Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective

Michael Keane[1] and Richard Rogerson[2]

[1]Department of Economics, University of Oxford, Oxford OX1 3UQ, United Kingdom; email: michael.keane@nuffield.ox.ac.uk

[2]Woodrow Wilson School, Princeton University, Princeton, New Jersey 08544; email: rdr@princeton.edu

Annu. Rev. Econ. 2015. 7:89–117

First published online as a Review in Advance on February 26, 2015

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi: 10.1146/annurev-economics-080614-115601

JEL codes: D91, E24, J22

Keywords

life cycle, taxes, employment, human capital

Abstract

The response of aggregate labor supply to various changes in the economic environment is central to many economic issues, including the optimal design of tax policies. Whereas the earlier literature often concluded that aggregate labor supply elasticities were small, we argue that recent work using structural models and micro data credibly supports large aggregate elasticities. We focus on three issues. First, earlier analyses abstracted from several key features, including human capital accumulation, that severely negatively biased estimates of a key preference parameter. 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 preference parameters for aggregate labor supply. Third, structural estimation of models that feature choice along the extensive margin using micro data typically finds large responses.

1. INTRODUCTION

This review deals with an issue that is extremely important for a wide range of economic issues—the magnitude of the elasticity of aggregate labor supply with respect to transitory and permanent changes in wages. As is well known, this issue is highly controversial: There is a long-standing controversy driven by the fact that labor economists typically estimate relatively small labor supply elasticities from micro data, whereas macroeconomists who use representative agent models to study aggregate outcomes typically employ parameterizations that imply large aggregate labor supply elasticities.

In this article, we seek to reconcile this apparent controversy. A key point we wish to stress is that, in general, labor supply elasticities are neither a single number nor a primitive feature of preferences. Rather, labor supply responses (individual or aggregate) to a particular change in the economic environment will typically depend on features of technology and market structure, as well as preferences. And they will typically be heterogeneous, differing by worker characteristics such as age, gender, and skill level. In a dynamic setting, labor supply responses will generally change over time, as long- and short-run effects will differ. For these and other reasons, one cannot estimate a labor supply elasticity in one context and import it into other contexts.

A key implication is that it is important to adopt a framework in which the choice problems of individuals are explicitly formulated. We should seek to identify the underlying structural parameters that characterize these choice problems and use that information to infer elasticities, rather than try to explicitly estimate a number called the "labor supply elasticity" that is then applied across different contexts.[1]

In Section 2, we lay out a benchmark model, based on MaCurdy (1981), that represents a standard framework used by many economists to think about labor supply. An important feature of this simple model is that it generates a direct mapping from two preference parameters to standard elasticity concepts (Frisch, Hicks, and Marshall).

In Sections 3–5, we describe several extensions to the benchmark model that have been pursued over the past 15 years. All of these extensions break the simple link between preference parameters and labor supply elasticities, creating a situation in which elasticities are context dependent. The micro-macro controversy seems much less apparent (if not nonexistent) if one uses these extended models to view the data. Our conclusion is that much of the micro-macro controversy results from the fact that many economists continue to view the world through the lens of this benchmark model, which abstracts from these empirically important extensions.

2. BACKGROUND AND OVERVIEW

2.1. A Benchmark Model

We begin with a benchmark life-cycle model that serves to clarify the macro-micro labor supply controversy. Each period, a T-period lived individual is born with preferences

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

where c_a and h_a are consumption and hours worked at age a, respectively. There are four preference parameters: β, η, α, and γ. We strive to use these parameters consistently throughout

[1]Browning et al. (1999) also critique the practice of exporting parameter estimates across contexts.

the article. The individual has one unit of time each period and faces an exogenous productivity sequence, denoted by e_a, so that working h_a units of time at age a yields $e_a h_a$ units of labor services. There is a constant returns to scale aggregate production function $F(K_t, H_t)$, where K_t and H_t are aggregate capital and units of labor services, respectively. In steady state, these satisfy

$$H = \sum_{a=0}^{T} e_a h_a, K = \sum_{a=0}^{T} k_a,$$

where h_a and k_a are the steady-state life-cycle profiles for hours worked and capital holdings, respectively. Output can be used as consumption or investment, and capital depreciates at rate δ. We consider the following tax and transfer system: Labor earnings are taxed at the constant rate τ, and the resulting revenues fund a lump-sum transfer. To avoid issues of intergenerational redistribution, one assumes that the lump-sum transfer received by any generation is equal in present value to their tax payments.[2]

With infinitely lived agents, the steady-state interest rate is unaffected by this policy and equals $1/\beta - 1$. This need not hold in an overlapping generations economy. But because our interest is in the effects of taxes controlling for changes in other factors, such as interest rates, we assume that the steady-state interest rate is not affected by τ and equals $1/\beta - 1$. We note that there is always a government debt policy that would support this interest rate as a steady-state equilibrium. Constant returns to scale of F then imply that the wage per unit of labor services, denoted by w, is also independent of τ.

An individual thus solves the following problem in steady-state equilibrium:

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

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

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

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

$$\alpha h_a^{\frac{1}{\gamma}} = (1-\tau) \lambda e_a w. \tag{2}$$

Equation 1 implies that c_a is constant over the life cycle.[3] Taking logs of Equation 2 gives a simple version of the equation used by MaCurdy (1981) and others in their estimation exercises using micro data:

[2]Although highly stylized, the effects of this policy correspond to the Hicks elasticity, thereby providing a clean connection to the literature on elasticities.

[3]This implication is not consistent with micro data. Assuming age effects on the marginal utility of consumption, for example, would avoid this implication. Because our focus is on labor supply, we abstract from these possibilities.

$$\log h_a = b + \gamma \log e_a, \tag{3}$$

where $b = \gamma[\log \lambda + \log w + \log(1 - \tau) - \log \alpha]$ is constant for an individual over his or her life cycle in steady state. Because changes in log e_a are equivalent to changes in log wages for individuals over the life cycle, this equation provides a strategy for uncovering the preference parameter γ using individual panel data. As described below, one can also uncover the value of η.

Estimates of these preference parameters from micro data also allow one to infer aggregate effects of changes in τ. Equation 3 is a useful starting point but is not sufficient. The reason is that if we are comparing h_a across steady-state equilibria that correspond to different values of τ, then the value of λ will also differ. Hence, to determine the change in h_a, we need to also derive an expression for the change in λ.

To do this, note from Equation 2 that given the optimal value of h_0, the rest of the profile satisfies

$$h_a = \left[\frac{e_a}{e_0}\right]^{\gamma} h_0. \tag{4}$$

Total labor income is therefore proportional to h_0. Because the present value of the transfer received by each individual is equal to the present value of his or her own tax payments, in steady-state equilibrium, we have

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

As c_a is constant over the life cycle, its value is proportional to h_0 and w. Write this as $c_a = \bar{c} w h_0$. Equation 1 then implies

$$\log \lambda = -\frac{1}{\eta}[\log h_0 + \log w + \log \bar{c}]. \tag{5}$$

Using Equation 4, we have

$$\log \lambda = -\frac{1}{\eta}\left[\gamma \log \frac{e_o}{e_a} + \log h_a + \log w + \log \bar{c}\right]. \tag{6}$$

Given that w is independent of τ, Equation 3 implies

$$\log h_a = -\frac{\gamma}{\eta}\left[\gamma \log \frac{e_o}{e_a} + \log h_a + \log \bar{c} + \log w\right] + \gamma \log w + \gamma \log(1 - \tau) - \gamma \log \alpha + \gamma \log e_a. \tag{7}$$

Rearranging gives

$$\log h_a = \frac{\gamma}{\eta + \gamma}[(\eta - 1)\log w - \eta \log \alpha - \log \bar{c} - \gamma \log e_0] + \frac{\gamma\eta}{\eta + \gamma}\log(1 - \tau) + \gamma \log e_a. \tag{8}$$

Here, the coefficient on log e_a is the Frisch elasticity. This is the effect of life-cycle variation in wages. The coefficient on $\log(1 - \tau)$ is 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 Hicks elasticity is smaller than the Frisch, with equality as $\eta \to \infty$ (i.e., when utility is linear in consumption and there are no income effects).

Equation 8 implies that a change in τ causes hours to change proportionally at all ages. Because H is simply the sum of h_a, it follows that

$$\log H = B + \frac{\gamma\eta}{\eta+\gamma}\log(1-\tau), \tag{9}$$

where B is a constant. Macroeconomists often impose $\eta = 1$ so preferences are consistent with balanced growth. Then the coefficient on $\log(1-\tau)$ is purely a function of γ.

A key point in the context of this benchmark model is the tight connection between preference parameters estimated from micro data and the implied aggregate elasticity for a particular tax and transfer policy. Moreover, in this benchmark model, observing the response in steady-state hours worked by an individual at one particular age is sufficient to infer the aggregate response.

2.2. Micro Evidence Based on the Benchmark Model

The literature that uses micro data to estimate labor supply elasticities is vast, so we make no attempt to summarize it here.[4] Instead, we consider three of the most influential papers, MaCurdy (1981), Browning et al. (1985), and Altonji (1986), each of which estimates the intertemporal elasticity of substitution, or Frisch elasticity. Details of their approaches differ, but all involve regressing changes in hours on changes in wages. For example, MaCurdy (1981) uses the basic model described above extended to allow for heterogeneity and uncertainty to derive the change in hours equation:

$$\Delta\log h_{it} = \gamma\Delta\log w_{it}(1-\tau_{it}) - \gamma\log\beta(1+r_t) + \alpha\gamma\Delta X_{it} + \gamma\xi_{it} + \gamma\Delta\varepsilon_{it}. \tag{10}$$

The parameters α, β, and γ are as above, and the tax rate (τ_{it}) is allowed to vary across time and individuals. The X_{it} are controls 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 focused on three issues: First, the ξ_{it} will 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, but also 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 issues, all three of these influential papers instrument for wage changes, using variables that were presumably known at time $t-1$, although they differ somewhat in the choice of instruments, the choice of observed taste shifters, and the exact choice of the functional form for the labor supply function. For instance, MaCurdy (1981) uses polynomials in age and education to instrument for wages, exploiting that wages are known to follow an inverted U-shape over the life cycle, the shape of which varies with education. 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 1981, Browning et al. 1985, and Altonji 1986). These results have been quite influential in generating a consensus within the profession that the Frisch elasticity is small.

---

[4]Classic reviews of this literature include Hausman (1985), Pencavel (1986), Killingsworth & Heckman (1986), and Blundell & MaCurdy (1999). For more recent reviews, readers are referred to Meghir & Phillips (2010) and Keane (2011).

[5]MaCurdy (1981) does not allow for shocks to the marginal utility of wealth, but MaCurdy (1983) extends the analysis to allow for this. In terms of implementation, the only impact is that one needs to lag the instruments, which MaCurdy (1981) actually does, even though he does not incorporate uncertainty.

Because the Frisch elasticity is an upper bound on the Marshall and Hicks elasticities in the benchmark life-cycle labor supply model, a small estimated Frisch elasticity has contributed to the view that the Marshall and Hicks elasticities are also small. In fact, MaCurdy (1981) also shows that the results of estimating Equation 10 allow one to infer responses to permanent wage changes. Estimation of Equation 10 uncovers all parameters of the hours equation in levels,

$$\log h_{it} = \gamma \log\bigl(w_{it}(1-\tau_{it})\bigr) + \gamma \log \lambda_{i0} - \gamma \log \rho(1+r_t)^t + \alpha\gamma X_{it} + \gamma\varepsilon_{it}, \tag{11}$$

except for $\gamma \log \lambda_{i0}$, which is the individual specific constant (or fixed effect) in the levels equation (λ_{i0} is the marginal utility of wealth at $t = 0$). Thus, he backs out the value of $\gamma \log \lambda_{i0}$ in a second stage after estimating Equation 10 in the first stage. He can then in principle regress them on the whole set of life-cycle wages.[6] His estimates imply that a 10% (fully anticipated) increase in wages at all ages increases labor supply by only 0.8%—a very small effect.

2.3. Macroeconomic Models

Although 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 & Rapping (1969), many macroeconomists have argued that relatively large Frisch elasticities are required to account for labor market fluctuations over the business cycle.[8] Prescott (2004) shows that a relatively large labor supply elasticity is also required to rationalize trend changes in hours of work among G-7 economies since 1970.[9] In fact, in the infinitely lived stand-in household models that remain the norm in much of the macro literature, it is standard to assume that the period utility function is log linear in consumption and leisure. If one-third of available time is spent in market work, this implies a Frisch elasticity of 2.0.

2.4. Overview of the Review

The above discussion is meant to illustrate that, viewed from the perspective of the simple benchmark model described earlier, there is a strong tension between evidence based on micro studies and specifications commonly adopted in aggregate studies. The purpose of this article is to provide a deeper analysis of this tension. We concede up front that we have a clear opinion on this matter: We argue that evidence from studies such as MaCurdy (1981), Browning et al. (1985), and Altonji (1986) is fully consistent with a world in which aggregate labor supply elasticities are large.

There are several approaches one could follow. One could challenge the claim that the micro literature offers a clear consensus on labor supply elasticities. And as with any empirical work, one could criticize the studies that find small labor supply elasticities on their own terms. That is, one could accept the basic empirical framework (e.g., Equation 10) but question the implementation. Specifically, one could question the instruments for wages, the controls for tastes for work, the

[6] Of course, MaCurdy (1981) observes wages only for his 10-year sample period, not the whole life cycle. To deal with this problem, he fits a quadratic life-cycle wage profile for each person using 10 years of data and regresses the estimated values of γ log λ_{i0} on the individual specific parameters of this profile. Using the coefficient on the wage profile intercept, he can determine how an upward shift in the whole wage profile affects γ log λ_{i0}, and hence labor supply.

[7] In his Nobel lecture, Prescott (2006) argues that large labor supply elasticities are important to reconcile 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.

[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).

functional forms for the labor supply function, measurement of wages, taxes, and so on. As Keane (2011) explores these two approaches in depth, we do not pursue them here.

We instead focus on three other issues. The first, pursued in Section 3, questions fundamental assumptions of the empirical framework of Equation 10. Specifically, we describe several important features absent from the benchmark model that may have led prior studies to understate labor supply elasticities: human capital accumulation (Imai & Keane 2004), credit constraints (Domeij & Floden 2006), uninsurable wage risk (Low 2005), and optimization frictions (Chetty 2012).

The second approach, pursued in Section 4, questions whether standard micro data estimates are relevant for determining aggregate labor supply responses. The key issues here are the extensive margin, population heterogeneity, and aggregation. Chang & Kim (2006), Rogerson & Wallenius (2009), and Erosa et al. (2014) show that labor supply could appear quite inelastic based on standard micro estimates, yet be quite elastic in the aggregate.

The third (and related) approach, pursued in Section 5, highlights that most micro empirical work finding small elasticities focuses on adjustment along the intensive margin. In contrast, we show that the relatively small micro data literature that allows for an extensive margin consistently finds large extensive margin elasticities, both for men and for women. This literature includes the early contributions of Cogan (1981) and Heckman & MaCurdy (1980) and the more recent contributions of Eckstein & Wolpin (1989), Kimmel & Kniesner (1998), Keane & Moffitt (1998), Keane & Wolpin (1997, 2000, 2001, 2010), French (2005), and Blundell et al. (2013).

In each section, we emphasize how the strong link between preference parameters estimated from micro data and the implied aggregate labor supply elasticity for a particular tax and transfer policy is broken. Section 6 concludes.

3. MICRO EVIDENCE BASED ON EXTENSIONS OF THE BASIC MODEL

The standard life-cycle labor supply model abstracts from several features, including human capital accumulation, credit constraints, uninsurable wage risk, and optimization frictions. Recent papers argue that failure to include these factors has 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. Here we review this work.

3.1. Human Capital Accumulation

The classic MaCurdy (1981) life-cycle model assumes that wages evolve exogenously, precluding the possibility that workers acquire human capital via learning-by-doing or on-the-job investment. Two early papers that include human capital accumulation are Heckman (1976) and Shaw (1989). Heckman (1976) studies a model with on-the-job investment in which workers are paid only for the time they spend on productive work (not the time they spend learning). Thus, a worker that devotes any time to investment has a measured wage rate (i.e., earnings divided by total hours at work) less than his or her true productivity. His estimates imply that productivity exceeded wages by over 50% for young workers, with this gap largely vanishing for workers in their forties. Shaw (1989) includes learning-by-doing in a life-cycle model. The return to work now consists of the wage and the value of the human capital generated as a by-product. Her estimates also imply that the return to an hour of work substantially exceeds the observed wage for young workers, but much less so for older workers.

These models share two key properties: (*a*) The observed wage is less than the true price of time for young workers, and (*b*) the observed wage grows more quickly than the price of time over the life cycle. Notably, neither Heckman (1976) nor Shaw (1989) directly assesses the implication of these properties for estimates of preference parameters and labor supply responses. This issue is addressed by Imai & Keane (2004), who argue that abstracting from human capital accumulation would downwardly bias estimates of γ.

To illustrate the logic, assume wages evolve according to

$$w_{t+1} = \left(1 + \kappa \sum_{j=1}^{t-1} h_{t-j}\right) w_1, \tag{12}$$

where $\kappa > 0$, and w_1 is the individual's wage when first entering the labor market. A unit increase in h_t raises w_t by κw_1 in all future periods.[10] In this model, the return to an hour of work, which Imai & Keane call the opportunity cost of time (OCT), consists of the current after-tax wage plus the expected present value of increased (after-tax) earnings in all future periods. Imai & Keane refer to this second component as the human capital term. The optimality condition for an interior solution equates the marginal rate of substitution between consumption and leisure to the OCT. Assuming the utility function from the benchmark model of Section 2, this gives

$$\frac{\alpha h_t^{\frac{1}{\gamma}}}{c_t^{-\frac{1}{\eta}}} = w_t(1-\tau_t) + E_t \sum_{j=0}^{T-t} \frac{\kappa w_1 h_{t+1+j}(1-\tau_{t+1+j})}{(1+r)^{1+j}}. \tag{13}$$

A model without human capital equates the marginal rate of substitution to the after-tax wage itself. The human capital term creates a wedge between the OCT and the after-tax wage. Importantly, this wedge declines with age owing to the shrinking time horizon for recouping returns to human capital investment.[11]

Figure 1 displays (stylized) life-cycle profiles for male wages and earnings, in addition to the OCT and human capital curves described above. The wage rate exhibits the familiar hump shape found in many studies (i.e., wages grow rapidly early in the life cycle, peak in the forties, and then decline).[12] Annual work hours also have a hump shape but with much less curvature (see, e.g., the descriptive regressions in Pencavel 1986). Graphically, the OCT curve is the vertical sum of the wage and human capital curves. Because the human capital curve declines with age (owing to shrinking in the remaining horizon and possibly decreasing returns to accumulating human capital), the OCT curve is much flatter than the wage curve.

The intuition for why ignoring human capital downwardly biases estimates of γ is now straightforward. Viewed through the lens of a MaCurdy (1981)–type model with exogenous wages, the relatively slow growth in hours relative to wages over the first half of the life cycle can only be rationalized if workers are very unwilling to substitute labor intertemporally, implying a small value for γ.[13] In contrast, in Imai & Keane (2004), it is the slope of the hours curve relative

[10]Heckman (1976), Shaw (1989), and Imai & Keane (2004) all assume much more complex human capital production functions than reflected in Equation 12. For instance, they all allow for complementarity between human capital and hours of work. This captures the empirical regularity that returns to work experience are greater for high-skilled workers.

[11]Like Heckman (1976) and Shaw (1989), Imai & Keane (2004) find that the OCT exceeds the wage by 50–100% for workers in their twenties but is only modestly greater than the wage for workers in their forties.

[12]Note that the wages of more educated workers tend to grow more and peak later.

[13]The male labor supply literature typically ignores older males to avoid dealing with corner solutions (i.e., retirement). Estimates of γ are thus roughly equal to the ratio of the slope of the hours curve to the wage curve prior to approximately age 50. Empirically, this ratio is approximately 0.25–0.30, which explains the bunching of Frisch elasticity estimates in that range (see Keane 2011 for more details).

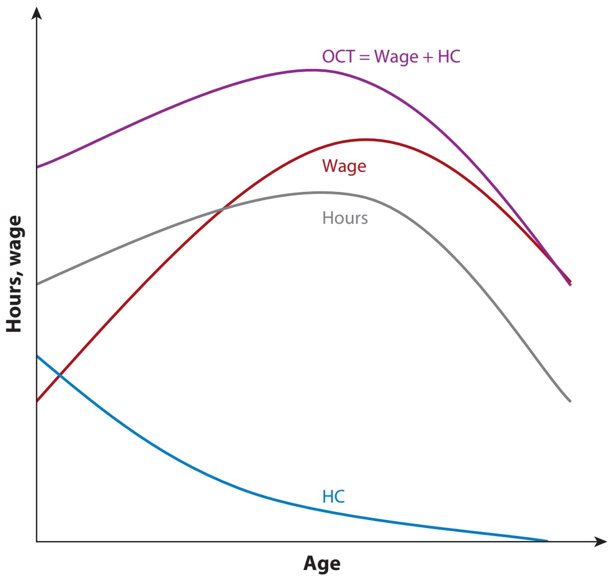

Figure 1

Hours, wages, and price of time over the life cycle. Human capital (HC) denotes the return to an hour of work experience, in terms of increased present value of future wages. The opportunity cost of time (OCT) is Wage + HC. Figure reproduced from Keane & Rogerson (2012).

to the OCT (rather than the wage) that matters for estimating γ, thus implying a much larger estimate of γ. Indeed, Imai & Keane estimate that $\gamma = 3.8$.[14]

Importantly, and in contrast to MaCurdy (1981), human capital breaks the direct link between γ and the Frisch elasticity. Indeed, simulations reported in Imai & Keane imply that the elasticity of hours with respect to an anticipated transitory wage change is 0.30 at age 20, much less than one might have expected, given that $\gamma = 3.8$. Intuitively, a young worker is not much inclined to reduce labor supply in response to a transitory wage cut because of the negative implications for future human capital, implying that human capital dampens the response to temporary wage changes.

Another key prediction of the human capital model is that labor supply elasticities with respect to (anticipated) transitory wage changes increase steadily with age. The estimates in Imai & Keane (2004) imply that this elasticity increases to roughly 0.44 at age 30, 0.66 at age 40, 1.1 at age 50, and 2.0 at age 60. Why does this elasticity increase with age? The answer is that the gap between γ and this elasticity is determined by the gap between the wage and OCT. As shown in **Figure 1** and noted above, this gap (i.e., the human capital term) decreases with age. Intuitively, the dampening effect of human capital decreases as a worker ages, and human capital accumulation becomes less important.

It follows that any downward bias in labor supply elasticity estimates due to ignoring human capital should be minor for older individuals. Consistent with this, it is interesting that French (2005), in a study of retirement behavior, estimates a large labor supply elasticity for 60-year-olds in the Panel Study of Income Dynamics (PSID) (i.e., approximately 1.3).

[14]Imai & Keane's sample consisted of white males from the National Longitudinal Survey of Youth 1979 (NLSY79), ages 20–36. As the focus is on labor supply, they were required to have finished school.

Human capital also has distinctive implications for the impact of permanent changes in wages or taxes. In the benchmark model of Section 2, the Hicks and Marshall elasticities are $\gamma\eta/(\eta+\gamma)$ and $\gamma(\eta-1)/(\eta+\gamma)$, respectively, and as $\gamma > \gamma\eta/(\eta+\gamma) > \gamma(\eta-1)/(\eta+\gamma)$, we have that Frisch > Hicks > Marshall. Because adding human capital alters the mapping from preference parameters to elasticities, this implication no longer necessarily holds. For example, given the Imai & Keane (2004) estimates of $\gamma = 3.8$ and $\eta = 1.3$, the benchmark model would imply a Hicks elasticity of approximately 1.0. However, using the Imai & Keane model to simulate a permanent tax increase (with proceeds distributed lump sum) yields a Hicks elasticity of only 0.64 at age 20.[15]

Two aspects of this result are striking. First, for young workers, the human capital mechanism substantially reduces the Hicks elasticity relative to what the benchmark model would imply given these utility parameters. Second, this is more than twice the impact of an anticipated temporary wage change for a worker at age 20, which as noted above is 0.30. This contradicts a strong prediction of the benchmark model (i.e., Frisch > Hicks) and the broader conventional wisdom in economics that temporary price changes have larger effects on demand than do permanent ones (owing to intertemporal substitution).

How can permanent tax changes have larger short-run effects than anticipated transitory tax changes? From Equation 13, a transitory tax affects only the current after-tax wage, which is just one component of the OCT. But a permanent tax also affects the human capital term, giving a larger overall effect on the OCT. Intuitively, with endogenous wages, a permanent tax increase creates an extra work disincentive, as it reduces both the current wage and the return to human capital investment.

However, the income effect of a permanent tax is a force working in the opposite direction. Keane (2009) notes the possibility that a permanent tax change (either compensated or uncompensated) could have a larger effect on current labor supply than a temporary one could. Using a simple two-period model, he shows that this occurs if and only if the return to work experience (the human capital effect) is large enough relative to the income effect. Whether effects of permanent tax changes can exceed those of transitory changes is thus an empirical question.

In the Imai & Keane model, elasticities with respect to transitory tax changes grow with age. They start to exceed elasticities with respect to permanent compensated tax changes once workers are in their forties. This is consistent with the result in Keane (2009); as workers age, the return to human capital investment falls, so the human capital effect can no longer dominate the income effect.[16]

As with temporary tax changes, the effects of permanent tax changes on current labor supply differ greatly depending on a worker's age when the tax is implemented. For instance, for workers in their twenties, thirties, and forties, the elasticities with respect to permanent compensated (surprise) tax increases are in the range of 0.45–0.64. But for workers in their fifties and sixties, these Hicks elasticities grow substantially, reaching 0.84 at age 50 and 2.0 at age 60. This growth occurs because, for older workers, human capital concerns no longer dampen the Hicks elasticity.

It is also of interest to ask how permanent changes in taxes affect labor supply, not just in the current period, but also over the rest of the life cycle. Simulating a permanent (compensated) 5% tax increase that takes effect at age 25 in the Imai & Keane model yields hours reductions of 2.7% at age 25, 5.1% at age 45, and 19.3% at age 60.

[15]Similarly, the utility parameters imply a Marshallian elasticity of 0.2, but the model implies only 0.1. We thank Susumu Imai for providing us with these permanent tax simulations, which were not given in the original Imai & Keane (2004) article.

[16]Although theoretically possible, simulations of the Imai & Keane model do not show Marshallian elasticities exceeding Frisch elasticities at any age. The income effect of such taxes is too strong.

This increasing labor supply response with age occurs for two reasons. First, as a worker ages, the after-tax wage makes up a larger fraction of the OCT, so a given tax has a larger direct effect.[17] Second, the higher tax reduces the rate of human capital accumulation, creating a feedback loop: If a worker reduces labor supply at time t, he or she has less human capital at time $t + 1$, causing him or her to work even less at time $t + 1$, leading to a lower wage at $t + 2$, and so on. In the above simulation, a worker's pretax wage is reduced by 1.0% at age 40, 3.6% at age 55, and 7.5% at age 60 relative to the baseline, so eventually, the pretax wage reduction due to the tax increase is greater than the tax increase itself.

An important implication, 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. Behavioral responses to tax changes alter the life-cycle wage path itself. Or, as noted by Imai & Keane (2004), in the human capital model, there is simply no such thing as an exogenous wage change.[18] It follows that observed labor supply responses to exogenous changes in tax rates have no clear structural interpretation.

Lastly, we consider the elasticity of lifetime hours of work. Simulating the impact of different permanent tax regimes in the Imai & Keane model yields uncompensated (Marshallian) and compensated (Hicks) elasticities of approximately 0.4 and 1.3, respectively. These values are quite large compared to typical estimates from models without human capital and exceed the values implied by the benchmark model given the same utility function parameters, which are 0.2 and 1.0. Although human capital dampens labor supply responses to transitory tax changes, it magnifies the impact of permanent changes.

3.2. Borrowing Constraints

In a model with credit constraints, reallocation of hours across time may require reallocating consumption across time, and the willingness to substitute labor intertemporally may be limited by the willingness to reallocate consumption.[19] Technically, the Frisch elasticity, defined as the change in hours in response to a change in the wage, holding the marginal utility of consumption fixed, no longer exists; 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. Nevertheless, the more general concept of an intertemporal elasticity of substitution in labor supply still exists.

Domeij & Floden (2006) argue that credit constraints may explain why researchers obtain low estimates of the intertemporal elasticity of substitution when estimating equations like Equation 10. Consider a worker who experiences a temporary negative wage shock. Absent credit market frictions, he or she would reduce hours today and either borrow against future income or run down current wealth to smooth consumption. But if borrowing is not possible, and the person has little or no wealth, the only way to smooth consumption is by increasing current labor supply. Thus, borrowing constraints may actually reverse the sign of the labor supply response, at least for

[17] This is the same reason that effects of transitory tax increases are greater for older workers.

[18] This implies that nothing short of full structural estimation of the joint labor supply/human capital investment process is adequate for estimating preferences and structural labor supply elasticities.

[19] 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. Thus, a positive association between consumption and labor income might suggest that individuals are credit constrained when in fact they are not (see Heckman 1974 for a discussion of this issue).

workers with low wealth. If such workers are prevalent in the data, it will attenuate the estimated hours response to wage changes.

Domeij & Floden argue that credit constraints are important in the US economy and that many households hold little wealth. That many households hold little wealth is well established empirically (see Deaton 1991 or Díaz-Giménez et al. 1997). Whether households are truly credit constrained is more difficult to determine. Indeed, the literature on testing for the existence of credit constraints is rather controversial, and there is no consensus on whether they are quantitatively important.[20] Here, like Domeij & Floden, we simply assume their existence and examine their implications for labor supply elasticities.

Domeij & Floden (2006) assume the same period utility function as in the benchmark model of Section 2, but the flow budget equation is now

$$A_{it} = (1+r)[A_{it} + w_{it}h_{it} - c_{it}], \quad A_{it} \geq 0. \tag{14}$$

The stochastic process for wages is

$$\log w_{it} = \psi_t + z_{it}, \text{ where } z_{it} = \rho z_{it-1} + \varepsilon_{it}. \tag{15}$$

MaCurdy's (1981) instrumental variable (IV) procedure to estimate γ in Equation 10 does not require one to specify a particular wage process. However, once we introduce extensions such as human capital 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, ϕ_{it} is zero when optimal assets are positive, but it is positive if the optimal asset level is negative (i.e., the nonnegativity constraint binds). The marginal utility of consumption evolves according to

$$\Delta \log \lambda_{it} = \log \beta(1+r_t) - \frac{\phi_{it-1}}{\lambda_{it-1}} + \xi_{it}, \tag{16}$$

and Equation 10 becomes

$$\Delta \log h_{it} = \gamma \Delta w_{it}(1-\tau_{it}) - \gamma \frac{\phi_{it-1}}{\lambda_{it-1}} - \gamma \log \beta(1+r_t) + \alpha\gamma\Delta X_{it} + \gamma\Delta\varepsilon_{it}. \tag{17}$$

The term $\phi_{it-1}/\lambda_{it-1}$ can be interpreted as an omitted variable in the conventional IV estimation method. Higher expected wage growth from $t-1$ to t tends 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} is positively correlated with expected wage growth. Higher expected wage growth from $t-1$ to t also 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.

Moreover, as is evident from Equation 17, the term $\phi_{it-1}/\lambda_{it-1}$ has a negative effect on hours growth. Intuitively, when people are liquidity constrained (i.e., $\phi_{it-1} > 0$), they tend to work more than they would if they could borrow against future income. Thus, $\phi_{it-1}/\lambda_{it-1}$ is positively

[20]Notable papers in this literature include Zeldes (1989), Jappelli (1990), Keane & Runkle (1992), Hubbard et al. (1995), and Keane & Wolpin (2001).

correlated with expected wage growth and negatively correlated with hours growth. Hence, its omission will lead to downward bias in estimates of γ.[21]

Domeij & Floden (2006) set $\{\rho, \sigma_\varepsilon, \sigma_\psi\} = \{0.90, 0.21, 0.34\}$ based on estimates in Floden & Linde (2001). The values of ρ and σ_ψ imply a high degree of persistence in wages. To assess the impact of credit constraints, it is crucial to use a reasonable value for the variance of transitory wage shocks. The value $\sigma_\varepsilon = 0.21$ is plausible if a large fraction of observed wage variation results from measurement error. They set $\gamma = 0.50$, so that without credit constraints, the Frisch elasticity would be 0.5.[22]

Domeij & Floden simulate data from the model and apply the MaCurdy (1981) IV estimation procedure to these data. Using the full sample, they estimate $\gamma = 0.23$. Restricting the sample to observations with positive assets, they obtain $\gamma = 0.44$. And further restricting the sample to observations with assets above the sample mean, they obtain $\gamma = 0.50$. This suggests that credit constraints can substantially reduce estimates of γ.

Unfortunately, Domeij & Floden (2006, p. 250) fall into an interpretation error when they state that "ignoring liquidity constraints ... the estimated elasticity is then 0.23 ... [while] the true elasticity is 0.50." As we emphasize above, once one extends the basic life-cycle model to include features such as human capital or credit constraints, there is no longer a direct mapping from regression coefficients in Equation 10 to the preference parameter γ or the Frisch elasticity. Thus, what they should conclude is that the failure to take credit constraints into account can lead to downward biased estimates of γ.

Domeij & Floden also report estimates of the intertemporal elasticity using PSID data on male household heads. Using the full sample and a MaCurdy (1981)–type IV procedure, they obtain an elasticity of 0.42. But restricting the sample to households with liquid wealth equal to at least one month's income, the estimate increases to 1.28. These results are again consistent with the idea that the preference parameter γ is considerably larger than prior estimates suggest. Another implication is that labor supply responses may be much more elastic for higher-wealth 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.

However, the standard errors on their PSID estimates are quite large. A 95% confidence interval takes it from near zero to two, and a formal test would probably not reject equality of the estimates in the two samples. This is a manifestation of the weak instrument problem (i.e., the age/education polynomials they use as instruments do not predict wage changes very well).

A related paper is by Low (2005), who explores the implications of uninsurable wage risk for the life-cycle path of labor supply. Young workers know that the average life-cycle wage path has a hump shape like that in **Figure 1** but also perceive that there is considerable idiosyncratic uncertainty about the extent of life-cycle wage growth they will experience and that there is no insurance against this uncertainty. Hence, if there is a strong precautionary motive, young workers will not choose to borrow against expected future income to finance higher consumption. Furthermore, young workers will have an incentive to work relatively long hours despite low wages to

[21]A subtle point is that the credit constraint variable $\phi_{it-1}/\lambda_{it-1}$ is endogenous in Equation 17. It is not correlated with the error component ξ_{it} because ξ_{it} is by definition a surprise not known at $t-1$. However, $\phi_{it-1}/\lambda_{it-1}$ can be correlated with the change in tastes for work $\Delta\xi_{it}$, as these may be expected at $t-1$. If a worker expects his or her tastes for work to increase from $t-1$ to t (e.g., the worker is recovering from an illness), then he or she 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 to estimate Equation 17.

[22]These parameter values, along with $\eta = 2/3$, match the feature of the US data that the bottom 40% of households hold only 1.4% of wealth.

build up a buffer stock of assets that serves as self-insurance against the potential adverse shocks to life-cycle wage growth.

The essential idea of Low's (2005) model can be seen in his figure 6. In a simulation in which workers have certainty about the wage path, hours rise steeply over the life cycle as wages increase. However, with the introduction of uncertainty, hours are much higher at young ages, and hours growth is greatly attenuated.

A researcher who looked at data generated from Low's model using MaCurdy (1981)–type methods to estimate equations like Equation 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. Now, 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 should induce workers to work substantially fewer hours when young. Unfortunately, Low did not use his model to explore such policy simulations.[23]

Although closely related, Low (2005) and Domeij & Floden (2006) focus on distinct mechanisms; in the former, young workers do not borrow against future income because they do not want to, whereas in the latter they do not borrow because they cannot. This illustrates why credit constraints are difficult to identify empirically: They generate behavior that looks very similar to that generated by several other mechanisms: a strong precautionary motive, complementarity of consumption and hours, time-varying tastes for work/consumption, etc.

3.3. Optimization Frictions

Chetty (2012) argues that abstracting from fixed costs of adjusting labor supply may also downwardly bias labor supply elasticities. These fixed costs may be time costs (e.g., paperwork, finding a new job with different hours), psychic costs (e.g., doing the mental calculations to reoptimize when tax rates change), or information costs (e.g., people may not notice small tax changes).

Unlike the papers discussed above, Chetty does not actually solve or estimate an extension of a basic labor supply model that incorporates them. Rather, he attempts to bound the magnitude of the bias in elasticity estimates that might be attributed to ignoring fixed costs. Basically, he asks, Assuming that when taxes change, people do not adjust labor supply if the resultant welfare gain is less than some small fraction (δ) of consumption (where that fraction represents the fixed cost), what is the implied bias in conventional elasticity estimates?[24]

Chetty (2012) argues that elasticity estimates are likely to be biased downward, perhaps substantially. This result stems from an asymmetry in how adjustment costs affect behavior when elasticities are high versus low. If the elasticity is large, then the objective function is fairly flat in the vicinity of optimal hours, so a sizeable departure from optimal hours causes only a small welfare loss. So, even if labor supply elasticities are large, we may observe small labor

[23]Alonso-Ortiz & Rogerson (2010) study an infinitely lived agent model with precautionary savings and find that this type of effect is large. That is, social insurance has a particularly large negative effect on the labor supply of low-productivity workers with low asset holdings.

[24]Although 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.

supply responses to taxes provided there are small adjustment costs. In contrast, if labor supply elasticities are truly small, adjustment costs provide no mechanism that would lead us to infer they are large.

To proceed, assume a simple quasi-linear utility function,

$$U_i = wh_i(1-\tau_i) - \frac{\alpha}{1+\frac{1}{\gamma}} h_i^{1+\frac{1}{\gamma}}. \tag{18}$$

As there are no income effects, the Marshall, Hicks, and Frisch elasticities are equivalent. Optimal hours are

$$h_t^* = \left[\frac{(1-\tau)w}{\alpha}\right]^{\gamma}, \tag{19}$$

and utility evaluated at the optimum is

$$U\left(h_t^* \middle| \tau_t\right) = \frac{1}{1+\gamma}\left[\frac{1}{\alpha}\right]^{\gamma}\left[(1-\tau_t)w\right]^{1+\gamma}. \tag{20}$$

Consider a change in $(1-\tau)$. The impact on utility can be decomposed into the direct effect of the change, holding h fixed, plus the effect induced by the behavioral response of changing h:

$$U\left(h_{t+1}^* \middle| \tau_{t+1}\right) - U\left(h_t^* \middle| \tau_t\right) = \left[U\left(h_t^* \middle| \tau_{t+1}\right) - U\left(h_t^* \middle| \tau_t\right)\right] + \left[U\left(h_{t+1}^* \middle| \tau_{t+1}\right) - U\left(h_t^* \middle| \tau_{t+1}\right)\right]. \tag{21}$$

From Equation 18, the first term on the right-hand side is just $wh_t^*\Delta(1-\tau)$, the change in c holding h fixed. The second term, the hours adjustment term, is a second-order effect that can be ignored in the case of small tax changes. From Equation 20, we have that $\left(dU(h_t^*|\tau_t)\right)/\left(d(1-\tau_t)\right) = wh_t^*$, so $\left(d^2U(h_t^*|\tau_t)\right)/\left(d(1-\tau_t)^2\right) = \gamma wh_t^*/(1-\tau_t)$. Thus, using a Taylor series approximation, we have that, to second order,

$$U\left(h_{t+1}^* \middle| \tau_{t+1}\right) - U\left(h_t^* \middle| \tau_t\right) = wh_t^*\Delta(1-\tau_t) + \frac{1}{2}\gamma\frac{wh_t^*}{(1-\tau_t)}\Delta(1-\tau)^2. \tag{22}$$

Now assume a worker will not adjust hours if the utility gain is less than a fraction δ of consumption:

$$U\left(h_{t+1}^* \middle| \tau_{t+1}\right) - U\left(h_t^* \middle| \tau_{t+1}\right) = \frac{1}{2}\left|U''\left(h_t^*\right)\right|\left(h_{t+1}^* - h_t^*\right)^2 < \delta wh_t^*(1-\tau_t). \tag{23}$$

With quasi-linear utility, one obtains $|U''(h_t^*)| = \alpha(1/\gamma)(h_t^*)^{\frac{1}{\gamma}-1}$. Assuming 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:

$$\frac{h_{t+1}^* - h_t^*}{h_t^*} < [2\gamma\delta]^{1/2}. \tag{24}$$

Chetty (2012) uses this bound to derive bounds on elasticities. 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 $\pm[2\gamma\delta]^{1/2}/\Delta\log(1-\tau)$, assuming workers start at an optimum at time t.[25] As the change in tax rates appears in the denominator, the bounds are wider for smaller tax changes.[26]

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 were the same in all cases. Specifically, Chetty argues that estimates from contexts in which wages or taxes changed relatively little might be expected to generate elasticity estimates 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.

A good illustration of this point is Chetty's analysis of MaCurdy (1981), assuming $\delta = 0.01$. Even though MaCurdy (1981) estimates an intensive margin elasticity of only 0.15, his estimate is consistent with a structural elasticity as large as $\gamma = 1.20$. 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. As another example, Blundell et al. (1998) estimate labor supply elasticities for employed married women, by exploiting UK tax rate variation from 1978 to 1992 and find a compensated elasticity of 0.20. Chetty (2012) derives bounds on their estimate that range from essentially 0 to 2.04. 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 also applies his methodology to the data used in Prescott (2004) to identify labor supply elasticities based on differential aggregate hours and tax rate changes between the United States and the United Kingdom from 1979 to 1996. Here the bounds are 0.42–2.14. 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 examines in the sense of generating a relatively large lower bound.[27]

3.4. Summary

Adding empirically plausible features to the simple benchmark model of Section 2 can drastically alter the mapping from coefficients in a prototypical regression equation such as Equation 10 to underlying preference parameters. The literature suggests that accounting for human capital, liquidity constraints, uninsurable wage risk, and adjustment costs causes conventional econometric methods to understate both the willingness of workers to substitute leisure intertemporally (as captured by the preference parameter γ) and the implied labor supply responses to changes in wages or taxes.

Our presentation considers each of these extensions to the benchmark framework in isolation to emphasize the distinctive economic forces in each case. It is of interest for future work to consider these extensions jointly, both to evaluate how they interact and to assess their relative importance. Integrating human capital accumulation, credit constraints, and precautionary savings motives into the benchmark life-cycle model seems both natural and straightforward. Integrating the optimization frictions in Chetty (2012) into a structural life-cycle model is less

[25] If that assumption is not invoked (as is the case in Chetty 2012), the width of the bounds doubles.

[26] Because the bounds depend on the elasticity γ itself, we obtain an implicit equation, which Chetty (2012) 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.

[27] Of course, other factors may have shifted labor supply in these countries over the sample period. Just as estimates from micro data face several econometric issues, omitted factors is a key issue for estimates using aggregate data.

straightforward, as this would require taking a stand on the exact nature of the optimization costs. In some cases, the effects we have studied may partially offset each other; for example, if human capital accumulation concerns lead younger workers to work more hours, then the impact of credit constraints may be less relevant. Based on existing work, it is our view that the inclusion of human capital accumulation as a way to account for life-cycle changes in wages is of paramount importance for the analysis of labor supply.

4. AGGREGATE LABOR SUPPLY IN MODELS WITH EXTENSIVE MARGIN ADJUSTMENT

The previous section summarizes how extensions of the benchmark model can influence estimates of the preference parameter γ. This section summarizes a class of models in which the estimate of γ from micro data has little or no influence on the value of aggregate elasticities. The key feature of this class of models is the presence of an operative extensive margin.[28]

4.1. Indivisible Labor Models

The starting points for our discussion are the indivisible labor papers by Hansen (1985) and Rogerson (1988), who study homogeneous agent models in which all adjustments at the individual level were assumed to occur at the extensive margin (i.e., the intensive margin was fixed by assumption). Specifically, individuals had preferences given by

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

but the choice of h_t was restricted to zero or $\hat{h}$. A key result was that, assuming a set of markets sufficiently rich to decentralize optimal allocations, aggregate allocations in this economy were identical to those that would emerge from an economy with a representative household that made all labor supply adjustment at the intensive margin but had preferences given by

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

where α is a constant.[29] This equivalence result says that the economy behaves as if there were a representative household choosing hours along the intensive margin but with an infinite Frisch elasticity. Importantly, this is independent of the function $v(1 - h)$ that described the true preferences of individuals in the economy, implying a disconnect between estimated parameters of $v()$ and aggregate behavior.

One issue with this result is that it assumes identical households. Both Cho (1995) and Mulligan (2001) demonstrate theoretically that the implication of an infinite Frisch elasticity for aggregate labor supply is not robust to including heterogeneity. More generally, the Frisch elasticity for

[28]Heckman (1984) was an early proponent for the development of labor supply models that include an extensive margin to better understand aggregate labor supply.

[29]Early derivations of this result assumed that individuals could trade employment lotteries in equilibrium. Ljungqvist & Sargent (2007, 2008) show that "time averaging" is a perfect substitute for lotteries if an individual has access to credit markets.

aggregate labor supply would depend on the nature and extent of heterogeneity. A key issue was to assess the implications of empirically relevant sources of heterogeneity.

A quantitative analysis of this issue was undertaken by Chang & Kim (2006).[30] They consider an aggregate model in which labor supply is indivisible but also assume that individuals are subject to idiosyncratic shocks and face incomplete markets for credit and insurance. Households consist of a male and a female, with household preferences given by

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

where c_t is household consumption, and h_{mt} and h_{ft} are hours worked by the male and female household member, respectively. Each individual can only supply zero or $\hat{h}$ units of labor in any period. Individual productivity, denoted by x_t, is stochastic and follows the stochastic process

$$\log x_{jt+1} = \rho_j \log x_{jt} + \varepsilon_{jt+1},\ j = m, f. \tag{25}$$

The process is the same for all individuals of a given gender, and innovations are independently and identically distributed across individuals. A worker of productivity x_t has labor earnings $w_t x_t \hat{h}$ if working, where w_t is the wage per efficiency unit of labor.

The production side of the economy is standard: A Cobb-Douglas aggregate production function uses capital and efficiency units of labor; output can be used as either investment or consumption; and capital depreciates at a constant rate δ.

Each period, there are competitive markets for capital and labor services, as well as output. There are no markets for insurance against idiosyncratic shocks, so as in Aiyagari (1994) and Huggett (1993), individuals accumulate capital to self-insure. Household capital holdings cannot go below $\overline{a}$.[31]

Chang & Kim (2006) calibrate this model and show that it does a reasonable job of capturing cross-sectional heterogeneity in earnings and wealth.[32] Over time, individuals move between employment and nonemployment. The authors do not ask whether the model produces empirically reasonable patterns for these transitions, but recent work has shown this to be the case.[33]

Chang & Kim proceed to study the properties of individual and aggregate labor supply in their calibrated model. First, they consider a sample of 50,000 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 with positive hours in each year:

$$\log h_{it} = \gamma(\log w_{it} - \log c_{it}) + \varepsilon_{it}. \tag{26}$$

[30]Additional issues are explored in Chang & Kim (2007) and An et al. (2009).

[31]There are no markets for employment lotteries in this economy.

[32]As is well known, the model cannot capture the concentration of wealth in the upper 1% of the wealth distribution.

[33]Chang et al. (2011, 2014) examine this in a slightly more general model. Krusell et al. (2010, 2011) further extend this analysis by including search frictions and considering movements between employment, unemployment, and out of the labor force, although they consider single individual households.

They obtain estimates of γ equal to 0.41 and 0.78 for males and females, respectively. The key finding is that standard labor supply regressions on individual data generated by the model yield relatively small estimates of the labor supply elasticity parameter for men, although a moderate estimate for women.

Second, Chang & Kim subject the economy to an AR(1) aggregate technology shock, simulate the economy for 30,000 quarters, compute aggregates, and run the regression in Equation 26 using aggregate time-series data. The resulting estimate for γ is now 1.08.

Third, they consider a stand-in household model with preferences of the form

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

where h_t is now allowed to take on any value in the interval [0, 1]. Assuming the same process for aggregate technology shocks, Chang & Kim find that a $\tilde{\gamma}$ of approximately 2 generates fluctuations in aggregate hours that are the same as in the heterogeneous agent economy. That is, using a stand-in household model to mimic the business cycle statistics for the heterogeneous agent economy requires a value of $\tilde{\gamma}$ that is roughly five times as large as the estimate based on individual data for male workers.

In summary, the presence of empirically reasonable heterogeneity in this model does indeed have a dramatic effect on the implied aggregate elasticity, lowering it from infinity to approximately 2. Nonetheless, the model still delivers a substantial disconnect between the labor supply elasticity obtained using standard methods on micro data for continuously employed individuals and the aggregate elasticity. And although the aggregate elasticity is not infinite, the value of 2 is still large.

Related work has examined how aggregate hours in this framework react to the simple tax and transfer program studied in Section 2. Alonso-Ortiz & Rogerson (2010) use a single-agent household version of the Chang & Kim model and find a large response in aggregate hours, in fact, somewhat larger than what is implied by a stand-in household model with a Frisch elasticity of 2. Ljungqvist & Sargent (2007, 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. Although they find that aggregate hours respond similarly to what is found in models that abstract from human capital accumulation, human capital accumulation has distinct predictions for the identities of which individuals choose not to work as the tax and transfer program is expanded. Assessing the extent to which these predictions match what is found in the data is an open issue.

4.2. Models with Intensive and Extensive Margin Adjustment

The above models exogenously imposed that all labor supply adjustments occur along the extensive margin. How important is the extreme assumption that all adjustments take place along the extensive margin? To answer this question, we next explore the aggregate properties of models that feature adjustments along both the intensive and extensive margins. We begin by describing the analysis in Rogerson & Wallenius (2009), which generalizes the model in Prescott et al. (2009). This model can also be viewed as embedding a simplified version of French (2005) into a general equilibrium setting.

Consider an individual with the length of life normalized to one and preferences

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

where $c(a)$ is consumption at age a, and $h(a)$ is time devoted to market work at age a.[34] Individual productivity varies over the life cycle and is denoted by $e(a)$.

Following Prescott et al. (2009), the key feature of the model is a nonconvexity in the mapping from time devoted to work to the resulting labor services: When a worker of age a devotes h units of time to market work, it generates labor services of $\max\{h - \overline{h}, 0\}e(a)$.[35] With $\overline{h} > 0$, the model can generate "retirement" as an endogenous outcome, in the sense of a worker who switches from full-time work to no work despite continuous changes in fundamentals.

Let w be the constant wage rate per unit of labor services. Assuming complete credit markets and a zero interest rate, one finds that the present value budget equation for each individual is

$$\int_0^1 c(a)da = w\int_0^1 \max\left\{h(a) - \overline{h}, 0\right\}e(a)da. \tag{27}$$

Rogerson & Wallenius (2009) consider this single-agent problem in the context of a steady-state equilibrium of an overlapping generations model. For their quantitative work, they adopt $u(c) = \log c$, $v(h) = \alpha\left(\left(h^{1+1/\gamma}\right)/(1+1/\gamma)\right)$ and assume that life-cycle productivity $e(a)$ is piecewise linear. They consider values of γ ranging from 0.10 to 2.00 and in each case calibrate the model to match some key properties of life-cycle labor supply.[36]

Three key results emerge. First, if one uses the micro data from the model to run a regression of the form

$$\log(h(a)) = b_0 + \tilde{\gamma}\log\left(w^h(a)\right) + \varepsilon(a), \tag{28}$$

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: The wage per unit of time $w^h(a)$ moves more over the life cycle than does the underlying exogenous productivity profile $e(a)$.

Second, the response of aggregate hours to the permanent tax and transfer policy change considered in Section 2 is to first order independent of the value of γ. Increasing the tax rate from 30% to 50% decreases aggregate hours by approximately 20% for all values of γ in the range of 0.10–2.00.

Third, although γ has virtually no effect on the change in aggregate hours, it determines how the change in aggregate hours is decomposed into changes in working life versus changes in hours worked while employed. When $\gamma = 2.00$, the downward shift in the hours profile accounts for over 60% of the total decrease in hours, whereas when $\gamma = 0.10$ this figure is less than 5%. If a researcher used the benchmark model from Section 2 to interpret steady-state differences in aggregate hours worked across two Rogerson-Wallenius economies with a low value of γ that were identical except for the scales of the tax and transfer system, he or she would infer a value of γ more

[34]Rogerson & Wallenius (2009) abstract from discounting to facilitate an analytic characterization.

[35]More generally, one could consider specifications in which the marginal wage is a function of the length of the work week. French (2005) considers both specifications.

[36]Wallenius (2013) includes endogenous human capital accumulation as in Imai & Keane (2004) and shows that it better matches the life-cycle profiles for wages and hours. For expositional reasons, we focus here on the simpler model.

than an order of magnitude larger than the true underlying value of γ. The reason is that the response in aggregate hours includes a response on the extensive margin, and the implied value of γ must proxy for adjustment along both margins.[37]

To summarize, in this life cycle economy with operative intensive and extensive margins, labor supply elasticities estimated on micro panel data using workers with positive hours are not particularly relevant in predicting the aggregate effects of permanent changes in taxes. Moreover, the aggregate elasticity is large.[38]

Because the model of Rogerson & Wallenius (2009) is somewhat stylized, one may also be concerned about robustness to allowing for richer and more realistic empirical specifications. Erosa et al. (2014) go quite far in assessing this. Specifically, they extend the Rogerson-Wallenius model along many dimensions to match a wide variety of features of wages and hours worked for males between the ages of 25 and 61. Their analysis allows for multiple sources of heterogeneity (both idiosyncratic shocks as in Chang & Kim 2006 and fixed effects), multiple nonconvexities (fixed utility costs of working in addition to nonconvex earnings), time aggregation, search frictions, and measurement error in wages. Although these features do matter for the empirical properties of the model and its ability to replicate the salient features of the data, the conclusions are broadly similar. Their calibrated model generates a Hicks elasticity of 0.44. Although this is smaller than the values of Rogerson & Wallenius (2009), who found values in the vicinity of 0.75, the Erosa et al. study does not include the youngest and oldest workers, two groups that are central to the large Hicks elasticities in Rogerson & Wallenius.

Whereas Rogerson & Wallenius consider only how aggregate hours respond to permanent changes in taxes, Erosa et al. (2014) also solve for the aggregate labor supply elasticity in response to a purely temporary unanticipated wage change and obtain a value of 1.75. One result of interest relates to their simulation of the temporary tax holiday that occurred in Iceland in the late 1980s. Their model does reasonably well in capturing the responses found in the Icelandic data, in contrast to what Chetty et al. (2013) find based on a simulation of the much simpler model of Rogerson & Wallenius (2009).

In another recent paper, Chang et al. (2014) embed the nonlinear earnings function of Prescott et al. (2009) into a heterogeneous agent model like Chang & Kim (2006) to assess how heterogeneity and the value of the individual preference parameter γ jointly influence the aggregate labor supply elasticity. A key message is that the presence of an operative intensive margin plays an important role in influencing how heterogeneity affects the aggregate labor supply elasticity, so that abstracting from the intensive margin can have important consequences even if the value of γ is quite small.

4.3. Summary

In this section we argue that the connection between the individual preference parameter γ and various aggregate labor supply elasticities is much more complex than suggested by the benchmark model in Section 2. We highlight the importance of accounting for labor supply adjustment along the intensive and extensive margin and, in particular, how an operative extensive margin breaks the tight link between the preference parameter γ and aggregate labor supply responses.

---

[37]Wallenius (2011) provides another context in which this issue arises. She considers a simpler version of the Imai & Keane (2004) model but includes fixed costs and therefore an endogenous retirement decision. She infers 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 do Imai & Keane, her model gives similar responses for aggregate hours owing to the presence of the extensive margin. Loosely speaking, the estimated value of γ in Imai & Keane (2004) captures the response along both margins.

[38]Kitao et al. (2009) and Ljungqvist & Sargent (2014) argue that the Rogerson & Wallenius (2009) model contains too much responsiveness on the retirement margin and that many individuals are not at an interior solution with respect to retirement.

Interestingly, with the exception of Ljungqvist & Sargent (2007, 2008) and Wallenius (2011), none of the papers in this literature has considered adjustment along the intensive and extensive margin in a model that also includes an explicit human capital accumulation decision, and none of these analyses included an assessment of business cycle responses. Analyzing business cycle responses in a model that features human capital accumulation as well as operative intensive and extensive margins is an important area for future research. As noted in Section 3, the human capital model alone does not necessarily generate large aggregate labor supply responses to transitory shocks, even for large values of γ, as the human capital accumulation channel dampens the labor supply response, especially for younger workers. We conjecture that adding an operative extensive margin to the human capital model will increase the aggregate labor supply responses to transitory shocks in these models as well.

5. ADJUSTMENT ON THE EXTENSIVE MARGIN: EVIDENCE FROM MICRO DATA

The macro literature surveyed in the previous section essentially shows that one can reconcile a relatively large aggregate labor supply elasticity with the small estimates of γ from micro-econometric studies if there is sufficient movement along the extensive margin. A key issue in the micro-macro labor supply controversy is thus whether micro studies estimate large responses on the extensive margin. Here we survey the literature on this issue.

5.1. Early Work on Structural Models of Participation

To study female labor supply, in which nonparticipation is prevalent, Heckman & MaCurdy (1980, 1982) modify the utility function in MaCurdy (1981) to

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

Although this generates a reservation wage for participation (the marginal disutility of work is not zero at full leisure), optimal hours are a continuous function of wages, implying that we should observe some women who work very low hours if the wage distribution is continuous.

In fact, few women are observed to work small positive hours. To match this pattern, Cogan (1981) introduces fixed costs of work into a static labor supply model, generating what he calls a "reservation hours" level. Specifically, consider the quasi-linear utility function

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

With w, Y, and m denoting the wage rate, nonlabor income, and the fixed (monetary) costs of working, respectively, utility as a function of hours worked is

$$U(h) = (wh + Y - m) + \alpha\frac{(\overline{H} - h)^{1-\frac{1}{\gamma}}}{1 - \frac{1}{\gamma}}. \tag{29}$$

Optimal hours conditional on working are

$$h^* = \overline{H} - \left(\frac{w}{\alpha}\right)^{-\gamma}. \tag{30}$$

Working is optimal if $U(h^*) > U(0)$, which reduces to

$$h^* = \overline{H} - \left(\frac{w}{\alpha}\right)^{-\gamma} > \frac{m}{w} + \frac{1}{w}\frac{\alpha}{1-\dfrac{1}{\eta}}\left[\overline{H}^{1-\frac{1}{\gamma}} - \left(\frac{w}{\alpha}\right)^{\gamma-1}\right] = h_{\mathrm{R}} > 0. \tag{31}$$

Equation 31 implies that a person works only when optimal hours exceed the reservation hours level h_{R}.

Cogan (1981) proposes joint estimation of the conditional labor supply function (Equation 30), an offer wage function, and the reservation hours function (Equation 31) using a sample of married women aged 30–34 from the 1967 NLS Mature Women survey, roughly half of whom worked. Cogan estimates that fixed costs are substantial (about 28% of average annual earnings).

Cogan's estimates imply that a 10% increase in the offer wage to the typical nonworking woman would not induce her to work, but that a 15% increase would induce her to work 1,327 h. An additional 15% wage increase would induce an extra 180 h of work (or 13.6%).

Importantly, labor supply can appear to be elastic or inelastic, depending on the initial status of the person and the magnitude of wage change considered. As in the indivisible labor models surveyed in the previous section, behavior is not summarized by a small set of preference parameters or elasticities, and simulation of the full model is required to assess labor supply responses.

In fact, this property emerges in any generalization of the standard static labor supply model, which dispenses with a linear budget constraint, such as including welfare benefits and progressive taxation. Indeed, the literature on tax-transfer program effects on labor supply 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 & Hausman 1978; Hausman 1980, 1985; Moffitt 1983).[39]

Keane & Moffitt (1998) and Keane (1995) illustrate this point. They model labor supply of single mothers in the United States in the early 1980s, a group that was eligible for Aid to Families with Dependent Children (AFDC) benefits and food stamps and faced large fixed costs of work (childcare), generating a complex nonlinear budget constraint. The AFDC program taxed the earnings of welfare recipients very heavily. Historically, the AFDC tax rate varied between 50% and 100%, but little labor supply response was observed, so that labor supply appeared inelastic for single mothers. Simulations of the Keane-Moffitt model were consistent with this: Even massive reductions in the AFDC tax rate led to little increase in labor supply by AFDC recipients.

However, the Keane-Moffitt model predicts that labor supply of single mothers is very responsive to wage and childcare subsidies, so in this context, labor supply appears to be very elastic. This reinforces one of the main points emphasized in Section 1 about treating elasticities as primitive parameters to be exported across studies. Indeed, the predictions of the model are consistent with the experience of the mid- to late 1990s, when a major shift in welfare policy toward wage and childcare subsidies led to dramatic labor supply increases among single mothers (see Keane 2011, pp. 986–89, or Fang & Keane 2004 for more details).

[39]For instance, as noted by Hausman (1980, p. 161), "structural econometric models which make labor force participation a function of . . . wages, income transfer levels and the tax system can attempt to answer questions such as the effect of lowering the marginal tax rates on labor force participation. The more traditional reduced form models which do not explicitly parameterize the tax system will be unable to answer such questions."

5.2. Life-Cycle Models with a Participation Margin

Kimmel & Kniesner (1998) extend the basic MaCurdy (1981) framework to include fixed costs. Rather than structurally estimating the model's primitives, they estimate a life-cycle labor supply equation analogous to Equation 11 jointly with a participation decision rule and an offer wage function:

$$\log h_{it} = f_{hi} + \gamma_I \log w_{it} + \alpha_h Z_{it} + \varepsilon_{hit}, \tag{32}$$

$$P(h_{it} > 0) = F\left(f_{pi} + \tilde{\gamma}_P \log w_{it} + \alpha_p Z_{it}\right), \tag{33}$$

where f_{hi} captures the marginal utility of wealth, along with any fixed effects in tastes for work; f_{pi} captures these and any individual heterogeneity in the fixed costs of work; and F is a cumulative distribution function. Kimmel & Kniesner assume that F is normal, giving a probit model.

There are now two elasticity concepts of interest: γ_I is the conventional Frisch elasticity of labor supply conditional on employment (i.e., the elasticity on the intensive margin), and γ_P, defined by,

$$\gamma_P = \frac{\partial \log P(h_{it} > 0)}{\partial \log w_{it}} = \tilde{\gamma}_P \frac{F'(\cdot)}{F(\cdot)}, \tag{34}$$

is a Frisch participation elasticity.

Kimmel & Kniesner (1998) estimate this model using data on 2,428 women from the Survey of Income Program Participation (SIPP), 68% of them married. The data cover nine periods (May 1983 to April 1986). They find that $\gamma_I = 0.66$ and $\gamma_P = 2.39$. Let $H = P\hat{h}$ be average hours in the population, where $\hat{h}$ is the average hours of the employed and P is the percentage employed. Then

$$\frac{\partial \log H}{\partial \log w} = \frac{\partial \log P}{\partial \log w} + \frac{\partial \log \hat{h}}{\partial \log w} = 0.66 + 2.39 = 3.05. \tag{35}$$

They also obtain results for men and find that $\gamma_I = 0.39$ and $\gamma_P = 0.86$ so that $\gamma_I + \gamma_P = 1.25$.

In summary, the participation elasticity is much larger than the hours elasticity for both women and men, and the overall elasticity is quite a bit larger for women than for men.[40] 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.

Kimmel & Kniesner (1998) avoid full solution of agents' dynamic optimization problem by relying on a participation condition (Equation 33) derived from the first-order condition for hours evaluated at $h = 0$. More complex life-cycle models that include features such as human capital and credit constraints cannot be handled so simply, and estimation of such models requires a full-solution structural approach. This means (*a*) solving the dynamic optimization problem faced by agents and (*b*) finding parameter values such that the model generates behavior similar to observed behavior (by some metric).

Eckstein & Wolpin (1989) were the first to adopt a fully structural approach to estimating female labor supply. Their model includes work decisions on the extensive margin and human capital accumulation through work experience. Subsequent work has extended this analysis to

[40] Studies of female labor supply that abstract from the extensive margin typically find modest elasticities. For instance, Blundell et al. (1998) estimate a life-cycle labor supply model for employed married women using the UK Family Expenditure Surveys from 1978 to 1992. The United Kingdom reduced tax rates substantially over the period, and the paper aimed to exploit that variation to identify labor supply elasticities. Their estimates of compensated and uncompensated wage elasticities at the mean of the data were a modest 0.20 and 0.17, respectively.

include other important life-cycle decisions, such as marriage (van der Klaauw 1996) and fertility (Francesconi 2002). The most comprehensive modeling effort to date is by Keane & Wolpin (2007, 2010). They extend earlier work to include marriage, fertility, school attendance, part-time work, and welfare participation as choices. Simulations of their model imply a "long-run" labor supply elasticity in response to permanent wage changes of approximately 2.8.

This "long-run" elasticity has a very different interpretation from elasticities reported in the more conventional labor supply literature. First, it measures how a person born into a higher wage (or lower tax) regime would be affected once he or she reaches adulthood. Second, aside from labor supply, the simulation allows for adjustments along several other dimensions. For example, if a wage increase causes a woman to work more in the current period, she will not only have more human capital in the next period, but her expected number of children is also reduced, thereby further enhancing labor supply in the next period, and so on. This is analogous to the feedback effect of labor supply on future wages that we see in Imai & Keane (2004), but for women, there are additional sources of dynamics, such as fertility, that are likely to be important. Conventional labor supply studies that treat fertility as given are likely to understate long-run responses to permanent wage/tax changes.

Finally, Blundell et al. (2013) develop a dynamic model of female labor supply that incorporates asset and human capital accumulation and that endogenizes education. Using UK data, they find substantial Frisch elasticities (0.90 and 0.45 on the extensive and intensive margins) and substantial Marshallian elasticities (0.50 and 0.38 on the extensive and intensive margins). The latter are evaluated in the year a permanent unexpected tax change is implemented, so they may understate the long-run effects implied by the model.

There are fewer structural estimation papers for males that incorporate the extensive margin as it has generally been viewed as a less important factor for men, given their high participation rate. However, research suggests that the extensive margin is important for males who are young, near retirement, or members of minority groups. For instance, as noted above, French (2005) finds high labor supply elasticities for older men and attributes this to the extensive margin becoming more important as they approach retirement.

In a series of papers, Keane & Wolpin (1997, 2000, 2001) study 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. Keane & Wolpin (2000) report a simulation in which the rental price of skill for blacks is increased to the same level as whites, approximately a 6% increase. The percent of blacks who are employed at age 30 increases from 83.8% to 90.7%. Thus, the implied elasticity with respect to a permanent (uncompensated) wage increase for black males at age 30 is approximately 8.2/6, or 1.4.

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 survey of the micro labor supply literature by Blundell & 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 intertemporal) labor supply elasticities are quite large (i.e., in the range of 1–2). In this review, we describe a relatively new literature 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 focuses on the micro perspective. In the basic life-cycle labor supply model of MaCurdy (1981), the only source of dynamics is borrowing/saving. Several authors have considered extensions of this model to include other potentially important sources of dynamics, such as human capital and credit constraints. Some work has also allowed for corner solutions in labor supply. This work has shown that if the true model (or data-generating process) contains such mechanisms, but the data are 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 of aggregation in the presence of the extensive margin and worker heterogeneity. This literature has shown that small (intensive margin) elasticities at the individual level are consistent with large elasticities at the aggregate level.

These two literatures share one key point in common. In the basic life-cycle model of MaCurdy (1981), there is a direct link between individual-level preference parameters and labor supply elasticities at the aggregate level. All the extensions to the basic model that we describe above break that direct link. This is not to say that individual preference parameters no longer matter. But, in general, labor supply elasticities also depend on many other aspects of the economic environment: the wage process, the functioning of credit markets, the technology of job search/hours adjustment, the production technology (i.e., 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 generally require structural modeling of the complete economic environment. Given the difficulty of such exercises, it is tempting to resort to an experimental approach of cataloging responses to observed tax changes. But in our view, this would be misguided. As shown above, 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. 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 described above can credibly support a view that at the macro level, compensated elasticities are in the range of 0.50–1.0 and intertemporal elasticities are in the range of 1–2, as 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, multiple mechanisms 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 & 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 for policy evaluation. As 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.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Raj Chetty, Christian Haefke, and the reviewing editor for comments. Rogerson thanks both the NSF and the Korea Science Foundation (WCU-R33-10005) for financial support. Keane thanks the Australian Research Council (grants FF0561843 and FL110100247).

LITERATURE CITED

Aiyagari R. 1994. Uninsured idiosyncratic risk and aggregate saving. *Q. J. Econ.* 109:659–84

Alonso-Ortiz J, Rogerson R. 2010. Taxes, transfers and employment in an incomplete markets model. *J. Monet. Econ.* 57:949–58

Altonji J. 1986. Intertemporal substitution in labor supply: evidence from micro data. *J. Polit. Econ.* 94:S176–215

An S, Chang Y, Kim S. 2009. Can a representative agent model represent a heterogeneous agent economy? *Am. Econ. J. Macroecon.* 1(2):29–54

Benhabib J, Rogerson R, Wright R. 1991. Homework in macroeconomics: household production and aggregate fluctuations. *J. Polit. Econ.* 99:1166–87

Blomquist S. 1983. The effect of income taxation on the labor supply of married men in Sweden. *J. Public Econ.* 22:169–97

Blundell R, Costas Dias M, Meghir C, Shaw J. 2013. *Female labour supply, human capital and welfare reform.* Unpublished manuscript, Yale Univ., New Haven, CT

Blundell RW, Duncan A, Meghir C. 1998. Estimating labor supply responses using tax policy reforms. *Econometrica* 66:827–61

Blundell RW, MaCurdy TH. 1999. Labor supply: a review of alternative approaches. In *Handbook of Labor Economics*, Vol. 3, ed. O Ashenfelter, D Card, pp. 1559–695. Amsterdam: North-Holland

Browning M, Deaton A, Irish M. 1985. A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica* 53:503–44

Browning M, Hansen L, Heckman J. 1999. Micro data analysis and general equilibrium models. In *Handbook of Macroeconomics*, Vol. 1, ed. JB Taylor, M Woodford, pp. 543–633. Amsterdam: North-Holland

Burtless G, Hausman J. 1978. The effects of taxation on labor supply: evaluating the Gary Income Maintenance Experiment. *J. Polit. Econ.* 86:1103–30

Chang Y, Kim S-B. 2006. From individual to aggregate labor supply: a quantitative analysis based on a heterogeneous agent macroeconomy. *Int. Econ. Rev.* 47:1–27

Chang Y, Kim S-B. 2007. Heterogeneity and aggregation in the labor market: implications for aggregate preference shifts. *Am. Econ. Rev.* 97:1939–56

Chang Y, Kim S-B, Kwan K, Rogerson R. 2011. Interpreting labor supply regressions in a model of full- and part-time work. *Am. Econ. Rev.* 101:476–81

Chang Y, Kim S-B, Kwan K, Rogerson R. 2014. *Individual and aggregate labor supply in a heterogeneous agent economy with intensive and extensive margins*. Unpublished manuscript, Princeton Univ., Princeton, NJ

Chetty R. 2012. Bounds on elasticities with optimization frictions: a synthesis of micro and macro evidence on labor supply. *Econometrica* 80:969–1018

Chetty R, Guren A, Manoli D, Weber A. 2013. Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities. In *NBER Macroeconomics Annual 2012*, ed. D Acemoglu, J Parker, M Woodford, pp. 1–56. Chicago: Univ. Chicago Press

Cho J-O. 1995. Ex-post heterogeneity and the business cycle. *J. Econ. Dyn. Control* 19:533–51

Cogan J. 1981. Fixed costs and labor supply. *Econometrica* 49:945–64

Deaton A. 1991. Saving and liquidity constraints. *Econometrica* 59:1221–48

Díaz-Giménez J, Quadrini V, Ríos-Rull V. 1997. Dimensions of inequality: facts on the US distributions of earnings, income and wealth. *Q. Rev. Fed. Reserve Bank Minneap.* 21(2):3–21

Domeij D, Floden M. 2006. The labor supply elasticity and borrowing constraints: why estimates are biased. *Rev. Econ. Dyn.* 9:242–62

Eckstein Z, Wolpin KI. 1989. Dynamic labor force participation of married women and endogenous wage growth. *Rev. Econ. Stud.* 56:375–90

Erosa A, Fuster L, Kambourov G. 2014. *Towards a micro-founded theory of aggregate labor supply*. Unpublished manuscript, Univ. Toronto

Fang H, Keane M. 2004. Assessing the impact of welfare reform on single mothers. *Brookings Pap. Econ. Act.* 35(1):1–116

Floden M, Linde J. 2001. Idiosyncratic risk in the United States and Sweden: Is there a role for government insurance? *Rev. Econ. Dyn.* 4:406–37

Francesconi M. 2002. A joint dynamic model of fertility and work of married women. *J. Labor Econ.* 20:336–80

French E. 2005. The effects of health, wealth and wages on labour supply and retirement behaviour. *Rev. Econ. Stud.* 72:395–427

Hansen G. 1985. Indivisible labor and the business cycle. *J. Monet. Econ.* 16:309–37

Hausman J. 1980. The effect of wages, taxes and fixed costs on women's labor force participation. *J. Public Econ.* 14:161–94

Hausman J. 1985. Taxes and labor supply. In *Handbook of Public Economics*, Vol. 1, ed. K Arrow, M Intriligator, pp. 213–64. Amsterdam: North-Holland

Heckman J. 1974. Life cycle consumption and labor supply: an explanation of the relationship between income and consumption over the life cycle. *Am. Econ. Rev.* 64:188–94

Heckman J. 1976. Estimates of a human capital production function embedded in a life-cycle model of labor supply. In *Household Production and Consumption*, ed. N Terlecky, pp. 227–64. New York: Columbia Univ. Press

Heckman J. 1984. Comments on the Ashenfelter and Kydland papers. *Carnegie-Rochester Conf. Ser. Public Policy* 21:209–24

Heckman J, MaCurdy T. 1980. A life-cycle model of female labor supply. *Rev. Econ. Stud.* 47:47–74

Heckman J, MaCurdy T. 1982. Corrigendum on a life cycle model of female labor supply. *Rev. Econ. Stud.* 49:659–60

Hubbard R, Skinner J, Zeldes S. 1995. Precautionary saving and social insurance. *J. Polit. Econ.* 103:360–99

Huggett M. 1993. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *J. Econ. Dyn. Control* 17:953–69

Imai S, Keane M. 2004. Intertemporal labor supply and human capital accumulation. *Int. Econ. Rev.* 45:601–42

Jappelli T. 1990. Who is credit constrained in the US economy? *Q. J. Econ.* 105:219–34

Keane M. 1995. A new idea for welfare reform. *Fed. Reserve Bank Minneap. Q. Rev.* 19:2–28

Keane M. 2009. *Income taxation in a life cycle model with human capital*. Work. Pap., Univ. New South Wales, Sydney

Keane M. 2011. Labor supply and taxes: a survey. *J. Econ. Lit.* 49:961–1075

Keane M, Moffitt R. 1998. A structural model of multiple welfare program participation and labor supply. *Int. Econ. Rev.* 39:553–89

Keane M, Rogerson R. 2012. Micro and macro labor supply elasticities: a reassessment of conventional wisdom. *J. Econ. Lit.* 50:464–76

Keane M, Runkle D. 1992. On the estimation of panel data models with serial correlation when instruments are predetermined but not strictly exogenous. *J. Bus. Econ. Stat.* 10:1–9

Keane M, Wolpin K. 1997. The career decisions of young men. *J. Polit. Econ.* 105:473–522

Keane M, Wolpin K. 2000. Eliminating race differences in schooling attainment and labor market success. *J. Labor Econ.* 10:614–52

Keane M, Wolpin K. 2001. The effect of parental transfers and borrowing constraints on educational attainment. *Int. Econ. Rev.* 42:1051–103

Keane M, Wolpin K. 2007. Exploring the usefulness of a non-random holdout sample for model validation. *Int. Econ. Rev.* 48:1351–78

Keane M, Wolpin K. 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. *Int. Econ. Rev.* 51:851–92
Killingsworth M, Heckman J. 1986. Female labor supply: a survey. In *Handbook of Labor Economics*, Vol. 1, ed. O Ashenfelter, R Layard, pp. 103–204. Amsterdam: North-Holland
Kimmel J, Kniesner TJ. 1998. New evidence on labor supply: employment vs. hours elasticities by sex and marital status. *J. Monet. Econ.* 42:289–301
Kitao S, Ljungqvist L, Sargent T. 2009. *A life cycle model of transatlantic employment experiences*. Unpublished manuscript, New York Univ.
Krusell P, Mukoyama T, Rogerson R, Sahin A. 2010. Aggregate labor market outcomes: the roles of choice and chance. *Quant. Econ.* 1:97–128
Krusell P, Mukoyama T, Rogerson R, Sahin A. 2011. A three state model of worker flows in general equilibrium. *J. Econ. Theory* 146:1107–33
Ljungqvist L, Sargent T. 2007. Do taxes explain European employment? Indivisible labor, human capital, lotteries and savings. In *NBER Macroeconomics Annual 2006*, ed. D Acemoglu, K Rogoff, M Woodford, pp. 181–246. Cambridge, MA: MIT Press
Ljungqvist L, Sargent T. 2008. Taxes, benefits, and careers: complete versus incomplete markets. *J. Monet. Econ.* 55:98–125
Ljungqvist L, Sargent T. 2014. Career length: effect of curvature of earnings profile, earnings shocks and social security. *Rev. Econ. Dyn.* 17:1–20
Low H. 2005. Self-insurance in a life cycle model of labor supply and savings. *Rev. Econ. Dyn.* 8:945–75
Lucas R, Rapping L. 1969. Real wages, employment and inflation. *J. Polit. Econ.* 77:721–54
MaCurdy T. 1981. An empirical model of labor supply in a life-cycle setting. *J. Polit. Econ.* 89:1059–85
MaCurdy T. 1983. A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty. *Int. Econ. Rev.* 24:265–89
McDaniel C. 2011. Forces shaping hours of work in the OECD, 1960–2004. *Am. Econ. J. Macroecon.* 3(4):27–52
Meghir C, Phillips D. 2010. Labour supply and taxes. In *Dimensions of Tax Design: The Mirrlees Review*, ed. T Besley, R Blundell, M Gammie, J Poterba, pp. 202–74. New York: Oxford Univ. Press
Moffitt R. 1983. An economic model of welfare stigma. *Am. Econ. Rev.* 73:1023–35
Mulligan CB. 2001. Aggregate implications of indivisible labor. *B.E. J. Macroecon.* 1(1):4
Ohanian L, Raffo A, Rogerson R. 2008. Long-term changes in labor supply and taxes: evidence from OECD countries, 1956–2004. *J. Monet. Econ.* 55:1353–62
Pencavel J. 1986. Labor supply of men: a survey. In *Handbook of Labor Economics*, Vol. 1, ed. O Ashenfelter, R Layard, pp. 3–102. Amsterdam: North-Holland
Prescott EC. 2004. Why do Americans work so much more than Europeans? *Fed. Reserve Bank Minneap. Q. Rev.* 28(1):2–13
Prescott EC. 2006. Nobel Lecture: the transformation of macroeconomic policy and research. *J. Polit. Econ.* 114:203–35
Prescott E, Rogerson R, Wallenius J. 2009. Lifetime aggregate labor supply with endogenous workweek length. *Rev. Econ. Dyn.* 12:23–36
Rogerson R. 1988. Indivisible labor, lotteries and equilibrium. *J. Monet. Econ.* 21:3–16
Rogerson R. 2008. Structural transformation and the deterioration of European labor market outcomes. *J. Polit. Econ.* 116:235–59
Rogerson R, Wallenius J. 2009. Micro and macro elasticities in a life cycle model with taxes. *J. Econ. Theory* 144:2277–92
Shaw K. 1989. Life-cycle labor supply with human capital accumulation. *Int. Econ. Rev.* 30:431–56
van der Klaauw W. 1996. Female labour supply and marital status decisions: a life cycle model. *Rev. Econ. Stud.* 63:199–235
Wallenius J. 2011. Human capital and the intertemporal elasticity of substitution of labor: How large is the bias? *Rev. Econ. Dyn.* 14:577–91
Wallenius J. 2013. Social security and cross-country differences in hours: a general equilibrium analysis. *J. Econ. Dyn. Control* 37:2466–82
Zeldes S. 1989. Consumption and liquidity constraints: an empirical investigation. *J. Polit. Econ.* 97:305–46