the 0.02 to 0.13 range are credible while all those in the 0.27 plus range are flawed.

103 It is important to note that the compensated and uncompensated elasticities reported in Table 7.6 are not the traditional Marshallian and Hicks elastictities. Instead they are generalisations of these formulas that apply for the dynamic case with human capital, as given in Keane (2009b).

104 The Hicks elasticity of 0.81 in Table 7.6 is also fairly close to the value of 0.66 that I obtained when simulating a tax increase with the proceeds distributed lump sum in the Imai-Keane multi-period model.

105 At that point I didn't note that the authors were specifically referring to the Hicks elasticity, as I had not yet defined the different elasticity concepts.

Table 7.7 Summary of Elasticity Estimates

Authors of study	Year	Marshall	Hicks	Frisch
Static models				
Kosters	1969	–0.09	0.05	–
Ashenfelter-Heckman	1973	–0.16	0.11	–
Boskin	1973	–0.29	0.12	–
Hall	1973	n.a.	0.45	–
8 British studies[a]	1976–83	–0.16	0.13	–
8 NIT studies[a]	1977–84	0.03	0.13	–
Burtless-Hausman	1978	0.00	0.07–0.13	–
Wales-Woodland	1979	0.14	0.84	–
Hausman	1981	0.00	0.74	–
Blomquist	1983	0.08	0.11	–
Blomquist-Hansson-Busewitz	1990	0.12	0.13	–
MaCurdy-Green-Paarsch	1990	0.00	0.07	–
Triest	1990	0.05	0.05	–
van Soest-Woittiez-Kapteyn	1990	0.19	0.28	–
Ecklöf-Sacklén	2000	0.05	0.27	–
Dynamic models				
MaCurdy	1981	0.08[b]	–	0.15
MaCurdy	1983	0.70	1.22	6.25
Browning-Deaton-Irish	1985	–	–	0.09
Blundell-Walker	1986	–0.07	0.02	0.03
Altonji[c]	1986	–0.24	0.11	0.17
Altonji[d]	1986	–	–	0.31
Bover	1989	0.00	–	0.08
Altug-Miller	1990	–	–	0.14
Angrist	1991	–	–	0.63
Ziliak-Kniesner	1999	0.12	0.13	0.16
Pistaferri	2003	0.51[b]	–	0.70
Imai-Keane	2004	0.20[e]	0.66[e]	0.30–2.75[f]
Ziliak-Kniesner	2005	–0.47	0.33	0.54
Average		0.03	0.30	0.83

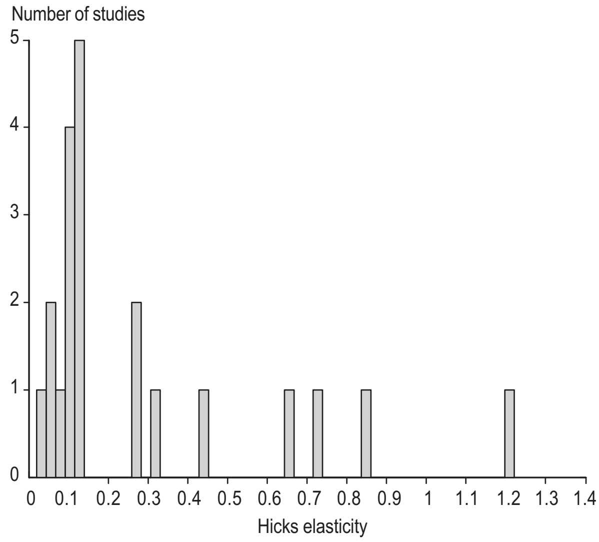
Notes: Where ranges are reported, the mid-point is used to take means. a = average of the studies surveyed by Pencavel (1986). b = effect of surprise permanent wage increase. c = using MaCurdy Method 1. d = using first difference hours equation. e = approximation of responses to transitory wage increase based on model simulation. f = age range. n.a. denotes not available.

I think such a position would be untenable, particularly as we can point to important flaws in all the studies in the 0.02–0.13 range (just as in all empirical work).[106]

The notion that there is consensus on a low Hicks elasticity may stem in part from a popular perception (which is a misconception) that the piecewise-linear budget constraint methods developed by Burtless and Hausman (1978), Wales and Woodland (1979) and Hausman (1980, 1981) have been discredited, and that the high estimates all come from use of these methods. But as I have discussed, a careful reading of literature suggests that this is not the case. These methods have sometimes produced low estimates of the Hicks elasticity, while alternative methods have sometimes produced high estimates. There isn't an obvious connection between the methods adopted and the result obtained.

Indeed, as the careful study by Eklöf and Sacklén (2000) showed, divergent results across studies can be much better explained by the data used than by the particular empirical methods employed. In particular, they find that studies that use 'direct wage measures' (i.e. a question about one's wage rate per unit of time, such as hourly or weekly or monthly) tend to get much higher estimates of labour supply elasticities than studies that use 'ratio wage measures' (i.e. annual earnings divided by annual hours).

Figure 7.5 Distribution of Hicks Elasticity of Substitution Estimates



Note: This figure contains a frequency distribution of the twenty-one estimates of the Hicks elasticity of substitution discussed in this chapter.

106 For example, Kosters (1969) does not account for endogeneity of wages, Ashenfelter and Heckman (1973) do not account for taxes, MaCurdy, Green and Paarsch (1990) and Triest (1990) use ratio wage measures that would lead to denominator bias, Blundell and Walker (1986) do not instrument for full income, and so on.

This is because the denominator bias inherent in taking the ratio biases the wage coefficient in a negative direction.

This pattern can be seen quite clearly in Table 7.7. Specifically, of the eight studies that obtain 'large' values for the Hicks elasticity (i.e. those in the 0.27 plus range), six use a direct wage measure (Hall 1973; Hausman 1981; van Soest, Woittiez & Kapteyn 1990; MaCurdy 1983[107]; Eklöf & Sacklén 2000; Ziliak & Kneisner 2005), one works with shares to avoid ratios (Wales & Woodland 1979), and one models the measurement error process to take denominator bias into account in estimation (Imai & Keane 2004).

So far I have argued that the existing literature supports an estimate of the Hicks elasticity of at least 0.30, and perhaps higher if one puts more weight on studies that have used direct wage measures.[108] However, a second point I have stressed is that the failure of prior literature to account for human capital has almost certainly caused downward bias in estimates of labour supply elasticities. The effect of human capital is to dampen the response of younger workers to changes in their wage rates. This is because, for them, the wage is a relatively small part of the opportunity cost of time. It is also very important to consider the return to work experience. This notion is quite intuitive: young workers are often willing to work long hours at relatively low wages in order to increase their chances of advancement, and hence their future wages.

Aside from leading to downward bias in estimates of wage elasticities, human capital also has important implications for tax policy. The conventional wisdom is that a temporary tax increase (or wage reduction) should have a larger effect on current labour supply than a permanent increase. This is because of inter-temporal substitution, that is, workers can substitute their labour towards periods when wages are high. But once we consider human capital it becomes possible that a permanent tax increase can have a larger adverse effect on current labour supply than a temporary tax increase. This is because the permanent tax increase reduces not just the current wage but also the return to human capital investment. It reduces one's future reward for current work.

Another key point that I have emphasised is that the 'labour supply elasticities' reported in the labour supply literature are in most cases hypothetical objects. Specifically, authors typically report elasticities relevant for changes in hypothetical straight line budget constraints. This tells us something about the shape of people's indifference curves for consumption versus leisure. But in the real world, budget constraints tend to be non-convex at the low end (i.e. effective tax rates of over 100 per cent due to fixed costs of work and welfare benefit withdrawal rates) and progressive at the middle to high end. Reported elasticity estimates do not reveal how people will respond to specific tax and transfer program changes given real world tax systems—determining this response requires model simulation.

In this regard, I gave a specific example where large changes in the benefit withdrawal rate of a welfare program would have essentially no effect on labour supply; perhaps leading researchers to conclude that program recipients were unresponsive to economic incentives such as after-tax wage rates. But I showed in this same example that the labour supply of recipients would be extremely responsive to small changes in their pre-tax wage rates, changes in fixed costs of work, or the provision of work bonuses. I argued that this example was not merely academic but is, in fact, an accurate description of why the US welfare caseload dropped so dramatically in the mid to late 1990s after being so seemingly immune to changes in program rules over the previous thirty years.

References

Altonji, J 1986, 'Intertemporal substitution in labor supply: evidence from micro data', *Journal of Political Economy*, vol. 94, June (part 2), pp. S176–215.

Altug, S & Miller, R 1990, 'Household choices in equilibrium', *Econometrica*, vol. 58, pp. 543–70.

Angrist, J 1991, 'Grouped-data estimation and testing in simple labor-supply models', *Journal of Econometrics*, vol. 47, pp. 243–66.

Ashenfelter, O & Heckman, J 1973, 'Estimating labor supply functions', in G Cain & HW Watts (eds), *Income maintenance and labor supply*, Markham, Chicago, pp. 265–78.

Bajari, P & Hortacsu, A 2005, 'Are structural estimates of auction models reasonable?—evidence from experimental data', *Journal of Political Economy*, vol. 113, no. 4, pp. 703–41.

Ben-Porath, Y 1967, 'The production of human capital and the life cycle of earnings', *Journal of Political Economy*, vol. 75, August (part 1), pp. 352–65.

Bernal, R & Keane, M 2009, *Child care choices and children's cognitive achievement: the case of single mothers*, working paper, University of Technology Sydney.

Blomquist, S 1983, 'The effect of income taxation on the labour supply of married men in Sweden', *Journal of Public Economics*, vol. 22, no. 2, pp. 169–97.

——1985, 'Labor supply in a two-period model: the effect of a nonlinear progressive income tax', *Review of Economic Studies*, vol. 52, pp. 515–24.

——& Hansson-Busewitz, U 1990, 'The effect of taxes on male and female labor supply in Sweden', *Journal of Human Resources*, vol. 25, no. 3, pp. 317–57.

Blundell, R & MaCurdy, TH 1999, 'Labor supply: a review of alternative approaches', in O Ashenfelter & D Card (eds), *Handbook of labor economics*, vol. 3, North-Holland, Amsterdam, pp. 1559–695.

——Pistaferri, L & Preston, I 2001, *Partial insurance and consumption dynamics*, working paper, University College London.

——& Walker, I 1986, 'A life cycle consistent empirical model of family labour supply using cross-sectional data', *Review of Economic Studies*, vol. 53, no. 4, pp. 539–58.

Boskin, MJ 1973, 'The economics of labor supply functions', in G Cain & HW Watts (eds), *Income maintenance and labor supply*, Markham, Chicago, pp. 265–78.

107 In the Denver experiment, workers were asked a direct question about their wage rate every month. MaCurdy (1993) is a bit vague about how he constructed his wage measure, but from his description I believe he took an average of the answers to these monthly questions over twelve months to get an annual wage.

108 There are thirteen studies in total that I can determine with some confidence either used direct wage measures or made some attempt to deal with the denominator bias problem. In addition to the eight cited above, these include Burtless and Hausman (1978), Blomquist (1983), Blomquist and Hansson-Busewitz (1990), Blundell and Walker (1986), and Ziliak and Kneisner (1999). The average Hicks elasticity among this group is 0.41.

Bover, O 1989, 'Estimating intertemporal labour supply elasticities using structural models', *The Economic Journal*, vol. 99, December, pp. 1026–39.

Brewer, M, Saez, E & Shepard, A 2008, 'Means-testing and tax rates on earnings', prepared for the Report of a Commission on Reforming the Tax System for the 21st Century, chaired by Sir James Mirrlees, Institute for Fiscal Studies, London.

Browning, M, Deaton, A & Irish, M 1985, 'A profitable approach to labor supply and commodity demands over the life-cycle', *Econometrica*, vol. 53, no. 2, pp. 503–44.

Burtless, G & Hausman, J 1978, 'The effects of taxation on labor supply: evaluating the Gary income maintenance experiment', *Journal of Political Economy*, vol. 86, pp. 1103–30.

Colombino, U & del Boca, D 1990, 'The effect of taxes on labor supply in Italy', *Journal of Human Resources*, vol. 25, no. 3, pp. 390–414.

Deaton, A 1982, 'Model selection procedures, or, does the consumption function exist?', in G Chow & P Crosi (eds), *Evaluating the reliability of macroeconomic models*, John Wiley, New York, pp. 43–65.

Diamond, P 1998, 'Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates', *American Economic Review*, vol. 88, no. 1, pp. 83–95.

Eckstein, Z & Wolpin, KI 1989, 'The specification and estimation of dynamic stochastic discrete choice models', *Journal of Human Resources*, vol. 24, pp. 562–98.

Eklöf, M & Sacklén, H 2000, 'The Hausman-MaCurdy controversy: why do the results differ across studies?', *Journal of Human Resources*, vol. 35, no. 1, pp. 204–20.

Erdem, T, Keane, M, Oncu, S & Strebel, J 2005, 'Learning about computers: an analysis of information search and technology choice', *Quantitative Marketing and Economics*, vol. 3, no. 3, pp. 207–46.

Fang, H & Keane, M 2005, 'Assessing the impact of welfare reform on single mothers', *Brookings Papers on Economic Activity*, vol. 1, pp. 1–116.

French, E 2005, 'The effects of health, wealth and wages on labour supply and retirement behaviour', *Review of Economic Studies*, vol. 72, pp. 395–427.

Geweke, J & Keane, M 2000, 'An empirical analysis of male income dynamics in the PSID: 1968–1989', *Journal of Econometrics*, vol. 96, pp. 293–356.

——& Keane, M 2001, 'Computationally intensive methods for integration in econometrics', in JJ Heckman & EE Leamer (eds), *Handbook of econometrics*, vol. 5, Elsevier Science BV, pp. 3463–568.

Goeree, JK, Holt, CA & Palfrey, TR 2003, 'Risk averse behavior in generalized matching pennies games', *Games and Economic Behavior*, vol. 45, no. 1, pp. 97–113.

Haley, W 1973, 'Human capital: the choice between investment and income', *American Economic Review*, vol. 63, December, pp. 929–44.

Hall, R 1973, 'Wages, income & hours of work in the US labor force', in G Cain & H Watts (eds), *Income maintenance and labor supply*, Markham, Chicago, pp. 102–62.

Hausman, J 1980, 'The effect of wages, taxes and fixed costs on women's labor force participation', *Journal of Public Economics*, vol. 14, no. 2, pp. 161–94.

——1981, 'Labor supply', in H Aaron & J Pechman (eds), *How taxes affect economic behavior*, Brookings Institution, Washington DC.

——1985a, 'The econometrics of nonlinear budget sets', *Econometrica*, vol. 53, pp.1255–82.

——1985b, 'Taxes and labour supply', in K Arrow & M Intriligator (eds), *Handbook of public economics*, vol. 1, pp. 213–64.

Heckman, J 1974, 'Shadow prices, market wages and labor supply', *Econometrica*, vol. 42, no. 4, pp. 679–94.

——1976, 'Estimates of a human capital production function embedded in a life-cycle model of labor supply', in N Terlecky (ed.), *Household production and consumption*, Columbia University Press, New York, pp. 227–64.

——& MaCurdy, T 1980, 'A life-cycle model of female labor supply', *Review of Economic Studies*, vol. 47, no. 1, pp. 47–74.

——& MaCurdy, T 1986, 'Labor econometrics', in Z Griliches & M Intriligator (eds), *Handbook of econometrics*, vol. 3, pp. 1917–77.

Hotz, VJ & Miller, RA 1993, 'Conditional Choice Probabilities and the Estimation of Dynamic Models', *Review of Economic Studies*, vol. 60, pp. 497–529.

Imai, S & Keane, M 2004, 'Intertemporal labor supply and human capital accumulation', *International Economic Review*, vol. 45, no. 2, pp. 601–42.

Keane, M 1983, 'Some tests of rational expectations using survey data', unpublished undergraduate thesis, Massachusetts Institute of Technology.

——1993, 'Nominal contracting theories of unemployment: evidence from panel data', *American Economic Review*, vol. 83, no. 4, pp. 932–52.

——1995, 'A new idea for welfare reform', *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 19, no. 2, pp. 2–28.

——2009a, 'Structural vs atheoretic approaches to econometrics', *Journal of Econometrics*, forthcoming.

——2009b, *Income taxation in a life cycle model with human capital*, working paper, University of Technology Sydney.

——2009c, 'Simulated maximum likelihood estimation based on first order conditions', *International Economic Review*, vol. 50, no. 2, May, pp. 627–75.

——& Moffitt, R 1998, 'A structural model of multiple welfare program participation and labor supply', *International Economic Review*, vol. 39, no. 3, pp. 553–89.

——& Runkle, D 1992, 'On the estimation of panel data models with serial correlation when instruments are predetermined but not strictly exogenous', *Journal of Business and Economic Statistics*, vol. 10, no. 1, pp. 1–9.

——& Wolpin, K 1994, 'The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence', *Review of Economics and Statistics*, vol. 76, pp. 648–72.

——& Wolpin, K 1997, 'The career decisions of young men', *Journal of Political Economy*, vol. 105, no. 3, pp.473–522.

——& Wolpin, K 2001, 'The effect of parental transfers and borrowing constraints on educational attainment', *International Economic Review*, vol. 42, no. 4, pp. 1051–103.

Killingsworth, M & Heckman, J 1986, 'Female labor supply: a survey', in O Ashenfelter & R Layard (eds), *Handbook of labor economics*, vol. 1, North-Holland, Amsterdam, pp. 103–204.

Kosters, M 1969, 'Effects of an income tax on labor supply', in A Harberger & M Baily (eds), *The taxation of income from capital*, Brookings Institution, Washington DC.

MaCurdy, T 1981, 'An empirical model of labor supply in a life-cycle setting', *Journal of Political Economy*, vol. 89, pp. 1059–85.

——1983, 'A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty', *International Economic Review*, vol. 24, no. 2, pp. 265–89.

——Green, D & Paarsch, H 1990, 'Assessing empirical approaches for analyzing taxes and labor supply', *Journal of Human Resources*, vol. 25, no. 3, pp. 415–90.

McCallum, B 1976, 'Rational expectations and the estimation of econometric models: an alternative procedure', *International Economic Review*, vol. 17, pp. 484–90.

Meghir, C & Phillips, D 2008, 'Labour supply and taxes', prepared for the Report of a Commission on Reforming the Tax System for the 21st Century, chaired by Sir James Mirrlees, Institute for Fiscal Studies, London..

Mirrlees, J 1971, 'An exploration in the theory of optimal income taxation', *Review of Economic Studies*, vol. 38, pp. 175–208.

Moffitt, R 1983, 'An economic model of welfare stigma', *American Economic Review*, vol. 73, no. 5, pp. 1023–35.

Pencavel, J 1986, 'Labor supply of men: a survey', in O Ashenfelter & R Layard (eds), *Handbook of labor economics*, vol. 1, North-Holland, Amsterdam, pp. 3–102.

Pistaferri, L 2003, 'Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply', *Journal of Labor Economics*, vol. 21, no. 3, pp. 729–54.

Saez, E 2001, 'Using elasticities to derive optimal income tax rates', *Review of Economic Studies*, vol. 68, pp. 205–29.

———Slemrod, J & Giertz, S 2009, *The elasticity of taxable income with respect to marginal tax rates: a critical review*, mimeo, University of California, Berkeley.

Sargent, T 1978, 'Estimation of dynamic labor demand schedules under rational expectations', *Journal of Political Economy*, vol. 86, no. 6, pp. 1009–44.

Shaw, K 1989, 'Life-cycle labor supply with human capital accumulation', *International Economic Review*, vol. 30, May, pp. 431–56.

Skinner, J 1987, 'A superior measure of consumption from the panel study of income dynamics', *Economics Letters*, vol. 23, no. 2, pp. 213–16.

Triest, R 1990, 'The effect of income taxation on labor supply in the United States', *Journal of Human Resources*, vol. 25, no. 3, pp. 491–516.

van Soest, A, Woittiez, I & Kapteyn, A 1990, 'Labor supply, income taxes and hours restrictions in the Netherlands', *Journal of Human Resources*, vol. 25, no. 3, pp. 517–58.

Wales, T & Woodland, A 1979, 'Labour supply and progressive taxes', *Review of Economic Studies*, vol. 46, no. 1, p. 83–95.

Ziliak, JP & Kniesner, TJ 1999, 'Estimating life-cycle labor supply tax effects', *Journal of Political Economy*, vol. 107, no. 2, pp. 326–59.

——& Kniesner, TJ 2005, 'The effect of income taxation on consumption and labor supply', *Journal of Labor Economics*, vol. 23, no. 4, pp. 769–96.