NBER WORKING PAPER SERIES

RECONCILING MICRO AND MACRO LABOR SUPPLY ELASTICITIES: A STRUCTURAL PERSPECTIVE

Michael P. Keane
Richard Rogerson

Working Paper 17430
http://www.nber.org/papers/w17430

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2011

Rogerson thanks both the NSF and the Korea Science Foundation (WCU-R33-10005) for financial support. Keane thanks the Australian Research Council (grant FF0561843). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective
Michael P. Keane and Richard Rogerson
NBER Working Paper No. 17430
September 2011
JEL No. E24,J22

ABSTRACT

The response of aggregate labor supply to various changes in the economic environment is central to many economic issues, especially the optimal design of tax policies. This paper surveys recent work that uses structural models and micro data to evaluate the size of this response. Whereas the earlier literature on this issue often concluded that aggregate labor supply elasticities were small, recent work has identified three key reasons that the aggregate elasticity may be quite large. First, earlier estimates abstracted from several key features, including human capital accumulation, leading to estimates that are dramatically negatively biased. Second, failure to understand that aggregate labor supply adjustments can occur along both the hours per worker and employment margins has led economists to misinterpret the implications of previous estimates for aggregate labor supply. Third, structural estimation of responses along the extensive (i.e., employment) margin are typically quite large.

Michael P. Keane
School of Economics
University of New South Wales
Sydney NSW 2052
Australia
Michael.Keane@uts.edu.au

Richard Rogerson
Woodrow Wilson School of Public
and International Affairs
323 Bendheim Hall
Princeton University
Princeton, NJ 08544
and NBER
rdr@princeton.edu

1. Introduction

This survey deals with an issue that is extremely important for a wide range of applied issues–the magnitude of aggregate labor supply responses to various changes in the economic environment. In addition to being a very important issue, it is also well known to be quite controversial. In particular, there is a long-standing controversy driven by the fact that on the one hand, researchers who look at micro data typically estimate relatively small labor supply elasticities, while on the other hand, researchers who use representative agent models to study aggregate outcomes typically employ parameterizations that imply relatively large aggregate labor supply elasticities.

The objective of this survey will be to clarify the issues related to this apparent controversy. A key point that we wish to stress from the outset is that in general labor supply elasticities are neither a single number nor a primitive feature of preferences. We believe that one important source of confusion in the literature is the idea that one can estimate a labor supply elasticity in one context and import this elasticity into other contexts. While this may be warranted in some contexts under specific assumptions, in general this is problematic. At a broad level, one can ask what happens to different dimensions of labor supply (individual or aggregate) in response to some change in the economic environment. This response will typically depend on the features of preferences, technology, market structure, and of course on the nature of the change to the economic environment. For example, in the context of taxes, one must specify if the tax changes are temporary or permanent, anticipated or unanticipated, and what the tax revenues are used for. Moreover, whereas in a static setting one might represent labor supply by a single number (say hours worked), in a dynamic setting, labor supply is necessarily a higher dimensional object, specifying hours worked at each age. In general there is no single number that can describe the change in labor supply to a wide class of changes in the economic environment. The change today may be different than the change in future periods. We might be interested in the current change or the change over the entire future.

A key message from this discussion is that it is important for economists to adopt a framework in which the choice problem of an individual is explicitly formulated, and where the parameters that characterize this choice problem are the key parameters that determine the response of various components of labor supply in response to different changes. This message is similar in spirit to that of the Handbook of Macroeconomics chapter by Browning, Hansen and Heckman (1999). Economists should be seeking to identify the underlying structural parameters of these choice problems and then use that information to infer elasticities, rather than trying to explicitly estimate something called a labor supply elasticity that is then applied across different situations.

Having set this as the framework for our analysis, we explore two key issues of interest. First, what is the mapping from a given set of observations to the underlying preference parameters of interest? Second, what is the relationship between parameters that describe the response of a given individual at a particular point in time to the response at the aggregate level? Our main conclusion here is that recent work forces us to rethink some aspects of labor supply and labor supply elasticities that many have taken for granted for some time.

In the next section we provide some background material that serves to highlight the apparent controversy regarding labor supply elasticities estimated from micro data and the aggregate elasticities implicit in many aggregate models. Specifically, we lay out a benchmark model that represents a standard framework used by many economists to think about these issues. An important feature of this benchmark model is that it provides a clear empirical strategy for using micro data to uncover the values of key preference parameters. It also provides a clear mapping from these preference parameters to the effect of various changes in the economic environment on aggregate labor supply. We use this benchmark model as a vehicle to present the controversy between micro and macro labor supply elasticities.

Because it will be easier to describe the specific contributions of this survey once this benchmark model has been laid out, we postpone a more complete outline of the survey until the next section. Put very briefly, having laid out the benchmark model, we describe several extensions that have been pursued over the last decade or so and argue that the controversy seems much less apparent (if not non-existent) when one uses these extended models to view the data. Our conclusion is that much of the controversy is due to the fact that economists generally continue to view the world through the lens of this benchmark model, which abstracts from these empirically important features.

2. Background and Overview

In this section we present a framework that serves to highlight the apparent tension between evidence from micro studies and common practice in modelling aggregate phenomena. Having presented this model we then use it as a vehicle to describe the nature of the controversy, and provide an outline for the remainder of this survey.

2.1. A Benchmark Model

In this section we present a benchmark model that connects elasticity estimates from panel data with those from aggregate data and serves to highlight the apparent controversy. In each period a T-period lived household is born with preferences given by:

$$\sum_{a=0}^{T} \beta^a [\frac{1}{1-\frac{1}{\eta}} c_a^{1-\frac{1}{\eta}} - \frac{\alpha}{1+\frac{1}{\gamma}} h_a^{1+\frac{1}{\gamma}}]$$

where a denotes age and c_a and h_a are consumption and hours worked at age a respectively. There are four key preference parameters: β, η, α, and γ. In what follows we will strive to use these parameters consistently throughout the survey.[1] The individual is endowed with one unit of time during each period of life, and faces an exogenous sequence of productivity over his life cycle. Specifically, an

[1] Readers familiar with the literature will know that there is some variation in the literature regarding the parameterizion of this class of utility functions, with some researchers preferring to write the exponents on consumption and hours worked as $1-\eta$ and $1+\gamma$. We note here that the reader should be careful in moving between results as reported here and in the original contributions that we survey since creating uniform notation in this survey necessarily means that what we refer to as γ in our survey may actually show up as $1/\gamma$ in one of the original contributions.

individual of age a has productivity e_a, so that if he supplies h_a units of time at age a it results in $e_a h_a$ units of labor services. The production side of the economy is the standard one associated with the neoclassical growth model: there is a constant returns to scale aggregate production function $F(K_t, H_t)$, where H_t is aggregate units of labor services. In steady state, H and K will be given by:

$$H = \sum_{a=1}^{T} e_a h_a, \; K = \sum_{a=1}^{T} k_a$$

where h_a and k_a are the steady state life cycle profiles for hours worked and capital holdings. Output can be used as either consumption or investment and capital depreciates at rate δ. We consider a tax and transfer system with the following properties: labor earnings are taxed at the constant rate τ and the resulting revenues are used to fund a lump-sum transfer. In order to abstract from issues of intergenerational redistribution, we assume that the lump-sum transfer received by any generation is equal in present value to the tax revenues that they pay.[2]

If we considered this type of labor tax and transfer system in an infinitely lived agent model, the steady state interest rate would be unaffected and would exactly offset the discount factor that is part of the preference specification, i.e., we would have $\frac{1}{1+r-\delta} = \beta$, where r is the rental rate on capital. Neither of these properties necessarily hold in an overlapping generations economy. Since our primary interest is in the effects of taxes controlling for changes in any other

[2] This simple tax and transfer system has little connection with the features of actual tax and transfer systems that we observe in reality. While analysis of it is perhaps of limited direct interest in terms of policy analysis, it is very useful as a vehicle to exposit the central issues that we address. The reason for this is that the effects of this type of policy correspond to the Hicks elasticity and hence provide a clean connection to the literature on elasticities.

factors, such as interest rates, in order to make the analysis more transparent we will assume that the steady state interest rate is not affected by the labor tax rate and transfer program and is equal to the same value as in the steady state of the infinitely lived agent economy, i.e., that the rental rate on capital net of depreciation must exactly offset the individual's discounting. We note that there is always a government debt policy that would support this interest rate as a steady state equilibrium. If the steady state interest rate is independent of the tax rate, the fact that F satisfies constant returns to scale implies that the wage per unit of labor services will also be independent of the tax rate. Denote this value by w. The above comments imply that in steady state, a newly born individual solves the following problem[3]:

$$\max_{c_a,\, h_a} \sum_{a=1}^{T} \beta^a \left[\frac{1}{1-\frac{1}{\eta}} c_a^{1-\frac{1}{\eta}} - \frac{\alpha}{1+\frac{1}{\gamma}} h_a^{1+\frac{1}{\gamma}}\right]$$

$$\text{s.t.} \quad \sum_{a=1}^{T} \beta^a c_a = \sum_{a=1}^{T} (1-\tau)\beta^a e_a h_a w + T$$

Letting λ denote the Lagrange multiplier on the budget equation, we have the following first order conditions:

$$c_a^{-\frac{1}{\eta}} = \lambda \tag{2.1}$$

$$\alpha h_a^{\frac{1}{\gamma}} = (1-\tau)\lambda e_a w \tag{2.2}$$

Equation (2.1) implies that c_a will be constant over the life cycle.[4] Taking logs

[3] We assume that the individual is born with zero capital, so capital holdings over time do not appear in the present value budget equation.

[4] The constancy of consumption over the life cycle is not consistent with micro data. There

of equation (2.2) gives a simple version of the equation used by MaCurdy (1981) and others in their estimation exercises using micro data:

$$\log h_a = b + \gamma \log e_a \tag{2.3}$$

where $b = \gamma[\log \lambda + \log w + \log(1 - \tau) - \log \alpha]$ is a constant from the perspective of following an individual over their life cycle in steady state. Recall that changes in $\log e_a$ are equivalent to changes in log wages for individuals over the life cycle. It follows that this equation provides a strategy for uncovering the preference parameter γ using individual panel data. As we describe in more detail below, one can also use this information to uncover the value of η.

Next we illustrate via one simple example how the values of these preference parameters estimated from micro data allow one to infer the consequences of changes in the aggregate economic environment on aggregate hours of work. Specifically, we consider the effects of a change in the scale of the tax and transfer program, i.e., a change in τ.

Equation (2.3) is a useful starting point, but note that it is not sufficient to determine the response in h_a to a change in the scale of the tax and transfer system. The reason for this is that if we are comparing h_a across the two steady state equilibria that correspond to different values of τ then it is also the case that the value of λ will differ across the two steady states. Hence, in order to solve for the change in h_a we need to also derive an expression for the change in λ.

are various ways to avoid this implication, for example by allowing for age effects on the marginal utility of consumption. Because our focus here is on the implications for labor supply we abstract from these possibilities.

To solve for the change in λ we proceed in several steps. First, note from equation (2.2) that given the optimal value of h_0, we can compute the rest of the optimal profile from:

$$h_a = [\frac{e_a}{e_0}]^\gamma h_0. \tag{2.4}$$

Total labor income is therefore proportional to h_0. Given that the present value of the transfer received by each individual is equal to the present value of their own tax payments, in steady state equilibrium we have:

$$\sum_{a=1}^{T} \beta^a c_a = \sum_{a=1}^{T} \beta^a e_a h_a w.$$

Recalling that c_a is constant over the life cycle, it follows that the optimal choice of c_a is proportional to h_0 and w. Write this as $c_a = \bar{c} w h_0$. Equation (2.1) then implies:

$$\log \lambda = -\frac{1}{\eta}[\log h_0 + \log w + \log \bar{c}] \tag{2.5}$$

Using equation (2.4) we have:

$$\log \lambda = -\frac{1}{\eta}[\gamma \log \frac{e_o}{e_a} + \log h_a + \log w + \log \bar{c}] \tag{2.6}$$

Given our assumption that the interest rate and hence w are not affected by the change in τ, equation (2.3) implies:

$$\log h_a = -\frac{\gamma}{\eta}[\gamma \log \frac{e_o}{e_a} + \log h_a + \log \bar{c} + \log w] + \gamma \log w + \gamma \log(1-\tau) - \gamma \log \alpha + \gamma \log e_a. \tag{2.7}$$

Rearranging gives:

$$\log h_a = \frac{\gamma}{\eta+\gamma}[(\eta-1)\log w - \eta\log\alpha - \log\bar{c} - \gamma\log e_0] + \frac{\gamma\eta}{\eta+\gamma}\log(1-\tau) + \gamma\log e_a. \tag{2.8}$$

For future reference it is of interest to note the different coefficients on $\log(1-\tau)$ and $\log e_a$. The coefficient on $\log(1-\tau)$ is always smaller than the coefficient on $\log e_a$, with equality holding in the limit as η goes to infinity, i.e., when utility is linear in consumption and there are no income effects. The effect of life cycle variation in wages represents the Frisch elasticity, whereas the coefficient on taxes in this expression represents the Hicks elasticity. A key distinction between the two is that the Frisch elasticity holds the marginal utility of consumption constant, whereas the Hicks does not.

The implication from equation (2.8) is that a change in τ leads to a parallel shift in the hours profiles, i.e., hours at all ages change by the same percentage. Since H is simply the sum of the h_a, it follows that:

$$\log H = B + \frac{\gamma\eta}{\eta+\gamma}\log(1-\tau) \tag{2.9}$$

where B is a constant. Macroeconomists often impose $\eta = 1$ in order that preferences be consistent with balanced growth, in which case the coefficient on $\log(1-\tau)$ is purely a function of γ.

The result that we want to stress for this one policy exercise in the benchmark model is the strong connection between preference parameters estimated from micro data and the implied aggregate elasticity for a particular tax and transfer policy. Moreover, note that in this benchmark model it is also the case that observing the response in steady state hours worked by an individual at one particular age is sufficient to infer the aggregate response. For future reference we note that all of the above results go through untouched if we impose that all individuals must retire exogenously at some given age T_R.

2.2. Micro Evidence Based on the Benchmark Model

The empirical literature that uses micro data to estimate labor supply elasticities is vast, so we make no attempt to summarize it here. Classic reviews of this literature include Hausman (1985), Pencavel (1986), Killingsworth and Heckman (1986) and Blundell and MaCurdy (1999). For more recent reviews we refer the reader to Meghir and Phillips (2008) and Keane (2010). We feel it a fair statement to say that, based on this literature, the majority of the economics profession has come to the conclusion that labor supply elasticities are small; and, in particular, that labor supply is not very responsive to tax changes. (This statement is certainly accurate for male labor supply; there is less consensus for female labor supply).

Rather than summarize the whole literature, we consider three of the most influential papers: MaCurdy (1981), Browning, Deaton and Irish (1985) and Altonji (1986). All three papers attempt to estimate the intertemporal elasticity of substitution, or Frisch elasticity, using micro data. Details of their approaches differ, but all involve regressing hours changes on wage changes. For example, MaCurdy (1981) uses the same basic model described above extended to allow for heterogeneity and uncertainty to derive the hours change equation:

$$\Delta \log h_{it} = \gamma \Delta \log w_{it}(1-\tau_t) - \gamma \log \beta(1+r_t) + \alpha\gamma\Delta X_{it} + \gamma\xi_{it} + \gamma\Delta\varepsilon_{it} \quad (2.10)$$

The parameters α, β, and γ are all as above, and we allow for the tax rate (τ_{it}) to vary across time and individuals. The X_{it} are control variables for exogenous shifts in tastes for work, the ε_{it} represent unobserved taste shocks, and ξ_{it} represents the surprise part of the change in the marginal utility of wealth (or of consumption) from $t-1$ to t.[5] The literature has focussed on three issues: First, the ξ_{it} will in general be correlated with wage changes to the extent that wage changes are not fully anticipated at $t-1$. Second, tastes for work may be correlated with wages (e.g., those with a higher taste for work may also work harder or acquire more skills, and, conversely, a higher taste for work can lower the after-tax wage by pushing one into a higher tax bracket). Third, the wage is presumably measured with considerable error.

To deal with these problems, all three of these influential papers instrument for wage changes, using variables that were presumably known at time $t-1$. For instance, MaCurdy (1981) uses polynomials in age and education, exploiting the fact that wages are known to follow an inverted U-shape over the life cycle, the shape of which varies with education. The three papers noted above differ somewhat in the choice of instruments, the choice of observed taste shifters and the

[5] MaCurdy (1981) did not allow for shocks to the marginal utility of wealth, but MaCurdy (1983) shows how to extend the previous analysis to allow for this. In terms of implementing the estimation procedure, the only impact is that one needs to lag the instruments, which MaCurdy (1981) actually did, even though he had not incorporated uncertainty.

exact choice of the functional form for the labor supply function. Nevertheless, all three obtain very small estimates of γ, the Frisch elasticity (the preferred estimates being 0.15, 0.09 and 0.31, respectively, for MaCurdy, Browning et al, and Altonji). These results have been quite influential in generating a consensus within the profession that the Frisch elasticity is small.

Given the theoretical result that the Frisch elasticity is an upper bound on the Marshall and Hicks elasticities in the life-cycle labor supply model, the finding of a small Frisch elasticity has also contributed to the view that the Marshall and Hicks elasticities (which are relevant for estimating responses to permanent tax changes) are small as well. Furthermore, MaCurdy (1981) noted that the results of estimating (2.10) could be used to infer the response to permanent wage changes. This is possible because estimation of (2.10) uncovers all parameters of the hours equation in levels,

$$\log h_{it} = \gamma \log(w_{it}(1-\tau_{it})) + \gamma \log \lambda_{i0} - \gamma \log \rho(1+r_t)^t + \alpha\gamma X_{it} + \gamma\varepsilon_{it} \quad (2.11)$$

except for $\gamma \log \lambda_{i0}$, which is the individual specific constant (or "fixed effect") in the levels equation (where λ_{i0} is the marginal utility of wealth at $t = 0$). Thus, MaCurdy (1981) backs out the value of $\gamma \log \lambda_{i0}$ in a second stage after estimating (2.10) in the first stage. Given these constants, MaCurdy can, in principle, regress them on the whole set of life-cycle wages.[6] His estimates imply that a 10% (fully

[6] Of course, MaCurdy only observes wages for his 10 year sample period – not the whole life-cycle. To deal with this problem, he fits a life-cycle wage profile for each person using 10 years of data. He then regresses the estimated values of $\gamma \log \lambda_{i0}$ on the individual specific parameters of this (assumed quadratic) profile. Using the coefficient on the wage profile intercept, MaCurdy can determine how an upward shift in the whole wage profile would affect $\gamma \log \lambda_{i0}$, and hence

anticipated) increase in wages at all ages would increase labor supply by only 0.8% - a very small effect.

As noted earlier, these sorts of results have lead to a majority view in the profession that labor supply elasticities are quite small. This majority view 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 labor supply .... With some exceptions, the profession has settled on a value for this elasticity close to zero... In models with only a labor-leisure choice, this implies that the efficiency cost of taxing labor income ... is bound to be low as well."

2.3. Macroeconomic Models

While the view that labor supply elasticities are small is clearly the majority position among microeconomists, this view is less well accepted among macroeconomists.[7] Beginning with Lucas and Rapping (1969), many macroeconomists have argued that relatively large Frisch elasticities are required in order to account for the size of labor market fluctuations over the business cycle.[8] Prescott (2004) shows that a relatively large labor supply elasticity is also required to rationalize the low frequency changes in hours of work among G-7 economies since

labor supply.

[7] In his Nobel lecture, Prescott (2004) argues that relatively large labor supply elasticities are important in reconciling various aggregate observations.

[8] Benhabib et al (1991) show that intratemporal substitution between home and market production can also contribute to a large elasticity for hours of market work.

1970.[9] In fact, in the infinitely lived stand-in household models that remain the norm in much of the literature on aggregate economic issues, it is standard to adopt specifications in which the period utility function is assumed to be log linear in consumption and leisure. Assuming that time spent in non-market work (leisure) is roughly twice as large as time spent in market work, this specification corresponds to assuming a Frisch elasticity of around 2.0.

The main point that we want the reader to take away from this brief overview is that viewed from the perspective of the simple benchmark model described earlier, there appears to be a strong tension between evidence based on micro studies and specifications commonly adopted in aggregate studies.

2.4. Overview of the Survey

The purpose of the present survey is to shed light on the reasons for the discrepancy between micro and macro views of labor supply. We will concede up front that we have a clear opinion on this matter: Our position is that the view that estimates based on micro data rule out large aggregate elasticities is flawed. In particular, our objective is to present the case that empirical evidence like that found by MaCurdy (1981), Browning et al (1985) and Altonji (1985) – as well as many other similar studies in the micro literature – is fully consistent with a world where aggregate labor supply elasticities are in fact large.

There are four main approaches one could take to this problem. The first, which we will not pursue, is to directly take issue with the claim that the mi-

[9] Ohanian et al (2008) extend this finding to a larger set of countries and a longer time period. See also Rogerson (2008) and McDaniel (2011).

cro literature reaches a clear consensus on labor supply elasticities. For example, Keane (2010) surveys 21 of the best known studies that estimate the Hicks elasticity for males. He notes that the studies seem to bifurcate into a low group vs. a high group, with thirteen producing estimates in a tight range from 0.02 to 0.13, while eight studies produce estimates in the 0.27 to 1.22 range. There is an odd gap between 0.13 and 0.27, with no studies falling in that range. Hence, he argues, even among labor economists there is a non-negligible minority who find relatively large elasticities using conventional methods. Keane (2010) also discusses the reasons that much of the profession seems to discount the sizeable minority of studies that find large elasticities, and discusses whether these reasons are valid. We refer the reader to his survey for further details.

Second, as with any empirical work, one could criticize the micro empirical studies that find small labor supply elasticities on their own terms. That is, one could accept the basic empirical framework (as illustrated by equation (2.10)) but question the implementation. Specifically, one could question the instruments used for wages, the controls used for tastes for work, the functional forms used for the labor supply function, the measurement of wages, taxes, etc. For instance, as noted earlier, it is quite common in the life-cycle labor supply literature to estimate equations like (2.10) using polynomials in age and education as instruments for wage changes. But these instruments are quite weak, leading to very large standard errors on the estimates of the Frisch elasticity. Thus, while the point estimates imply small elasticities, conventional confidence intervals are consistent with rather large elasticities. Again, we will not take this tack here (and we refer the reader to the surveys cited earlier for further details).

The third approach, which we do adopt here, involves questioning key assumptions of the standard micro labor supply model described earlier. We will describe a number of potentially important omitted features of the standard model that may have led prior studies to understate both the value of the preference parameter γ and the implied labor supply responses to specific changes in tax rates. For example, Imai and Keane (2004) argue that estimates of γ are severely downward biased by the failure to account for human capital. They found that estimating a model where wages grow with work experience yielded rather large estimates of γ. Other potentially important omitted factors that we will look at are credit constraints (Domeij and Floden (2006)), and uninsurable wage risk (Low (2005)).

The fourth approach, which we also take, involves questioning whether the standard labor supply model described earlier is even relevant for determining labor supply responses at the macro level. The key issues here are the extensive margin, population heterogeneity and aggregation. For instance, as noted earlier in the benchmark model, macro labor supply responses are determined by the individual preference parameter γ. Rogerson and Wallenius (2009) show that the close link between this individual preference parameter and aggregate labor supply responses is broken when one accounts for labor supply choices on the extensive margin. In particular, the aggregate responses can be much larger than would be implied by the estimated value of γ when viewed through the lens of the benchmark model described above. And Chang and Kim (2006) show that in a population with heterogeneous productivities and incomplete markets in which all adjustment occurs at the extensive margin, the labor supply response at the aggregate level is determined by the distribution of reservation wages and is unrelated to the underlying individual preference parameter γ. It is thus possible that labor supply could be quite inelastic on the intensive margin at the individual level, yet quite responsive in the aggregate.

The fourth approach just described emphasizes the role of adjustment along the extensive margin in terms of reconciling small elasticities estimated from micro data with large elasticities at the aggregate level. Key to this reconciliation is that each of the three early studies noted above (MaCurdy, Browning et al and Altonji), as well as most of the other structural analyses based on micro data, implicitly focus on adjustment along the intensive margin. If adjustment along the extensive margin plays such an important role, a key issue is to assess the size of this response using micro data. We close the survey by considering the results from micro data analyses that allow for an extensive margin, due to say fixed costs associated with market work (Cogan (1981)).

3. Micro Evidence Based on Extensions of the Basic Model

As we noted earlier, the basic life-cycle labor supply model of MaCurdy (1981) abstracts from a number of potentially important features of the economic environment. These include human capital, credit constraints, uninsurable wage risk, adjustment costs and optimization errors. In this section we discuss a number of recent papers that argue that the failure to account for these features may have led prior micro empirical studies to underestimate the value of the preference parameter γ and hence the responsiveness of labor supply to changes in wages or taxes. We note that all of the estimates described in this section are based on adjustment along the intensive margin.

3.1. Human Capital Accumulation

The classic MaCurdy (1981) life-cycle model assumes that wages evolve exogenously over the life-cycle. That is, it rules out the possibility that workers may acquire human capital via on-the-job investment in skills, or through learning by doing. Heckman (1976) considered a model with on-the-job investment in skills, where workers are only paid for the fraction of the day they engage in productive work (i.e., not the time they spend learning). He noted that in this type of model a worker's measured wage rate (i.e., earnings divided by total hours at work) would be less than his/her true productivity. His estimates implied that for young workers productivity exceeded the wage by as much as 54%, while for workers in their forties the divergence had largely vanished.

Shaw (1989) extended the MaCurdy (1981) framework to include human capital investment of the learning-by-doing variety. Analogous to Heckman (1976), her estimates implied that the return to an hour of work substantially exceeded the observed wage for young workers, due to the fact that a substantial part of the return came in the form of learning that augmented future wages. Again, this divergence narrowed for older workers. However, neither Heckman (1976) nor Shaw (1989) directly considered the impact of ignoring human capital on estimates of preference parameters and labor supply responses.

Imai and Keane (2004) argued that ignoring human capital would lead to a downward bias in estimates of γ. To illustrate the key point, assume that wages are given by the simple equation

$$w_{t+1} = (1 + \kappa \sum_{j=1}^{t-1} h_{t-j}) w_1 \tag{3.1}$$

where $\kappa > 0$ and w_1 represents a person's initial wage (or skill endowment) upon first entering the labor market. Given this simple functional form, a one unit increase in h_t raises the wage by κw_1 in all future periods. This, in turn, raises earnings by $\kappa w_1 h_{t+j}$ for $j = 1...T - t$.[10]

In a model with human capital, the return to an hour of work, which we will refer to as the opportunity cost of time (OCT), consists of the after tax wage plus the expected present value of increased (after-tax) earnings in all future periods obtained by working an extra hour at time t. We will refer to this additional "return to work experience" term as the "human capital term" (HC).

Of course, the optimality condition for an interior solution for hours equates the marginal rate of substitution (MRS) between consumption and leisure to the OCT. Using the same utility function as in the benchmark model of Section 2 we

[10] In fact, Heckman (1976), Shaw (1989) and Imai and Keane (2004) all assume a much more complex wage process than that given in equation (3.1). They all allow for complementarity between human capital and hours of work in the human capital production function. In particular, the Imai-Keane 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. They achieve this both by allowing for complementarity between human capital and hours, and by letting the parameters of the human capital production function differ by education level. The simple wage equation used here helps to clarify the key points, as it leads to a simple expression in equation (**??**).

obtain:

$$\frac{\alpha h_t^{\frac{1}{\gamma}}}{c_t^{-\frac{1}{\eta}}} = w_t(1-\tau_t) + E_t \sum_{j=0}^{T-t} \frac{\kappa w_1 h_{t+1+j}(1-\tau_{t+1+j})}{(1+r)^{1+j}} \tag{3.2}$$

A model without human capital would equate the MRS to the after tax wage itself, ignoring the human term on the right hand side of equation (3.2).

To gain intuition for the effect of ignoring the human capital term, consider Figure 1, which presents a stylized (but fairly accurate) picture of how male wages and hours move over the life-cycle.[11] The wage rate exhibits the typical "hump shape" over the life-cycle observed in many studies for males (i.e., wages start out low when a person is young, grow rapidly early in the life-cycle, peak in the 40s, and then decline)[12]. The curve representing annual hours of work also has a hump shape but with much less curvature (see, e.g., the descriptive regressions presented by Pencavel (1986)).

As noted earlier, the typical study in the male labor supply literature uses equations similar to (2.10) to estimate γ, the intertemporal elasticity of substitution. That is, it simply regresses hours growth on wage growth (along with controls for changing tastes for work). To deal with endogeneity of wages, it instruments for wages primarily using polynomials in age and education. These instruments are chosen precisely because they capture the hump shape of the life-cycle wage path shown in Figure 1, so predicted wages based on these instruments closely

[11] 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. Note that wage and hours patterns for women are rather different, as both wages and hours tend to flatten out or drop in the 30s, presumably due to fertility.

[12] The details of this pattern differ by education level (i.e., wages of more educated workers tend to peak later).

track the typical life-cycle wage path depicted in the figure. Thus, by regressing hours on predicted wages, one essentially uncovers the relative slope of the hours and wage curves in Figure 1. 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 wages is small (i.e., much less than 1.0).[13]

The third line in Figure 1 represents the return to human capital investment, the second term on the right-hand-side of equation (3.2), which we denote "HC." The estimates in Imai and Keane (2004) imply that at age 20 this human capital return is actually slightly larger than the wage itself, which is why in Figure 1 the HC curve is drawn as starting slightly higher than the wage curve. Of course, the human capital investment return declines with age, both because of diminishing returns to human capital and because the worker approaches the end of the planning horizon T.[14]

Figure 1 also plots the opportunity cost of time (OCT) which equals the wage plus the human capital return to an hour of work. It is obtained as the vertical sum of the wage and HC curves. Because the HC term falls as the wage increases, the OCT curve is much flatter than the wage curve. This basic pattern is common to the Heckman (1976), Shaw (1989) and Imai and Keane (2004) estimates (although the relative slopes of the two curves differ across the studies).

Imai and Keane (2004) estimate the intertemporal elasticity of substitution

[13] In other words, both wages and hours have a humped shape over the life-cycle, but the hump in wages is much more pronounced. This apparently weak response of hours to wages leads conventional methods (which ignore human capital) to infer a low value for the preference parameter γ.

[14] The Imai and Keane estimates imply that by age 36 the human capital return is only 25% as large as the wage. Thus, the HC curve is drawn as falling with age.

using a model that takes into account the HC term in (3.2). That is, they look at how hours respond to changes in the OCT, rather than the wage. Intuitively, their procedure amounts to taking the ratio of the slope of the hours line to the OCT line in Figure 1 (in contrast to the conventional procedure that amounts to taking the ratio of the slope of the hours curve to the wage curve). As the OCT line is much flatter than the hours line, this will produce a much larger estimate of the responsiveness of labor supply to the price of time.

They estimate their model using white males from the NLSY79. The men in their sample are aged 20 to 36 and, as the focus of the paper is solely on labor supply, they are required to have finished school. Imai and Keane (2004) estimate that $\gamma = 3.8$. In a model without human capital, this would imply a much higher willingness to substitute labor intertemporally than in almost all prior studies for men (see MaCurdy (1983) for an exception).

A key point, however, is that once human capital is incorporated into the life-cycle model, there is no longer a simple direct mapping from the preference parameter γ to the response in hours to a transitory wage change. As we discussed in Section 2, conditional on a value of η, MaCurdy (1981) could summarize the effects of permanent and transitory tax changes in his model by the single parameter γ. In the Imai-Keane model, the situation is much more complicated. This is illustrated in Table 1, which simulates the effects of permanent and transitory tax increases using their model.[15] Note that effects of transitory tax increases were

[15] Note that since we are assuming a constant returns to scale production function these tax experiments have no impact on the pre-tax wage rate.

reported in Imai and Keane (2004).[16] The simulations of the effects of permanent tax changes are new.[17]

Table 1

Effects of an Unexpected 5% Tax Increase on Hours Worked (in %) in Imai-Keane

Age	Transitory	Permanent	
		Uncompensated	Compensated
20	−1.5	−0.7	−3.2
25	−1.8	−0.6	−2.7
30	−2.2	−0.6	−2.4
35	−2.6	−0.1	−2.3
40	−3.2	−0.7	−2.3
45	−3.8	−1.0	−2.8
50	−4.7	−2.3	−4.2
60	−8.6	−9.4	−10.5

The table reports the effects of an unexpected five percentage point increase in the tax rate, assuming that there is no change in any transfer payments received by the individual. In the column labelled "transitory," the tax increase only applies for one year at the indicated age. For example, at age 20, a temporary 5% tax increase reduces hours by 1.5%. This implies a labor supply elasticity with respect to transitory tax changes of approximately 0.3. This figure is far smaller than one

[16] The only difference is that Imai and Keane (2004) reported effects of 2% tax increases, while here we have used their model to simulate 5% tax increases.

[17] We thank Susumu Imai for providing us with these simulations.

might expect, given the estimate of $\gamma = 3.8$. But, as Imai and Keane noted, the effect of transitory taxes increases substantially with age. For instance, at age 60, a temporary 5% tax increase reduces hours by 8.6%, implying an elasticity of roughly 1.7.

Why does the labor supply elasticity increase with age in this model? And why is the effect of transitory tax increases so small at young ages (despite the high value of γ)? The reasons are as follows: At young ages the wage is only a relatively small part of the opportunity cost of time (i.e., a large part of the return to work comes in the form of increased future wages).[18] A temporary tax increase alters the current after-tax wage in equation (3.2), but it has no direct effect on the human capital component of the OCT. Indeed, to the extent that a temporary tax increase causes workers to plan to work fewer hours in the current period and more hours in future periods, the human capital return component to work may actually increase – as the expected present value term in (3.2) contains hours in all future periods.

Unfortunately, Imai and Keane (2004) did not use their model to simulate responses to permanent tax increases. These are arguably more interesting from a public finance point of view. To fill this gap, Keane (2009) uses the Imai-Keane model to simulate the impact of an unexpected permanent 5% tax rate increase (starting at age 20 and lasting through age 65) on labor supply over the entire working life. If the revenue is simply thrown away, the model implies that average hours of work (from ages 20 to 65) drops from 1992 per year to 1954 per

[18] According to Imai and Keane (2004)'s estimates, at age 20 the wage is less than half of the opportunity cost of time, but by age 40 the wage is 84% of the opportunity cost of time.

year, a 2% drop. If the revenue is redistributed as a lump sum transfer, labor supply drops to 1861 hours per year, a 6.6% drop. The former figure implies an uncompensated (Marshallian) elasticity of roughly 0.4, while the latter figure implies a compensated (or Hicks) elasticity with respect to permanent tax changes of roughly 1.3. Both these values are quite large compared to ones typically obtained in models without human capital.[19]

Keane (2009) also uses the Imai-Keane estimates to calibrate a simple two-period equilibrium version of their model. Not surprisingly, given the large value of the Hicks elasticity that the model implies, he finds that the welfare costs of labor income taxation are much larger than more conventional estimates would suggest.

It is notable that the effects of permanent tax changes on current labor supply differ greatly depending on a worker's age at the time the tax is implemented. This is shown in the last two columns of Table 1. In the table, the permanent tax increase takes effect (unexpectedly) at the indicated age and lasts until age 65. In the uncompensated case the revenues are thrown away, while in the compensated case the proceeds of the tax (in each year) are distributed back to agents in lump sum fashion. One notable finding is that compensated effects are much larger than uncompensated, implying that income effects are substantial. And both compensated and uncompensated effects of tax increases are much larger at older ages. For instance, as we see in Table 1, for workers in their 20s, 30s and 40s, the

[19] When discussing transitory tax increases we assumed that the revenues were thrown away. This has virtually no effect on the results since the revenue generated from a transitory tax is sufficiently small that the income effects are too small to make much difference for the elasticity calculation (as any extra income is spread over the whole remaining life).

compensated effects of a 5% permanent tax increase on annual hours worked range from -2.3% to -3.2%. But for workers in their 50s and 60s these magnitudes grow substantially.

One striking result in Table 1 is that, for younger workers, permanent tax increases have larger effects on current labor supply than do transitory tax increases. For instance, consider a 5% tax increase that takes place at age 25. If it is perceived as transitory, hours are reduced by 1.8%. But if the tax increase is perceived as permanent and the proceeds are distributed lump sum, hours fall by 2.7%. So at age 25, the permanent tax effect is about 50% greater. By the mid-30s, the effects of permanent and transitory tax cuts are roughly equal. Only in the 40s do effects of transitory tax cuts become somewhat larger.

These findings contradict a strong prediction of the standard life-cycle model (without human capital) that transitory tax changes should have larger effects on current labor supply than permanent ones.[20] They also contradict the broader conventional wisdom in economics that temporary price changes will have larger effects on demand than permanent ones.

Why can permanent tax changes have larger short run effects than transitory tax changes once human capital is introduced? The reason for this phenomenon was already mentioned when we discussed why the effects of transitory tax changes appear small in this model given the estimate of γ. Specifically, as we see in equation (3.2), a transitory tax increase only directly reduces the current after-

[20] This is equivalent to the statement that the Frisch elasticity should exceed the Hicks, which should in turn exceed the Marshallian (see Blundell and MaCurdy (1999)). The exception is when there are no income effects, in which case the three are equal.

tax wage, which is just one component of the OCT. A permanent tax increase, on the other hand, also reduces the expected present value of future after-tax earnings, so it also affects the human capital term in the OCT.

Keane (2009) pointed out the possibility that a permanent tax change could have a larger effect on current labor supply than a temporary tax change in a life-cycle model once human capital is introduced. Using a simple two-period version of the Imai-Keane model, he clarifies the condition under which this phenomenon may occur. Specifically, there are two opposing forces at work: On the one hand, a permanent tax increase has a larger effect on the current value of time, because it lowers the future returns to work experience (the human capital effect). On the other hand, a permanent tax increase also has a larger income effect. As Keane (2009) shows, permanent tax changes will have larger effects than transitory changes if the return to work experience (the human capital effect) is large enough relative to the income effect.

So far, we have discussed how the current effects of tax changes differ depending on the age of the individual at the time of the change. Next, we examine how the effect of a given tax change differs as an individual ages. This is described in Table 2.

Table 2

Effects of an Unexpected 5% Permanent Tax Increase in Imai-Keane (all changes in %)

	Age 25			Age 30			Age 35		
Age	Hours	Wage	Assets	Hours	Wage	Assets	Hours	Wage	Assets
25	−2.7								
30	−2.9	−0.4	+19.8	−2.4					
35	−3.2	−0.7	+26.3	−2.7	−0.3	+12.4	−2.3		
40	−3.8	−1.0	+14.5	−3.3	−0.6	+8.5	−2.7	−0.2	+3.2
45	−5.1	−1.3	+6.9	−4.4	−0.9	+4.3	−3.8	−0.5	+1.9
50	−7.9	−2.0	+2.6	−7.0	−1.4	+1.5	−6.2	−1.0	+0.5
55	−13.3	−3.6	−0.4	−12.2	−2.9	−0.8	−11.0	−2.3	−1.2
60	−19.3	−7.5	−3.0	−18.4	−6.6	−3.0	−17.4	−5.8	−3.0
65	−29.2	−11.6	−3.8	−28.1	−10.7	−3.6	−26.9	−9.7	−3.5

The Table considers a permanent (compensated) 5% tax increase that takes effect at either age 25, age 30 or age 35. The table reports how this tax increase alters a person's labor supply at 5-year intervals from age 25 to age 65. For instance, suppose the 5% tax increase goes into effect (unexpectedly) when the worker is 25. Then, at age 25, his labor supply is reduced by 2.7% (the same figure as we reported in Table 1). But, as the worker grows older, the negative effect on labor supply increases substantially. For instance, by age 45 it is −5.1%, and by age 60 it is −19.3%. Thus, in response to a permanent tax increase, workers not only reduce labor supply, but also shift their lifetime labor supply out of older ages towards younger ages.

Why does the effect of a permanent tax increase grow with age? There are two reasons. The first is the same reason that effects of transitory tax increases are greater for older workers. As a worker gets older, the after-tax wage makes up a larger fraction of the OCT, so a given tax has a larger direct effect. The second reason involves the dynamics of the human capital model. To the extent that a worker reduces his labor supply at time t, he will have less human capital at time $t + 1$. This causes the worker to work even less at time $t + 1$, leading to a lower wage at $t + 2$, etc..

This "snowball" effect of a permanent tax increase on after-tax wages is also evident in Table 2. At first, tax effects on human capital are modest, but they grow substantially with age. For instance, we see that if a 5% tax increase is instituted when a worker is 25, then by age 40 his wage is reduced by only 1.0%, but by age 55 his wage is reduced by 3.6%, and by age 60 the reduction is 7.5%. So, eventually, the pre-tax wage reduction due to the tax increase is greater than the tax increase itself.

An important implication of these results, emphasized in Keane (2009), is that in a model with human capital, changes in taxes cannot be viewed as a source of exogenous variation in after-tax wages for the purpose of identifying labor supply elasticities. The behavioral responses induced by tax changes feed back and alter the life-cycle wage path itself. Or, as asserted by Imai and Keane (2004), in the human capital model there is simply no such thing as an exogenous wage change.[21] It follows that elasticities estimated from quasi-experimental evidence

[21] Keane (2009) discusses the implications of this statement for estimation of labor supply elasticities – i.e., nothing short of full structural estimation of the joint labor supply/human

on responses to exogenous changes in tax rates are not easily interpretable in terms of underlying preference parameters.

Finally, Table 2 also reports how asset accumulation over the life-cycle responds to permanent tax changes. The basic pattern is that, upon implementation of the tax, savings first increase, while later, savings fall. For example, given a 5% permanent tax increase at age 25, a worker responds so as to increase his assets by 26.3% at age 35, but ends up with assets that are 3.8% lower at the age of retirement (age 65).[22] In other words, a permanent tax increase reduces consumption in the short run by more than the amount of the tax, not only because labor supply falls, but also because the savings rate increases. However, we see in Table 2 that the magnitude of the increase in savings following the tax increase, as well as the drop in assets at retirement, are less if the tax is implemented when the worker is older.

The reason a permanent tax increase generates more saving in the short to medium run is precisely the "snowball" effect of the tax on wage growth described earlier. Given that a tax increase reduces labor supply, a worker knows that his rate of wage growth has been reduced. So the asset response pattern is as one would expect – young workers consume less today if their perceived life-cycle wage path is flattened. This influence of the perceived life-cycle wage path on current consumption is a central issue in the papers by Domeij and Floden (2006) and Low (2005) that we will discuss in the next two sections.

---

capital investment process is adequate for estimating labor supply elasticities.

[22] Note that the Imai and Keane (2004) model includes a motive to carry assets into retirement, both to finance retirement and to leave bequests.

An important question about the Imai-Keane model can be stated as follows: Suppose this model did in fact generate the data. Would a researcher using methods like those in MaCurdy (1981), Browning et al (1985) and Altonji (1986) to study data generated by the model conclude that the intertemporal elasticity of substitution is small?[23] Imai and Keane (2004) addressed this question by (i) simulating data from their model, and (ii) applying instrumental variable methods like those in MaCurdy (1981) and Altonji (1986) to obtain estimates of γ in equations similar to (2.10). In conducting this exercise, they obtain an estimate of .325 (standard error = .256) using the MaCurdy (1981) approach, and .476 (standard error = .182) using the Altonji (1986) approach. Thus, the Imai-Keane model generates life-cycle histories that, when viewed through the lens of models that ignore human capital, do imply rather small values for the Frisch elasticity.

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. Thus, a negative correlation between hours changes and wage changes in the raw data is perfectly consistent with the high willingness to substitute labor intertemporally over the life cycle that we see in Tables 1-2.

What reconciles these prima facie contradictory observations is, of course, the divergence between the OCT and the wage in a model with human capital. In particular, Imai and Keane (2004) estimate that from age 20 to 36 the mean of

[23] We thank an anonymous referee for suggesting the importance of this question.

the opportunity cost of time increases by only 13%. In contrast, the mean wage rate increases by 90% in the actual data, and 86% in the simulated data. Thus, the wage increases about 6.5 times faster than the OCT. These figures imply that conventional methods of calculating γ will understate it by a factor of roughly 6.5.[24]

3.2. Borrowing Constraints

In a model with credit constraints, consumption and labor supply decisions no longer separate. Reallocation of hours across time requires reallocating consumption across time as well, and so an individuals' willingness to substitute labor intertemporally is potentially limited by their willingness to reallocate consumption over time.[25] Technically, the Frisch elasticity, defined as the change in hours in response to the change in the wage, holding the marginal utility of consumption fixed, no longer exists. That is, any reallocation of hours to the current period and away from other periods will reduce the marginal utility of consumption in the current period while increasing it in other periods. Still, while the Frisch elasticity

[24] It is interesting that French (2005), in a study of retirement behavior, also obtains a rather large labor supply elasticity for 60 year olds 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. In French (2005), however, the reason elasticities are greater for older workers is that for them the extensive margin is more important, in the sense that more of them are close to indifferent between working and not working. We discuss the importance of the extensive margin below.

[25] The decisions about consumption and hours also do not separate if hours and consumption are complements in utility (see MaCurdy (1983)). If the degree of complementarity is great enough, consumption will closely track hours. The positive association between consumption and labor income might suggest that individuals are credit constrained when in fact they are not. See Heckman (1974b) for a discussion of this issue.

concept no longer applies, the more general concept of an intertemporal elasticity of substitution in labor supply still applies.

Domeij and Floden (2006) argue that the existence of credit constraints may explain why researchers tend to obtain low estimates of the intertemporal elasticity of substitution when estimating MaCurdy (1981)-type equations like (2.10). They consider a model in which a household may save, but faces a non-negativity constraint on assets. Given this environment, consider the situation of a worker who is hit by a temporary negative wage shock. If borrowing were possible, the person would reduce hours in the current period and borrow against future income to smooth consumption. Similarly, even if borrowing were not possible, if the person had a stock of wealth then he/she could run it down to smooth consumption. But if borrowing is not possible, and the person has little or no wealth, then he/she can only smooth consumption by actually increasing labor supply in the current period. Thus, the non-negativity constraint on assets may actually reverse the expected sign of the intertemporal labor supply elasticity, at least for the segment of consumers with low wealth holdings. Thus, if such workers are prevalent in the data, it will attenuate the estimated hours response to wage changes.

Domeij and Floden (2006) also argue that credit constraints are important in the U.S. economy, and that a large fraction of households hold little wealth. That a large fraction of U.S. households hold little wealth appears to be well established empirically – see Deaton (1991) or Diaz-Gimenez, Quadrini and Rios-Rull (1997). The latter paper finds that the bottom 40% of the U.S. wealth distribution own only 1.4% of the capital stock.

Whether households are credit constrained is obviously a much more difficult question; i.e., whether households hold little wealth is "simply" a matter of measurement, while assessing whether households are credit constrained involves making inferences from their behavior. Indeed, the literature on testing for existence of credit constraints is rather controversial, and it is difficult to claim there is a clear consensus on whether credit constraints are quantitatively important for describing consumption and/or labor supply behavior. Domeij and Floden (2006) appeal to the work of Japelli (1990) on the number of households who report being rejected for credit in the Survey of Consumer Finances, as providing evidence for the importance of credit constraints. Some other notable papers in this literature include Zeldes (1989), Keane and Runkle (1992), Hubbard, Skinner and Zeldes (1995) and Keane and Wolpin (2001). It is beyond the scope of the present survey to assess the large literature on credit constraints and consumption, so, like Domeij and Floden (2006), we will simply assume their existence and examine their implications for labor supply elasticities.

The model that Domeij and Floden (2006) use to assess the impact of credit constraints consists of three components. The first is a period utility function as in our benchmark model in Section 2:

$$U_{it}(c_{it}, h_{it}) = \frac{c_{it}^{1-\frac{1}{\eta}}}{1-\frac{1}{\eta}} - \frac{\alpha_{it}}{1+\frac{1}{\gamma}} h_{it}^{1+\frac{1}{\gamma}}$$

The individual discounts future utility at rate β. The second is the asset accumulation constraint:

$$A_{it} = (1+r)[A_{it} + w_{it}h_{it} - c_{it}], \;\; A_{it} \geq 0 \tag{3.3}$$

And the third component is the stochastic process for wages:

$$\log w_{it} = \psi_t + z_{it} \text{ where } z_{it} = \rho z_{it-1} + \varepsilon_{it} \tag{3.4}$$

An important point to note is that the MaCurdy (1981) procedure to estimate the preference parameter γ does not require one to specify a particular wage process. Estimation of preference parameters using (2.10) only requires specification of a set of instruments that are (i) correlated with anticipated wage changes, (ii) uncorrelated with surprise wage changes, (iii) uncorrelated with tastes for work, and (iv) uncorrelated with measurement error in wages. However, as we have already seen, once we introduce extensions to the basic life-cycle model, such as human capital accumulation or credit constraints, it becomes necessary to specify the complete model, including the wage process.

Let ϕ_{it} denote the marginal utility of borrowing for person i at time t. Of course $\phi_{it} = 0$ when optimal assets are positive, but it is positive if the (nominal) optimal asset level is negative (so the non-negativity constraint is binding). In this case the marginal utility of consumption evolves according to:

$$\Delta \log \lambda_{it} = \log \beta(1+r_t) - \frac{\phi_{it-1}}{\lambda_{it-1}} + \xi_{it} \tag{3.5}$$

and hence (2.10) becomes:

$$\Delta \log h_{it} = \gamma \Delta w_{it}(1-\tau_{it}) - \gamma \frac{\phi_{it-1}}{\lambda_{it-1}} - \gamma \log \beta(1+r_t) + \alpha\gamma\Delta X_{it} + \gamma \Delta \varepsilon_{it} \quad (3.6)$$

The additional term $\phi_{it-1}/\lambda_{it-1}$ can be interpreted as an omitted variable in the conventional IV estimation method. As Domeij and Floden (2006) point out, a higher expected wage increase from period $t-1$ to t will tend to increase the marginal utility of borrowing at time $t-1$. That is, ceteris paribus, a steeper future wage profile increases one's desire to borrow against future income to finance current consumption. Thus, ϕ_{it-1} will be positively correlated with expected wage changes. Also, a higher expected wage increase from $t-1$ to t increases the worker's perceived wealth, and this reduces the marginal utility of consumption at time $t-1$. Thus, the entire term $\phi_{it-1}/\lambda_{it-1}$ is positively correlated with expected wage growth.

As is evident from equation (3.6), the term $\phi_{it-1}/\lambda_{it-1}$ also has a negative direct effect on hours growth. That is, in periods when people are liquidity constrained (i.e., the non-negativity constraint is binding and $\phi_{it-1} > 0$), they will tend to work more than they would if they could borrow against future income. Thus, the omitted variable $\phi_{it-1}/\lambda_{it-1}$ is positively correlated with expected wage growth and negatively correlated with hours growth. Hence, its omission, as in the conventional estimation procedure, will lead to a downward bias in the wage growth coefficient γ.[26]

[26] While Domeij and Floden (2006) do not discuss this issue, it is worth noting that the credit

To proceed, Domeij and Floden (2006) calibrate their model using the parameter values $\{\rho, \sigma_\varepsilon, \sigma_\psi\} = \{0.90, 0.21, 0.34\}$ obtained from estimates in Floden and Linde (2001). Note that the values of ρ and σ_ψ imply a high degree of individual persistence in wages due to autoregressive errors and time invariant individual heterogeneity. To assess the impact of credit constraints it is crucial to use a reasonable value for the variance of transitory wage shocks. The value of σ_ε $= 0.21$ is low compared to what one typically sees in raw wage data. It is consistent with the idea that a substantial fraction of observed wage variation is actually measurement error. In the utility function they set $\eta = 2/3$ and $\gamma = .50$.[27] In a model without credit constraints, this would of course imply a Frisch elasticity that is also equal to 0.5.

Domeij and Floden (2006) then follow a procedure similar to Imai and Keane (2004): They simulate data from the model, and apply MaCurdy (1981) and Altonji (1986) estimation procedures to this simulated data. As they are using simulated data they can make the estimates as precise as desired, obviating any need to consider standard errors. Using the full sample, which has roughly $200,000$ observations, their estimate of γ is 0.23. But if they restrict the sample to the roughly $157,000$ observations with positive assets, they obtain 0.44. And if they

constraint variable $\phi_{it-1}/\lambda_{it-1}$is endogenous in equation (3.6). It should not be correlated with the error component ξ_{it}, because ξ_{it} is by definition a surprise not yet revealed at time $t-1$. However, $\phi_{it-1}/\lambda_{it-1}$ can be correlated with the change in tastes for work $\Delta\varepsilon_{it}$, as these may be expected at $t-1$. If a worker expects his tastes for work to increase from $t-1$ to t (e.g., he is recovering from an illness) then he would want to borrow more at $t-1$. Thus, even if $\phi_{it-1}/\lambda_{it-1}$ could be measured (and some authors have attempted this by including proxies for credit constraints) it must be instrumented for to estimate (3.6).

[27] These parameter values replicate the statistic from the U.S. data, noted earlier, that the bottom 40% of households hold only 1.4% of wealth.

further restrict the sample to roughly $60,000$ observations with assets above the sample mean, they obtain 0.50.[28] These figures suggest that credit constraints can substantially reduce estimates of the utility function parameter γ.

Unfortunately, Domeij and Floden (2006) fall into an interpretation error when they go on to state that "... ignoring liquidity constraints ... the estimated elasticity is then 0.23 ... [while] the true elasticity is 0.50." As we emphasized when discussing Imai and Keane (2004), once one extends the basic life-cycle model to include features like human capital or credit constraints, there is no longer a direct mapping from the preference parameter γ to the intertemporal labor supply elasticity. Thus, what they should be concluding is that the mapping from regression coefficients to preference parameters is affected by the presence of credit constraints, and failing to take this into account will lead to mistaken inference regarding the true value of the underlying preference parameters.

Domeij and Floden (2006) also report estimates of the intertemporal elasticity using PSID data on male household heads. As they need to use wealth data to attempt to ascertain if households are credit constrained, they can only use the 1984, 1989 and 1994 waves (which collected wealth information). When they use the full sample (1277 observations) and instrument for wage changes using a polynomial in age and education, they obtain an elasticity of 0.42 (standard error 0.25). But when they restrict the sample to households with liquid wealth equal to at least one month's income, the estimate increases to 1.28 (standard error

[28] Finally, if they use data on the $68,000$ households with assets below 10% of the mean, they obtain a wage coefficient of -0.09. This is consistent with the example given at the start of this section to the effect that a household with little wealth may actually increase hours in response to a wage reduction.

1.15).[29]

These results are consistent with the idea that credit constraints substantially dampen the intertemporal labor supply response. One implication is that the preference parameter γ may be considerably larger than prior estimates suggest. And for some policies it may be the preference parameter γ, not the intertemporal elasticity, that matters. For example, to predict the impact of a policy that relaxes credit constraints, we would need to know γ itself. Unfortunately, Domeij and Floden (2006) do not explore implications of their calibrated model for such policies.

Another implication is that labor supply responses may be much more elastic for higher income workers. As such workers make up a disproportionate share of the tax base, the elasticity of revenue with respect to taxes may substantially exceed that of labor supply. (See Saez, Slemrod and Giertz (2009) for a survey of the recently emerging literature on the elasticity of taxable income with respect to tax rates).

However, before drawing any conclusions, a number of caveats are in order. Notice that the standard errors on the Domeij-Floden PSID estimates are so large that it is hard to draw any clear conclusions about the intertemporal elasticity. A 95% confidence interval takes it from near zero to two, and a formal test would

[29] These estimates are reported in Table 6 column 4 of Domeij and Floden (2006). We feel this column gives their most reliable estimates. In other specifications, they use an alternative wage change measure as an additional instrument, following a procedure suggested by Altonji (1986). However, this procedure is only valid under the assumption of perfect foresight, as otherwise any wage change measure is correlated with the surprise term ξ_{it} in the hours change equation. Altonji's procedure is only intended to deal with measurement error (i.e., using one noisy wage change measure to instrument for another).

probably not reject equality of the estimates in the full sample vs. the high asset sample. This is another manifestation of the weak instrument problem (i.e., age/education polynomials do not predict wage changes well) that we noted earlier.

A related paper is that by Low (2005). He explores the implications of uninsurable wage risk for the life-cycle path of labor supply. In his model wages are assumed to follow an exogenous stochastic process. The key idea in his model is that workers know that the typical life-cycle wage path has a hump-shape like that in our Figure 1. However, young workers also perceive that there is considerable uncertainty about the extent of life-cycle wage growth that they will experience personally – and they cannot insure against this uncertainty.

Hence, to the extent that there is a strong precautionary motive, workers will not choose to borrow against expected future income to finance higher consumption when they are young. Furthermore, workers will have an incentive to work relatively long hours when they are young, despite low wages. By doing so they build up a buffer stock of assets that serves as self-insurance against the potential adverse outcome that wage growth over the life cycle turns out to be much less than expected.

The essential idea of Low's model can be clearly seen by looking at his Figure 6. In a simulation where workers have certainty about the wage path, hours rise steeply over the life-cycle as wages increase (see panel b). Also, workers go heavily into debt to finance higher consumption when young, and pay off this debt in their 40s and 50s when the wage profile peaks (see panel c). However, the introduction of uncertainty changes behavior substantially. Under uncertainty, hours are much higher at young ages, and hours' growth is greatly attenuated. Indeed, hours now follow a mildly curved upside-down U-shape over the life-cycle similar to that observed in the data. As for assets, workers no longer go into debt in their 20s, and they begin to accumulate assets in their 30s.

A researcher who looked at data generated from Low's model using MaCurdy (1981) type methods to estimate equations like (2.10) would again conclude that the value of the preference parameter γ was quite small. Hence, if the insurance mechanism that Low describes is quantitatively important, it is again the case that the preference parameter γ may be considerably larger than prior estimates suggest. And again, for some policies it may be the preference parameter γ, not the intertemporal elasticity, that matters. For example, consider a policy that enhanced social insurance, such as more generous insurance against unemployment or health risks (or against any other outcomes that might lead to negative wage shocks in middle age). In Low's model this might be expected to induce workers to work substantially fewer hours when young. But a model with a low value of γ would not generate that outcome. Unfortunately, Low (2005) did not use his model to explore any policy simulations.

Of course, while the qualitative results that are generated by Low's model stem directly from uninsurable wage risk, the quantitative results hinge on the specification of the wage process (in particular the degree of uncertainty) and preferences (in particular the strength of the precautionary motive). Low (2005) uses the period utility function:

$$U(c_{it}, l_{it}) = \frac{(c_{it}^{\theta} l^{1-\theta})^{1-\frac{1}{\eta}}}{1-\frac{1}{\eta}}$$

where l_{it} is now leisure of individual i in period t. In this specification η plays multiple roles. First, it determines the degree of curvature of the utility function in the consumption/leisure composite. As η decreases, the degree of curvature increases. More curvature implies, loosely speaking, that consumers are less willing to substitute utility across periods. That is, if hours are to be set high in a period, consumers desire to set consumption high as well (to compensate). Thus, a lower η dampens the intertemporal elasticity of substitution in labor supply. Second, and for the same reason, a lower η also increases the precautionary savings motive (as consumers have more of an incentive to insure against fluctuations in w_{it}).

In his baseline model Low sets $\eta = 1/2.2$, and $\theta = .4$, based on estimates from Attanasio and Weber (1995). He assumes a wage process that includes a deterministic quadratic in age with a peak at age 50 (so as to match the inverted U-shape in our Figure 1), as well as permanent and transitory wage shocks. Based on estimates from Meghir and Pistaferri (2004), the standard deviation of permanent shocks is set to 0.18 and that of transitory shocks is set to 0.17. These values seem plausible in light of other existing estimates.

The difference between Low (2005) and Domeij and Floden (2006) is that in the former paper workers do not borrow against future income when young because they do not want to, while in the latter they do not borrow when young because they cannot. This illustrates the point we made at the beginning of this section about why credit constraints are difficult to identify empirically: they generate behavior that looks very similar to behavior generated by several other mechanisms: a strong precautionary motive, complementarity of consumption and hours, time varying tastes for work/consumption, etc.. Ignoring any one of these mechanisms can potentially lead studies based on equation (2.10) to obtain biased estimates of preferences.

3.3. Optimization Frictions

Chetty (2010) offers a different explanation for why the small estimated labor supply elasticities in many studies may be biased. In particular, he considers the implications of fixed costs of adjusting labor supply for estimates of labor supply elasticities. He argues that fixed adjustment costs may arise for a number of reasons. Such costs may be features of the technology (e.g., one may have to deal with various organizational details to adjust one's work hours or search for a new job at a different firm). Alternatively, these "fixed costs" may be a reduced form representation of mental phenomena such as optimization errors or psychic costs (i.e., a cost of doing the necessary mental calculations to re-optimize when tax rates change). A related idea is that not all tax changes are "salient" (i.e., people ignore them) because any gains from adjusting to them are too small to be concerned with.

Chetty's approach to the problem of adjustments costs is rather different from papers we have discussed so far in that he does not actually solve or estimate an extension of a basic labor supply model that incorporates fixed costs of adjustment.

Rather, he attempts to bound the magnitude of the bias in labor supply elasticity estimates that might reasonably be attributed to ignoring fixed costs. The basic question that Chetty asks is this: Suppose that when taxes change, people do not find it optimal (or cannot be bothered) to adjust their labor supply if the resultant loss in welfare is less than some small fraction (δ) of consumption, where that fraction represents the fixed cost.[30] In that case, what is the bias in conventional estimates of labor supply elasticities (that ignore costs of adjustment) likely to be?

One might well expect that adjustment costs would not lead to any particular bias in labor supply elasticities. If people often do not bother to respond to small tax changes, it also follows that when they do respond they will occasionally make very large adjustments – e.g., a small tax change may follow a series of prior small changes that have left the person rather far from his/her optimal hours point, and one additional small change may then induce a large jump in hours.

But the somewhat surprising conclusion that Chetty draws is that elasticity estimates are likely to be biased downward. This stems from an asymmetry in how adjustment costs affect behavior when elasticities are high vs. low: If the labor supply elasticity is large it means the objective function is fairly flat in the vicinity of optimal hours. Thus, the converse of the labor supply elasticity being large is that a fairly large departure from optimal hours will lead to only a small

[30] While the idea of consumers not putting in the mental effort to adjust when the gains would be small has intuitive appeal, it runs into a logical lacuna in practice: One has to calculate the gain in the first place to determine if it would be small. So, if one has to do the mental effort anyway, why not adjust? This is not to say that mental effort is not important, only that it is difficult to model formally.

welfare loss. So if labor supply elasticities are large, we may easily observe small labor supply responses to taxes provided there are small adjustment costs. In contrast, if labor supply elasticities are small, then small adjustment costs provide no mechanism that would cause us to infer they are large.

To give an extreme example, Keane (2010) surveys 21 well-known studies that estimate the Hicks elasticity, and finds that 13 produce very small estimates (near zero) while 8 produce fairly large estimates (0.30 or above). In Chetty's framework, such a distribution of estimates would be unsurprising if the "true" elasticity were large. But it would be difficult to rationalize if the "true" elasticity were small. Thus, a researcher faced with these estimates would conclude the "true" elasticity is probably fairly large. Of course, this example is extreme because it ignores all the differences between the studies (i.e., it assumes that the studies represent iid draws from a distribution of possible estimation outcomes). One could still rationalize a small elasticity by arguing that all the studies that produced large elasticities were flawed in some way. But this example does clearly illustrate Chetty's basic idea

To proceed, we assume a simple quasi-linear utility function[31]:

$$U_i = wh_i(1-\tau_i) - \frac{\alpha}{1+\frac{1}{\gamma}} h_i^{1+\frac{1}{\gamma}} \tag{3.7}$$

As there are no income effects, the Marshall, Hicks and Frisch elasticities are

[31] Chetty shows how the results can be extended to the case of utility functions that are more commonly used in the literature. But it is much easier to exposit the ideas in the quasi-linear case.

equivalent. The optimal level of hours in this model is simply:

$$h_t^* = [\frac{(1-\tau)w}{\alpha}]^\gamma \tag{3.8}$$

and utility evaluated at the optimum is:

$$U(h_t^*|\tau_t) = \frac{1}{1+\gamma}[\frac{1}{\alpha}]^\gamma[(1-\tau_t)w]^{1+\gamma} \tag{3.9}$$

Now, consider a change in the after-tax rate $(1-\tau)$. The impact on utility can be decomposed into the direct effect of the change in the tax rate holding the tax rate fixed, plus the part induced by the behavioral response of changing labor supply:

$$U(h_{t+1}^*|\tau_{t+1}) - U(h_t^*|\tau_t) = [U(h_t^*|\tau_{t+1}) - U(h_t^*|\tau_t)] + [U(h_{t+1}^*|\tau_{t+1}) - U(h_t^*|\tau_{t+1})] \tag{3.10}$$

From (3.7), the first term on the right hand side is obviously just $wh_t^*\Delta(1-\tau)$, the increase in consumption holding labor supply fixed. Note from (3.9) that $\frac{dU(h_t^*|\tau_t)}{d(1-\tau_t)} = wh_t^*$. Thus, the second term on the right hand side of (3.10) – the hours adjustment term – is a second order effect that can be ignored for purposes of calculating effects of small tax changes (of course, this is just a simple application of the envelope theorem).

The idea that utility gains from adjusting hours are only second order is the key to Chetty's idea that agents will not make these adjustments if there are small adjustment costs. The utility gains from hours adjustments are captured by the second (and higher) order terms of the Taylor series approximation.[32] As $\frac{d^2U(h_t^*|\tau_t)}{d(1-\tau_t)^2} = \gamma w h_t^*/(1-\tau_t)$ we have that to second order:

$$U(h_{t+1}^*|\tau_{t+1}) - U(h_t^*|\tau_t) = wh_t^*\Delta(1-\tau_t) + \frac{1}{2}\gamma\frac{wh_t^*}{(1-\tau_t)}\Delta(1-\tau)^2 \qquad (3.11)$$

Dividing the second order term by consumption, we get that the utility gain from adjusting hours to a change is approximately $\frac{1}{2}\gamma[\frac{\Delta(1-\tau)}{(1-\tau_t)}]^2$. Thus, for example, if $\gamma = 1$ and the tax rate falls from 33% to 30%, the utility loss from failing to adjust hours is only about 0.1% of initial consumption. In contrast, for a larger tax cut from 40% to 30% (a 25% cut) the utility loss is about 3% of consumption, as the second order effects become important.

However, for any assumed values of γ and δ (the adjustment cost as a fraction of consumption) one can work out whether people would adjust to a particular tax change (assuming they start from optimal hours). Chetty then looks at the U.S. Tax Reform act of 1986 (TRA86), which flattened the progressive tax system by reducing rates in a way that was biased toward the high end. Assuming $\gamma = .5$, he finds that the cost of failing to adjust hours in response to TRA86 was generally less than 1% of consumption for people earning $\$100,000$ or less, but that the loss grew to 4% of consumption at the $\$200,000$ level. Thus, Chetty argues that an adjustment cost of roughly $\delta = 1\%$ could rationalize the empirical finding that middle income workers had little response to TRA86 (see Gruber

[32] Note that the second order term in the Taylor series approximates the second term on the right hand side of (3.10) – the hours adjustment term – but is not equivalent to it. The divergence will become greater for very large tax changes where the higher order terms matter.

and Saez (2002), Saez (2004)) while high income workers had large responses (see Auten and Carroll (1999), Saez (2004)).

Now suppose we require that the utility loss from failing to adjust hours satisfies the condition that it is less than the fraction δ of consumption:

$$U(h^*_{t+1}|\tau_{t+1}) - U(h^*_t|\tau_{t+1}) = \frac{1}{2}|U''(h^*_t)|(h^*_{t+1} - h^*_t)^2 < \delta w h^*_t(1-\tau_t) \tag{3.12}$$

Given the utility function in (3.7), we have that $|U''(h^*_t)| = \alpha(1/\gamma)(h^*_t)^{\frac{1}{\gamma}-1}$. Utilizing this and the assumption that hours were at their optimal level at t, we obtain a bound on the maximum percentage deviation of hours at $t+1$ from their optimal level:

$$\frac{h^*_{t+1} - h^*_t}{h^*_t} < [2\gamma\delta]^{1/2} \tag{3.13}$$

This expression clarifies the point about asymmetry we made earlier. If γ is small, then, for a given δ, hours must stay close to their optimal level. Thus, it is unlikely we will see a large hours response to a small tax change. But if γ is large we can see large deviations of hours from the optimum, so it is plausible to see negligible responses to moderate tax changes.

Chetty uses the bound on hours changes to derive bounds on elasticities, given observed hours responses to tax changes. As the estimated elasticity is the observed percent change in hours divided by the percentage change in $(1-\tau)$, it is clear that the observed elasticity in a study may depart from the true one by plus or minus $[2\gamma\delta]^{1/2}/\Delta\log(1-\tau)$, assuming workers start at an optimum at time t.[33]

[33] If that assumption is not invoked (as is the case in Chetty (2010)), the width of the bounds doubles.

As the change in tax rates appears in the denominator of the bounds, they will be wider for smaller tax changes.[34]

This argument suggests that estimates of labor supply elasticities in different contexts could lead to a range of estimated elasticities even if the true underlying elasticity was the same in all cases. Specifically, he argues that estimates from contexts in which wages or taxes changed relatively little might be expected to generate elasticity estimates that were biased toward zero, so that the true value of the preference parameter γ might be significantly larger than the estimated coefficient on wages or taxes. In this sense, his conclusion is similar in spirit to those of Imai and Keane (2004), Domeij and Floden (2006) and Low (2005) that we have discussed previously. However, one difference with Imai and Keane, for example, is that while they estimate the features of the human capital accumulation technology jointly with the preference parameters of interest, Chetty simply considers the implications of various values of δ that are chosen without reference to the data.[35] He goes on to argue that this same logic can help us understand why elasticity estimates based on aggregate data are typically larger than those based on micro data. The reason for this is that the variation in taxes in aggregate studies is often larger than in micro studies.

A good illustration of this point is Chetty's analysis of MaCurdy (1981), assuming $\delta = 0.01$. As he states "... even though MaCurdy (1981) estimates an

[34] Note that, as the bounds depend on the elasticity γ itself, we actually obtain an implicit equation, which Chetty solves to obtain an explicit expression for the bounds. In this expression, the square of the percentage change in the tax rate appears in the denominator.

[35] Chetty et al (2011) consider a specific friction–search costs–and pursue a more structural approach that is consistent with the ideas in Chetty (2010).

intensive margin elasticity of only 0.15, his estimate is consistent with a structural elasticity as large as $\gamma = 5.63$. The reason is that MaCurdy's estimates are identified from changes in wage rates of approximately 10%, which is not big enough to overcome small frictions."

Similarly, the Blundell et al (1998) study identified labor supply elasticities for employed married women, by exploiting UK tax rate variation from 1978-92. They obtained a modest value for the compensated elasticity of 0.20, but Chetty (2010) derives bounds on their estimate that range from essentially zero to 2.54. In general, the message is that if we admit the possibility of small adjustment costs, then the best known micro data studies that have estimated small (intensive margin) elasticities do not actually rule out large elasticities (although they do not rule them in either!).

Chetty (2010) also applies his methodology to the data in Prescott (2004)'s analysis by identifying labor supply elasticities based on differential aggregate hours and tax rate changes between the U.S. and the UK from 1979-1996. Here the bounds are 0.42 to 2.14. They are tighter because the relative tax changes were quite large. In fact, this turns out to be one of the most informative studies that Chetty examined in the sense of generating a relatively large lower bound.[36]

Another implication that Chetty notes is that estimated elasticities may depend on the length of time that has elapsed since the change in taxes. One reason for this is that adjustment costs might exhibit random fluctuations, so that the

[36] Of course, a key issue here is that other factors may have also shifted labor supply in the US vs. the UK over the sample period. More generally,just as estimates from micro data face several econometric issues, the issue of omitted factors is a key issue for estimates based on aggregate data.

more time that has passed, the more likely it is that an individual has had a low realization of these costs and has therefore adjusted labor supply.

Finally, Chetty examines several studies that look at the elasticity of taxable income (ETI) with respect to tax rates. Since Feldstein (1995) a large literature has grown up around this topic in public finance. The basic idea is that taxable income may be quite responsive to tax rates even if labor supply is not. This is because people have mechanisms to shield income from taxes, but these mechanisms require effort, and higher taxes increase the optimal level of effort to devote to tax avoidance (or income shifting). An essential idea from the optimal tax literature is that if income is more elastic with respect to the tax rate then the optimal rate is lower. Thus, even if labor supply is quite inelastic, it may nevertheless be optimal to have low tax rates if taxable income is highly elastic.

We do not discuss this ETI literature further here, primarily because the point of the literature is conceptually quite different from our focus. In effect, the ETI literature is an attempt to reconcile an argument against high tax rates with the "inconvenient" evidence of low labor supply elasticities – by arguing, as Feldstein (1995) did, that the ETI is nevertheless large. In contrast, we seek to argue that the existing evidence on labor supply can be reconciled with a world where labor supply elasticities are in fact large. In particular, if the Hicks labor supply elasticity is actually large then one can obviously construct an argument that welfare costs of taxation are high without having to resort to arguments about tax avoidance.[37]

[37] A second reason we do not discuss the ETI literature is that it has not yet come to any clear conclusions about the size of the ETI. For instance, the survey by Saez, Slemrod and

3.4. Summary

The main conclusion of this section is that when one adds empirically plausible features to the simple benchmark model of Section 2, the mapping from coefficients in a prototypical regression equation such as equation (2.10) to underlying preference parameters can be drastically altered. Additionally, the implications of a particular value of the preference parameter γ for how individual and aggregate labor supply respond to various tax policies is also affected. Some pieces of conventional wisdom may even be overturned. For example, with human capital accumulation, the hours responses to temporary shocks can be smaller than the hours responses to permanent shocks. With credit constraints, the immediate response to a temporary decrease in wages may even be to increase hours.

4. Aggregate Labor Supply In Models with Extensive Margin Adjustment

In section 2 we embedded a simple life cycle labor supply problem into a standard aggregate model. A key property of that benchmark model is that a single preference parameter, γ, played a key role in determining how individual life cycle labor supply as well as aggregate labor supply respond to specific changes in the

Giertz (2009) states: "Estimates of the long-run elasticity of taxable income are plagued by extremely difficult issues of identification, so difficult that we believe that there are no convincing estimates of the long-run elasticity of reported taxable income to changes in the marginal tax rate." Furthermore, Slemrod and Kopczuk (2000) argue that, even in models where the labor supply elasticity is a primitive parameter of preferences, the ETI will not be. For instance, a reform that broadens the tax base (like TRA86) would lower the ETI by making income shifting more difficult. This non-invariance makes any structural interpretation of ETI estimates quite problematic.

economic environment. In particular, we showed explicitly in the context of a simple tax and transfer policy that the individual and aggregate elasticities were identical functions of this one preference parameter. Key to this result was the fact that in this benchmark model, the tax and transfer policy induces a uniform shift of the life cycle labor supply profile, so that knowledge of the shift in labor supply at any one point in the life cycle is a sufficient statistic for the change in both lifetime and aggregate labor supply. The life cycle labor supply problem in this benchmark model is consistent with the earlier estimation exercises of MaCurdy (1981) and others. The previous section studied how various extensions to this simple life cycle model influence the estimated values of the key preference parameter and its implication for various labor supply elasticities. In this section we take up a different issue which also relates to the connection between the value of this particular preference parameter and the responsiveness of aggregate labor supply to various changes in the economic environment.

The key feature of the models discussed in this section relative to the benchmark model from Section 2 will be the presence of an extensive margin. As a historical note, it is of interest to note that more than 25 years ago in his joint discussion of one micro and one macro paper on labor supply, Heckman (1984) called for the development of labor supply models that featured both intensive and extensive margins in order to have a unified theory capable of reconciling individual and aggregate features of labor supply.

The starting point for our discussion are the papers by Hansen (1985) and Rogerson (1988), who studied homogeneous agent models in which all adjustment at the individual level was assumed to occur at the extensive margin, i.e., the intensive margin was fixed by assumption. Specifically, it was assumed that individuals had preferences over consumption and hours of work given by:

$$\sum_{t=0}^{\infty} \beta^t [u(c_t) + v(1 - h_t)]$$

The assumption that all adjustment occurs at the extensive margin was captured by the constraint that the individual's choice of h_t had to lie in the finite set $\{0, \hat{h}\}$. A key early result in this literature was that assuming a set of markets sufficiently rich to decentralize optimal allocations, the aggregate allocations in this economy were identical to those that would emerge from an economy in which there was a representative household that made all labor supply adjustment at the intensive margin but had preferences given by:

$$\sum_{t=0}^{\infty} \beta^t [u(c_t) - \alpha h_t]$$

where α is a constant. The importance of this equivalence result is that this representative household behaves as if they have a Frisch elasticity equal to infinity, and that this is true independently of the function $v(1-h)$ that described the true preferences of individuals in the economy. It should be clear to the reader that this result potentially creates a serious disconnect between micro data estimation exercises in which researchers estimate parameters of $v()$, and the associated implications for aggregate behavior.

However, there are a few issues concerning this result that are worth noting.

First, because the choice set of individuals in this economy is not convex, the early derivations of this result assumed that individuals in the economy could trade lotteries in the decentralized equilibrium. That is, individuals could sell a lottery in which they work with probability π and do not work with probability $1-\pi$. To the extent that we do not observe workers and firms trading these types of lotteries, one might question the relevance of this result if this feature of the decentralization is essential. However, subsequent work has argued that lotteries are not essential to this result. Ljungqvist and Sargent (2006, 2008) argue that "time averaging" is a perfect substitute for lotteries if an individual has access to credit markets. The basic idea is that if there are many time periods, then working each period with probability π is equivalent to working with certainty during a fraction π of the periods. In the equilibrium with lotteries, the individual receives a smooth stream of compensation. In the equilibrium with "time averaging", the compensation profile is no longer smooth, since the individual only receives income in those periods in which he or she works. But as long as the individual has access to credit markets, all that matters is the present value of the compensation profile. Ljungqvist and Sargent establish this formally in an environment with finite lifetimes, continuous time and no discounting. Krusell et al (2008) extend this result to the case of infinitely lived agents in discrete time with discounting.

Browning et al (1999) raised another issue concerning the relevance of the above equivalence result. Specifically, they argued that in order to be empirically relevant, a model of choice along the extensive margin should be able to capture the movements of individual workers into and out of employment. With this in mind they suggested that the lottery equilibria in Hansen (1985) had the counterfactual implication that all workers were equally likely to be employed next period, independently of their current employment status. While this critique does apply to the specific equilibrium that Hansen studied, it turns out that equilibrium does not impose any restrictions on the nature of individual transitions. One way to see this is to note that in the time averaging equilibrium that Ljungqvist and Sargent (2006, 2008) studied, all that is required is that individuals spend a certain fraction of their life in employment. That is, any profile for employment that implies the same total labor supply over an individual's lifetime is consistent with individual optimization. However, one could amend the Browning et al critique and argue that a defect of the Rogerson and Hansen model is that it does not impose any discipline on individual employment histories. Given the many empirical regularities that have been documented, this could be interpreted to imply that the model is missing some important features. To the extent that this is the case, the possibility arises that the result would not be robust to the inclusion of these additional features.

Related to this, Cho (1995) and Mulligan (2001) showed that the implication of an infinite Frisch elasticity for aggregate labor supply was not robust to allowing for certain kinds of heterogeneity. More generally, the Frisch elasticity for aggregate labor supply would depend on the nature and extent of heterogeneity.

Finally, another issue of interest is that the simple aggregate models of Hansen and Rogerson do not speak to the issue of reconciling the features of life cycle labor supply with properties of aggregate labor supply. If one is looking for a unified theory of labor supply at the individual and aggregate level then one clearly wants to be able to address life cycle observations in these models.

In the next two subsections we consider models which have addressed these issues.

4.1. Chang and Kim (2006)

We begin by considering the paper by Chang and Kim (2006). This paper considers an aggregate model in which individuals are subject to idiosyncratic shocks, face incomplete markets for credit and insurance, and in which all labor supply adjustment occurs at the extensive margin. Additionally, it considers households that consist of a male and a female member, with household preferences given by:

$$\sum_{t=0}^{\infty} \beta^t [2\log(.5c_t) - \alpha_m \frac{h_{mt}^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - \alpha_f \frac{h_{ft}^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}]$$

where c_t is household consumption, h_{mt} is hours worked by the male household member and h_{ft} is hours worked by the female household member. As noted, it is assumed that individuals can only supply 0 or $\hat{h}$ units of labor in any period. In steady state the wage per efficiency unit of labor services will be constant and denoted by w, but individual productivity will be stochastic. If x_t denotes labor productivity for a worker in period t then he or she will have labor earnings equal to $wx_t\bar{h}$ if working. The labor productivity of each member is assumed to follow a stochastic process:

$$\log x_{jt+1} = \rho_j \log x_{jt} + \varepsilon_{jt+1},\ j = m, f \tag{4.1}$$

where the innovations are assumed to be normally distributed with mean zero and standard deviation σ_j. Innovations are iid across time and across individuals.

The production side of the economy is standard, with a Cobb Douglas aggregate production function that uses capital and labor services. Capital depreciates at the constant rate δ and output can be used as either investment or consumption.

The market structure is as follows. In each period there are competitive markets for capital and labor services as well as output. Individuals are allowed to have negative holdings of capital (i.e., be in debt) but capital holdings cannot go below $\bar{a}$. There are no markets for insurance against idiosyncratic shocks, but as in Aiyagari (1994) and Huggett (1993), individuals can accumulate capital to self insure. Note that there are no markets for employment lotteries in this economy.

Chang and Kim focus on the steady state equilibrium of a calibrated version of their model. Table 2 in their paper provides information on all of the calibrated parameter values. Here we focus on a few key details. First, the period length is chosen to be one quarter. The stochastic processes for both male and female idiosyncratic shocks are quite persistent ($\rho_m = .948$, $\rho_f = .925$) and quite variable ($\sigma_m = .269$, $\sigma_f = .319$). The constants α_j are calibrated so as to match the employment to population ratios for both males and females. Total time endowment is normalized to one and $\hat{h}$ is set to 1/3.

Chang and Kim present various statistics on the distribution of earnings and wealth to argue that their model does a reasonable job of capturing the amount of heterogeneity in the data along these dimensions.[38] Because all households are

[38] As is well known, the model cannot capture the extreme concentration of wealth in the upper one percent of the wealth distribution.

the same except for the realizations of the idiosyncratic shocks, the steady state distribution of households across outcomes is also the same as the time series averages for a given household. Since the model is calibrated to match the economy wide employment to population ratios for both males and females, it follows that each individual will only spend a fraction of their life in employment. Over time, individuals will move between spells of employment and nonemployment. While Chang and Kim do not address the issue of whether this model produces empirically reasonable patterns for these transitions, recent work by Krusell et al (2010, 2011) argues in a slightly more general version of this model that this is the case.

Chang and Kim then use the steady state equilibrium of their calibrated economy as a laboratory to consider the properties of individual and aggregate labor supply. The first exercise that they carry out is the following. They consider a sample of 50000 households in the steady state, simulate their histories for 120 quarters and then aggregate the observations to annual frequencies. In the spirit of Altonji (1986) they run a panel regression of the following form using individuals who have positive hours in each year:

$$\log h_{it} = \gamma(\log w_{it} - \log c_{it}) + \varepsilon_{it} \tag{4.2}$$

They do this separately for both men and women. For their benchmark calibration they obtain estimates of γ equal to .41 and .78 for males and females respectively. The key finding here is that if one runs standard labor supply regressions on individual data generated by the model, one will obtain relatively small estimates of the labor supply elasticity parameter for men, and a larger estimate for women.

They then consider the aggregate labor supply elasticity in their model. To assess this they carry out the following experiment in the spirit of Kydland and Prescott (1982). They assume that the economy is subject to an AR(1) aggregate technology shock, simulate the economy for 30000 quarters, compute aggregates and run the regression in equation (4.2) using aggregate time series data. The resulting estimate for γ is now 1.08. In a second exercise they consider a stand-in household model with preferences of the form:

$$\sum_{t=0}^{\infty} \beta^t [\log(c_t) - \tilde{\alpha} \frac{h_t^{1+\frac{1}{\tilde{\gamma}}}}{1+\frac{1}{\tilde{\gamma}}}]$$

where h_t is now allowed to take on any value in the interval $[0, 1]$. They consider various values of $\tilde{\gamma}$ and in each case recalibrate the model so as to match the same aggregate targets. Assuming the same process for aggregate technology shocks they compute standard business cycle moments for both this economy and the previous economy. The business cycle statistics from the stand-in household model are most similar to those from the heterogeneous agent economy when $\tilde{\gamma}$ is set equal to 2. From the perspective of this exercise we see that if one wanted to use a stand-in household model to mimic the business cycle statistics for the heterogeneous agent economy, one would have to adopt a value of $\tilde{\gamma}$ that is roughly five times as large as the estimate based on individual data for male workers. The presence of empirically reasonable heterogeneity in this model does indeed have a dramatic effect on the implied aggregate elasticity, lowering it from infinity down to around 2. Nonetheless, the key point is that the value of 2 is still large.

Related work has examined the extent to which this framework influences how aggregate hours work react to a simple tax and transfer program like the one studied in Section 2. Alonso-Ortiz and Rogerson (2010) use a single agent household version of the Chang and Kim model and find that the response in aggregate hours is large, in fact, somewhat larger than what one finds for a stand-in household model with a Frisch elasticity of 2. Ljungqvist and Sargent (2006, 2008) consider a model in which individuals have finite lives, are subject to a stochastic learning-by-doing technology, and face a discrete labor supply choice. While they do not use their model to assess elasticity estimates from micro data, they do find that the response of aggregate hours in their model is similar to what is found in models that abstract from human capital accumulation altogether. However, the model with human capital accumulation has very different predictions for the identities of which individuals choose not to work as the tax and transfer program is expanded.

In subsequent work, Chang and Kim (2007) use a version of their model described above that assumes single agent households in order to study additional properties of business cycles. While not of direct relevance to the issues that we focus on, it is interesting to note that their model addresses some earlier shortcomings of the stand-in household model commonly used in aggregate analyses. Specifically, as noted by many authors, beginning with Mankiw et al (1985), the aggregate data are not consistent with the static first order condition implied by consumer optimization. That is, observed values for hours, consumption and wages (or alternatively labor productivity) are not consistent with the marginal rate of substitution being equated to the real wage rate or labor productivity. The subsequent literature has labelled the discrepancy in this first-order condition as the "labor wedge".[39] Chang and Kim (2007) show that in their model with aggregate technology shocks, if one tries to interpret the resulting aggregate data using a stand-in household model, one will generate substantial movements in the labor wedge over the business cycle. An et al (2009) further show that this model can reconcile the types of results found in Mankiw et al (1985).

To summarize, the key finding from Chang and Kim (2006) is that in the steady state of their model, the Frisch labor supply elasticity estimated from micro data is not the same as the one estimated from aggregate data. Moreover, there is no connection between the elasticity from micro data and the preference parameter γ, since by construction the value of γ is irrelevant in their model. The contribution of this work is to establish that the earlier results of Hansen (1985) and Rogerson (1988) continue to be quantitatively relevant in a setting that features empirically relevant sources of heterogeneity and a plausible market structure, thereby providing a better bridge between analyses of individual and aggregate labor supply.

One limitation of the Chang and Kim analysis is that it precludes adjustment along the intensive margin. In their regressions involving micro data, all of the variation in their measure of the intensive margin (i.e., annual hours worked) comes from extensive margin adjustment during the year. In the data we know

[39] See for example, the papers by Parkin (1988), Bencivenga (1992), Hall (1997), Gali, Gertler and Lopez-Salido (2002), Mulligan (2002), Chari, Kehoe and McGrattan (2004) and Cole and Ohanian (2004) who note this property in various contexts.

that there is some "true" adjustment along the intensive margin. In the next subsection we discuss models that allow for a continuous choice of hours along the intensive margin. Chang et al (2011) consider a model that lies in between these two alternatives. Specifically, they assume that workers must choose among three work options: no work, part-time work, or full-time work. In this model, there is adjustment along the intensive margin, i.e., workers moving between full-time and part-time work, but this adjustment is discrete. Chang et al (2011) find that standard labor supply regressions do not uncover the true value of γ in this environment, even when focusing on the choice of intensive margin for continuously employed workers. Additionally, the value of γ seems not to matter for aggregate responses in total hours.[40]

4.2. Rogerson and Wallenius (2009)

We now turn to a second class of models that considers the relation between individual and aggregate labor supply elasticities in environments that feature an extensive margin at the individual level. Unlike the previous models in which all adjustment occurred at the extensive margin, this class of models will allow for adjustment along both margins. We focus on the analysis in Rogerson and Wallenius (2009), which is in turn a generalization of the model in Prescott et al (2009). This model can also be viewed as embedding a simplified version of

[40] Another case of interest is when the choice of hours along the intensive margin features an element of coordination. In this case, individuals may not be free to adjust hours along the intensive margin in response to idiosyncratic shocks, whereas we might observe adjustment along the intensive margin in response to aggregate changes in the economic environment. See Chetty et al (2011a) and Rogerson (2011) for further discussion of this issue.

French (2005) into a general equilibrium setting.

4.3. Model

This model of life cycle labor supply emphasizes two key dimensions of lifetime labor supply: the fraction of life that an individual devotes to employment and the fraction of time devoted to market work in those periods in which the individual is employed. It is convenient to formulate the problem in continuous time to make the choice of what fraction of life to spend in employment a continuous one. Consider an individual with length of life normalized to one who has preferences defined by:

$$\int_0^1 [u(c(a) - v(h(a))]da$$

where $c(a)$ is consumption at age a, $h(a)$ is time devoted to market work at age a, $u(\cdot)$ gives the utility flow from consumption and $v(\cdot)$ gives the disutility flow from working. Note that the individual does not discount future utility flows in this specification. Although we will not present the analytics of the model here, this serves to simplify the analytic characterization of the solution to the individual's maximization problem. The steady state interest rate is assumed to equal zero, so that these two factors will be offsetting as is standard in many macroeconomic models with infinitely lived agents.

To generate variation along the intensive margin when working, the productivity of an individual's time is assumed to vary systematically over the life cycle and is denoted by $e(a)$.[41] The wage rate per unit of labor services is assumed

[41] We note an interesting issue that arises with a specification in which this productivity

to be constant over time and equal to w. The individual faces complete markets for borrowing and lending and as noted above the interest rate on borrowing and lending is set equal to zero. In a "standard" model the present value budget equation would be given by:

$$\int_0^1 c(a)da = w \int_0^1 h(a)e(a)da$$

The key innovation of this model is to follow Prescott et al (2009) by adding a nonconvexity to the mapping from time devoted to work to the resulting labor services. In particular, when a worker of age a devotes h units of time to market work, the resulting supply of labor services is given by $g(h)e(a)$, where for ease of exposition $g(h)$ is assumed to take the form:

$$g(h) = \max\{h - \bar{h}, 0\}.$$

The key property of this specification is that the relation between total labor earnings and hours devoted to market work is convex. More generally, one could

process is exogenous and credit markets are complete. In the data, wages are not symmetric over the life cycle, in the sense that wages at the end of the life cycle are much higher than wages at the beginning of the life cycle. If wages are exogenous and markets for borrowing and lending are perfect, this creates a problem for a model that includes an endogenous retirement decision. The reason for this is that there is an incentive for individuals to avoid working in the early part of life in order to avoid the low wages during this period, and to instead work more at the later part of the life cycle when wages are higher. Wallenius (2009) develops a version of the model studied here that features endogenous human capital accumulation as in Imai and Keane (2004) and shows how it can match the life cycle profile for both wages and hours. To maintain tractability, rather than include a human capital accumulation decision, here we follow Rogerson and Wallenius (2009) and abstract from trying to match the actual profile of wages over the life cycle.

consider specifications in which the marginal wage is a function of the length of the workweek.[42] If $\bar{h} = 0$ then the individual labor supply problem is completely standard, but cannot generate "retirement" as an endogenous outcome, in the sense of a worker who has labor supply that switches from full time work to no work at a point in time.[43] The new present value budget equation is now:

$$\int_0^1 c(a)da = w \int_0^1 \max\{h(a) - \bar{h}, 0\} e(a)da \tag{4.3}$$

The above discussion has thus far only described a single agent decision problem. Rogerson and Wallenius consider this single agent problem in the context of a steady state equilibrium of an overlapping generations model. At the risk of trivializing the general equilibrium considerations, but with the gain of transparency, assume a small open economy in which the real interest rate is exogenously fixed at zero, and an aggregate production function that is linear in labor services with marginal product normalized to one.[44] If the price of output is normalized to one, the equilibrium wage rate w per efficiency unit of labor must also equal unity. Assuming a new generation of identical individuals with total mass equal to one is born at each instant, in steady state a new-born household will maximize lifetime

[42] French (2005) considers both specifications in his empirical work. Specifically, he assumes that labor earnings as a function of hours devoted to market work are given by $w(h - \bar{h})^{1+\theta}$, where $\theta > 0$. For the issues that we discuss here it does not matter whether one focuses on the case $\bar{h} > 0$, $\theta = 0$ or $\bar{h} = 0$, $\theta > 0$, or $\bar{h} > 0$, $\theta > 0$. The key point is that nonconvexities introduced either via $\bar{h}$ or θ can lead to discontinuities in labor supply and the endogenous creation of an operative extensive margin.

[43] More precisely, if $\bar{h} = 0$ the model cannot generate discontinuous adjustment along the hours worked margin in response to continuous changes in the underlying economic environment.

[44] The small open economy assumption is not essential. In this model one can always specify a government debt policy that will support a steady state equilibrium with a zero interest rate.

utility subject to the present value budget equation (4.3).

Rogerson and Wallenius use this model to assess the quantitative consequences of the simple tax and transfer scheme introduced in Section 2. That is, consider a proportional tax on labor earnings that is used to finance a uniform lump-sum transfer to all individuals. For their quantitative work they assume $u(c) = \log c$, $v(h) = \alpha \frac{h^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$ and that life cycle productivity $e(a)$ is piecewise linear.

Given these functional forms, the key issue that Rogerson and Wallenius investigate is how the parameter γ matters for properties of the life cycle profile and how this profile responds to changes in the scale of the tax and transfer policy. For each of several values of γ they choose values for the model's other parameters so as to match three targets: fraction of life spent in employment, peak hours worked over the life cycle and wage changes over the lifecycle. Some care needs to be taken in matching up wages in the model with wages in the data. In the model, the wage per unit of labor services, which was denoted by w, is equal to unity at all points in time. But wages in the data are measured as labor earnings per hour of work, and because of the nonconvexities in the $g(h)$ function, wages per unit of time, denoted by w^h, are not equal to unity.[45] A tax rate on labor earnings of .3 is assumed when calibrating the model, which corresponds to the average effective tax on labor income in the US in recent years. Having calibrated the model, Rogerson and Wallenius then examine what happens to equilibrium hours if the tax rate is increased to .5, which corresponds to the average effective

[45] If the fixed time cost $\bar{h}$ were intepreted as a commuting cost then this effect would not be present. While this intepretation is not relevant for the main results reported below, the assumption that $\bar{h}$ represents time at work does have interesting implications for the connection between standard labor supply regressions and the underlying preference parameter γ.

tax on labor income in several economies in continental Europe in recent years.[46]

4.3.1. Micro Elasticities On the Intensive Margin

Before reporting the effects of the change in tax and transfer policies, it is of interest to examine some features of individual labor supply in the calibrated benchmark economies. Given a value of γ and the calibration procedure just described, the model will generate a life cycle profile for hours worked, $h(a)$, and hourly wages, $w^h(a)$. Rogerson and Wallenius generate a panel life cycle data set for hourly wages and hours worked by choosing 67 equally spaced values during the period of life in which hours are positive, running from 0 to .66 and evaluating the two functions $h(a)$ and $w^h(a)$ at these points. Note that all of the data points in the sample are times at which individuals are employed. They run the regression:

$$\log(h(a)) = b_0 + \tilde{\gamma}\log(w^h(a)) + \varepsilon(a) \tag{4.4}$$

The resulting parameter estimate $\tilde{\gamma}$ is the micro labor supply elasticity for individuals in the model, viewed through the lens of the standard model described in Section 2.

Table 3 shows the estimated values of $\tilde{\gamma}$ for the benchmark calibrated model for four different values of $\gamma : .1, .5, 1,$ and 2.

[46] Several authors have produced estimates of effective tax rates for various countries, including Mendoza et al (1994), Prescott (2004) and McDaniel (2006). While there are small differences in methodology across studies, the 20% differences between the US and countries such as Belgium, France, Germany and Italy is a robust finding.

Table 3

Estimated Micro Elasticities			
$\gamma = 2$	$\gamma = 1$	$\gamma = .5$	$\gamma = .10$
1.29	.59	.28	.05

The table shows that lower values of γ are associated with lower estimated elasticities. Interestingly, however, the estimated value of $\tilde{\gamma}$ is only about half as large as the true underlying value of γ. The reason for this discrepancy is the nonlinearity of the earnings function in hours. In particular, the nonlinearity of g implies that higher hours worked imply higher hourly wage rates, so that the wage $w^h(a)$ moves more over the life cycle than does the underlying exogenous productivity profile $e(a)$. If one were to run the micro labor supply regression using the exogenous productivities $e(a)$ instead of the wage profile $w^h(a)$ then the regression coefficient would be much closer to the true value of γ.

4.3.2. Aggregate Elasticities

For each of the four different calibrated economies (one for each of the four values for γ in Table 3), Rogerson and Wallenius consider what happens to the steady state hours profile if the tax rate on labor income is increased from .3 to .5, assuming that the proceeds continue to fund a uniform lump-sum transfer to all individuals subject to a balanced budget constraint at each point in time. With the given functional forms, one can show that such a tax causes a proportional shift in the hours profile, conditional on being employed. It follows that one can summarize the shift in the hours profile by simply reporting the shift in peak hours worked, which we denote by h^P. For each economy Rogerson and Wallenius compute the values of aggregate hours (H), fraction of life spent in employment (f), and peak hours worked over the life cycle (h^P), all relative to the values in the benchmark calibrated economy with $\tau = .3$. Table 4 reports their results.

Table 4

Relative Outcomes for $\tau = .5$

γ	H	f	h^P
2.00	.777	.857	.856
1.00	.784	.825	.918
0.50	.788	.808	.956
0.10	.790	.794	.991

Several features are worth noting. First, the implied change in aggregate hours worked is large in all four cases–more than 20%. Second, despite the dramatic differences in estimated micro labor supply elasticities in the four economies–a factor 25 difference between the highest and lowest–the changes in aggregate hours worked are essentially constant across the four different economies. Third, although the value of γ has virtually no effect on the change in aggregate hours worked, it has very significant effects on how the change in aggregate hours is broken down into changes in working life versus changes in hours worked while employed. In analyzing this decomposition, note that the relative change in h^P is a measure of the change in total hours due to changes in the h profile holding f constant, since as noted earlier, the hours profile shifts proportionately, and for a given f, a proportionate shift in the profile shifts aggregate hours by the same amount. However, it is not true that a shift in f leads to a proportionate shift in aggregate hours, since as f decreases the marginal employment episodes that are lost represent fewer hours of work. In any case, when $\gamma = 2.00$ the downward shift in the hours profile accounts for over 60% of the total decrease in hours, while when $\gamma = .10$ this downward shift accounts for less than 5% of the shift.

The above results indicate that in this life cycle economy with operative intensive and extensive margins, micro labor supply elasticities estimated from workers with positive hours are not particularly relevant in predicting the aggregate effects of permanent changes in taxes. The key feature of the economy that is responsible for this is the nonconvex mapping from time spent working to labor services, which in turn gives rise to the operative extensive margin in terms of life cycle labor supply. To understand this, consider an economy that is identical to the one described above except that the function $g(h)$ is now assumed to be the identity function, i.e., that $\bar{h} = 0$. Figure 2 illustrates how this will influence the findings.

In this figure, the top line shows a stylized life cycle productivity profile. The two solid lines indicate the life cycle profile for hours worked in the case of $\bar{h} = 0$ and $\bar{h} > 0$. As the picture shows, if $\bar{h} > 0$ then the model can generate outcomes in which hours worked are concentrated in the period of life in which productivity is highest. In particular, hours worked are not continuous in productivity. In contrast, if $\bar{h} = 0$, it is optimal for the individual to smooth hours worked across time, although hours of work will be higher when productivity is higher. But in this case hours vary continuously with productivity. The two dashed lines indicate the effects of higher taxes on hours of work in the two cases. If $\bar{h} > 0$, then the hours worked profile shifts down and the reservation productivity level shifts up, so that individuals spend a lower fraction of their life in employment.[47] In the case of $\bar{h} = 0$, the only effect is a downward shift in the hours profile. In both cases the extent of the downward shift of the hours profile is very strongly related to the micro labor supply elasticity. Because this downward shift is the only effect when $\bar{h} = 0$, it turns out that there is a strong relationship between micro and macro elasticities in this case, as shown in Section 2.

However, the issue is more severe than simply being that the micro elasticity only captures one piece of the aggregate adjustment in hours in the case when $\bar{h} > 0$. The results in Table 4 show that the smaller is the part that the micro elasticity captures, the larger is the part that it does not capture, i.e., the lower the value of γ, the larger is the response on the extensive margin. The important message to take away from this is that adjustment along the intensive and extensive margins are not independent of each other; changes in parameters that influence the extent of adjustment along the intensive margin will necessarily change the extent of adjustment along the extensive margin as well.[48]

[47] Here we did not present any analytic results. Rogerson and Wallenius (2007) shows analytically that an increase in the scale of the tax and transfer system leads to a reduction in lifetime labor supply along both the intensive and extensive margin.

[48] An important clarification should be noted here. Conceptually, there is no fundamental connection between an individual's willingness to substitute work along the intensive marginover the life cycle and their willingness to subsitute work along the extensive margin over the life cycle. In particular, one can choose parameters for preferences and technology such that both elasticities are small. The results in Rogerson and Wallenius reflect that fact that they impose a particular set of functional forms and some additional moment restrictions. Specifically, as they change the value of γ they also change other parameters of the model so as to continue to match a given set of moments.

Rogerson and Wallenius go on to ask what a researcher might infer if they used the benchmark model from Section 2 to interpret steady state differences in aggregate hours worked across two economies that were identical except for different scales of the tax and transfer systems. The answer is that they could infer that the value of γ is more than an order of magnitude larger than the true underlying value of γ. The reason is that whereas in the model the key role that γ plays is to determine the response along the intensive margin, the response in aggregate hours also includes a possibly large response on the extensive margin.[49] If one tries to infer γ from aggregate data, the implied value of γ must proxy for adjustment along both margins.[50]

Rogerson and Wallenius (2009) carry out their analysis in a very stylized economy. Subsequent work has examined richer versions of this framework. While the results described above come from a very simple aggregate model, Wallenius (2009) has analyzed similar issues in a more standard aggregate framework. Specifically, she studies a discrete time overlapping generations version of the standard neoclassical growth model that also allows for endogenous human capital accumulation as in Imai and Keane (2004). The above conclusions continue

[49] Kitao et al (2009) and Ljungqvist and Sargent (2010) have argued that the model of Rogerson and Wallenius contains too much responsiveness on the retirement margin and argue that one should adopt a specification in which individuals are not at an interior solution with respect to retirement.

[50] Wallenius (2011) provides another context in which this issue arises. She considers a simpler version of Imai-Keane which features fixed costs and therefore an endogenous retirement decision. She uses this framework to infer preference parameters consistent with the average life cycle profiles for wages and hours along the intensive margin. Although she obtains a substantially smaller value of γ than Imai and Keane, her model gives similar responses in aggregate hours worked due to the fact that there is an extensive margin an addition to the intensive margin. Loosely speaking, in Imai and Keane the estimated value of γ is capturing the response along both margins.

to hold in this more elaborate framework. She also considers a much richer set of policies, that include modeling specific details of how social security programs vary across countries.

Erosa et al (2011) extend the Rogerson and Wallenius model along many dimensions in order to better match a wide variety of features of male labor supply over the life cycle. For example, their analysis allows for multiple sources of heterogeneity (both idiosyncratic shocks as in Chang and Kim (2006), as well as fixed effects), explicitly considers time aggregation effects, has a serious treatment of measurement error in wages, and considers different sources of nonconvexities. While the properties of their model are broadly consistent with those in Rogerson and Wallenius, they find that the details that they introduce are important determinants of aggregate labor supply responses.[51]

4.4. Summary

The key conclusion from this section is that in models that feature adjustment along the extensive margin, the preference parameter γ need no longer play a key role in determining the response of aggregate hours of work to changes in the economic environment. In the model of Chang and Kim, all adjustment occurs along the extensive margin, and the value of the parameter γ is completely irrelevant in determining the responsiveness of aggregate hours. In the analysis of

[51] One result of interest relates to their simulation of the the tax holiday that occured in Iceland in the late 1980s. In their model the response along the extensive margin in response to this type of temporary tax change matches what was observed in Iceland during their one year tax holiday. This is in contrast to what was found in Chetty et al (2011b) based on a simulation of the much simpler model of Rogerson and Wallenius (2009).

Rogerson and Wallenius, the value of the parameter γ is important in influencing how the change in aggregate hours is broken up into changes along the intensive and extensive margins, but to first order is irrelevant in determining the response of total hours to changes in the scale of a simple tax and transfer program.

5. Adjustment Along the Extensive Margin: Evidence from Micro Data

The previous section suggests that an important source of reconciliation regarding small estimates of γ from micro data vis-a-vis the relatively large assumed aggregate labor supply elasticities in many macro models is that the seminal papers from the micro labor supply literature focus almost exclusively on adjustment along the intensive margin. If a large part of the aggregate adjustment occurs along the extensive margin, then there need not be any conflict. But this suggests that a key empirical issue is to assess the responsiveness of labor supply along the extensive margin in micro data. In this section we survey the relatively young literature that treats this issue in the context of structural models.[52]

It is important to note an important issue with this objective relative to the

[52] A recent paper by Chetty et al (2011b) surveys the literature on quasi-experimental evidence on the elasticity of the extensive margin response in different settings and compares those estimates with the implications of a parameterized version of the model in Rogerson and Wallenius (2009). Interestingly, they find that responses to permanent changes are similar whereas responses to transitory changes are much larger in the Rogerson-Wallenius model than in the quasi-experimental studies. Understanding the source of this discrepancy is an important topic for future work in this area. We already noted that the more elaborate model of Erosa et al (2011) does reconcile the observed responses in Iceland. We note that these exercises do not allow for human capital accumulation, and as pointed our earlier, this can have first order effects for how hours respond to tax changes.

previous literature that estimated the responsiveness along the intensive margin. In the simple benchmark model introduced in Section 2, the responsiveness along the intensive margin was intimately related to the preference parameter γ, thereby creating a very focused objective for researchers who wanted to estimate the determinants of labor supply responses from micro data using structural methods. Of course, in more elaborate models such as those that we discussed in Section 3, the relationship is more complicated, although it remains true that the preference parameter γ remains important. In contrast, even in very simple models, responsiveness along the extensive margin is not captured by a single preference parameter. Whether a given individual responds along the extensive margin will depend upon how close they are to some threshold that determines the point at which it becomes optimal to switch discretely from working to not working. The aggregate response is thereby intimately related to the determinants of these thresholds and the distribution of individuals around these thresholds.

This creates a much more diffuse objective for empirical researchers who want to use micro data to structurally estimate the responsiveness of a given population of individuals along the extensive margin. For example, there are many potential sources of heterogeneity and relatively little is known about how different sources of heterogeneity might matter. In the previous section it was common to assume heterogeneity in market wage rates, but one could also, for example, have heterogeneity in the fixed costs associated with market work. The results of Rogerson and Wallenius discussed previously also suggest that one cannot in general disentangle the estimation of responses along the intensive margin from the estimation of responses along the extensive margin.[53] It follows that all of the issues discussed previously about obtaining accurate estimates of the preference parameter γ are still relevant in the context of estimating responses along the extensive margin.

5.1. Early Work on Structural Models of Participation

All of the extensions to the basic life cycle model that we discussed in Section 3 continue to generate interior solutions for optimal hours, so wage changes only affect labor supply on the intensive margin. In order to study labor supply of women, for whom non-participation in the labor force is prevalent, Heckman and MaCurdy (1980, 1982) stayed within the MaCurdy (1981) framework, but adopted the alternative utility specification:

$$U_{it}(c_{it}, h_{it}) = \nu_{it}\eta^{-1}c_{it}^{1-\frac{1}{\eta}} + \alpha_{it}\gamma^{-1}(H_{\max} - h_{it})^{1-\frac{1}{\gamma}}.$$

Unlike the specification that we have used previously, this specification generates a reservation wage. The reason for this is that with this specification the marginal disutility of work is not zero at full leisure. The offer wage must exceed this value in order for an agent to choose to work. But this model maintains an

[53] The reverse is also true. Wallenius (2011) argues that incorporating an extensive margin to capture endogenous retirement into a model that features human capital accumulation as in Imai and Keane (2004) can have important implications for the value of γ that is consistent with the standard life cycle profile for hours worked. However, this need not have any implications for the overall responsiveness of labor supply. In response to a permanent change in labor taxes used to fund a lump-sum transfer, a lower value of γ does imply less response on the intensive margin, but the response along the extensive margin due to the endogenous retirement margin can largely offset this effect.

important feature of the MaCurdy (1981) framework: (notional) optimal hours are a continuous function of the offer wage. This means that, for wages slightly above the reservation wage, a worker will choose to work a small number of hours.

An implication of this model is that given a continuous distribution for underlying primitives, it predicts a continuous distribution for hours worked, and in particular that one should observe women who work a very small number of hours. Instead, the hours distribution is bimodal, with some people not working at all, while those who do generally work a fairly large number of hours (e.g., 25 to 45 per week). It is precisely this observation that motivated Cogan (1981) to consider a departure from the standard text-book model of labor supply. It is exactly this observation that motivated the specifications of Rogerson (1988) and Hansen (1985) that we studied previously.

To generate this pattern, Cogan introduced fixed costs of work into a static labor supply model. He showed that, given fixed costs, the model generates not only a reservation wage, but also "reservation hours." That is, when the offer wage passes the reservation wage, optimal wage hours jump from zero to a substantial positive value.

To be specific, consider the simple quasi-linear utility function:

$$u(c,h) = c + \alpha \frac{(\bar{H} - h)^{1-\frac{1}{\gamma}}}{1 - \frac{1}{\gamma}}$$

Letting w denote the wage rate, Y be non-labor income and F be the fixed (monetary) costs of working, utility as a function of hours worked can be written as:

$$U(h) = (wh + Y - F) + \alpha\frac{(\bar{H} - h)^{1-\frac{1}{\gamma}}}{1 - \frac{1}{\gamma}} \tag{5.1}$$

Optimal hours conditional on working are then:

$$h^{*} = \bar{H} - (\frac{w}{\alpha})^{-\gamma} \tag{5.2}$$

In the absence of fixed costs the reservation wage would be simply:

$$h^{*} > 0 \text{ if } \bar{H} - (\frac{w}{\alpha})^{-\gamma} > 0, \text{ i.e., if } w > \alpha\bar{H}^{-\frac{1}{\gamma}} \tag{5.3}$$

However, as Cogan (1981) points out, it is not appropriate to use marginal conditions to determine the participation decision rule in the presence of fixed costs. Instead, we must compare the utilities conditional on working and not working. The decision rule for working is $U(h^{*}) > U(0)$, which can be expressed as:

$$h^{*} = \bar{H} - (\frac{w}{\alpha})^{-\gamma} > \frac{F}{w} + \frac{1}{w}\frac{\alpha}{1 - \frac{1}{\eta}}[\bar{H}^{1-\frac{1}{\gamma}} - (\frac{w}{\alpha})^{\gamma-1}] = h_R > 0 \tag{5.4}$$

It is instructive to compare (5.3), which simply says a person begins to work when desired hours are positive (i.e., $\bar{H} - (\frac{w}{\alpha})^{-\gamma} > 0$), with (5.4), which says a person will begin to work only when optimal hours exceed the reservation hours level h_R. If fixed costs are substantial, then reservation hours may be substantial, and we will not observe people working a small number of hours.[54]

[54] It is notable that both costs of working (F) and tastes for work (β) enter the participation

Cogan (1981) went on to show that ignoring fixed costs could lead to severe bias in estimates of female labor supply functions. To explain why, we need to take a slight detour to discuss estimation of labor supply functions in the presence of non-participation. Given non-participation, a person's market wage rate is typically not observed. But the classic paper by Heckman (1974a) developed a method for estimating labor supply functions when wages are only observed for workers. In his framework, the labor supply equation is estimated jointly with a wage equation by maximum likelihood. The unobserved wages of the non-participants are treated as latent variables, and they are integrated out of the likelihood. To estimate labor supply behavior in the presence of fixed costs, Cogan (1981) proposed extending the Heckman (1974a) approach to estimate a three equation system, consisting of a labor supply function, an offer wage function and a reservation hours function as captured by equation (5.4).

Cogan (1981) applied his approach to data on married women aged 30 to 34 taken from the 1967 National Longitudinal Survey of Mature Women. In this sample, 898 wives worked and 939 did not. The labor supply and reservation hours functions both include the wife's education and age, number of young children, and husband's earnings. Cogan estimated that fixed costs are substantial (about 28% of average annual earnings), and that a young child raises fixed costs by about a third. Cogan's labor supply function implies a Marshallian elasticity of 0.89 at the mean of the data, and a Hicks elasticity of 0.93.

equation, while only β enters the labor supply equation. Hence, it is possible that a variable like young children could affect fixed costs of working but not tastes for work. Then, it would affect the participation decision but not labor supply conditional on participating.

However, Cogan also shows that these elasticities are rather meaningless in this context. As he notes, a 10% increase in the offer wage to the average non-working woman in the sample would not induce her to enter the labor market. But a 15% increase would induce her to jump to over $1,327$ hours. However, an additional 15% wage increase would "only" induce a further increase of 180 hours (or 13.6%). [Note: this is still a rather large increase, consistent with a Marshallian elasticity of $13.6/15 = 0.90$].

It should be noted that any generalization of the standard labor supply model which dispenses with a linear budget constraint will break the close link between preference parameters and labor supply elasticities (or responses) that characterize that model.[55] Aside from fixed costs, other leading examples of departures from linearity are welfare benefits (which play a role symmetric with fixed costs if grants are paid to unemployed workers), progressive taxation, and the tax-transfer system more generally.

Indeed, the literature on tax-transfer program effects on labor supply, which had to deal with the problem of the non-linear budget constraints that such programs create, recognized early on that, in this context, utility function parameters were no longer tightly linked with any particular elasticity concept (see, e.g., Blomquist (1983), Burtless and Hausman (1978), Hausman (1980, 1985), and Moffitt (1983).[56] Thus, labor supply could appear to be "elastic" or "inelastic,"

[55] We noted one instance of this in the previous section where the presence of a non-linear budget equation in Rogerson and Wallenius (2009) implied a substantial discrepancy between the coefficient on log wages in a standard labor supply regression and the preference parameter γ.

[56] For instance, as noted by Hausman (1980), "Structural econometric models which make labor force participation a function of ... wages, income transfer levels and the tax system can

depending on the type of budget constraint shift one considered.[57]

To illustrate, Figure 3 presents a budget constraint that is similar to the old AFDC program in the US for single mothers. It incorporates (i) a fixed welfare grant G that is taxed away at a 100% rate with earned income, and (ii) a fixed cost of working FC. The resulting constraint goes through points a, b, c, d, and e. This non-convex constraint is in contrast to the linear wage line through the $h = 0$ point. The indifference curve is drawn in such a way that utility is maximized at point a, where $h = 0$.

The figure illustrates the effect of a drop in the program tax rate from 100% to 50%. This shifts the budget line from the solid line bc to the dotted line bd. Notice that this substantial increase in after-tax wages induces no labor supply response – the person continues to locate at zero hours. Nevertheless, we can see that a small additional tax cut would cause the person to jump to full-time work. This is similar to the pattern found by Cogan (1981) when he used his model to

attempt to answer questions such as the effect of lowering the marginal tax rates on labor force particpation. The more traditional reduced form models which do not explicitly parameterize the tax system will be unable to answer such questions." Or, as noted by Blomquist (1983), "A change in the gross wage rate, nonlabor income, or parameters of the tax system changes the whole form of the budget set ... the elastcities presented above should therefore not be used to calculate [their] effects ..."

[57] A particularly striking example is Blomquist and Hansson-Busewitz (1990), who model labor supply of 602 married men (aged 25-55) in Sweden, using data from the 1980 Level of Living Survey. They model these men as making optimal hours choices subject to the progressive tax structure (in a static framework). Using their utility function estimates, they plot both the "structural" labor supply equation that would obtain if people maximized utility subject to a linear budget constraint, and the "reduced form" equation that gives desired hours as a function of wages and the existing tax structure. Strikingly, while the "structural" labor supply curve has a positive Marshallian elasticity throughout, the reduced form supply curve is backward bending for wage rates above 26 SEK per hour. This compares to an average after-tax rate of 14.83 SEK. Thus, an analysis that fails to account for progressive taxation could easily conclude labor supply is backward bending beyond a certain point, when this is only a feature induced by the tax system, not by underlying preferences.

simulate responses to wage changes.

An even more interesting point is that a small increase in the wage rate would cause the person in Figure 3 to jump from 0 to 40 hours of work per week (by slightly raising point d) even at the initial 100% tax rate. Thus, a large effect of a small wage increase is consistent with a scenario where substantial reductions in the tax rate (e.g., from 100% to 50%) have no effect whatsoever. The implication is that, with a non-linear budget constraint, wage increases and tax reductions of the same magnitude may have very different effects.

Given a budget constraint like that in Figure 3, a researcher given historical data that contained variation in tax rates over the 100% to 50% range, who estimates what Blundell and MaCurdy (1999) call a "prototype empirical specification" (see their equation 4.30), might well conclude that labor supply of program participants is highly inelastic. Historically this is roughly what happened: years of tinkering with the AFDC tax rate in attempts to create work incentives had little effect, leading to a conventional wisdom that labor supply was "inelastic" for single mothers. Thus, many observers were taken completely by surprise when a change in policy in the mid-1990s, toward 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 dramatic labor supply increases for this group.

Notably, however, Keane and Moffitt (1998) and Keane (1995), had modelled labor supply behavior of single mothers taking into account the full complexity of the AFDC budget constraint (as well as the Foodstamp program and fixed costs of work). Simulations of their model implied that AFDC tax rate reductions would have little effect, but that labor supply of single mothers would be quite sensitive to wage subsidies, EITC and fixed cost of work subsidies (or work bonuses).

5.2. Life-Cycle Models with a Participation Margin

We now turn to the topic of introducing fixed costs of work and the extensive margin into dynamic life-cycle models. Kimmel and Kniesner (1998) appear to have been the first to extend the basic MaCurdy (1981) and Heckman and MaCurdy (1980, 1982) framework to include fixed costs, though they did not structurally estimate the model's primitives. Specifically, they estimate a life-cycle labor supply equation analogous to (2.11) jointly with a participation decision rule and an offer wage function. We can write the system as:

$$\log h_{it} = f_{hi} + \gamma_I \log w_{it} + \alpha_h Z_{it} + \varepsilon_{hit} \tag{5.5}$$

$$P(h_{it} > 0) = F(f_{pi} + \tilde{\gamma}_P \log w_{it} + \alpha_p Z_{it}) \tag{5.6}$$

Here (5.5) is a Frisch labor supply function where the fixed effect f_{hi} captures the marginal utility of wealth (consumption), along with any fixed effects in tastes for work. Equation (5.6) gives the probability of participation and F is a cumulative distribution function (which Kimmel and Kniesner (1998) assume to be normal, giving a probit model for participation). The fixed effect f_{pi} in the probit model captures not just the marginal utility of wealth and tastes for work, but also individual heterogeneity in the fixed costs of work.

In this framework γ_I is the conventional Frisch elasticity of labor supply conditional on employment (i.e., the elasticity on the intensive margin). But we now introduce a Frisch participation elasticity given by:

$$\gamma_P = \frac{\partial \log P(h_{it} > 0)}{\partial \log w_{it}} = \tilde{\gamma}_P \frac{F'(\cdot)}{F(\cdot)} \tag{5.7}$$

Kimmel and Kniesner (1998) estimate this model using data on 2428 women from the Survey of Income Program Participation (SIPP), 68% of them married. The data were collected from May 1983 to April 1986, giving 9 periods of data.

The estimates imply a Frisch intensive margin elasticity of 0.66 for employed women, and a Frisch participation elasticity of 2.39. Let average hours in the population be given by $H = P\hat{h}$ where $\hat{h}$ is average hours of the employed and P is the percentage employed. Then we have:

$$\frac{\partial \log H}{\partial \log w} = \frac{\partial \log P}{\partial \log w} + \frac{\partial \log \hat{h}}{\partial \log w} = .66 + 2.39 = 3.05 \tag{5.8}$$

Kimmel and Kniesner (1998) also obtain results for men, and find $\gamma_I = 0.39$ and $\gamma_P = 0.86$ so that $\gamma_I + \gamma_P = 1.25$.

Thus, the results suggest that: (i) the participation elasticity is much larger than the hours elasticity for both women and men, and (ii) the overall elasticity is quite a bit larger for women than men. These results strongly suggest that failure to account for participation decisions may lead one to substantially underestimate the overall responsiveness of labor supply to wage changes. This is supportive of the mechanics in the model of Rogerson and Wallenius (2009) discussed in the previous section.

There is a strong conventional wisdom in the economics profession that labor supply elasticities are greater for women than for men. Granting that this is correct, an interesting question is whether it arises because (a) the extensive margin is more relevant for women (i.e., because they have a lower participation rate) or (b) because of a host of other potential explanations, involving differences in preferences and/or constraints that women face.

Interestingly, studies that estimate labor supply elasticities for employed women typically find low elasticities similar to those found for men. Consider two of the best known papers: Blundell and Walker (1986) estimated a life-cycle model of labor supply for employed married women. They obtained an (average) Frisch elasticity of only 0.033, a Hicks elasticity of only 0.009, and a Marshallian elasticity of -0.197 (at the mean of the data). More recently, Blundell, Duncan and Meghir (1998) estimated a life-cycle labor supply model for employed married women using data from UK Family Expenditure Survey 1978 to 1992. UK tax rates were reduced substantially over the period, and the basic idea of the paper was to exploit this variation to help identify labor supply elasticities. Their estimates of the compensated and uncompensated wage elasticities at the mean of the data were a modest 0.20 and 0.17, respectively. Thus, it appears that using methods that account for the participation decision are important in finding large labor supplies elasticities for women.

We next turn to the literature that has structurally estimated life-cycle models that include the participation decision. It is worth noting that the approach to estimating life cycle models developed by MaCurdy (1981), as well as the extensions to accommodate the extensive margin implemented by Heckman and MaCurdy (1980, 1982) and Kimmel and Kniesner (1998), avoid having to fully solve agents' dynamic optimization problem by dealing exclusively with (i) the first order condition for an interior solution for hours and (ii) a reservation wage condition that can be derived from the first order condition – i.e., whether the wage exceeds the MRS evaluated at zero hours. But life-cycle models that include the participation decision along with other mechanisms that extend the basic life-cycle model (e.g., human capital, credit constraints) cannot be handled so simply. Estimation of such rich models requires a full-solution structural approach.

This structural approach requires (i) solving the dynamic optimization faced by agents and (ii) finding parameter values for preferences such that the model generates behavior that is by some metric similar to the behavior observed in the data. For reviews of this literature, see, e.g., Keane and Wolpin (2009) and Rust (1996).

The first paper to adopt a full solution approach to modelling female labor supply was Eckstein and Wolpin (1989). Their model included work decisions on the extensive margin and human capital accumulation through work experience. Indeed, only the extensive margin is operative in the model (agents must choose to work either full-time or not at all), and they estimated it using married women in the NLS Mature Women's cohort. Subsequently, the female labor supply literature has extended Eckstein and Wolpin (1989) to include other important life-cycle decisions. Van der Klaauw (1996) extends Eckstein-Wolpin to make marriage a choice, while Francesconi (2002) makes fertility a choice. All three of these papers find that labor supply is highly responsive for women on the extensive margin, with uncompensated elasticities with respect to permanent wage changes in the 3 to 5 range.[58]

The most comprehensive modelling effort to date is Keane and Wolpin (2007, 2010). They extend earlier work to include a part-time work option. And marriage, fertility, school attendance and welfare participation are all included as choices.[59] The model is estimated using data from the NLSY79 cohort (women aged 14 to 21 in 1979 who attain a maximum age of 33 by 1991). In experiments where they permanently increase the offer wage by 5%, Keane and Wolpin (2010) find a wage elasticity of roughly 2.8.

It is important to note, however, that the elasticities generated by these dynamic structural models are rather different from ones we are used to seeing reported in the more conventional labor supply literature. That is, aside from labor supply, they also allow (depending on the study) some combination of experience, fertility, marriage and education to adjust to wage changes. Thus, they measure "long run" or cumulative responses.

For instance, say a tax cut causes a woman to work more in the current period. This means not only that she will have more human capital in the next period, but also that her expected number of children is reduced. Both the human capital

[58] The focus of these papers is not on labor supply elasticities per se, so they do not report elasticities directly. The figures in the text are our own calculations based on simulations reported in the papers. See Keane (2010) for more details.

[59] It is not feasible to solve such a complex model analytically. Thus, Keane and Wolpin utilize approximate solution methods developed in Keane and Wolpin (1994).

and fertility effects further enhance labor supply in the next period, and so on. This is identical to the "snowball" effect of a permanent wage or tax change that we referred to when discussing the Imai and Keane (2004) model for men. But for women additional sources of dynamics, like fertility, are likely to be important. This means that conventional labor supply studies that treat fertility as given are likely to understate long run responses to permanent wage/tax changes.

Finally, Keane and Wolpin (2010) also find that wage elasticities are inversely proportional to skill, with the highest skilled women having an elasticity of only 0.6, while lowest skilled have an elasticity of 9.2. These differences are indicative of the importance of the extensive margin. The high skilled women have a very high participation rate in the baseline, so there is little scope for them to adjust. In contrast, for the low skilled women, only about a third are working in the baseline simulation, and this increases to roughly 50% with wage increase. Clearly, a large segment of the low skilled women are close to indifferent between working and not working, and a small wage increase can shift a large number across the margin. This is consistent with our earlier discussion of labor supply effects of welfare programs, which are very relevant for the low skilled group.

There are many fewer papers in the male labor supply literature that consider the extensive margin (an exception being the Kimmel and Kniesner paper discussed earlier). This is because it has generally been viewed as a less important factor for men, because of their high participation rate. However, research suggests that the extensive margin is much more important for young males, males near retirement, and minority groups. For instance, as we noted earlier, French

(2005) finds high labor supply elasticities for older men, and attributes this to the extensive margin becoming more important as they approach retirement and the participation rate falls.

In addition, Keane and Wolpin (1997, 2000, 2001) wrote a series of papers on the career decisions of young men. Their models allow for work decisions (on the extensive margin), along with schooling and occupation choices, all of which influence the evolution of human capital. Unfortunately, as they focus on education/occupation choices, they did not simulate the labor supply responses implied by their models. The best we can do to assess this is to look at Keane and Wolpin (2000), which estimated the same model of career choice on both blacks and whites in the US. The only parameters allowed to differ between the two groups were the initial distribution of skill types at age 16, and the rental price of skill (which presumably captures labor market discrimination).[60] Keane and Wolpin (2000) report a simulation where they increase the rental price of skill for blacks up to the same level as whites. This implies roughly a 6% increase in the wage rate.[61] At age 30 it causes the percent of blacks who are employed to increase from 83.8% to 90.7%, an 8.2% increase. Thus, the implied elasticity with respect to a permanent (uncompensated) wage increase for black males at age 30 is roughly $8.2/6 = 1.4$.

[60] Keane and Wolpin (2000) could not reject the hypothesis that other parameters (e.g., tastes for work) were the same for blacks and whites. They argued that the differences in skill distributions between the two races at age 16 was likely due to differences in human capital investment at younger ages (e.g., the quality of childcare, pre-school and primary schools, the home environment, etc.).

[61] Actually, to achieve equality the rental price was increased by 8% in the white collar occupation and 5% in the blue collar occupation. As roughly twice as many of the employed blacks were in blue collar, the average rental price increase is roughly 6%.

It is also of interest to revisit Chetty's (2010) analysis of the implications of optimization frictions in the context of estimating labor supply responses along the extensive margin. Our earlier analysis of this issue assumed that workers are always at an interior solution, and so implicitly is only applicable to choice along the intensive margin. Chetty (2010) goes on to argue that adjustment costs would not lead one to understate labor supply elasticities on the extensive (participation) margin. Consider extending (3.7) to include fixed costs of work (F), non-labor income (Y) and welfare and/or unemployment benefits (B) for those who do not work:

$$U(h_t, \tau_t) = wh_t(1-\tau_t) + Y - \frac{\alpha}{1+\frac{1}{\gamma}} h_t^{1+\frac{1}{\gamma}} - F \cdot I[h_t > 0] + B \cdot I[h_t = 0] \quad (5.9)$$

Now consider a person who is indifferent between participating and not participating in the labor force. Letting h_t^* denote optimal hours conditional on working, we have that

$$U(h_t^*, \tau_t) = U(0, \tau_t) \text{ implies } wh_t^*(1-\tau_t) - \frac{\alpha}{1+\frac{1}{\gamma}} h_t^{*1+\frac{1}{\gamma}} - F = B \quad (5.10)$$

Now, consider a reduction in the tax rate, leading to an increase in $(1-\tau_t)$. The person's utility says fixed if he/she remains at $h_{t+1} = 0$. However, if he/she begins to work, the utility gain can again be decomposed into two parts: (i) the gain from beginning to work, but holding hours fixed at the old optimum of $h_{t+1} = h_t^*$, and (ii) the gain from adjusting hours to the new optimum implied by the lower tax rate. The gain in utility, and consumption, from beginning to work is simply $wh_t^*\Delta(1-\tau_t)$. This gain is a first order function of the change in the tax rate. As a fraction of $t+1$ earnings/consumption (given that $h_{t+1}=h_t^*$), it is simply equal to the percentage change in the tax rate.

Thus, Chetty argues, adjustments costs (as a percentage of earnings) would need to be as large as the percentage change in the tax rate for people near the participation margin not to adjust to a tax cut by starting to work. This in turn, means existing estimates of elasticities on the extensive margin (that ignore adjustment costs) are not likely to be seriously biased.

There is one qualification regarding this result. Chetty's analysis does not consider welfare/unemployment benefits B or non-labor income Y (he assumes consumption in the non-working state is zero). Given the existence of non-work benefits, the consumption gain from beginning to work (at hours level $h_{t+1}=h_t^*$), expressed as a fraction of time t consumption, is $wh_t^*\Delta(1-\tau_t)/(Y+B)$. Even though this gain is first order in taxes, it can be arbitrarily small, depending on how large B and Y are (or how small w is). Implicitly, Chetty is allowing the individual to make suboptimal choices that are small when measured by consumption loss as a percent of potential labor earnings. But it may be more natural to ask if the consumption loss is small as a percent of total consumption (i.e., potential labor earnings plus non-labor income).

Whether this issue matters depends on the empirical relevance of the situation in which an individual has consumption that is substantially larger than (potential) labor earnings. For most individuals this is probably not the case.

To the extent that labor earnings are the dominant source of income for most individuals, even a hand-to-mouth consumer would have consumption no larger than labor earnings. For the truly wealthy, say those in the top 1% of the wealth distribution, non-labor income may be larger relative to labor income and as a results consumption may be large relative to labor income. But these individuals are presumably of little relevance for estimating participation rate elasticities in most studies.

There are, however, two cases of possible interest where consumption may be relatively large compared to potential labor earnings. The first is the situation of multi-member households where the second earner has a much lower wage rate than the primary earner. In evaluating the labor supply decision of the second earner it is possible that the bias due to adjustment frictions might become relevant. The second case is that of a single mother who is eligible for benefits that are large relative to potential labor earnings. If the benefits are only received in the event that the individual does not work, then the issue is not relevant, in the sense that for such an individual there is no incentive to work and hence the labor supply decision is not really relevant. But, if the individual is eligible for such things as food stamps, housing subsidies and/or medicaid even when working, then it is certainly possible that consumption is much larger than labor earnings. Hence, it is possible that biases associated with optimization frictions continue to be relevant for low income single mothers.

5.3. Summary

The literature on estimating extensive margin elasticities in dynamic structural models is relatively young. However, based on the existing studies, there appears to be a very consistent pattern of high estimated labor supply elasticities for women at the extensive margin, as well as for males who have relatively low participation rates (i.e., the young, the old and minorities).

6. Conclusion

Based on the last major survey of the micro labor supply literature by Blundell and MaCurdy (1999), it is fair to say that the consensus view among labor economists was (and still is) that labor supply elasticities are small. In contrast, macroeconomists generally work with equilibrium models in which Hicks (or compensated) and Frisch (or inter-temporal) labor supply elasticities are quite large (i.e., in the range of 1 to 2). In this survey we have described a relatively new literature – which, with a few notable exceptions, has emerged since Blundell and MaCurdy (1999) – that seeks to reconcile these conflicting micro and macro views on labor supply.

This literature can be viewed as consisting of two branches. The first branch focuses on the micro perspective. In the basic life-cycle labor supply model (i.e., MaCurdy (1981)) the only source of dynamics is borrowing/saving. A number of authors have considered extensions of this model to include other potentially importance sources of dynamics, such as human capital accumulation, borrowing constraints, precautionary saving (given future wage uncertainty), decisions on the extensive margin and/or labor supply adjustment costs. This work has shown that if the true model (or data generating process) contains such mechanisms, but the data is viewed through the lens of the basic model, then estimates of labor supply elasticities will tend to seriously understate their true values.

The second branch focuses on the macro perspective. This literature emphasizes issues associated with aggregation in the presence of the extensive margin and worker heterogeneity. This literature has shown that small (intensive margin) elasticities at the individual level can be consistent with large elasticities at the aggregate level. In some cases, the value of the preference parameter γ, which was the focus of much of the early literature, is virtually irrelevant for the response of aggregate hours to specific changes in the economic environment.

Both of these literatures share one key point in common, however. In the basic life cycle model of MaCurdy (1981) there is a direct link between parameters of individual level preferences and the Hicks and Frisch elasticities at the aggregate level. All the extensions to the basic model that we have described break that direct link. This is not to say that individual preference parameters no longer matter. But, in general, labor supply elasticities are not only a function of preference parameters but also of all other aspects of the economic environment as well: This includes the wage process, the functioning of credit markets, the technology of job search/hours adjustment, the production technology (in particular how productivity varies with hours), and so on.

In this complicated world, estimation of individual preferences alone is not adequate to model labor supply. Predicting the effects of changes in wages and/or taxes and transfers will, in general, require structural modelling of the complete economic environment. Given the difficulty of such exercises, it may be tempting to resort to an "experimental" approach of just cataloguing responses to observed tax changes. But in our view this would be misguided. As we have shown, even in simple models, changes in after-tax wages can have very different effects on labor supply, depending on the source of the change and/or slight differences in its magnitude. Thus, it is very difficult to generalize from historical episodes to predict how people would respond to a new policy change. The failure of most of the profession to predict the consequences of the U.S. welfare reform of the mid 90s is an excellent example of this problem. An even more basic point is that, even if we could predict labor supply responses to hypothetical changes in public policy simply by extrapolation from historical episodes, we cannot evaluate the welfare consequences of policies without a model of the economic structure.

In our view, the literature we have described can credibly support a view that compensated and inter-temporal elasticities at the macro level fall in the range of 1 to 2 that is typically assumed in macro general equilibrium models. Indeed, the problem that confronts us now is that the reconciliation is, in a sense, too easy. That is, we have described multiple mechanisms that can achieve the desired reconciliation. Of these, which are actually the most relevant? In our view, answering this question will require building models with multiple mechanisms, and seeing how well they explain multiple aspects of behavior – not just labor supply, but also schooling, occupational choice, savings, etc.. (The work by Keane and Wolpin (2001, 2010) is an example of this type of strategy). Obviously this is a large (and daunting) program for future research. But it is important to realize that simply being able to reconcile aggregate labor supply responses with observations from micro data is not in itself sufficient. As we have described, the specific mechanism(s) used to achieve the reconciliation will lead to different implications regarding welfare effects of policies, even if those policies generate similar labor supply responses.

Finally, we offer some conjectures on how such an ambitious research program might proceed. As we have seen, models with human capital and/or the extensive margin can generate large labor supply elasticities. In our view it would be hard to argue that work experience does not augment wages, or that the production technology along with fixed costs of work does not constrain workers' choices of working hours. The empirical evidence that experience augments wages and that workers rarely choose to work a small number of hours is quite convincing. Thus, we strongly suspect that human capital and the extensive margin will be key components of future labor supply models.

On the other hand, the importance of other mechanisms we have discussed – i.e., liquidity constraints, precautionary saving, fixed costs of adjusting hours – seems more speculative. As we discussed, the evidence on liquidity constraints and the strength of the precautionary motive is controversial. This is largely because both these mechanisms lead to behavior that looks similar, as well as being similar to the behavior generated by other mechanisms like complementarity between hours and consumption, age varying tastes, etc.. And the obvious problem with the costs of adjustment mechanism is that – unlike say, the extent to which wages rise with work experience – it is hard to know what plausible values for costs of adjustment are.

This is not to say we dismiss the importance of these other mechanisms. Rather, our point is that they are relatively subtle (i.e., hard to identify), and we strongly suspect it will not be possible to credibly pin down their importance using data on wages and labor supply alone. This brings us back to the program we advocated earlier of building models with multiple mechanisms, and seeing how well they explain multiple aspects of behavior. While models with human capital and/or the extensive margin can generate large labor supply elasticities, we may need mechanisms like liquidity constraints or adjustments costs to explain more subtle aspects of savings/consumption behavior, occupational choice, fertility, etc.

References

[1] Aiyagari, R. (1994). "Uninsured Idiosyncratic Risk and Aggregate Saving." *Quarterly Journal of Economics* 109, 659-684.

[2] Alonso-Ortiz, J., and Rogerson, R. (2010). "Taxes, Transfers and Employment in an Incomplete Markets Model." *Journal of Monetary Economics* 57, 949-958.

[3] An, S., Chang, Y., and Kim, S.. (2009). "Can a Representative Agent Model Represent a Heterogeneous Agent Economy?," *American Economic Journal: Macroeconomics* 1, 29-54.

[4] Altonji, Joseph. (1986). "Intertemporal Substitution in Labor Supply: Evidence from Micro Data." *Journal of Political Economy* 94, S176-S215.

[5] Attanasio, O. and Weber, G. (1995). "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* 103, 1121-1157.

[6] Auten, G. and Carroll, R., (1999). "The Effect of Income Taxes on Household Behavior." *Review of Economics and Statistics* 81, 681-693.

[7] Bencivenga, V. R.. (1992). "An Econometric Study of Hours and Output Variation with Preference Shocks." *International Economic Review* 33, 449-71.

[8] Benhabib, J., Rogerson, R., and Wright, R. (1991). "Homework in Macroeconomics: Household Production and Aggregate Fluctuations," *Journal of Political Economy* 99, 1166-1187.

[9] Blomquist, Soren (1983). "The Effect of Income Taxation on the Labour Supply of Married Men in Sweden." *Journal of Public Economics* 22, 169-197.

[10] Blomquist, Soren, and Hansson-Busewitz, Urban. (1990). "The Effect of Taxes on Male and Female Labor Supply in Sweden." *Journal of Human Resources* 25, 317-57.

[11] Blundell, Richard and MaCurdy, Thomas H. (1999). "Labor Supply: A Review of Alternative Approaches." in O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, Vol. 3, North-Holland, Amsterdam, pp. 1559-1695.

[12] Blundell, Richard W., and Walker, Ian. (1986). "A Life Cycle Consistent Empirical Model of Family Labour Supply Using cross-Sectional Data." *Review of Economic Studies* 53, 539–58.

[13] Blundell R.W., Duncan, A.,and Meghir, C. (1998). "Estimating Labour Supply Responses Using Tax Policy Reforms." *Econometrica* 66, 827-861.

[14] Browning, M., Deaton A. and Irish, M. (1985). "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle." *Econometrica* 53, 503-44.

[15] Browning, M., Hansen, L., and Heckman, J. (1999). "Micro Data Analysis and General Equilibrium Models," in John B. Taylor and Michael Woodford (eds.), Handbook of Macroeconomics, Vol. 1, North-Holland, Amsterdam, pp. 543-633.

[16] Burtless, Gary and Jerry Hausman (1978), "The Effects of Taxation on Labor Supply: Evaluating the Gary Income Maintenance Experiment." *Journal of Political Economy* 86, 1103-1130.

[17] Chang, Y., and Kim S., (2006). "From Individual to Aggregate Labor Supply: A Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy." *International Economic Review* 47, 1-27.

[18] ____________________. (2007). "Heterogeneity and Aggregation in the Labor Market: Implications for Aggregate Preference Shifts." *American Economic Review* 97, 1939-1956.

[19] Chang, Yongsung, Kim, Sun-Bin, Kwan, Kyooho, and Rogerson, Richard. (2011). "Interpreting Labor Supply Regressions in a Model of Full- and Part-time Work." *American Economic Review* 101, 476-481.

[20] Chari, V.V., Kehoe P. J. and McGrattan E. R. (2004). "Business Cycle Accounting." *NBER Working Paper* #10351.

[21] Chetty, Raj. (2010). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." mimeo, Harvard University.

[22] Chetty, Raj, Friedman, John, Olsen, Tore, and Pistaferri, Luigi. (2011a). "Adjustment Costs, Firm Responses and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics* 126.

[23] Chetty, Raj, Guren, Adam, Manoli, Day, and Weber, Andrea. (2011b). "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticties? A Meta-Analysis of Extensive Margin Elasticities." mimeo, Harvard University.

[24] Cho, Jang-Ok. (1995). "Ex-post Heterogeneity and the Business Cycle." *Journal of Economic Dynamics and Control* 19, 533-551.

[25] Cogan, J. (1981). "Fixed Costs and Labor Supply." *Econometrica* 49, 945-964.

[26] Cole, H. and Ohanian, L. E. (2002). "The Great U.S. and U.K. Depressions through the Lens of Neoclassical Theory." *American Economic Review* 92, 28-32.

[27] Deaton, A. (1991). "Saving and Liquidity Constraints." *Econometrica* 59, 1221-1248.

[28] Diaz-Gimenez, J., Quadrini, V., and Rios-Rull, V. (1997). "Dimensions of Inequality: Facts on the US Distributions of Earnings, Income and Wealth." *Quarterly Review of the Federal Reserve Bank of Minneapolis*, 3-21.

[29] Domeij, D., and Floden, M. (2006). "The Labor Supply Elasticity and Borrowing Constraints: Why Estimates Are Biased," *Review of Economic Dynamics* 9, 242-262.

[30] Eckstein, Z., and Wolpin, K.I. (1989). "Dynamic Labor Force Participation of Married Women and Endogenous Wage Growth." *Review of Economic Studies* 56, 375-90.

[31] Erosa, A., Fuster, L., and Kambourov, G. (2011). "Towards a Micro-Founded Theory of Aggregate Labor Supply." mimeo, University of Toronto.

[32] Feldstein, Martin. (1995). "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy* 103, 551-557.

[33] Floden, M., and Linde, J. (2001). "Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance?" *Review of Economic Dynamics* 4, 406-437.

[34] Francesconi, M. (1998). A Joint Dynamic Model of Fertility and Work of Married Women." *Journal of Labor Economics* 20, 336-380.

[35] French, E. (2005). "The Effects of Health, Wealth and Wages on Labour Supply and Retirement Behaviour." *Review of Economic Studies* 72, 395-427.

[36] Gali, J., Gertler M. and Lopez-Salido, J. D. (2002). "Mark-ups, Gaps, and the Welfare Cost of Business Fluctuations." *NBER Working Paper* #8850.

[37] Gruber, J., and Saez, E.. (2002). "The Elasticity of Taxable Income: Evidence and Implications." *Journal of Public Economics* 84, 1-32.

[38] Hall R. E. (1997). "Macroeconomic Fluctuations and the Allocation of Time." *Journal of Labor Economics* 15, 223-250.

[39] Hansen, G. (1985). "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics* 16, 309-337.

[40] Hausman, Jerry. (1980). "The Effect of Wages, Taxes and Fixed Costs on Women's Labor Force Participation." *Journal of Public Economics* 14, 161-194.

[41] _____________. (1985b). "Taxes and Labour Supply." in K. Arrow and M. Intriligator (eds.), The Handbook of Public Economics, Vol. 1, North-Holland, Amsterdam, 213-264.

[42] Heckman, James. (1974a). "Shadow Prices, Market Wages and Labor Supply." *Econometrica* 42, 679-694.

[43] ______________. (1974b). "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle." *American Economic Review* 64, 188-194.

[44] ______________. (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 (New York: Columbia University Press). 227-264.

[45] _______________. (1984). "Comments on the Ashenfelter and Kydland Papers." *Cargnegie-Rochester Conference Series on Public Policy* 21, 209-224.

[46] Heckman, James and MaCurdy, Thomas. (1980). "A Life-cycle Model of Female Labor Supply." *Review of Economic Studies* 47, 47-74.

[47] ____________________________________. (1982). "Corrigendum on a Life Cycle Model of Female Labor Supply." *Review of Economic Studies* 49, 659-660.

[48] Hubbard, R., Skinner, J., and Zeldes,S. (1995). "Precautionary Saving and Social Insurance." *Journal of Political Economy* 103, 360-399.

[49] Huggett, M.. (1993). "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies." *Journal of Economic Dynamics and Control* 17, 953-969.

[50] Imai, Susumu and Keane, Michael. (2004). "Intertemporal Labor Supply and Human Capital Accumulation." *International Economic Review* 45, 601-642.

[51] Keane, Michael. (1995). "A New Idea for Welfare Reform." *Federal Reserve Bank of Minneapolis Quarterly Review* 19, 2-28.

[52] _____________. (2009). "Income Taxation in a Life Cycle Model with Human Capital. " Working paper, University of New South Wales.

[53] _____________. (2010). "Labor Supply and Taxes." Working paper, University of New South Wales.

[54] Keane, Michael and Runkle, David. (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* 10, 1-9.

[55] Keane, Michael and Moffitt, Robert. (1998). "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review*, 39, 553-589.

[56] Keane, Michael, and Wolpin, Kenneth. (1994). "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *Review of Economics and Statistics* 76, 648-72.

[57] __________________________________. (1997). "The Career Decisions of Young Men." *Journal of Political Economy* 105, 473-522.

[58] __________________________________. (2001). "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." I*nternational Economic Review* 42, 1051-1103.

[59] __________________________________. (2007): "Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation." *International Economic Review* 48, 1351-1378.

[60] ________________________________. (2009). "Applications of Discrete Dynamic Programming Models. Plenary Address to Society of Economic Dynamics." *Review of Economic Dynamics* 12, 1-22.

[61] ________________________________. (2010). "The Role of Labor and Marriage Markets, Preference Heterogeneity, and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women." *International Economic Review*, 51,851-892.

[62] Killingsworth, Mark and Heckman, James. (1986). "Female Labor Supply: A Survey." in O. Ashenfelter and R. Layard (eds.), Handbook of Labor Economics, Vol. 1, North-Holland, Amsterdam, pp. 103-204.

[63] Kimmel J. and Kniesner T.J. (1998). "New Evidence on Labor Supply: Employment vs. Hours Elasticities by Sex and Marital Status." *Journal of Monetary Economics* 42, 289-301.

[64] Kitao, Sagiri, Ljungqvist, Lars, and Sargent, Thomas. (2011). "A Life Cycle Model of Transatlantic Employment Experiences." mimeo, New York University.

[65] Krusell, P., Mukoyama, T., Rogerson, R., and Sahin, A. (2008). "Aggregate Implications of Indivisible Labor, Incomplete Markets and Labor Market Frictions." *Journal of Monetary Economics* 55, 961-979.

[66] ____.
(2010). "Aggregate Labor Market Outcomes: The Roles of Choice and Chance." *Quantitative Economics* 1, 97-128.

[67] ____.
(2011). "A Three State Model of Worker Flows in General Equilibrium." *Journal of Economic Theory* 146, 1107-1133.

[68] Kydland, F., and Prescott, E. (1982). "Time to Build and Aggregate Fluctuations." *Econometrica* 50, 1345-1370.

[69] Ljungqvist, L., and Sargent, T.. (2006). "Do Taxes Explain European Employment? Indivisible Labor, Human Capital, Lotteries and Savings." *NBER Macroeconomics Annual* 21, 181-246.

[70] _________________________. (2008). "Taxes, Benefits, and Careers: Complete Versus Incomplete Markets." *Journal of Monetary Economics* 55, 98-125.

[71] _________________________. (2010). "Curvature of Earnings Profile and Career Length." mimeo, New York University.

[72] Low, Hamish. (2005). "Self-insurance in a Life Cycle Model of Labour Supply and Savings." *Review of Economic Dynamics* 8, 945-975.

[73] Lucas, Robert, and Rapping, Leonard. (1969). "Real Wages, Employment and Inflation." *Journal of Political Economy* 77, 721-754.

[74] MaCurdy, Thomas. (1981). "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy* 89, 1059-1085.

[75] ________________. (1983). "A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty." *International Economic Review* 24, 265-289.

[76] ________________. (1985). "Interpreting Empirical Models of Labor Supply in an Intertemporal Framework with Uncertainty." In J. J. Heckman and B. Singer ed., Longitudinal Analysis of Labor Market Data, Cambridge University Press: Cambridge.

[77] Mankiw, G., Rotemberg, J., and Summers, L. "Intertemporal Substitution in Macroeconomics." *Quarterly Journal of Economics* 100, 225-251.

[78] McDaniel, Cara. (2006). "Effective Tax Rates for 15 OECD Countries: 1950-2003." mimeo, Arizona State University.

[79] ______________ . (2011). "Hours of Work in the OECD 1960-2003: Driving Forces and Propagation Mechanisms." Forthcoming, American Economic Journal: Macroeconomics.

[80] Meghir, Costas, and Phillips, David. (2008). "Labour Supply and Taxes." in Report of a Commission on Reforming the Tax System for the 21st Century, James Mirrlees chair.

[81] Mendoza, E., Razin A., and Tesar L.. (1994). "Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption." *Journal of Monetary Economics* 34, 297-323.

[82] Mulligan C. B. (2001). "Aggregate Implications of Indivisble Labor." *The B.E. Journal: Advances in Macroeconomics* 1.

[83] _____________. (2002). "A Century of Labor-Leisure Distortions." *NBER Working Paper 8774.*

[84] Moffitt, Robert. (1983). "An Economic Model of Welfare Stigma." *American Economic Review* 73, 1023-1035.

[85] Ohanian, L., Raffo, A., and Rogerson, R.. (2008). "Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956-2004." *Journal of Monetary Economics* 55, 1353-1362.

[86] Parkin M.. (1988). "A Method for Determining Whether Parameters in Aggregative Models are Structural." *Carnegie-Rochester Conference Series on Public Policy* 29, 215-52.

[87] Pencavel, John. (1986). "Labor Supply of Men: A Survey." in O. Ashenfelter and R. Layard (eds.), Handbook of Labor Economics, Vol. 1, North-Holland, Amsterdam, pp. 3-102.

[88] Prescott, Edward C. (2004). "Why Do Americans Work So Much More than Europeans?" *Federal Reserve Bank of Minneapolis Quarterly Review* 28, No. 1, 2-13.

[89] _________________. (2006). "Nobel Lecture: The Transformation of Macroeconomic Policy and Research." *Journal of Political Economy* 114, 203-235.

[90] Prescott, E., Rogerson, R.,and Wallenius, J. (2009). "Lifetime Aggregate Labor Supply with Endogenous Workweek Length." *Review of Economic Dynamics* 12, 23-36.

[91] Rogerson, Richard. (1988). "Indivisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics* 21, 3-16.

[92] _________________. (2008). "Structural Transformation and the Deterioration of European Labor Market Outcomes." *Journal of Political Economy* 116, 235-259.

[93] _________________. (2011). "Individual and Aggregate Labor Supply in a Model With Coordinated Working Times." Forthcoming, *Journal of Money, Credit and Banking.*

[94] Rogerson, R., and Wallenius, J. (2007). "Micro and Macro Elasticities in a Life Cycle Model with Taxes." NBER Working Paper #13017.

[95] _____________________________. (2009). "Micro and Macro Elasticities in a Life Cycle Model with Taxes." *Journal of Economic Theory* 144, 2277-2292.

[96] Rust, John. (1996). "Numerical Dynamic Programming in Economics." Chapter 14 in H.M. Amman, D.A. Kendrick and J. Rust (eds.), Handbook of Computational Economics. North Holland: Amsterdam, pp. 619-729.

[97] Saez, Emmanuel. (2004). "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications." *Tax Policy and the Economy* 18, 117-184.

[98] Saez, E., Slemrod J., and Giertz, S. (2009). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." Mimeo, University of California Berkeley.

[99] Shaw, Katherine. (1989). "Life-cycle Labor Supply with Human Capital Accumulation." *International Economic Review* 30, 431-56.

[100] Slemrod, Joel B. and Kopczuk, Wojciech. (2000). "The Optimal Elasticity of Taxable Income." NBER Working Paper #7922.

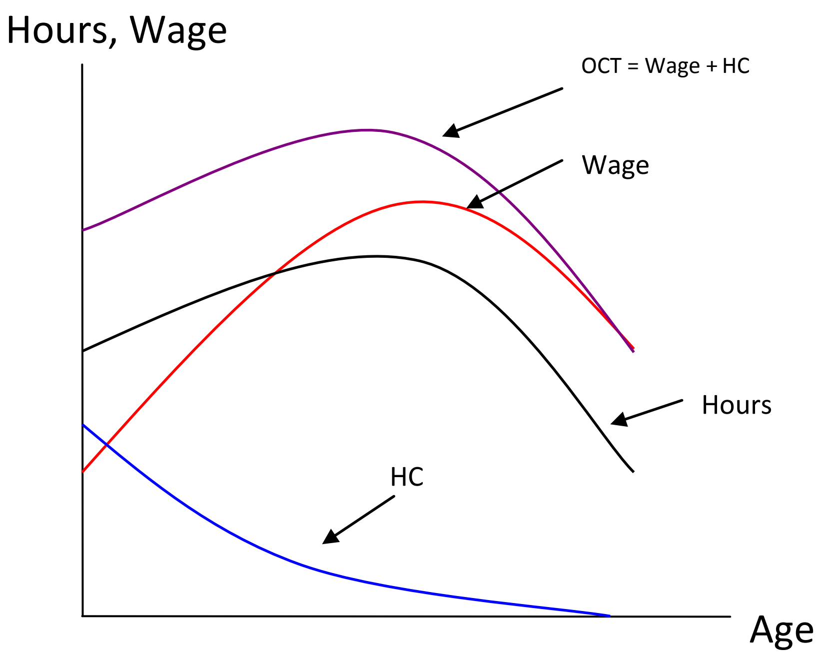
[101] Van der Klaauw, W. (1996). "Female Labour Supply and Marital Status Decisions: A Life Cycle Model." *Review of Economic Studies* 63, 199-235.

[102] Wallenius, J. (2009). "Life Cycle Labor Supply and Retirement in a Model with Endogenous Human Capital Accumulation." mimeo, Arizona State University.

[103] ___________. (2011). "Human Capital and the Intertemporal Elasticity of Substitution of Labor." Forthcoming, *Review of Economic Dynamics*.

[104] Zeldes, S. (1989). "Consumption and Liquidity Constraints: An Empirical Investigation." *Journal of Political Economy* 97, 305-346.

Figure 1: Hours, Wages and Price of Time over the Life-Cycle



Note: 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.

Figure 2: Effects of Taxes in the Rogerson-Wallenius Model

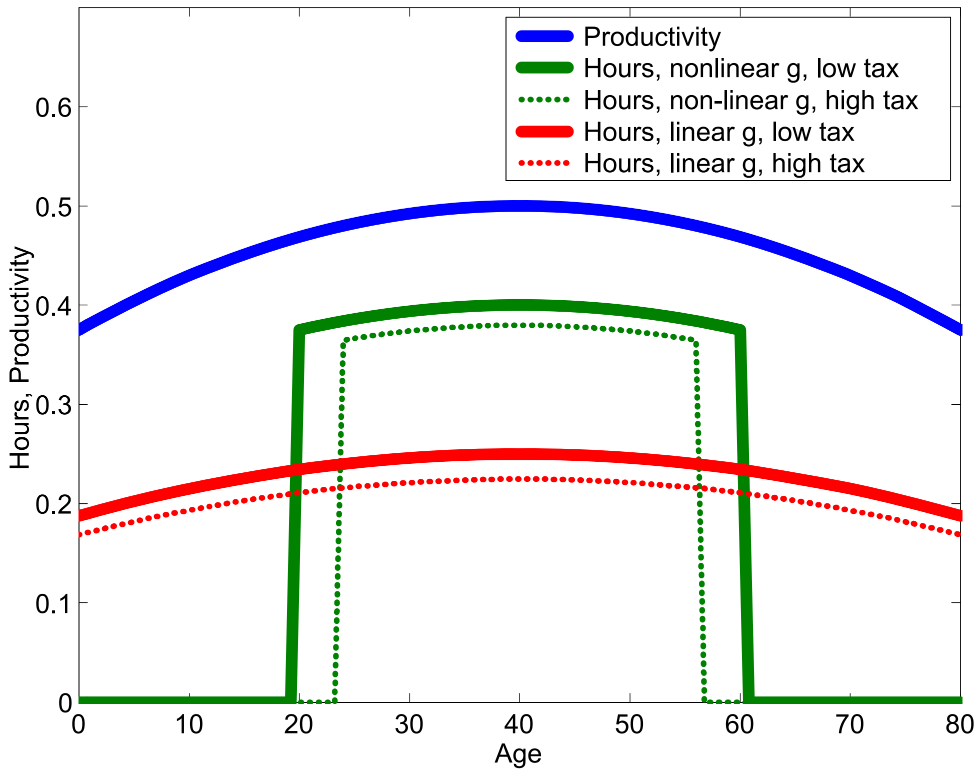
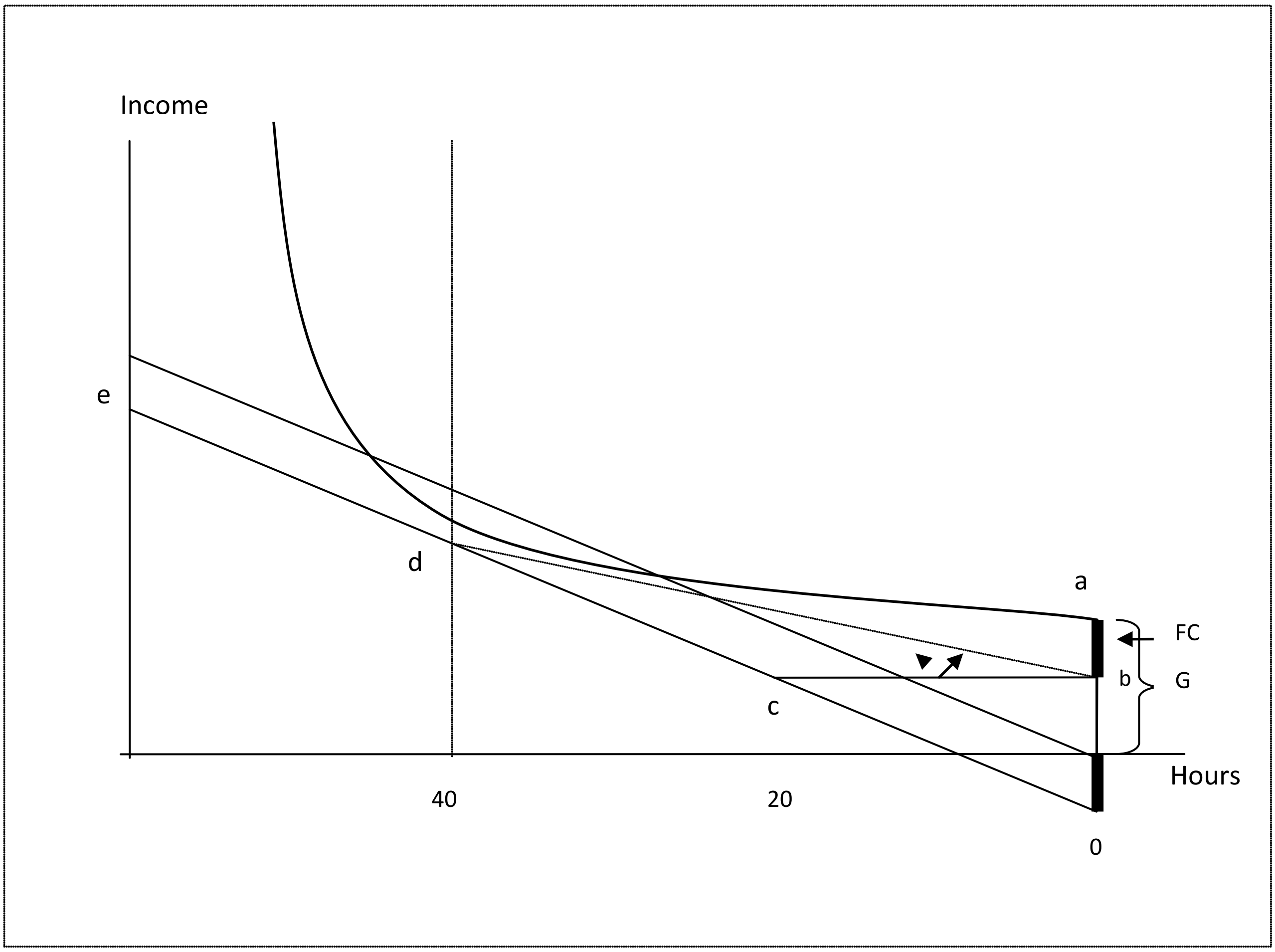


Figure 3: Labor Supply under an AFDC type Budget Constraint



Note: The wage line drawn through the zero hours point is not the relevant budget constraint, due to the AFDC grant (G), the fixed cost of working (FC) and the AFDC tax on earnings, which render the actual constraint non-convex. The actual budget constraint goes through a, b, c, d, e. The dotted line shows the shift in the budget constraint when the AFDC tax on earnings is reduced to 50%.