Volume Title: NBER Macroeconomics Annual 1998, volume 13

Volume Author/Editor: Ben S. Bernanke and Julio Rotemberg, editors

Volume Publisher: MIT Press

Volume ISBN: 0-262-52271-3

Volume URL: http://www.nber.org/books/bern99-1

Publication Date: January 1999

Chapter Title: Substition over Time: Another Look at Life-Cycle Labor Supply

Chapter Author: Casey B. Mulligan

Chapter URL: http://www.nber.org/chapters/c11246

Chapter pages in book: (p. 75 - 152)

Casey B. Mulligan
UNIVERSITY OF CHICAGO

Substitution over Time: Another Look at Life-Cycle Labor Supply

1. *Introduction*

Today macroeconomics relies almost exclusively on intertemporal models, models which are also quite common in other areas of economic inquiry. One important question that must be answered in the building of a dynamic model is "What are the possibilities for substitution over time?" This paper is about one potential margin of substitution, the substitution of work over time.

An analysis of the substitution of work over time dates at least back to Hicks (1939, Chapters XVI, XVII), and since Lucas and Rapping (1969) this margin of substitution has been included in a substantial number of macro models. The presence of this margin affects predictions for labor-market and output fluctuations in response to aggregate productivity shocks (Hall, 1980; Hansen, 1985), to changes in the timing of taxes (Judd, 1987), to changes in the timing of government spending (Aiyagari, Christiano, and Eichenbaum, 1992; Barro, 1981, 1987; Baxter and King, 1993), and to monetary shocks (Lucas, 1972), as well as many other questions of substantial interest. But since a series of studies by Altonji (1986), MaCurdy (1981), and Abowd and Card (1989) and analy-

Casey B. Mulligan is an assistant professor of economics at the University of Chicago, a Faculty Research Fellow of the NBER, and a John M. Olin Faculty Fellow. Mulligan gratefully acknowledges the careful research assistance of Song Han and comments by Gary Becker, Ben Bernanke, John Cawley, Bob Hall, Dan Hamermesh, Jim Heckman, Steve Levitt, Bob Lucas, Costas Meghir, Jacob Mincer, Derek Neal, Steve Pischke, Julio Rotemberg, Xavier Sala-i-Martin, University of Chicago Economics 203 and 337 students, and seminar participants at the University of Chicago, CEPR, MIT, NBER, Northwestern, the University of Rochester, and Vanderbilt. Jim Heckman and Lance Lochner generously provided their life-cycle estimates of worker-financed training time. The University of Chicago Center on Aging grant P20 AG12857 from the National Institute on Aging provided financial support.

ses of tax experiments [some of which are surveyed by Hausman (1985) and Pencavel (1986)], it has been widely believed that the substitution of work over time is a quantitatively unimportant margin. This belief and its origins can be seen in Pencavel's (1986) and Card's (1994) surveys, in the comments by Ashenfelter (1984), Mankiw (1989), and Plosser (1989), and in the calibration of macro models by Auerbach and Kotlikoff (1987, p. 50), Judd (1987), and many others exploring the consequences of some interesting and important policy changes.

This paper revisits the *life-cycle* evidence—evidence on low-frequency changes in wages and time worked with age. My goal here is not to reconcile the variety of life-cycle and non-life-cycle studies that are available [I do attempt this in another paper, Mulligan (1995)]. Nor is my goal to argue that the substitution of work over time is the single missing piece for every macroeconomic puzzle. Instead, I revisit life-cycle evidence to show that it is consistent with quantitatively important substitution of labor over time. Although I do not think that life-cycle evidence is necessarily the best evidence for getting at this issue, it is true that the widely cited studies based on the PSID and other micro data sets are basically life-cycle studies. I then show that the usual cohort data might be reasonably modified in several ways, including the aggregation of micro measures of labor supply, taking account of time spent training on the job, using different techniques for measuring work time, measuring marginal tax rates, and including older workers.

Although the usual cohort data are consistent with a lot of substitution over time, there are some problems with their construction. There is no obvious way to infer the wages of those who are not working at various points in the life cycle. Nor is it particularly plausible that tastes, health, and other relevant variables are independent of age or even uncorrelated with wages. Nor is it particularly obvious exactly what procedure should be used to obtain the "right" estimate from the usual cohort samples. These problems are enough to make one eagerly search elsewhere for anticipated wage changes and associated changes in work. My search led me to another life-cycle event that has not yet been studied: the termination of aid to families with dependent children (AFDC) at the 18th birthday of a family's youngest child.[1] Section 4 explains how this study can alleviate some of the apparent problems with a synthetic-cohort sample, that it is consistent with substantial substitution over time, and—perhaps surprisingly—that its estimates are similar to the synthetic-cohort estimates.

1. I have also been led to look at the anticipated wage changes associated with non-life-cycle events, including seasons, wars, and agricultural shocks (Mulligan, 1995).

2. *A Life-Cycle Model of Labor Supply*

2.1 AN OVERVIEW OF THE INTERTEMPORAL SUBSTITUTION HYPOTHESIS

The intertemporal substitution hypothesis (ISH) can be simply stated: workers intertemporally reallocate their work in response to changes in the relevant relative price. A statistical model that captures this idea is

$$E_t\left[\ln\frac{n^i_{t+1}}{n^i_t}\right] = -\sigma r_t + \sigma E_t\left[\ln\frac{w^i_{t+1}}{w^i_t}\right] + \epsilon^i_t \qquad (1)$$

where n^i_t is the labor supply of worker i at date t, w^i_t is worker i's date-t after-tax market value of time, and r_t is the real rate of return to savings between periods t and $t+1$. The term ϵ_t includes preference parameters and, in some models, a precautionary motive for working. $\sigma > 0$ is the intertemporal elasticity of substitution (IES). In many environments, labor-supply decisions are made sequentially in time rather than at the "beginning of time," so the law of demand more appropriately applies to *plans* for labor supply as a function of anticipated wages. E_t denotes date-t expectations of future variables and is therefore included in equation (1).

Notice that the model does not necessarily predict that workers will work the hardest when the conventionally measured "real wage" is the highest. There are two relevant forces at work. First, the relative price of leisure in any two periods depends not only on the ratio of the wages but also on an interest rate. Second, through the taste term ϵ_t, the model allows for impatience, prudence, or other reasons that workers may have different preferences for current and future leisure. Consider, for example, a perfectly flat time profile for wages. A worker will choose more leisure in the later periods if the interest rate is high enough, or more leisure in the earlier periods if he is impatient enough. Without data on the interest rate and preferences, it is necessary to estimate the trend component of the intertemporal labor allocation together with the responsiveness of work effort to incentives.

Beneath any notational complexities, the econometrics of estimating the elasticity σ is also quite straightforward: identify situations with different anticipated rates of wage growth, and measure the associated differences in anticipated labor-supply growth. The implementation of this can be quite challenging (e.g., how is anticipated wage growth measured?), but the thought experiment is simple enough. I (and many others in the empirical literature) do not intend to say or do anything more complicated than this, but a more complicated model is needed to be clear about the measurement of the variables of interest and interpretation of the parameter σ. My life-cycle model includes several complications necessary for understanding the data: (1) uncertainty about future wages, (2) various measurement errors in hours and wages, (3) discrete choices about labor-force participation at various points in time, (4) time aggregation in the measurement of labor supply, and (5) potentially nonlinear labor income tax schedules.

2.2 INDIVISIBLE LABOR

A person's lifetime includes many potential *work sessions*, equally spaced in time. For our purposes, it may be useful to think of a potential work session as a day or week, although a month is in some ways a more convenient definition for the AFDC application in Section 4.3. If work is to occur during a potential session, it must occur for exactly $\bar{n}$ units of time. This indivisibility of labor might be interpreted as the optimal bunching of labor in continuous time in the presence of fixed commuting costs and high-frequency fatigue effects (see Mulligan, 1998), in which case $\bar{n}$ is a function of the magnitude of the fixed cost and of the form of the fatigue effects, but not the other variables in the model such as wages, taxes, or tastes for leisure. w denotes the average product of labor for the session, so that $w\bar{n}$ is the total amount produced by a worker who chooses to work the session. w and $\bar{n}$ may vary over time and across states of nature.

2.3 HUMAN-CAPITAL ACCUMULATION

Two things can happen during a work session: goods production or the production of one's own human capital. When goods production occurs, the worker is paid $w\bar{n}$. When human-capital accumulation occurs, the worker may or may not be paid, depending on whether the firm is financing the training or not. If the firm does not finance the training, the worker does not have any earnings for that session but still values the training for the future earnings it produces. In this case, I let $w\bar{n}$ denote that valuation.[2]

In Section 3, I use two measures of time spent producing human capital. The first is time spent searching for a job. The second is time at work spent learning new things.

2.4 TIME AGGREGATION

It will be assumed that the econometrician observes only time-aggregated measures of the relevant variables, so it is convenient to use

2. Human-capital accumulation can be valued more than $w\bar{n}$ without changing the interpretation of my results, although see my discussion of time aggregation.

two indices t and k to identify a potential work session. t indexes the time-aggregated periods, or *time intervals,* and varies for 0 to T during a consumer's lifetime. k indexes the potential *sessions* within any particular time interval and varies from 1 to K for each t. Thus there are a total of $K(T+1)$ potential sessions in a consumer's lifetime.

2.5 TAX AND BENEFIT SYSTEM

During any potential session, a consumer receives government transfer payments in the amount $b_{t,k}$ (the consumer is a net taxpayer in the case $b < 0$). Benefits are determined according to the formula

$$b_{t,k} = \max\left\{\underline{b}_{t,k}, \bar{b}_{t,k} - R_{t,k}\max\left[w_{t,k}n_{t,k}l_{t,k} - d_{t,k}, 0\right]\right\}, \tag{2}$$

where $w_{t,k}$ is the average product of time during the kth session of time interval t (time which may be spent either in goods or human-capital production), $n_{t,k}$ is time worked, $l_{t,k}$ is an indicator variable for goods production or firm-financed human-capital production, $l_{t,k}w_{t,k}n_{t,k}$ is labor earnings, $d_{t,k}$ are deductions from earnings and *earnings disregards,* $R_{t,k}$ is the *benefit reduction rate* or *marginal labor income tax rate,* $\bar{b}_{t,k}$ is the maximum benefit available (and in some applications may be a function of family composition, asset holdings, and other variables), and $\underline{b}_{t,k}$ is the minimum benefit available. Benefits are reduced $R_{t,k}$ dollars for every dollar of net earnings, where net earnings are computed as gross earnings net of deductions.

Notice that earnings and deductions are not aggregated across sessions for the purposes of computing taxes and benefits. I note in the text where this assumption may be unrealistic and of some consequence for the results.

2.6 UNCERTAINTY, BUDGET CONSTRAINTS, AND UTILITY FUNCTIONS

I index a realization of a consumer's life history by ω. For each possible realization $\omega \in \Omega$, consumer choices of stochastic processes for consumption $c_{t,k}(\omega)$, work $n_{t,k}(\omega)$, and tax deductions $d_{t,k}(\omega)$ must satisfy a present-value budget constraint:

$$\begin{aligned}\sum_{t=0}^{T}\sum_{k=1}^{K} e^{-\rho(tK+k)}Q_{t,k}(\omega)\,[c_{t,k}(\omega) - w_{t,k}(\omega)n_{t,k}(\omega) - b_{t,k}(\omega) \\ + f_{t,k}(d_{t,k}(\omega),\omega) + H_{t,k}(\omega)w_{t,k}(\omega)\bar{n}_{t,k}(\omega)] = A_0, \end{aligned}\tag{3}$$

$$\ln Q_{t,k}(\omega) \equiv -\sum_{s=0}^{t-1}\sum_{l=1}^{K}[r_{s,l}(\omega) - \rho] - \sum_{l=1}^{k}[r_{t,l}(\omega) - \rho],$$

where $r_{t,k}$ is the *ex post* real return on a one-period bond purchased at date $t,k-1$, and $H_{t,k}$ is an indicator variable for purchases of self-financed human capital. In addition to (3), we have the constraints (2) and $n_{t,k}(\omega) \in \{0,\bar{n}_{t,k}(\omega)\}$ for all $t=0, \ldots ,T$, $k=1, \ldots ,K$, and $\omega \in \Omega$.

Deductions are costly. Deductions in the amount d cost $f_{t,k}(d,\omega)$. For each t,k,ω, the deduction cost function is nondecreasing and nonconcave in the amount deducted. There is no cost if no deductions are taken.

The revelation of information over time is modeled with the filtered probability space $(\Omega,\mathcal{F},F,\pi)$. Each state of nature ω has unconditional probability $\pi(\omega)$. The filtration $F=\{\mathcal{F}_{0,1}, \mathcal{F}_{0,2}, \ldots ,\mathcal{F}_{T,K}\}$ on Ω is assumed to be increasing, and the stochastic processes $w_{t,k}$, $\bar{n}_{t,k}$, $\gamma_{t,k}$, $r_{t,k}$, $\bar{b}_{t,k}$, $R_{t,k}$, and $\underline{b}_{t,k}$, are adapted to it.[3]

Consumption and work are assumed to evolve as if workers chose functions $c_{t,k}(\omega)$, $n_{t,k}(\omega)$, and $d_{t,k}(\omega)$ adapted to the filtration F with the objective of maximizing the expected value of an intertemporally and intratemporally separable utility function (4) subject to the constraints (2), (3), and $n_{t,k}(\omega) \in \{0,\bar{n}_{t,k}(\omega)\}$:

$$\begin{gathered}\sum_{\omega\in\Omega} \pi(\omega)\sum_{t=0}^{T}\sum_{k=1}^{K}e^{-\rho(tK+k)}\left[u\Big(c_{t,k}(\omega)\Big) - \gamma_{t,k}(\omega)v\Big(n_{t,k}(\omega)\Big)\right], \\ \rho > 0, \qquad v(0) \equiv 0, \qquad v_{t,k}(\omega) \equiv v\Big(\bar{n}_{t,k}(\omega)\Big) > 0.\end{gathered} \tag{4}$$

Notice that I do not treat $H_{t,k}$ and $I_{t,k}$ as choice variables and do not say exactly how sessions devoted to human-capital production translate into higher future wages, but the first-order conditions of the problem described above are among those of the larger problem that include optimal human-capital accumulation over the life cycle.[4]

2.7 WHEN TO WORK

When describing the decision to work, it is useful to define $\tau_{t,k}(\omega)$ as the implicit tax rate on work in session k of time interval t:

$$\begin{gathered}\tau_{t,k}(\omega) \equiv \min\left\{\frac{\bar{b}_{t,k}(\omega)-\underline{b}_{t,k}(\omega)}{w_{t,k}(\omega)\bar{n}_{t,k}(\omega)}, R_{t,k}(\omega) - \frac{R_{t,k}(\omega)d^{*}_{t,k}(\omega)-f_{t,k}\big(d^{*}_{t,k}(\omega),\omega\big)}{w_{t,k}(\omega)\bar{n}_{t,k}(\omega)}\right\} \le R_{t,k}(\omega), \\ d^{*}_{t,k}(\omega) \equiv \operatorname*{arg\,min}_{d\in[0,w_{t,k}(\omega)\bar{n}_{t,k}(\omega)]}\Big\{f_{t,k}(d,\omega) - R_{t,k}(\omega)d\Big\}.\end{gathered} \tag{5}$$

3. The random function $f_{t,k}$ should be measurable with respect to $\mathcal{F}_{t,k}$.

4. See Ghez and Becker (1975). The larger problem with $H_{t,k}$ and $I_{t,k}$ as choice variables has the constraint $H_{t,k}\bar{n}_{t,k} = I_{t,k}n_{t,k}$ for all t,k,ω.

If $\bar{b}_{t,k}$ is large enough (in what follows, I suppress the index ω), the implicit tax rate is related to the *benefit reduction rate* $R_{t,k}$. However, when positive deductions are optimal and the deduction cost function is strictly concave, the indivisibility of labor means that the implicit tax rate is strictly less than the benefit reduction rate. If $\bar{b}_{t,k}$ is small enough, $\tau_{t,k}$ is unrelated to $R_{t,k}$.

The first-order condition determining where $n_{t,k}$ equals 0 or $\bar{n}_{t,k}$ is

$$\gamma_{t,k}\, v_{t,k} \gtrless Q_{t,k}\, w_{t,k}\, \bar{n}_{t,k}\, (1 - \tau_{t,k})\, E_{t,k}\lambda_{t,k}, \tag{6}$$

where $E_{t,k}$ denotes date-t,k expectations and $\lambda_{t,k}$ is the lifetime "marginal" utility of wealth.[5] The expression (6) simply says that people work during those periods when the benefit exceeds the cost. The benefit of working is the discounted after-tax earnings (or the value of self-financed human-capital production) for session k of interval t, given by $Q_{t,k}w_{t,k}\bar{n}_{t,k}(1-\tau_{t,k})$, times the expected marginal utility of wealth, $E_{t,k}\lambda_{t,k}$. The cost is the disutility of work, $\gamma_{t,k}v_{t,k}$. Comparative static changes in $Q_{t,k}w_{t,k}(1-\tau_{t,k})$ can be interpreted as generating a substitution effect, while changes in $E_{t,k}\lambda_{t,k}$ can be interpreted as generating a wealth effect.

2.8 CONSUMPTION INSURANCE?

The problem described above presumes that insurance against surprises to wages, taxes, and interest rates is unavailable.[6] It is straightforward to allow for perfect consumption insurance by collapsing the series of budget constraints (3) into a single budget constraint that equates the expected present value of expenditures to the expected present value of resources. In this case, the lifetime marginal utility of wealth, $\lambda_{t,k}$, is no longer a random variable with respect to $\mathcal{F}_{t,k}$ (that is, $E_{t,k}\lambda_{t,k} = \lambda_{t,k}$). Otherwise the conditions for working session k of interval t is the same as (6).

2.9 INFORMATION LOST FROM TIME AGGREGATION

Given time-aggregated lifetime data on all of the exogenous variables of the model, we cannot make sharp predictions about time-aggregated labor supply without answering three questions:

5. Because of its discreteness, the labor-supply decision at any date and state has a discrete effect on expected remaining lifetime discounted utility of consumption, and $\lambda_{t,k}$ is the size of this effect per dollar of date-t,k after-tax earnings $w_{t,k}\bar{n}_{t,k}(1-\tau_{t,k})$. The value of $\lambda_{t,k}$ approaches the "marginal" utility as $w_{t,k}\bar{n}_{t,k}(1-\tau_{t,k})$ goes to zero. Mulligan (1998) also shows that $\lambda_{t,k}$ is literally a marginal utility when particular lotteries and "taste insurance" contracts are introduced as choice variables.
6. Although one of Mulligan's (1998) interpretations of equation (6) presumes that "taste insurance" or some other mechanism is available to compensate those who are unlucky enough to especially dislike work when wages are high.

1. How is information about future wages and interest rates revealed during a time interval?
2. What is a worker's market value of time at each potential session, including those sessions he did not work?
3. What is a worker's nonmarket value of time at each potential session?

I make three assumptions about this:

(A1) No information is revealed within time intervals: $\mathcal{F}_{t,k} = \mathcal{F}_{t,1}$ for all $k = 1, \ldots, K$.

(A2) The market value of time and the indivisibility of labor do not vary within time intervals: $Q_{t,k}w_{t,k}\bar{n}_{t,k}(1-\tau_{t,k}) = Q_{t,1}w_{t,1}\bar{n}_{t,1}(1-\tau_{t,1})$ and $\bar{n}_{t,k} = \bar{n}_{t,1}$ for all $k = 1, \ldots, K$

(A3) $\gamma_{t,k} = \gamma_t g_{t,k}$, where $\sigma \ln g_{t,k}$, $k = 1, \ldots, K$, are drawn according to the distribution function G with unbounded support. $\sigma > 0$ is a constant. $E[\sigma \ln g_{t,k}] = 0$ and $E[(\sigma \ln g_{t,k})^2] = 1$. Without loss of generality, the K draws are assumed to be independent.

Assumptions (A1) and (A2), although often only implicit, are extremely common in the labor-supply literature. (A3) produces increasing marginal disutility of work within a time interval, an assumption which appears in one form or another in the literature. To see mechanically how (A3) produces increasing marginal disutility, notice that, given (A1) and (A2), workers that work at all during a time interval will work during the low disutility sessions. Since sessions differ in their marginal disutility, workers require a higher wage to work a larger fraction of the time interval.[7] I use (A1)–(A3) to derive some empirical specifications, but comment further on their relevance as I discuss the empirical results.

In accordance with assumption (A2), I suppress the subscript k except where necessary for clarity.

2.10 CONSUMER HETEROGENEITY AND THE INDIVIDUAL LABOR-SUPPLY EQUATION

Consumers may differ in their realizations of the exogenous variables $w_{t,k}$ and $\gamma_{t,k}$. They are the same regarding the stochastic processes generating these realizations, the utility functions $u(c)$ and $v(n)$, the parameter σ, and the way in which information arrives (including the function G defined above).

Consider a group of consumers who are identical in terms of the two

7. Lucas (1970, p. 25) has exactly the specification (A3), except that he has a continuum of potential work sessions per time interval (as compared to my K).

random variables $\gamma_t v_t$ and $Q_t w_t \bar{n}_t(1-\tau_t)E_t\lambda_t$. We compute the group average N_t of each individual's time worked during the time interval t, including those who did not work at all during the interval:

$$G^{-1}\left(\frac{N_t}{K\bar{n}_t}\right) = \sigma \ln w_t(1-\tau_t) + \sigma \ln \bar{n}_t + \sigma \ln \frac{Q_t E_t \lambda_t}{\gamma_t v_t}. \tag{7}$$

This is a nonlinear version of MaCurdy's (1981) "λ-constant labor supply function," which may be surprising given MaCurdy's apparently different description of the labor-supply problem. In fact, the only difference is in the measurement of labor supply.

For a group of consumers who, for some integer $s > 0$, have the four random variables $\gamma_t v_t$, $Q_t w_t \bar{n}_t(1-\tau_t)E_t\lambda_t$, $E_t[\gamma_{t+s}v_{t+s}]$, $E_t[Q_{t+s}w_t\bar{n}_{t+s}(1-\tau_{t+s})\lambda_{t+s}]$ in common, we have a version of equation (1):

$$\begin{aligned} &E_t\left[G^{-1}\left(\frac{N_{t+s}}{K\bar{n}_{t+s}}\right) - G^{-1}\left(\frac{N_t}{K\bar{n}_t}\right)\right] \\ &= \sigma K \sum_{s'=1}^{s}(E_t r_{t+s'} - \rho) + \sigma E_t \ln \frac{w_{t+s}(1-\tau_{t+s})}{w_t(1-\tau_t)} + \sigma E_t \ln \frac{\bar{n}_{t+s}\gamma_t v_t}{\bar{n}_t \gamma_{t+s} v_{t+s}} \\ &\quad + \sigma(E_t \ln E_{t+s}\lambda_{t+s} - \ln E_t E_{t+s}\lambda_{t+s}). \end{aligned} \tag{8}$$

The only difference between an individual's labor supply plans and those of the group is the sampling error due to the fact that a single individual only samples K times from the distribution G during a time interval. The probability that a particular individual does not work at all during a particular time interval is

$$\left[1 - G\left(\sigma \ln \frac{w_t \bar{n}_t Q_t(1-\tau_t)E_t\lambda_t}{\gamma_t v_t}\right)\right]^K.$$

Thus extended intervals of participation and nonparticipation during a person's lifetime are evidence of either (a) large and persistent changes in tastes or the value of time or (b) a large willingness to substitute over time, σ.

The responsiveness of labor supply to wages depends on the level of labor supply in the model (8). Typically there will be little response of N when N is near 0 or $K\bar{n}$. This is even true in the special case that $g^{\sigma}_{t,k}$ is distributed uniformly on $[0,e]$ and the equation (1) obtains—a special case which gets a lot of attention in the literature—once the possibilities of corner solutions for N_t and N_{t+s} are taken into account.[8] The bulk of my analysis follows Altonji (1986), Ghez and Becker (1975), MaCurdy (1981), and others by focusing on this special case and ignoring possible corner solutions. My Section 4 returns to the more general model (8) and includes some analysis of how the responsiveness of labor supply might vary with its level.

The derivation of equation (8) reveals several points that are quite relevant for the empirical applications in this paper:

1. The market value of time, w_t, for the time interval t can, for those who have some earnings during the interval, be measured as $y_t/[N_t(1-h_t)]$, where h_t is the fraction of sessions worked that were devoted to self-financed human-capital production, and y_t is the total pretax labor earnings for the time interval.
2. σ measures the responsiveness of *anticipated* labor-supply changes to anticipated wage changes, not *ex post* labor changes to *ex post* wage changes.
3. The implicit tax on work has the *benefit reduction rate* or *marginal labor income tax rate* $R_{t,k}$ as its upper bound.
4. The derivation shows how time aggregation is related to the measurement of labor supply, wages, tax rates, and other variables.
5. The derivation shows how measured employment and hours are related.
6. Those who are not working during any particular work session are not a random sample of the population, but are those for whom work yields a greater disutility.
7. The derivation shows the effect of consumption insurance on comovements of wages and labor supply.

2.11 MACRO "EXPERIMENTS" TO BE CALIBRATED FROM LIFE-CYCLE DATA

Before reviewing empirical studies and proposing ways to improve them, something must be said about why life-cycle substitution is of interest for macroeconomics. One item of substantial interest is the response of *aggregate* labor supply to aggregate temporary shocks to the market value of time. Candidate shocks to the market value of time include productivity shocks such as those in Kydland and Prescott (1982), monetary shocks such as those in Lucas (1972), temporary government spending shocks such as those modeled by Hall (1980), Barro

8. See Smith (1977, p. 249) for a discussion, and Rogerson and Rupert (1991) for estimates of a labor-supply model with corners at year-round work.

(1981), and Baxter and King (1993), an income-tax cut that is phased in over time, or a tax or subsidy on savings. Computing the response to these shocks probably requires a general equilibrium model of which my model (8) would be one piece, but the parameter σ in my model is the most crucial ingredient. When we, for example, compare two equilibria (which differ, say, according to the processes generating policy or productivity shocks) which have different anticipated rates of growth of the after-tax discounted market value of time, σ will—up to an aggregation bias term—measure the cross-equilibria difference in anticipated rates of aggregate labor-supply growth. This paper uses various life-cycle data to estimate σ. Whether the world actually exhibits the temporary wage fluctuations predicted by these models is an interesting empirical question, but one beyond the scope of this paper.

Although very high-frequency nonseparabilities can be used to motivate my indivisible-labor model (see Mulligan, 1998), labor supply should be separable over time at low frequencies in order for the degree of intertemporal substitution found in life-cycle data to be the same as that applicable to higher-frequency temporary wage movements. If, for example, nonmarket capital were accumulated while a person was not working and that capital increased the marginal utility of leisure, then people would be more willing to substitute time over long periods than over short periods. Or if, as modeled by Kydland and Prescott (1982), extended nonmarket time lowered the marginal utility of leisure, then people would be more willing to substitute time over short periods than over long periods.

The model also has strong predictions for consumption which have been the subject of extensive testing in the literature (e.g., Friedman, 1957; Hall 1978, 1988; Shea, 1995). One area of testing relates to the responsiveness of consumption growth to interest rates and is intimately related to my analysis of labor-supply growth. However, variations in *ex ante* real interest rates that are uncorrelated with tastes and other relevant variables are even tougher to find than are variations in anticipated wage growth. Furthermore, my model is perfectly consistent with lots of substitution of work over time but little substitution of consumption (just set σ large and make u'' highly negative).

A second area of consumption testing relates to the predictions of the permanent income hypothesis for the effect of income shocks on consumption. But, because the consumption side of the model could easily be modified to include intertemporal consumption nonseparabilities (e.g., Becker and Murphy, 1988; Becker and Mulligan, 1997) or even consumption–leisure nonseparabilities (Heckman, 1974; Ghez and Becker, 1975), a variety of observed responses of consumption to income shocks are consistent with the hypothesis that labor supply is correlated with anticipated wage growth.

If workers in my data were literally unable to transfer resources across periods by borrowing, saving, or consuming assets, then the static model of labor supply would apply period by period (whatever "period" means). The sign and magnitude of the synthetic cohort correlation between work and wages would depend on offsetting "income" and "substitution" effects, and, assuming the income effect is positive, the regression (10) would underestimate σ in my synthetic cohort data.[9] As discussed below, σ would be overestimated in my AFDC samples, because the income and substitution effects are in the same direction.

3. *Synthetic-Cohort Samples*

3.1 INDIVIDUAL-PANEL AND COHORT CROSS-SECTION APPROACHES COMPARED

There are two approaches that have commonly been used to estimate σ (with pretty similar results), and each has its advantages. The first focuses on equation (1) and uses individual panel data. Anticipated labor-supply growth is measured as actual annual hours growth for those working at t and $t+1$ (as reported in the survey), and anticipated wage growth is estimated in a first-stage regression of the growth in actual average hourly earnings on a variety of variables presumed to be in the date-t information set. However, it is important to note that, among those variables that have been used to predict wage growth, functions of age are the best predictors. To see this, consider Altonji's (1986) prediction of average-hourly-earnings growth in a sample of 10,036 continuously married prime-aged man-years from the 1968–1981 waves of the PSID. With two socioeconomic indicators for parents, years of schooling of father and mother, age, a schooling quadratic, age interacted with a schooling quadratic, and year dummies as explanatory variables, his prediction equation has an R^2 of 0.0054 and a standard error of 0.254. In other words, the standard deviation of his predicted wage growth is 0.0187, which can be compared with the standard deviation of average-hourly-earnings growth of 0.037 in my CPS synthetic-cohort data for cohorts aged 25–60, and 0.087 for cohorts aged 25–79.[10] Hence, among

9. An exhaustive list of tests of "liquidity constraints" is beyond the scope of this paper, but I point out that an inability to transfer resources across periods would also imply extraordinary rates of return to schooling and OJT (on-the-job training), a prediction which is at odds with the empirical findings of Mincer (1974) and others.

10. It is interesting to note that, if the intertemporal substitution elasticity of hours were as large as one and the intertemporal model fitted Altonji's data perfectly, the standard

the variables that have been used in the literature to predict wage growth, age dummies have the vast majority of the explanatory power. This, of course, does not rule out the possibility that someday someone will use the PSID or other individual-panel data to identify quantitatively important anticipated wage changes that are not associated with age.

The second approach studies cross-sectional cohort ("synthetic cohort") specifications motivated by equation (7). First, we make the decomposition (9) and assume that the cross-sectional covariance is zero:

$$\sigma \ln \frac{Q_t^i\,(E_t\lambda_t^i)}{\gamma_t^i v_t^i} = \alpha_t a_t^i + \epsilon_t^i, \qquad \mathrm{cov}_t\Big(\epsilon_t^i, w_t^i(1 - \tau_t^i)\Big) = 0, \tag{9}$$

where i indexes individuals and a denotes a person's age. Second, assume that $g_{t,k}^a$ is distributed uniformly on $[0,e]$ (or, equivalently, integrate equation (1) over time], and obtain (10) by averaging (7) across consumers of the same age:

$$\ln N_t^a = \sigma \ln \tilde{w}_t^a + (1 + \sigma) \ln \bar{n}_t + \ln K + \alpha_t a_t + \epsilon_t^a + A_t^a, \tag{10}$$

where N_t^a is the cohort *arithmetic*-average annual hours worked, $\tilde{w}_t^a$ is the cohort *geometric*-average after-tax market value of time, and A_t^a is a within-cohort aggregation bias depending on the second and higher moments of the within-cohort distributions of ϵ_t and $\ln \tilde{w}_t$.[11] A very similar expression can be derived relating arithmetic averages of work hours to arithmetic averages of after-tax market values of time, but I use the geometric average $\tilde{w}_t^a$ because it can be decomposed in a straightforward way into geometric averages of tax factors, measured pretax wages, and proportional measurement errors.[12]

Treating $\bar{n}$ and K as constants in the cross section and assuming that the aggregation bias is uncorrelated with $\tilde{w}_t^a$ across cohorts, equation (10) is then estimated by least squares in the cross section of workers aggregated by cohort. Notice that age is included in the regression because—

deviation of log hours growth would be 0.0187. This is the difference between no wage growth and the growth of annual hours from 2000 to 2038.

11. $A_t^a = \ln \{1 + \sum_{i=2}^{\infty} [\mu_i^a(\epsilon) + \sigma^i \mu_i^a(\ln w)]/i!\}$, where $\mu_i^a(\epsilon)$ is the ith moment of the within-cohort distribution of ϵ, and $\mu_i^a(\ln \tilde{w})$ is the ith moment of the within-cohort distribution of $\ln \tilde{w}$.

12. Similar regression estimates of σ are obtained when arithmetic averages are used. Even if the aggregation bias term were correlated with $\tilde{w}_t^a$, an estimate of σ inclusive of that correlation may be the more relevant for macroeconomic forecasting because macro data are by definition, aggregated.

as mentioned in the above discussion of equation (1)—the ISH has nothing to say about the rate of growth of labor supply that would occur in the absence of growth of the after-tax value of time. Age is also included in the regression (10) because cohorts may differ in their time-t tastes or their expected lifetime marginal utility of wealth, $E_t\lambda_t$.

The slope parameter σ in equation (10) is a well-defined structural parameter in my life-cycle model. Indeed, contrary to the claims of Smith (1977, p. 249), σ is a parameter of an individual's preferences—it describes the amount of intertemporal heterogeneity in his marginal disutility of work (see also Lucas, 1970). σ also dictates the response of *aggregate* labor supply—including all its components (the fraction working sometime during the year, weeks worked conditional on working, weekly hours, etc.)—to temporary wage fluctuations, as well as the welfare implications of those fluctuations. Most important, the value of σ is the major determinant of the response of macro variables to temporary monetary, fiscal, technological, and other shocks.

Assumption (9) is crucial, so it deserves some interpretation. It says that those tastes for work and interest rates which cannot be explained by a linear term in age are uncorrelated with after-tax market values of time. Part of this is effectively a recursivity assumption—that, given wage growth, workers of all ages trade off working this period or next in the same way. But (9) is also an assumption about *cohort effects*.[13] Assumption (9) is violated, for example, when health affects tastes for work and health deteriorates at an increasing rate with age (i.e., recursivity is violated) or the detrended expected lifetime marginal utility of wealth, $E_t\lambda^a$, varies with date of birth in a way that is correlated with detrended $\tilde{w}_t^a$ (i.e., cohort effects do not follow a trend). A weaker assumption than (9) can be used when proxies for tastes or interest rates are included in the regression (10). The presence of cohort effects can also be tested by obtaining cohort cross-section estimates at different points in time. Mulligan (1995) does so and finds similar regression estimates of σ for 1979, 1980, and 1985 (which are in turn similar to the 1976 estimates reported here).

The cross-section approach has three important advantages. First, cross-section studies offer a number of interesting and highly relevant variables—such as measures of health, training, and alternative measures of time use—that are unavailable in panel studies. Second, cross-section samples are typically much larger than panel samples, so older workers may be more reliably studied and a much richer relationship

13. Depending on the interpretation of λ_t, (9) may also limit the length of a time period or the degree to which λ_t diminishes with lifetime wealth. See Mulligan (1998).

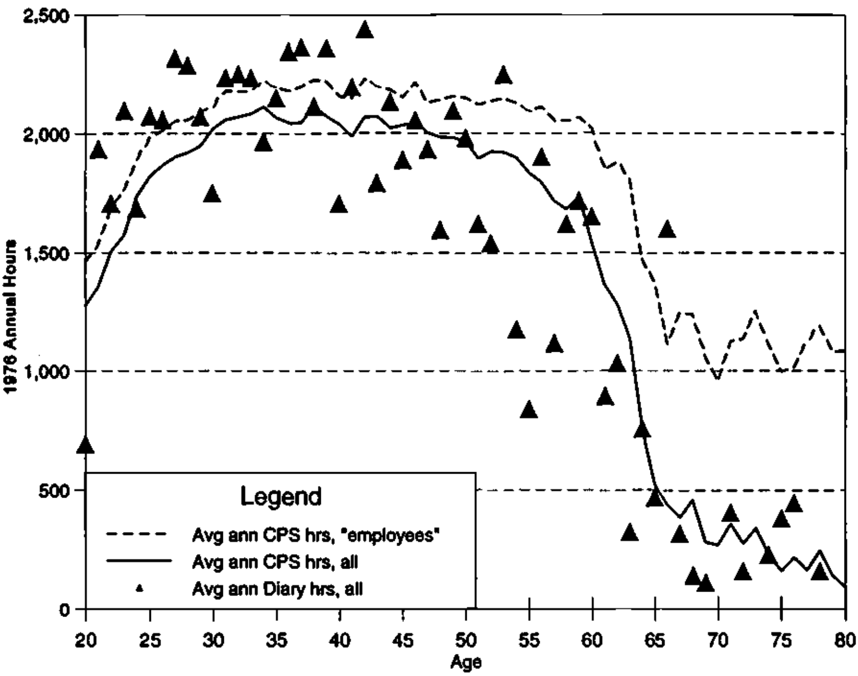
between age and wages can be accurately estimated in cross sections. Third, some of the disadvantages of cross-section data can be overcome by complementing the cross sections with some information from panel studies.

Another difference between individual-panel and cross-sectional cohort analyses is that the latter is subject to a *composition bias* due to the death of cohort members and (because in practice the market value of time can only be observed for those who work sometime during the year) the variation across age groups in the fraction and types of individuals who are employed during the year. However, I argue in Section 3.6 that cross-section cohort data can be supplemented with individual-panel data to correct for the composition bias.

Both individual-panel and cross-sectional cohort studies must make inferences about the market value of time when a man is not working. When someone is not working at instant t, it is typically assumed that the market value of time can be inferred from earnings during some other nearby period or from the earnings of a similar person who is working, or some combination of these. But it might be the case that a person's market value of time when not working is low compared to the time he does work or compared with the market value of time of apparently similar people who are working. I see no solution to this form of "selection bias" for the individual-panel and cohort studies in the literature or for my own cross-section cohort study, but point out that another advantage of my panel study of welfare mothers in Section 4 is that we can be confident that changes in the tax rules dominate any unobserved changes in the market value of time.

I begin by revisiting the cross-section cohort (synthetic cohort) samples, sticking with the same basic specification (10) but emphasizing the measurement of the key variables of interest. First, I include the employment margin in my measures of labor supply as suggested by my model. Second, I include income and social security taxes in the calculations of the value of time. Third, I assume that average hourly earnings misestimates the value of time in a way that is related to the amount of time training on the job. Fourth, I consider the possibility that hours worked as reported by employees to standard demographic surveys are systematically biased. Because many of the data needed to address these issues are only available in micro cross sections, I am necessarily constrained to construct cross-section cohort samples instead of individual-panel samples. But, fifth, I do use panel data to supplement the cross-section data and address a measurement problem that is peculiar to the latter—*composition bias*. I return also to an individual-panel sample later in the paper.

Figure 1 MALE HOURS WORKED BY AGE GROUP



Formally, I decompose cohort a's geometric-average after-tax market value of time, $\tilde{w}_t^a$, into five components:

$$\begin{aligned}\tilde{w}_t^a &= \hat{w}_t^a(1-\tau_t^a)\frac{1}{1-h_t^a}\nu_t^a B_t^a,\\ \nu^a &\equiv \frac{\hat{N}_t^a}{N_t^a},\end{aligned} \tag{11}$$

where

$\hat{w}_t^a$ = pretax CPS average hourly earnings,
$1-\tau_t^a$ = tax factor,
$1/(1-h_t^a)$ = correction for those reported hours that are self-financed OJT,
ν_t^a = correction for CPS hours reporting error,
B_t^a = composition bias.

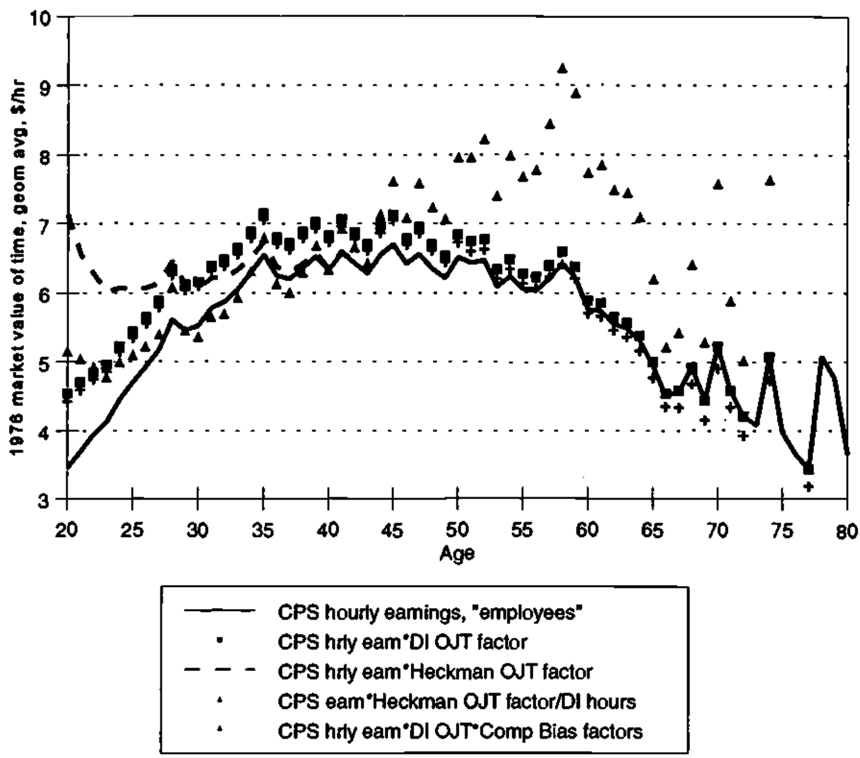
Below I show how $\hat{w}_t^a$, $1-\tau_t^a$, and B_t^a are computed as cohort geometric averages from micro-level data, and h_t^a is computed as an arithmetic average from micro-level data. N_t^a and $\hat{N}_t^a$ are computed as cohort arithmetic averages from separate micro data sets, and then their ratio is used to compute ν_t^a.

3.2 TOTAL LABOR SUPPLY IN A CROSS-SECTION OF COHORTS

Figure 1 displays hours worked for male cohorts aged 20–80 in 1976.[14] The solid line displays CPS average annual hours for all men in each of 61 age brackets. Because annual hours worked are zero for some men, this solid line lies below the dashed line, which displays average annual hours for only those men who worked positive hours in 1976.

The CPS cohort geometric mean of average hourly earnings (which are measured only for those working positive hours in 1976) are displayed as a solid line in Figure 2. Judging from the cross-sectional data dis-

Figure 2 MALE PRETAX MARKET VALUE OF TIME BY AGE GROUP



14. The cross section is the March 1977 CPS (ICPSR study 7784). Hours are measured as "usual weekly hours last year" times "weeks worked last year." Average annual earnings is measured as "earnings last year" divided by annual hours.

played in that figure, wages almost double between ages 20 and 40. CPS annual hours worked by "employees" shown in Figure 1 increase, but more modestly. The increase in average hours by all between ages 20 and 40 is more substantial, almost doubling. The data in Figures 1 and 2 suggest that both hours and wages fall dramatically between ages 55 and 80.

Using the 1960 Census 1/1000 Public Use Micro Sample (PUMS), Ghez and Becker (1975) study the hourly earnings of *employed* white men aged 22 to 65. They compute the average annual hours and average hourly earnings by age for the men in their sample. In a regression of each age group's log average hours on its age and its log average pretax average hourly earnings, they obtain a wage coefficient of 0.39.[15] When the log of *leisure* (5096 minus annual hours worked) is used as the dependent variable, they find a coefficient of 0.25. Because they study samples of employed men, measure annual hours and earnings, and isolate the wage variation that occurs with age, Ghez and Becker's specification is essentially the same as those of Altonji (1986) and MaCurdy (1981) (subject to the caveats mentioned in Section 3.1 above).

My cohort analysis begins with 1976 annual hours and earnings as reported in the 1977 CPS. In a regression of each age group's log average hours of employees on its age and its average log pretax average hourly earnings of employees, I report a wage coefficient of 0.37 in the first cell of Table 1 for a sample of men aged 25–55. Like Ghez and Becker's, my estimate of 0.37 is from a sample of only those employed sometime during the year.[16]

However, hours of employees are difficult to interpret in my life-cycle model and may be of limited interest for making predictions for the aggregate labor supply N_t^a. For example, the hours of employees are constant over time and equal to $\bar{n}$ in the special case of $K=1$, while the aggregate labor supply might be quite sensitive to temporary wage changes. When aggregate labor supply is measured as in the model—annual hours averaged across all men, including those who did not work during the year—I obtain a σ-estimate of 0.57.

Heckman (1993, pp. 116, 119) suggests that changes in labor supply

15. They smooth their data by taking three-year moving averages. Their more widely cited estimate of 0.45 is obtained when income and family size regressors are included. See their Table 3.5.

16. Our samples sizes are quite similar: they have 33,591 men and I have 34,654 men. I include all men regardless of race and do not smooth my cohort data by computing moving averages. Estimates of σ obtained with 3-year moving averages (not reported in this paper) are typically 20% larger than the corresponding estimates reported here. Ghez and Becker (1975) also include workers aged 56–65, but, since retirement ages have fallen over time and the importance of social security has grown over time (Costa, 1998, Chapter 2), my aged 25–55 sample is probably the better comparison.

Table 1 ESTIMATES OF σ, WITH OJT AND TAX FACTORS

		IES estimates			
		Ages 25–55		*Ages 24–64*	
OJT source	*Tax factor*	*Employees only*	*All men*	*Employees only*	*All men*
None	None	0.370	0.569	0.650	1.406
	Marginal A (MA)	0.388	0.588	0.665	1.462
	Marginal B (MB)	0.406	0.613	0.693	1.525
	Average A (AA)	0.366	0.555	0.661	1.407
	Average B (AB)	0.376	0.571	0.673	1.440
SRC	None	0.397	0.614	0.710	1.551
	Marginal A (MA)	0.424	0.646	0.786	1.695
	Marginal B (MB)	0.442	0.671	0.818	1.778
	Average A (AA)	0.425	0.635	0.764	1.590
	Average B (AB)	0.435	0.651	0.783	1.641
Heckman et al.	None	0.463	0.761	1.184	2.770
	Marginal A (MA)	0.500	0.784	1.348	3.163
	Marginal B (MB)	0.487	0.755	1.363	3.223
	Average A (AA)	0.512	0.767	1.265	2.829
	Average B (AB)	0.501	0.756	1.335	3.029

1. IES estimates are coefficients from regressions of log age-group hours on age and age-group workers' average log after-tax average hourly earnings. Earnings and hours (inclusive of OJT) from the 1977 CPS.
2. Marginal tax factor A is one minus the 1976 Federal individual income tax (IIT), and social security old age, survivors, disability, and hospital insurance (OASDI & HI) tax on an additional dollar of gross earnings minus—for those aged 62–71 and receiving social security—the implicit marginal tax of social security benefits.
3. Marginal tax factor B is marginal tax factor A plus the accumulation of social security wealth (SSW) asociated with an additional dollar of gross earnings.
4. Average tax factor A is the increment to 1976 IIT, OASDI & HI, and the implicit tax on social security benefits that would result from not working at all during 1976, as a fraction of 1976 gross earnings, plus one.
5. Average tax factor B is an age-weighted average of average tax factor A and marginal tax factor B.
6. The log of each tax factor is averaged across workers within each cohort to obtain a cohort tax factor. See Appendix A for more details of the computation of the four tax factors.
7. Employees are those men reporting positive work hours and average hourly earnings between $1 and $100 for calendar year 1976.
8. SRC OJT is cohort average self-financed on-the-job training, computing from the 1976 Time Use Study as indicated in the text.
9. Heckman *et al.* OJT is on-the-job training inferred from earnings panel data for young men. See Heckman, Lochner, and Taber (1998).

over the life cycle consist of changes at both the *intensive* and the *extensive* margin. This is not true in my model, in which *all* changes are at the extensive margin. However, my model, Heckman's discussion, and Coleman's [1984] and Alogoskoufos's (1987a,b) studies of business-cycle fluctuations agree that annual hours worked by those working positive hours is not the same as aggregate labor supply and that the difference between the two—the fraction of workers who work positive hours during a year—may also respond to wage fluctuations. And the larger elasticity estimates in my Table 1 for "all" as compared to "employees only" confirm Heckman's (1993) life-cycle conjecture.[17]

CPS estimates of σ are sensitive to the number of older age groups included. Table 1 reports estimates about twice as large for samples aged 24–64.[18] After introducing data on taxes and health, I return in Section 3.8 to some of the differences between young and older workers.

3.3 TAXES AND INCENTIVES TO WORK OVER THE LIFE CYCLE

One difference between CPS average hourly earnings and the after-tax market value of time, $w_t(1-\tau_t)$, is the labor-income-tax factor $1-\tau_t$. This factor can vary with age because of the progressivity of the federal individual-income-tax system and its dependence on marital status, because of the regressivity of social security payroll taxes, and because of the effect of age on the rate of accumulation of social security wealth. Four components of federal tax and benefit rules are included in my calculations of the tax factor:

1. Individual income taxes
2. Social security OAS, DI, and HI payroll taxes (employee component only)[19]
3. Implicit taxation by social security benefit formulae from earnings limits (applies only to men aged 62–71 who receive a social security benefit) and inadequate delayed retirement credits (applies only to men age 65–71)
4. The accumulation of social security wealth (applies to men under age 72)

The fourth component is the least straightforward, both because the rules are complicated and cohort-specific and because computation of the tax factor requires information about workers' expectations of future benefit formulae. I therefore report calculations with and without this fourth component.

17. Given the maximum feasible work during a time interval ($K\bar{n}$), the parameter K dictates in my model how aggregate labor-supply responses are partitioned between employment and employee hours. For $K=1$, employment is the entire response. For, say, $K=2$, employment is half of the response at the margin for a group with an 89% employment rate.
18. Depending on the specification, including the four age groups 20–23 or excluding some of the younger age groups affects estimates of σ, although not systematically in one direction or another. These results are available from the author.
19. It is assumed that the earnings data measured by the CPS are net of employer social security contributions.

The accounting period for individual income and social security taxes is a calendar year. If potential work sessions coincide with the tax accounting period (as they do in my model), then incorporating taxes into the estimation of σ is fairly straightforward [see equation (5)], although the CPS does not provide much information about the deductions that families might be taking from their individual income tax or the costs they bear in the acquisition of those deductions. I therefore assume that all families take the standard deduction and, according to equation (5), compare the taxes paid with taxes that would be paid if the man's earnings were set to zero. This produces an average-tax-rate measure of the tax factor.

Another implication of a long work session is that the decision to work before age 62 typically has a negligible effect on the accumulation of social security wealth. 1976 earnings affects social security benefit formulae by affecting the lifetime average of one's top index earnings years (AIME). If someone plans to work most of his prime-age years, then not working a long session in a particular year can only affect the AIME by dropping that year from the calculation and adding another year to the calculation.[20] I therefore exclude SSW from any calculations of the average tax rate.

If the tax accounting period includes multiple potential work sessions, incorporating taxes into the estimation of σ is significantly more complicated. In one extreme (and counterfactual) case, however, the appropriate tax factor is one minus the marginal tax rate on an additional dollar of earnings.[21] Shorter work sessions also imply that a work decision might have an important effect on social security wealth, so I include one computation of the marginal tax rate that includes the effect of work on SSW. Three of the four measures of the tax factor used in my analysis are:

MA. Marginal tax rate, PIA fixed (IIT, OASDI and HI payroll, phaseout of OA benefits)

MB. Marginal tax rate with accumulation of SSW (IIT, OASDI and HI payroll, phaseout of OA benefits, and accumulation of SSW)

AA. Average tax rate (IIT, OASDI and HI payroll, phaseout of OA benefits, and effect of retirement decision on OA benefits)

20. In the extreme case that a worker's earnings during the years he works grows at the same rate as the national index, this substitution of one year for another has zero effect on AIME and therefore zero effect on SSW.

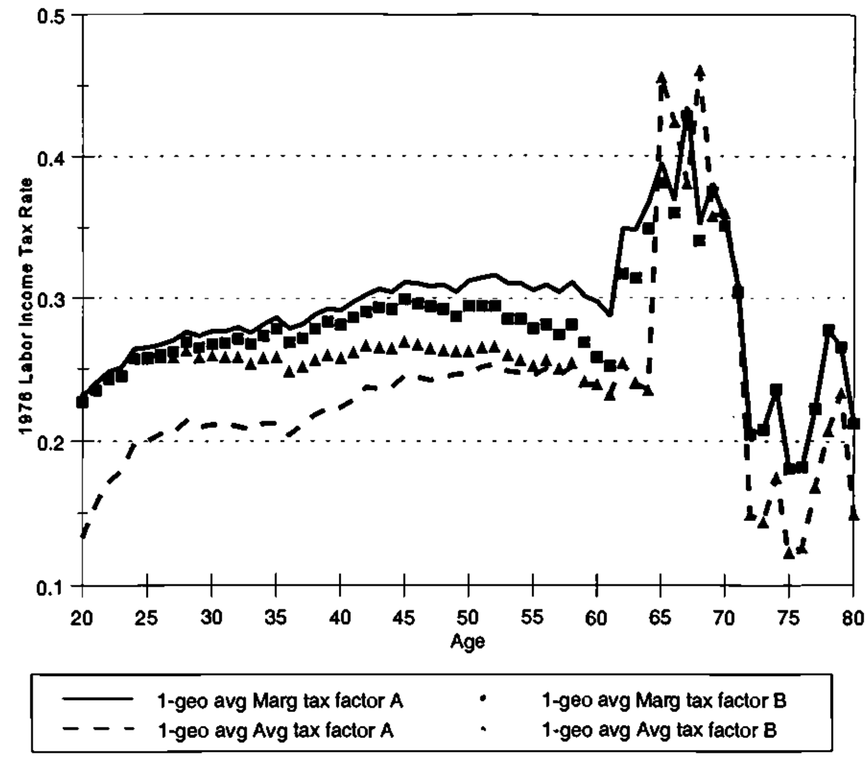
21. This special case requires that the marginal tax rate be a continuous function of earnings [otherwise one has to allow for the "kinky" behavior described by Hausman (1985)] and that each tax accounting period include very many potential work sessions.

My model supposes that K and $\bar{n}$ are the same for all age groups, but another interesting model might allow $\bar{n}$ to increase with age (while holding $K\bar{n}$ constant). The growth of $\bar{n}$ with age would, for example, explain why the employment rate appears to be (at least in CPS data) a relatively more important margin for the old than for the young, even before the "old" reach age 62. I therefore consider a fourth measure of the tax rate that averages the tax rates MB and AA with the weights depending on age:

AB. Age-weighted average of MB and AA

The four tax rates for each cohort (one minus the within-cohort geometric average of their corresponding tax factor) are displayed in Figure 3. With the exception of men aged 65–71, average rates (AA) are lower than marginal rates (MA and MB). But what is more relevant for estima-

Figure 3 LABOR INCOME TAX RATES BY AGE GROUP



tion of σ is the change in the tax rate with age, and we see that average rates (AA) rise more rapidly for prime-aged men, a difference which can be attributed to the fact that marginal rates are well above average rates for young men and taxable income is rising with age and, for the tax measure MB, the increased rate of accumulation of SSW with age.[22]

Rates of change of the four tax series differ most at ages 62–65. Here average tax rates remain relatively low because most men have not yet retired by age 61 and retirement at age 62 is not particularly encouraged or discouraged by social security benefit formulae. For ages 65–71, however, the social security disincentive for work in the following year is pretty large—over 90% of a year's benefits are lost by delaying retirement one year (Appendix A). Thus Figure 3 displays low cohort-average tax rates except for ages 65–71. Marginal tax rates increase a bit at age 62 because some of the men who do begin to take social security still earn above the earnings limit and are subject to an additional 50% marginal tax. As men age, more are in this situation, so cohort-average marginal tax rates continue to rise in Figure 3.

It is important to note that none of the series in Figure 3 fully capture the work disincentives of social security. One consequence of the social security benefit formulae is that some men aged 62–71 switch to part-time jobs to keep their earnings at or below the earnings limit. The part-time jobs have lower hourly pretax wages, so a man's acceptance of a part-time job—which shows up in Figure 2 as a lower pretax wage—is itself a tax even if he does not pay a dime of the 50% tax implicit in the social security benefit formulae. Thus there are a variety of ways one might partition the after-tax market value of time, $\tilde{w}_t^a$, into a pretax component and a tax factor; my Figures 2 and 3 are only one such way.

Estimates of σ derived from the various tax factors are shown in Tables 1 and 3. The sign and magnitude of the effect of including the tax factor depends on the method of its calculation. To the extent that detrended tastes for work do vary with age (even in a way that is uncorrelated with pretax wages), the tax accounting period differs from the period of labor indivisibility, and the tax system is progressive, $1-\tau_t$ is negatively correlated with ϵ_t and estimates of σ derived from the various tax factors are biased downwards.

3.4 ON-THE-JOB TRAINING OVER THE LIFE CYCLE

My empirical analysis so far treats a worker's productivity growth as exogenous. It is quite plausible that the increase in productivity of young

22. See Appendix A for some details of my calculations, and Feldstein and Samwick (1992) for a study of the accumulation of SSW.

men is due to the accumulation of skills over time and that the decrease (or slower growth) in old age is due to depreciation and low levels of skill accumulation. If time is an important component of skill accumulation and skill accumulation occurs on the job, then the correlation of skill accumulation with age leads to a problem with the interpretation of life-cycle wage estimates such as those as the solid line in Figure 2. The problem is that worker productivity is estimated as the ratio of earnings to hours spent at work but some of the hours at work are not spent producing and are not compensated by the employer. Younger men are presumably engaged in the most skill accumulation, so that their productivity is underestimated the most. As men age, the underestimation is mitigated as skill accumulation time falls. For this reason, the growth of the market value of time of young men is overstated.[23]

Few data are available on time spent training on the job, and even less on the question of who finances that training. A 1976 study of time use by the Survey Research Center (SRC) asked study participants two relevant questions[24]:

(i) "Do you feel you are learning skills on your job that could lead to a better job or to a promotion?" (IF YES) "Sometimes people learn these skills as part of their regular work, while others use time at work to learn skills that are *not* part of their regular job. About how many hours *per week* do you usually spend learning new things as *part* of your regular work?"

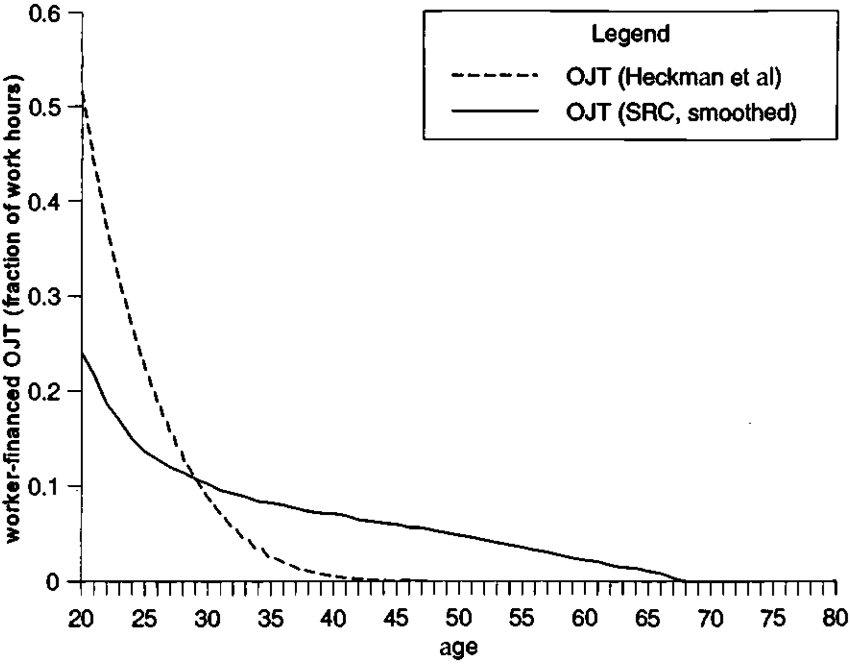
(ii) "And how many hours *per week* do you spend learning new things that are *not* part of your regular work?" (Stafford and Duncan, 1985, p. 284).

Under the assumption that 100% of the hours reported as a response to (ii) and 50% of the hours reported as a response to (i) are not compensated by employers, I display as a solid line in Figure 4 the total hours of worker-financed training time for each of 61 male age groups.

23. Jacob Mincer has made this point in several studies (including Mincer, 1974, 1977), distinguishing between average hourly earnings ("wages") and the market value of time ("capacity wages") and deriving implications for the life cycle pattern of hours and earnings. Ghez and Becker (1975, pp. 94, 100) and Heckman (1975, p. 228) also make this point in their studies of labor supply, but Heckman treats training time as unobservable, while Ghez and Becker correct for the bias only by including age squared in equation (10).

24. The 2406 respondents are the same 2406 respondents who filled out the time diaries that are the object of my study in Section 3.5 below, although the two OJT questions were part of the questionnaire and not part of the diary section of the study (my sample sizes are also smaller because I use only men and require that the necessary variables have valid codes).

Figure 4 ESTIMATES OF THE LIFE-CYCLE PATTERN OF WORKER-FINANCED ON-THE-JOB TRAINING



Rather than attempting to measure time spent training on the job, Heckman, Lochner, and Taber (1998) infer OJT time from the earnings growth they observe in a panel study of young men. Figure 4 displays their estimates as a dashed line. Because their method of measuring training time is so different from the SRC survey method, it is a nice complement to the SRC measure, although it should be noted that the Heckman–Lochner–Taber measure is derived from a model with exogenous labor supply.[25]

The 1976 SRC OJT data are first smoothed by regressing OJT on a quadratic in experience, and then the two OJT estimates are used to adjust the CPS estimates of the after-tax market value of time according to equation (11) and displayed in Figure 4. Both measures result in a substantially reduced rate of growth in the after-tax market value of

25. Heckman (1975, p. 255) reports estimates of training time derived from a model with endogenous labor supply that are very similar to the numbers in Heckman, Lochner, and Taber (1998). Heckman (1975, p. 251) also reports another set of estimates which are very much like the SRC numbers before age 30 and like the estimates in Heckman, Lochner, and Taber (1998) after age 30.

time, and the data of Heckman, Lochner, and Taber actually imply a decline in that value of time between the ages of 20 and 23. Not surprisingly, larger estimates of σ are reported in Table 1 when OJT time is used to adjust CPS average hourly earnings. The OJT numbers of Heckman, Lochner, and Taber typically make the biggest difference.[26]

Consistent with the models of Ben-Porath (1967) and Mincer (1974), my measurement of OJT implicitly assumes that human capital is accumulated through training time that cannot be used to produce current output. But an upward adjustment of the value of time that declines with age and hence flattens the profile of the life-cycle value of time can be derived from a number of other accumulation models, including those with learning by doing, signaling, or firm financing of some general training [see Rosen (1972) or Gibbons and Murphy (1992)].

3.5 HOURS-REPORTING BIAS

It has been argued by Juster and Stafford (1991, p. 496), Stafford and Duncan (1985), and others that life-cycle studies of labor supply are sensitive to the method of measuring labor supply. In particular, they claim that larger life-cycle changes in hours are found in time-diary data than in CPS-type survey data.

There are a variety of reasons one expects the time-diary estimates to be more accurate than measures based on CPS-type surveys. First, CPS respondents merely answer the questions "How many weeks did you work last year?" and "In the weeks you worked, how many hours did you usually work?"[27] whereas time-diary respondents are obligated to account for all of their time in a particular day or days. The stereotypical response of "40 hours" might be expected for the CPS respondent, but such a response by a time-diary respondent would create inconsistencies in his diary unless he actually worked 40 hours [although Pencavel (1986, p. 14) suggests that common reports of 40 are real and the result of legal restrictions]. Suppose, for example, that a time-diary respondent works more than 40 hours in a week. A response of "40 hours" would leave a hole in his schedule, which he would have to fill by fabricating a

26. Notice that log CPS hours are still used as the independent variable, as suggested by my model, where OJT yields disutility just like other work.

27. CPS respondents are also asked about hours worked in the week prior to the interview, but there is no corresponding earnings question (for example, there are many CPS men who did not work in the week prior to the interview but had substantial earnings in the prior calendar year). One might use log CPS cohort hours last week as the LHS variable in equation (10) and CPS average log hourly earnings last year as the RHS variable. Slightly higher IES estimates are found, but closer to those reported in the first two columns in Table 1 than to those in the last two columns [results available upon request; see also Mulligan (1995)].

story for what he was doing when he was actually working. Casual "40 hours" responses by CPS respondents would generate the appearance that hours do not fluctuate over the life cycle. A second reason to prefer the time diary is that an attempt is made to measure travel time, coffee breaks, lunches at work, and other activities done "at work" while not actually working. A third advantage of the diary data is that it measures time devoted to finding a job. According to my model, time spent searching for a job is work even though it is not compensated. A fourth reason that CPS annual hours are misreported is that they include time spent on paid vacation or sick leave. Fifth, there is some evidence that—aside from the distinction between weeks worked and weeks paid—workers make systematic errors in their responses to retrospective questions about weeks worked (Horvath, 1982).

I suggest four separate corrections for hours-reporting bias. The first and preferred method discards CPS hours all together and computes age-group hours from a study of time use. Figure 1 compares the reports of 1977 CPS respondents with results from a 1975–1976 SRC time-diary study by age group. 624 male respondents completed time diaries for three or four days between October 1975 and September 1976. Total minutes of "normal work," "work on a second job," and "unemployment activities" were summed over the diary days, weighted in such a way as to represent a seven-day synthetic week.[28,29] The sum does not include minutes spent on coffee breaks at work, eating lunch at work, or commuting. There is more idiosyncratic variability in the diary data, which is certainly due to (1) the smaller sample sizes used to compute age-group means and (2) the greater micro-level variation in diary hours. But it seems clear that, as compared to the CPS survey, somewhat more hours are measured for younger men by the diaries and substantially fewer hours for men nearing retirement age (1976 hours for men aged 50–64 are 1448 in the diary, and 1685 in the CPS). This second discrepancy with the CPS is consistent with Ruhm's (1990) study of *bridge jobs* (switches away from career occupation or industry) and, for those aged 62–64, his study of *partial retirement* (periods of employment separated by spells of retirement).

28. My sample is larger than that of Stafford and Duncan (1985) because, apparently, they restrict attention to people aged 64 or less with "regular work schedules," exclude "supplemental respondents," and use only one of the four waves of the 1975–1976 time study.

29. I assume that each respondent's three or four diary days are randomly chosen from the year (actually they're chosen from October, November, February, May, June, and September), so that the SRC's calculation of minutes per synthetic week is a calculation of minutes per representative week. I multiply minutes per synthetic week by 52/60 to get annual hours. Mechanically, the scaling up is done first by the SRC, who computes minutes per representative week, and then I multiply by 52 and divide by 60 to get hours per year.

Table 2 ESTIMATES OF σ CORRECTED FOR HOURS REPORTING ERRORS

Method	*Age group*	*Only hours corrected*	*Hours and wages corrected*[a]	*Addendum: no correction*[b]
Compute age-group hours	25–55	1.357	6.417	0.569
from time diary	24–64	1.991	4.797	1.406
Compression model, $\theta = 0.7$	25–55	0.729	0.815	0.569
	24–64	1.685	2.088	1.406
Liars model, $\theta = 0.3$	25–55	0.729	1.140	0.569
	24–64	1.685	6.412	1.406
Drop CPS men reporting	25–55	1.000	0.809	0.569
exactly 40 hr/wk	24–64	2.261	1.942	1.406

1. IES estimates are coefficients from regressions of log age-group hours on age and age-group workers' average log market value of time. "All" sample (workers and nonworkers) is used.
2. No tax or OJT factors are used.
[a]When diary hours are used to compute the wage (first two rows), log CPS wage instruments for the corrected wage.
[b]From Table 1.

Table 2 shows that the elasticity of hours (workers and nonworkers) with respect to the CPS hourly wage is about 1.4 for the 25–55-year-old sample and larger for samples that include older men. Elasticities of 2 are typical when the CPS hourly wage is corrected using the SRC OJT hours (see Table 3), and even larger when Heckman, Lochman, and Taber's (1998) numbers are used (not reported).[30]

Like those based on CPS hours, estimates of σ based on diary hours are sensitive to the number of older age groups included, and this sensitivity can be seen in Tables 2 and 3. The diary estimates are even more sensitive to the truncation of older workers in the 50–64 age range, because this is the largest discrepancy between CPS and diary hours.

If it is true that the life-cycle hours profile is misestimated with CPS data, then average hourly earnings are also likely to be misestimated. Ideally, one would like to discard the CPS hours data from the computation of average hourly earnings and replace them with diary data, but, due to the relatively small sample size, Figure 1 shows that the age-grouped diary data are fairly idiosyncratic. This means that average hourly earnings computed with diary data would be idiosyncratic and *in a way that is correlated with idiosyncratic errors in measured hours.* One ap-

30. As is obvious from Figure 1, there is substantially more sampling error when time-diary cohorts are used to estimate cohort hours. Not surprisingly, OLS standard errors for σ are larger, typically between 0.2 and 0.6, as compared to a typical range of 0.02 to 0.25 when CPS hours are used. Of course, serial correlation and a number of first-stage estimation errors mean that OLS standard errors overstate the precision of my estimates.

proach is to instrument for log (CPS earnings/diary hours) with log (CPS earnings/CPS hours) in the regression of diary hours on average hourly earnings. We see in the first two rows of Table 2 that doing so substantially increases point estimates of σ. However, because of the poor fit of the first-stage regression, the magnitude of the increase is quite sensitive to the tax factor, OJT factor, and sample used.

Diary age-group sample sizes are fairly small, with as few as 5 men and 18 days sampled in any single group aged 24–55.[31] The typical age group samples 12 men and 47 days. Hence, my second and third corrections for hours-reporting bias are of some interest, because they do not rely on grouping the diary study by age. Both methods assume a model of hours-reporting bias, calibrate the model by comparing histograms of reported hours for 365 diary men and 23,899 CPS men aged 25–55 report-

Table 3 ESTIMATES OF σ INCLUDING OLDER MEN, WITH TAX, OJT, HEALTH, AND COMPOSITION-BIAS CORRECTIONS

		IES Estimates[a]					
		1977 CPS: Health regressors?[c]			*1975–1976 time-diary hours, CPS average hourly earnings: Health regressors?*[c]		
Age group	*Tax factor*[b]	*No*	*Yes*	*Yes, AA*[d]	*No*	*Yes*	*woPC*[e]
24–64	None	1.488	1.238	0.984	2.088	1.167	1.897
	MA	1.538	1.270	1.147	2.132	1.262	1.898
	MB	1.604	1.317	1.225	2.181	1.251	1.888
	AA	1.509	1.135	0.944	2.128	1.058	1.749
	AB	1.542	1.191	1.052	2.173	1.117	1.829
24–79	None	2.953	2.799	1.027	2.923	1.139	1.766
	MA	2.672	2.685	1.168	2.438	1.186	1.695
	MB	2.668	2.597	1.363	2.413	1.110	1.583
	AA	2.193	1.991	1.366	1.916	0.780	1.167
	AB	2.198	1.982	1.425	1.924	0.777	1.159

1. OJT is cohort-average self-financed on-the-job training, computing from the 1976 Time Use Study as indicated in the text.
2. See text for details of computation of composition-bias correction of wages.
[a]Coefficients from regressions of log age-group hours on age and age-group workers' average log after-tax market value of time. "All" sample (workers and nonworkers) is used.
[b]See Notes to Table 1 and Appendix A for explanation.
[c]Health regressors are age-group averages of hours spent on personal care and four indicators of physical condition.
[d]Uses arithmetic-average rather than geometric-average CPS wage.
[e]Personal-care hours omitted.

31. Among groups aged 56–69, only the aged-63 group samples fewer days, 8.

ing positive hours, and infer the true hours elasticity from estimated CPS hours elasticities. The first model is a *compression model*, assuming that reported hours $\hat{n}$ is closer to some stereotypical number $\tilde{n}$ than is true hours n for those whose true hours are positive: $\ln \hat{n} - \ln \tilde{n} = \theta(\ln n - \ln \tilde{n})$. The second, the *liars model*, assumes that a fraction θ of those working positive hours report the stereotypical number $\tilde{n}$ while the rest report the truth. When conservatively calibrated to the cross-section of 365 diary men, both models imply substantially larger elasticities (see the middle four rows of Table 2). Details of these calculations can be found in Appendix B.

A critical review of the literature on time measurement is beyond the scope of this paper [see Juster (1986) for a favorable evaluation of diary studies, and Juster and Stafford (1991) for a literature review], but there may be other reasons to distrust time diaries, so my fourth method is based on the CPS data only. It discards those men reporting *exactly* 40 weekly hours (roughly half of the sample) before computing cohort averages and estimating cross-cohort regressions. The method implicitly assumes that all reporting 40 are liars and that the propensity to lie is unrelated to determinants of labor supply other than the wage. Estimates of 0.8 and 1.0 are found with this method for the 25–55 age group. Correcting wages in addition to hours increases estimates in the first six rows of the table, but this is not true with the fourth method.

3.6 OLDER WORKERS I: SOURCES OF WAGE VARIATION

It appears from the pretax wage series in Figure 2 and is obvious from the tax measures shown in Figure 3 that there is more age-related variation in the after-tax market value of time when older workers are included in the sample. All else the same, more variation in the after-tax market value of time might be expected to minimize underestimation of σ due to measurement and other errors. Because labor supply may not be as linear as suggested by my equation (10), it may also be desirable—from the point of view of making predictions for aggregate labor supply—to have an older sample with mean hours worked closer to the population mean.[32]

However, including older workers may increase the difficulty of inferring an age group's average market value of time, because relatively few work some time during the year (although see my discussion in the next subsection). Panel data complement the cross-section data in my Section 3.7 to make some progress on this problem.

32. The average annual hours worked for those aged 24 and over in the 1975–1976 Diary Study (including women) is 1165, which can be compared with 2047 for men aged 25–55 and 1677 for men aged 24–79.

Another problem is that older workers are more likely to hold part-time jobs for "noneconomic" reasons, and thus part-time jobs may pay less than full-time jobs (author's calculations from the March 1977 CPS). If the old-age movement to part-time jobs is in response to the declining labor productivity growth that occurs with age, my measures of wages overstate the declining rate of growth and may understate σ. However, an old-age movement to part-time employment may be in response to an exogenous change in preferences for work, so that the decline in wages is a response to a decline in labor supply (not the other way around) and σ is overestimated. One adjustment for this is to estimate the age-group average pretax market value of time from a sample of full-time workers only. Doing so produces estimates of σ which are very similar to those displayed in Tables 1–3, suggesting that these two biases are nearly offsetting.[33] It should be noted, however, that even the average hourly earnings of full-time workers in their sixties are as low as or lower than those of full-time workers aged 30.

3.7 COMPOSITION BIAS OVER THE LIFE CYCLE

Because employment rates vary with age, the average hourly earnings of a cohort's workers is an average over a sample that varies with age. In contrast, the implications of the theory have been derived for a sample which (ignoring mortality) is constant over time—the entire cohort. To the extent that the market value of time differs for employees and nonemployees, the average log hourly earnings of cohort a's workers, X^{*a}_t, is a biased estimate of the average log market value of time of the entire cohort, X^a_t. I refer to this bias as a *composition bias* and denote it by B^a_t:

$$\text{composition bias}^a_t = \ln B^a_t = (1 - \Pi^a_t)\,(X^{*a}_t - \tilde{X}^a_t),$$

where $\tilde{X}^a_t$ is the average log market value of time of those aged a who don't work at t. The composition-bias corrections are done separately for young and old men (age≤55 and age>55). For young men, PSID waves 1976–1978 (referring to hours and earnings for 1975–1977) are used to correct for the composition bias arising when the average hourly earnings of workers in 1976 are used to estimate the value of time of all

33. Full-time wage results—including those that make OJT, tax, composition bias, and hours reporting bias corrections and include workers as old as 79—are available upon request. Of course, these calculations do not rule out the possibility that the decline in the taste for work mainly reduces wages of full-time workers rather than increasing the propensity to take low-wage part-time jobs. It is also possible that full-time wage estimates understate σ because the movement to part-time employment is part of the implicit social security tax.

cohort members. Both PSID–SRC and SEO samples are used, but, by weighting each man with his 1976 family weight, SEO members are effectively downweighted for the purpose of computing various sample average log wages.

At the annual level of time aggregation, transitions to and from employment are often associated with changes in household status. For example, a young man is likely to form his own household when he goes from a full year of no employment to a year of some employment. I therefore utilize the individual files from the PSID (in addition to the family files), because the individual files include information about hours and incomes of men who are not heads of households. The individual files report an individual's taxable income and indicate whether that income includes any nonlabor income, but not how much. This is the only measure of labor income available for PSID men who are not heads of households, but, not surprisingly, taxable income includes asset income only for 2% of these men.[34] I therefore exclude those nonheads with asset income from my calculations of average log hourly earnings.

In order to estimate the average value of time of nonworkers, some assumptions are required. I assume that, conditional on age, the annual growth rate of a person's market value of time is uncorrelated with changes in his employment status. This assumption is true under the null hypothesis of a zero labor-supply elasticity and can be true with a nonzero elasticity, but is false under other conditions. If those who change employment status have a different growth rate of their market value of time, then my estimates of the composition bias are subject to a selection bias. There is no magic correction for selection bias, so I only note that it may be a problem and also point out that nearly every other study of labor supply—including those individual-panel studies of the hours margin in isolation from the employment margin [see my discussion of assumption (A2) in Section 2.9]—has this kind of selection bias. The selection bias arises in so many studies of labor supply because there is no person who works continuously—even the hardest-working man has some time when he is not working (e.g., evenings, weekends, sick days, holidays, vacations)—and the econometrician is typically unable to directly measure the market value of time during those nonwork periods.

34. Based on comparisons of individual taxable income, individual transfer income, and household-head labor income for men who are heads of households and report no asset income, it is clear that "taxable income" excludes transfer income even though some forms of transfer income are taxable by the IRS.

3.7.1 Young Men Over a three-year period, enough young men (95%) are employed at some time that we can obtain a reasonable estimate of the market value of time of those not employed in any particular year by looking at the earnings of those men in adjacent years. First, I compute the $(t-1,t)$ and $(t,t+1)$ average log hourly earnings growth for those employed at least 200 hours in each of the two relevant years. For those not employed at t but employed in an adjacent year, I estimate their market value of time at t by adding (or subtracting) the relevant growth rate to (or from) their average log hourly earnings when they were working at $t-1$ or $t+1$.[35]

Since employment rates are much greater than 50% for all young age groups, one might expect that the relative market value of time of those not employed is decreasing in the employment rate because those not employed are a more select sample in the high-employment age groups. In fact, the data—together with the estimation method outlined above—support this.

It is too much to use the relatively few available PSID observations to estimate the market value of time of those not employed as a *function* of age, so I compute instead the mean percentage difference between a nonworker's market value of time and the average value of time of workers his age—a 20% (in log points) difference in my sample.[36] Bils (1985) obtains a similar estimate of 20% for young men from the NLSY. Assuming that the 20% gap is independent of age, the percentage composition bias of my CPS value of time estimates is

$$\ln B_t^a = (1 - \Pi_t^a)\,0.20, \tag{12}$$

where Π_t^a is the fraction of cohort a employed at some time during year t.

If instead the relative market value of time of those not employed is decreasing in age (as the PSID data suggest), then my calculation (12) overstates the correlation between composition bias and employment rates, and I overstate the growth of the market value of time with age among young men. This error tends to reduce estimates of the labor-supply elasticity.

3.7.2 Old Men The annual employment status is much more persistent for old men, because it is typical for a man to retire and never return to employment again. One therefore might compute the composition bias for old men by accumulating the composition bias. However, this accu-

35. For those working in both adjacent years, I average the results of the two methods.
36. The median difference is 11%.

mulation requires many years of a panel data set large enough to estimate age-specific employment rates and retirement hazards for each cohort. Assuming that $B_{t-1}^{a} = B_{t}^{a-1}$, the composition bias can be computed according to

$$B_t^a = (X_t^{*a} - g_t^a) - X_{t-1}^{*a} + B_t^{a-1}, \qquad \text{age} > 55$$

The first term is the (log of) date-($t-1$) hourly earnings averaged across those working at t, which is decomposed in the parentheses into the date-t hourly earnings minus the growth rate between $t-1$ and t. The second term is estimated (log of) date-($t-1$) hourly earnings for those working at $t-1$. The third term is an estimate of the lagged bias, which differs from the actual lagged bias to the extent that retirement rates were different for the a and $a-1$ cohorts. Hence, the first two terms are the *increment to the composition bias.*

Another relevant complication for old men is that some of them change from full-time to part-time work and the part-time work tends to have lower hourly earnings. As discussed above, an important part of the difference between part-time and full-time wages by the elderly is the form in which they pay the implicit social security tax, so I do not want to count the retirement of a full-time worker as the exit of a relatively high-wage person unless his wage is higher than that of other high-wage workers. Yet another complication is that workers may receive a bonus, accumulated sick pay, or other extra earnings upon retirement that cannot be attributed to the work they did during the year prior to retirement. This would also make it appear that high-wage men are more likely to retire. Using the 1980 5% Census PUMS, I therefore compute the increment to the composition bias by imputing wages for sample men aged 55 and over according to the median average hourly earnings for full-time male workers aged 50–54 in the same schooling and two-digit occupation category, and then, cohort by cohort, comparing the log imputed wage for those retiring and those remaining in the labor force.[37] Those who continue to work typically earn 8% per hour more than those retiring in cohorts and 55–59, 4% more than those retiring in cohorts aged 60–69, and 6% more than those retiring in cohorts aged 70–79. In other words, low-wage men are more likely to retire, especially before age 60. This finding is consistent with the fact that social security rules encourage retirement most for low-earnings people and with the patterns of retirement by occupation and schooling documented by Costa (1998) and others.

37. The 1980 PUMS is used so as to have enough observations in each age–labor-force status cell and in each schooling–occupation cell.

Although the wage gap between workers and recent retirees is largest for cohorts aged 55–60, the increment to the composition bias is larger for cohorts aged 60 and over because the retirement hazard is so much larger. Figure 2 displays as crosses the age–wage profile corrected for OJT and composition biases. We see that correcting for composition bias lowers the estimated cohort market value of time (compare the crosses with the squares), especially for young and old men. Table 3 reports corresponding estimates of σ and, when compared with Tables 1 and 2, shows that correcting for composition bias in the aged-24–64 sample typically lowers estimates of σ, although the effect is quite small relative to the other adjustments I've made. Although not reported in the tables, the same is true for my estimates for the aged-25–55 and aged-24–79 samples.

3.8 OLDER WORKERS II: HEALTH AS AN INDEPENDENT DETERMINANT OF LABOR SUPPLY

Aging is associated with changes in physical capabilities, especially at older ages. Some of these changes affect labor productivity and, presuming my various measures of the market value of time are related to labor productivity, are useful sources of life-cycle wage variation. But aging may also affect the marginal disutility of work. Without a proxy for the marginal disutility of work, my basic labor-supply specifications attribute all of the age-related labor-supply changes (apart from a trend) to age-related wage changes.

I introduce several proxies for the marginal disutility of work in the specifications that include older men. One proxy is annual hours spent on *personal care:* dressing, bathing, toilet, trips to the doctor, helping another adult with personal care, sleeping, and napping. There are two reasons why one might expect this to be a good proxy for the marginal disutility of work. First, a sick and frail person who must spent extra time on these activities effectively has a shorter day to divide between work and leisure. Second, one might expect age-related changes in personal-care time to be correlated with other factors shifting the marginal disutility of work.

Men spend a substantial amount of time on personal care: 8 minutes per week on medical appointments, 268 minutes per week washing and dressing, 59 minutes per week on medical care for oneself or another adult, 2 minutes per week on other personal care, and 3455 minutes per week sleeping or napping.[38] The sum of these personal-care minutes is correlated with age, although the only subcategories with some positive correla-

38. These are averages for men aged 24–79.

tion are medical appointments, night sleep, and resting or napping (medical care for other adults is slightly negatively correlated with age).

I use four measures of physical capabilities from wave I of the Health and Retirement Survey (HRS) and the survey of Assets and Health Dynamics of the Oldest Old (AHEAD) to proxy for health status, although it is unfortunate that these health surveys occurred 15 years after my CPS and diary surveys were conducted.[39] Physical incapacities are likely to be associated with pain experienced during work activities and would thus be a reason why an older person might not work even if he were as productive as when young. Physical incapacity measures are also expected to be correlated with other determinants of the marginal disutility of work.

The first measure of physical (in)capabilities is the fraction of cohort affirmative answers to three questions:[40]

Is it very difficult or impossible for you to . . .
(i) . . . walk several blocks?
(ii) . . . climb a flight of stairs without resting?
(iii) . . . lift or carry weights over 10 pounds, like a heavy bag of groceries?

My second measure is the fraction of the cohort often troubled with pain that makes normal work difficult.[41] The third measure is the fraction of the cohort that had been an overnight patient in a hospital during the year prior to the interview.[42] The fourth is the *body mass index* [BMI = (weight in kilograms)/(height in meters)^2], which has been shown to be closely related to mortality, health, and labor-force participation [see Costa (1998, Chapter 4) for a review of the relevant literature]. Age-group average BMI and age-group average squared BMI are included in the cross-age-group regressions.

It is perhaps unsurprising that the inclusion of health measures does not substantially affect elasticity estimates, especially in the 24–64 age group. After all, health deteriorated relatively rapidly with age in the early part of the twentieth century, while gainful employment rates did

39. Neither HRS nor AHEAD surveyed many men aged 50 or less, so I assign all cohorts aged 24–50 the values of the health variables for the 51–52-year-old cohort.
40. Those respondents who say they cannot answer the question because they do not do the activity are assumed to be unable to do it.
41. AHEAD asks a slightly different question—whether pain "kept you from doing things you wanted to do" during the last 12 months. For the age groups that appear in both HRS and AHEAD, affirmative response rates to the AHEAD question are twice as high, so I cut AHEAD responses in half to make them comparable with HRS.
42. The HRS also includes measures of cognitive ability such as memory skills. I exclude these measures under the assumption that they are related to labor productivity but not to the marginal disutility of work.

not (Costa, 1998, Chapters 2, 4). For example, Costa (1998, Figure 2.3) reports gainful-employment rates of over 80% for white men aged 65 and over in 1880, 1900, 1910, and 1920 (as compared with 20 or 25% for the 1970s). At the same time she reports sizable majorities of veterans aged 65 and over in 1910 suffering from chronic musculoskeletal (67.7%), chronic digestive (84.0%), and chronic circulatory (90.1%) conditions and compares them with much lower rates for World War II veterans aged 65 and over in 1983 (47.2% with chronic musculoskeletal, 48.9% with chronic digestive, and 39.9% with chronic circulatory conditions). Costa also displays data from 1930 and 1992 showing a much steeper 1930 age gradient for risks of heart disease, arteriosclerosis, hypertension, and other chronic conditions. With an age–health gradient that is so small by historical standards, health probably should not explain much of the modern age–employment pattern. Furthermore, my measures of physical incapacity are not very high (e.g., only 13% of those aged 65 report that pain prevents them from working) and follow a pretty linear trend with age.

The only case in which the health variables make a substantial difference for the 24–64 group is when cohort hours are measured with the time diary. But this may be spurious, because diaries constrain that all uses of time—including the LHS variable (work hours) and one of the RHS variables (personal-care hours)—sum to 24 hours per day. So the final column of Table 3 reports wage elasticities from regressions including measures of physical incapacity but not personal-care hours.

While modern changes in health with age may not produce dramatic changes in the marginal disutility of work, this is not to say that health does not explain a lot of life-cycle labor supply through its effect on the wage. Pretax wages grow much less rapidly and even fall with age, which partly reflects some declining physical capacity with age and partly reflects human-capital investment decisions made in response to that declining physical capacity. Bartel and Taubman's (1979) study of four health conditions finds that, among those working during the year, adverse health reduces pretax wages twice as much as it reduces hours, and presumably some of this reduction in hours is a response to the wage. The U.S. House Ways and Means Committee (1996, Section 1, p. 5), Diamond and Mirlees (1978), and others suggest that even the implicit tax disincentives of the social security system are a response to the age–health relationship.

3.9 WITHIN-COHORT AGGREGATION BIAS

I estimate equation (10) by regressing log cohort average hours on age and average log after-tax market value of time. A within-cohort aggregation bias A_t^a, a function of the second and higher moments of the within-cohort distribution of ϵ_t and ln $\tilde{w}_t$, is an omitted variable in this regression. One check for the presence of an omitted-variable bias is to use log average after-tax market value of time as an independent variable. Doing so has little effect on estimates of σ (results available upon request) with the exception of the aged 24–79 CPS sample and, to a much lesser degree, the aged 24–64 CPS sample (see the third column of Table 3). An aggregation bias term is still an omitted variable, but may be weakly or even negatively correlated with A_t^a, so the similarity of results with average log wage and with log average wage suggests that aggregation bias is not serious.

Nor is it clear that purging aggregation bias from estimates of σ is especially interesting for macroeconomic forecasting. Shocks affecting the market value of time may also affect the distribution of ϵ_t and ln $\tilde{w}_t$ and do so in a way that is correlated with $\tilde{w}_t$ in much the same way it is in life-cycle data.

3.10 OVERVIEW OF RESULTS FROM SYNTHETIC-COHORT SAMPLES

Life-cycle data are consistent with a substantial willingness to substitute leisure over time. Hours worked by a cohort grow almost twice as fast relative to trend as that cohort's average after-tax market value of time. An important—but not the only—component of the cohort's labor-supply response is due to changes in its annual employment rate, while many previous studies of similar life-cycle data focus only on employee hours. Moreover, time-diary data suggest that the life-cycle changes in aggregate work hours are understated by CPS-type surveys. Training data suggest that life-cycle changes in the market value of time are overstated by CPS average hourly earnings. Another departure from previous studies is my consideration of behavior after age 60 or 65. Tables 1–3 review the six modifications I have made to life-cycle-based calculations of the IES. Using a sample of prime-aged working CPS males, I arrive at an estimate of 0.37 that is similar to that of Ghez and Becker (1975). Expanding the sample to include all prime-aged men increases the estimate to 0.57. Modifying this estimate by introducing older workers delivers estimates of 1.5 or higher. Instead, an elasticity of 1.36 is the result when aggregate labor supply is considered with time-diary data. If the time diaries are correct, young men slightly underreport and retirement-aged men substantially overreport annual hours in CPS-type surveys. Larger elasticity estimates are found even with CPS data when either a correction for self-financed on-the-job training or a correction for life-cycle changes in the marginal tax rate is made. Slightly smaller estimates are obtained after a correction for composition bias and an introduction of health regressors. The result of making all seven modifications—considering aggregate cohort labor supply, adding older workers, correcting for life-cycle changes in the marginal tax rate and training time, using diary data, and correcting for composition bias—produces elasticity estimates of 1 or 2. Even if one is dubious about some of my departures from previous studies of life-cycle data, it seems difficult to argue that the life-cycle data offer a powerful rejection of the ISH.

4. *A Life Event: The Termination of AFDC Benefits*

4.1 OVERVIEW AND COMPARISON WITH SYNTHETIC-COHORT SAMPLES

Clearly the value of time grows at different rates at different points of the life cycle. It is also clear that labor supply grows at different rates over the life cycle. But the discussion and analysis above shows that it is difficult to quantitatively determine how anticipated wage growth changes with age and to what degree labor supply responds to these changes as well as changes in the willingness to work. As an attempt to get better measures of anticipated changes in the value of time and corresponding changes in labor supply—changes which cannot be attributed to changes in the willingness to work or other unobservables—I look at a special life-cycle event, the termination of AFDC benefits at the 18th birthday of the youngest child.

4.2 TERMINATION OF AFDC AS A LIFE-CYCLE EVENT

For three reasons, the termination of AFDC benefits on the 18th birthday of the youngest child is an interesting life-cycle event. First, it is *fully anticipated.* Children can only get older or die, either of which eventually terminates AFDC benefits. Furthermore, since policy for the period 1970–1995 had been rather stable over time regarding the age at which benefits are terminated,[43] it is only reasonable for most of this period for families with youngest child aged 17 to anticipate termination of benefits. Since benefits are reduced to zero on the 18th birthday of the youngest child, the magnitude of the benefit reduction that a family with youngest child aged 17 might expect is easy for them to calculate: a reduction from current benefit levels to zero.

43. It can be argued that the 1996 "welfare reform" introduced substantial uncertainty about a particular family's future welfare eligibility. During the prior 25 years, the most important legislative changes were in 1981, changes which affected to some degree the magnitude of the AFDC tax on work (U.S. House Ways and Means Committee 1996, pp. 517–518).

Second, because AFDC benefits act as an implicit tax on the earnings of AFDC household heads, the event produces a quantitatively important change in the value of time. To a good approximation, AFDC and food-stamp benefits b_t for an *eligible* family at date t can be computed according to equation (2) with $\underline{b}_t = 0$, and $\bar{b}_t$ a function of calendar time, family size, state of residence, and other demographic variables. In the language of the House Ways and Means Committee, d_t are "expenses reasonably attributable to the earning of income" (including child-care costs, transportation costs, and payroll taxes) and "earnings disregards" and other deductions.[44] R_t is the "benefit reduction rate," approximately equal to 0.5 for the years 1970–1995. The benefit formula (2) thereby acts as a significant tax on work for pay, and represents a change in the value of time that is an order of magnitude larger than those teased out of synthetic cohort samples.[45]

If a work session were very short relative to the AFDC accounting period (one month) and the constraint $d_t \leqq w_t n_t$ did not bind, then labor supply might be viewed as perfectly divisible and the tax rate on an additional hour of work for otherwise eligible families would be $R_i(t)$ [see Barro and Sahasakul (1983) for a proof]. However, if participation in the labor force during an AFDC accounting period requires a discrete increase in hours, then the implicit tax rate $\tau_{t,k}$ on participation is less than $R_{t,k}$ and—to the extent that the length of a work session is similar to the AFDC accounting period—can be computed according to (5) for otherwise eligible families.[46]

While the change in the market value of time is large and anticipated, one disadvantage of a study of AFDC is that the magnitude of the change is difficult to compute exactly. An exact calculation requires information about earnings disregards, the costs of obtaining those disregards, the degree to which work is indivisible, tastes, the pretax market value of time, and other determinants of family eligibility. However, equation (5) shows that the benefit reduction rate is an upper bound on this implicit tax rate. Zero is a lower bound and applies to families whose family structure, asset holdings, or other characteristic makes them ineligible. The implicit tax rate is close to zero for a family with a high pretax market value of time.

44. U.S. House Ways and Means Committee (1996, Section 8, p. 390). This section also provides a discussion of the rules concerning deductions and how they changed over time.
45. Computing the implicit tax on work may be a difficult task (for an economist or for a welfare participant), but, since the families in my sample have youngest child aged 17, it is reasonable to expect that their experience with the welfare system will cause them to act as if they had pretty accurate estimates of the tax.
46. This formula ignores the effect of AFDC labor supply on asset holdings, which in turn affects $\bar{b}$. Mulligan (1997) shows that the effect is small.

Many studies of labor supply face the problem of estimating a person's market value of time when he or she is not working. A third advantage of studying the termination of AFDC is that we can plausibly argue that the change in the value of time associated with the termination of benefits dominates other changes over the relatively short section of the life cycle being studied. I argue that, while government policy produces an important change in the value of time, the 18th birthday of the youngest child is not associated with large and rapid changes in tastes, health, productivity, and other variables. However, it turns out that one disadvantage of this episode is that changes over time in the implicit tax rate cannot be calculated precisely.

Another disadvantage of my AFDC study is that I do not have OJT or diary hours measures. Indeed, hours reporting bias may be especially systematic before the termination of AFDC, since AFDC beneficiaries have an incentive to underreport their earnings (and hence their hours) to government agencies—although they may not have the same incentive to misreport to the PSID. If such PSID misreporting does occur, then I am likely to underestimate hours before age 18 and overestimate σ.

4.3 DATA DESCRIPTION AND TABULATION

Using the 1970–1990 waves of the PSID and including both the PSID–SRC and SEO samples, I extract all households with youngest child aged 17 and a wife or female head present. I add to these records information on the employment and family situation as well as some consumption expenditures of the wife or female household head two years earlier (when the child was 15) and two years later (when the child was 19). This main sample can be divided into two:

AFDC sample. AFDC income $>$ 0 in calendar year of the interview when the youngest child was aged 17.
Non-AFDC sample. AFDC income = 0 in calendar year of the interview when the youngest child was aged 17.

If the 18th birthday of the youngest child is not associated with large and rapid changes in tastes, health, productivity, and other variables, then we expect little change in the employment of non-AFDC women whose youngest child turns 18. Column (1) of Table 4 verifies this conjecture, showing a slight decrease in the fraction of women employed sometime during the year as their youngest child ages from 15 to 19. Perhaps this decrease is to be expected given that the women in the sample are typically in their forties and fifties and labor-force participation rates of women are declining slightly in this range [Sweet (1973,

Table 4 WOMEN'S MARKET WORK AS A FUNCTION OF AGE OF YOUNGEST CHILD[a]

	(1)	(2)	(3)	(4)	(5)
Fract. hours >0 at age 15	0.69	0.30	0.28	0.32	0.35
Fract. hours >0 at age 19	0.67	0.33	0.34	0.38	0.43
Annual hours, 15	1051	348	314	382	387
Annual hours, 19	1090	422	406	486	560
Log diff. annual hours	0.04	0.19	0.26	0.24	0.37
Annual h of workers, 15	1533	1146	1135	1192	1112
Annual h of workers, 19	1630	1283	1200	1288	1289
Log diff. ann. h of workers	0.06	0.11	0.06	0.08	0.15
Ann. h of continuous, 15[b]	1594	1264	1243	1363	1363
Ann. h of continuous, 19[b]	1679	1391	1374	1517	1517
Log. diff. ann. h of continuous	0.05	0.10	0.10	0.11	0.11
Sample	Non-AFDC	AFDC	AFDC	AFDC	AFDC
Subsample	All	All	Female head	Fem. hd., (AFDC 19) =0	Fem. hd., (AFDC 19) =0, hd. age<62
Sample size	1622	79	65	53	46

Source: Author's calculation using the PSID.
[a]Annual measures of mother's market work according to the age of the youngest child.
[b]Hours of mothers employed sometime during both of the years when the youngest was aged 15 or 19.

Table 1–4) and the author's own calculations using the 1981–1991 CPS outgoing rotation groups]. Conditional on working, annual hours worked for non-AFDC increase slightly.

Column (2) of Table 4 reports the fraction of women in the AFDC sample who were employed when the youngest child was 15 and the fraction of the same women who were employed four years later when the youngest child was 19. We see that the fraction working positive hours increases from 0.30 to 0.33. Considering that many of the families were not eligible for AFDC when the child was aged 15, the percentage change in the annual hours worked by this group is a substantial 0.19 log points. If we exclude the 14 male-headed AFDC households, the fraction increases from 0.28 to 0.34 and annual hours increase by 26%.

A few of the 65 age-17 female-headed AFDC households—12 to be exact—still receive AFDC payments after the youngest child turns 18. For all of these 12, a younger child entered the household and may be the reason for continued AFDC eligibility. If we exclude those households who do not appear to have eligibility terminated, the fraction working increases from 0.32 to 0.38 and annual hours increase by 24%.

Seven of the female household heads are old enough to be eligible for social security, although it is not clear that all of them would have had long enough work histories to be eligible. For those who do collect social security after their youngest child turns 18, the change in the incentives to work is different from those who do not. One way to separate the effect of social security is to delete all households with heads aged 62 or more when the child was 19. For those in this sample who have their AFDC eligibility terminated, annual hours increase 37%.

"Annual hours" reported in the third and fourth rows is an average across the same sample of women at two points in time, with women who do not work at all (who are 60–70% of the AFDC sample) counted as zeros. What about hours among those who do some work during the year? This is equal to annual hours of all divided by the fraction with positive hours and is reported in the sixth and seventh rows of Table 4. "Hours of workers" increase substantially in the AFDC sample, about as much as the fractions working some hours. Notice from the first two rows of the table that "hours of workers" is an average over two different samples at the two points in time, because the fraction of those working positive hours is different. Another interesting statistic is hours worked by those working in both years (annual hours of continuous workers), a statistic which is necessarily computed over the same sample at both points in time. We see even larger increases here, and it can be inferred from the table that those entering the labor force between ages 15 and 19 (i.e., those who have zero hours at 15 and positive hours at 19) work fewer hours than those who were already in the labor force and remain there.

Notice that, after the termination of benefits, the annual hours of continuous workers are quite similar to those of women in the non-AFDC sample. However, the annual hours and the fractions with positive annual hours in the AFDC and non-AFDC groups are still fairly different. Thus, although the termination of benefits has a substantial effect on work, it is clear that AFDC eligibility cannot explain all of the difference between the employment of the AFDC and non-AFDC samples.

4.4 TAX RATE AND ELASTICITY ESTIMATES

What do we learn about willingness to substitute work over time from this life-cycle event? An estimated intertemporal elasticity of substitution of work is one summary of the data which is comparable with other intertemporal studies of wage and work changes. This first requires an estimate of the change in the sample average implicit tax rate on work. I consider three estimates of the sample average τ before age 18: 0.2, 0.3, and 0.4.

At the micro level, the implicit tax rate τ is zero for families who are ineligible regardless of mother's earnings and can be bounded above by the benefit reduction rate (including the food-stamp program, 0.47 before 1981 and somewhat larger after) for otherwise eligible families. Even conditional on nonearnings determinants of eligibility, it is difficult to compute the implicit tax rate, because the tax on earnings is nonlinear, is determined by both AFDC and food-stamp rules, depends on earnings-related deductions and how they interact with other determinants of eligibility, depends on the costs of obtaining various deductions and earnings disregards, depends on the pretax market value of time and the degree to which work is indivisible, and changes with calendar time. Fraker, Moffitt, and Wolf (1985) compute average tax rates of 0.25 in 1971, 0.32 in 1979, and 0.70 in 1982, and even these are too high, because they ignore the costs of obtaining various deductions [shown as f in equation (5)]. Thus I view 0.5 as a conservative upper bound on the implicit tax rate τ for families satisfying eligibility requirements other than mother's earnings.

In order to compute a sample-average τ, we must compute the fraction of families satisfying eligibility requirements other than mother's earnings. We know that a family in my AFDC sample is eligible (or at least appears eligible to the welfare agency) when the youngest child is aged 17, because they receive AFDC income in that year. We also know that the 53 families in columns (4) and (5) of Table 4 are ineligible when the child is aged 19. What we do not know is exactly how many families are eligible when the child is aged 15. However, studies of AFDC mobility surveyed by the U.S. House Ways and Means Committee (1995, pp. 500–510) suggest that an important fraction—perhaps as much as one-half[47]—of those receiving AFDC at child age 17 would not be eligible at child age 15 because of changes in eligibility other than high earnings at that age. If so, then the sample-average age-15 implicit tax rate could eaisly be as small as 0.2. Thus I view 0.4 as a conservative upper bound on the sample-average τ.

Second, as shown in my model, we need to say something about the shape of the distribution of unobserved determinants of participation in

47. 80% of the sample of 46 families in column (5) receive AFDC sometime during the calendar year the youngest child was 15.

order to say something about how a, say, 50% wage change would affect participation rates in a group with a participation rate of 25% vs. a group with a rate of 75%. We make two assumptions in this regard—normal and uniform distributions. Then the estimated elasticity of participation is the percentage change in G^{-1} (Π_t) (where Π_t is the participation at child's age t) divided by the percentage change in the value of time (ln 0.6, ln 0.7, or ln 0.8). Table 5 reports elasticity estimates for a variety of work and tax rate estimates. An elasticity of 1 is fairly typical.

It is well known that AFDC participation involves movement into and out of the labor force [see the studies surveyed by the U.S. House Ways and Means Committee (1996)], and this fact alone strongly suggests that at least some people are willing to substitute work over time. It may even be the case that people select into the AFDC program on the basis of their willingness to substitute over time (σ) as well as income, tastes, and other variables. If so, then the elasticity estimates reported in Table 5 are not indicative of a typical person's willingness to substitute leisure over time, with the deviation determined by the importance of σ relative to income, tastes, and other variables.

AFDC is most relevant for women workers, and women may have a different willingness to substitute leisure over time [Mulligan (1995), for example, estimates larger σ's using synthetic cohorts of women]. One potential reason for a difference between women and men is that the woman is often the "secondary" and the man the "primary" household worker in a two-worker household. The bearing of children may

Table 5 INTERTEMPORAL ELASTICITY ESTIMATES

	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Fract. of full-year work at age 15[a]	0.17	0.17	0.19	0.19	0.19
Fract. of full-year work at age 19[a]	0.20	0.20	0.27	0.27	0.27
Implicit tax rate[b]	0.2	0.4	0.2	0.3	0.4
ΔG^{-1}, std lognormal	0.13	0.13	0.28	0.28	0.28
σ	0.60	0.26	1.25	0.78	0.55
ΔG^{-1}, std uniform	0.19	0.19	0.37	0.37	0.37
σ	0.86	0.38	1.66	1.04	0.73
Subsample	All		Female heads, (welf 19)=0, (hd age)<62		

[a]AFDC sample average annual hours divided by 2080.
[b]See text for computation.

be another source of differences. But neither of these differences is especially relevant for our AFDC sample, because 65 of the 79 AFDC households are headed by women and the youngest child for these women is 17.

Even if the AFDC sample were special with respect to willingness to substitute over time, this select sample might be especially interesting for business-cycle and other applications where the workers whose labor supply is changing over time may also be a select sample with respect to σ.

4.5 SAMPLE SELECTION BIAS

AFDC eligibility depends on employment status, so my AFDC sample is a select sample in two ways. First, as can be seen from the age-19 employment variables for the AFDC and non-AFDC samples, the AFDC sample selects adults whose lifetime propensity to work is low. Second, to the extent that the propensity to work varies over time in ways unrelated to the AFDC benefit formulae, my AFDC sample selects parents who happened to have a low propensity to work in the year their youngest child was 17 years old. For this reason alone, employment and earnings in the AFDC sample would tend to be higher during years the child was not 17.

However, the selection bias with respect to willingness and ability to work when the child was aged 17 is not particularly relevant to my calculations unless it introduces a differential selection bias with respect to ability and willingness to work at child ages 15 and 19. In an intertemporally separable model, it seems that the selection biases should be similar for aged-15 and aged-19 work measures.

This may not be the case if there is some willingness to substitute over time *and* labor supply is not intertemporally separable. To the extent that work (leisure) capital is accumulated by working (not working), the sample of adults not working much in the year prior to an anticipated increase in the market value of time (i.e., when the child is aged 17) is a sample of adults who, for one reason or another, do not anticipate working much following the increase. If instead, work (leisure) capital is accumulated by not working (working) as in Kydland and Prescott (1982), the sample of adults not working much in the year prior to an anticipated increase in the market value of time is a sample of adults who are "resting" in anticipation of an above-average amount of work in the following year or two. Thus intertemporal nonseparabilities introduce a differential selection bias on the aged-15 and aged-19 work variables, but without saying more about the form of the nonseparabilities, the direction of the differential bias cannot be determined.

While the tastes or nonmarket value of time of a typical woman might not change substantially at the 18th birthday of her youngest child, my AFDC sample might—because of AFDC rules regarding marital status—be one where tastes do change. Marriage decreases the likelihood of AFDC eligibility (Rosenzweig, 1995), so my AFDC sample may select women who, because of program incentives or for other reasons, are more likely to get married after their youngest child's 18th birthday. If marriage increases a woman's nonmarket value of time, then this is a force *discouraging* work after age 18, which would bias my estimates downward.[48]

4.6 INCOME EFFECTS

Since the termination of AFDC benefits at date t^* is fully anticipated, it cannot be a wealth or income effect in a model in which different periods' budget are somehow tied together. In such a model, the only determinant of the *relative* labor supply before and after t^* is the relative value of time (including, if appropriate, time discounting) and the rate at which resources can be transferred across periods.

If budgets for different periods were not tied together in any way—perhaps because borrowing and lending involve a substantial fixed cost or because individuals do not realize that such transactions are possible and advantageous—then, despite its anticipation, the termination of benefits would have an income effect as well as the usual substitution effect. Both effects would tend to promote work after AFDC. One way to test for this possibility is to see whether consumption falls at t^*. Using Skinner's (1987) proxy for nondurable consumption (a weighted average of food purchased at home, food purchased away from home, and rent or housing value), real family consumption decreases 8% ($250 a year, in 1967 dollars) from age 15 to 18 in the sample from column (5) of Table 4 with valid consumption indicators. Annual family food consumption (including food stamps) falls $100. Given that AFDC–ADC annual income falls by an average of $1121, this decrease is economically insignificant. But it is a decrease, and it should be noted that a decline in consumption together with an increase in work is not found in other life-cycle studies (Ghez and Becker, 1975; Heckman, 1974).

There are two reasons why even a large reduction in consumption at t^* might be consistent with the dynamic labor-supply view. First, consumption and leisure may not be separable. Second, consumption as I measure it is excessively sensitive to household size, since it measures rent

48. Note that there is not an income effect of marriage in the life-cycle model unless the expectations of marriage by women in my sample deviate from the fraction of them who actually get married.

and food expenditures. If, for example, part of the effect of AFDC were to keep teenagers living at home until their benefits terminated, at which time the teenager moved out, then household rent and food expenditures would fall even though each individual's standard of living might be unchanged. Unfortunatly, there are not enough households in my AFDC sample with stable composition to test this second hypothesis.

5. *Conclusions*

Following Mulligan (1998), I build a model of labor-force participation and time aggregation which, in reduced form, looks very much like the empirical models used by Lucas and Rapping (1969), MaCurdy (1981), and Altonji (1986). The model predicts that labor supply is substituted over time in response to *anticipated* wage changes, although the magnitude of the response depends on the parameter σ, which in theory can vary from 0 to ∞. The model is quite explicit about the measurement of wages and hours, including the treatment of on-the-job training, time spent searching for work, nonlinear tax rules, and the aggregation of various components of labor supply.

It might be argued that people cannot substitute over time because they are unable to choose their hours or because they cannot borrow or lend as freely as the model implies. Although my model does include the idea that people are less than free to choose their hours—labor is indivisible—it is true that the present-value budget constraints (3) are crucial for deriving the main empirical specifications. While the validity of the assumptions of the theory is open to debate, I have shown that a reasonable analysis of various age-related changes in wages and labor supply confirms one main prediction of the theory—labor-supply growth is positively correlated with anticipated wage growth.

In a broad sample of men, I find that age groups of men with 1% more age-related growth in their after-tax market value of time have 1–2% more age-related growth in group hours worked. This response is an order of magnitude greater than those found in other life-cycle studies of men. Equation (11) summarizes what I believe to be my essential departures from that literature: "labor supply" is defined to include the employment and all other margins, the market value of time is adjusted for taxes, average hourly earnings are adjusted for self-financed OJT, and CPS hours are adjusted for systematic reporting bias.

Because my inferences from the life-cycle data are so different from those made in previous studies, I expose my empirical analysis through a few simple graphs, difference estimators, and regressions. More efficient estimators are certainly available, but I want it to be clear that my point of departure from previous studies is not in the econometric details but in the very basic economic and statistical issues which are arguably of primary relevance to the problem.

Studies of individual-panel or synthetic-cohort data—mine included—cannot do much about the fact that the market value of time when not working is not observed. Someone who is not working at an instant in time may have a market value of time that is low compared with his average hourly earnings at some other point in time or with the average hourly earnings of an otherwise similar person who is working at that instant. My study of the termination of AFDC benefits upon the 18th birthday of the youngest child is an attempt to mitigate this bias, since I believe that the observable change in the after-tax market value of time dominates any unobserved changes. IES estimates from this study are quite similar to those found with the synthetic-cohort data.

Social security rules create another life-cycle event that, like the termination of AFDC benefits, is a quantitatively important anticipated change in the after-tax market value of time. This event has been included in my analysis of synthetic cohorts, although, for students of intertemporal substitution, retirement is worthy of its own study. And there have been a number of studies of the effects of social security rules on retirement decisions, with a variety of results. Cross-country studies (e.g., Gruber and Wise, 1997; Modigliani and Sterling, 1983) have typically enjoyed large and fairly obvious differences in benefit rules and have found large differences in retirement. Some time-series studies (e.g., Burtless, 1986; Hausman and Wise, 1985; Krueger and Pischke, 1992) have enjoyed less variation in benefit rules and found even smaller labor-supply responses.[49] However, there are two disadvantages of social security rules as a source of an exogenous change in the anticipated rate of wage growth. First, it is not at all clear that workers are fully aware of all the subtleties of social security rules and their changes over time, including those changes that are exploited by some of the time-series studies. Second, it can be argued that social security rules and their changes are less likely than other policies to be exogenous with respect to the situation of the people affected by the policy. The American Association of Retired Persons (AARP) is America's most powerful lobby (Birnbaum, 1997), and its preferences are certainly reflected in social security legislation. The endogeneity of social security policy might also explain the relatively large cross-country correlation between labor supply and social security rules. Nevertheless, I eagerly await a

49. Several time-series studies also utilize unanticipated policy changes, which are less relevant for the ISH.

study of retirement designed to measure the magnitude and importance of intertemporal substitution and perhaps even to overcome some of the disadvantages I've mentioned.

Another avenue for future life-cycle research may be to exploit some of the differences in wage growth across occupation and schooling categories. However, doing so is a difficult exercise if the issues I emphasize in this paper are important. It seems that the level and life-cycle rate of change of OJT varies across occupation and schooling categories [although no such variation can be detected in my fairly small data set—see Stafford and Duncan (1985)], so that the more rapid average hourly earnings growth in some categories cannot be fully attributed to more rapid growth in the market value of time. Also, the currently available time-diary data sets seem too small to correct for hours reporting biases separately by occupation and schooling categories.

I study synthetic cohorts of men mainly because there is a greater consensus in the previous literature about the lack of response of male labor supply to temporary wage fluctuations. But macro data are an aggregate of men and women, so a differential response by gender would be one reason why my synthetic-cohort estimates might not directly apply to macro modeling. I do study women in Section 4 and obtain similar estimates of σ, but it is unclear whether the AFDC women are more like the representative man or the representative woman. Like men, the AFDC women are typically heads of household and might for that reason have a lower σ than the representative woman. On the other hand, AFDC women are women. AFDC women might even have a higher σ than other women because a higher σ makes AFDC participation more attractive.

There is another reason that caution should be exercised in the application of my estimates to macro modeling. The OJT, reporting, composition, and other biases which are important in the life-cycle data may also be present in the macro data. There is no easy solution to this, since the biases may be of different magnitudes or even different signs in the two sources, but one can try to similarly correct macro data or rely mainly on earnings data (rather than hours and wage data) where we suspect some of the biases to be less important.

Although the life-cycle data suggest that the temporal pattern of work responds to the temporal pattern of the market value of time, the data do not necessarily support the hypothesis that all or even most of the variation in work across workers or over time for a given worker can be explained as a response to temporary wage fluctuations. In fact, my model allows for a number of other potential determinants of work through the indivisibility of labor as well as cross-sectional and intertemporal variation in the taste parameter γ. I only suggest that, if and when temporary fluctuations in the after-tax market value of time occur, aggregate labor supply will respond and respond in a quantitatively important way. And, regardless of the number and importance of other determinants of labor supply, my claim has important implications for macroeconomics.

Appendix A: Notes on Social Security Benefits and Work Incentives

Because social security benefits are a function of a worker's life history of earnings and labor-force participation, they can affect incentives to work at all ages. These incentives can be parsed into two categories: (1) the effect of work on earnings and the *primary insurance amount* (PIA) that will be used to compute benefits when retirement occurs, and (2) the effect of participation and earnings at age 62 and older on the fraction of the PIA to be received as a social security benefit. This appendix draws heavily on the U.S. House Ways and Means Committee (1996, Section 1) and Myers (1993, Chapter 2, 3).

A.1 WORK BEFORE AGE 62 AND THE PIA

A person's PIA is a concave function of his lifetime average earnings where each year's earnings is indexed to reflect nationwide wage growth (with the exception of earnings after age 60, which are unindexed), the average caps each year's earnings according to the payroll tax cap, and the average drops the lowest-indexed-earnings years. The second provision means that no additional benefits accrue for men earning above the cap. If the number of potential work sessions in a year (K) is small, the last provision is particularly relevant for workers who participate most of their lives, because the marginal participation decision has no affect on PIA. To see this, notice that, unless the labor-supply trend term is very different in magnitude from the rate of nationwide wage growth, the marginal participation decision occurs in a year with low indexed earnings conditional on working.[50] I refer to this as the *PIA fixed* case.

If K is large, the marginal work session will typically affect computation of the PIA, and this is the case analyzed by Feldstein and Samwick

50. Even if the marginal participation decision did not occur in a low-indexed-earnings year (conditional on working), nonparticipation would produce zero earnings, remove the year from the average, and add another year to the average, so the effect of the decision on the lifetime average would be limited to the difference between what might have been earned in the year in question and what was earned in the best year not included in the average.

(1992). For single men and for married men whose wives earn enough to have substantial social security benefits, they use the following formula to compute the effect of an additional dollar of earnings at age a on discounted expected social security benefits:

$$\frac{1}{35} e^{0.01 \max\{a_R - a - 5, 0\}} \frac{\text{PIA}}{\text{AIME}} \left(1 - \frac{\tau^{\text{IIT}}}{2}\right) S(a_R, a, r).$$

The first factor reflects the use of 35 years to compute lifetime-average indexed earnings (AIME). The second factor reflects the indexing, where a_R is the normal retirement age and a is age (Feldstein and Samwick assume individual and national earnings growth of 1%/yr). PIA/AIME depends on AIME; Feldstein and Samwick assume the median ratio of 0.32. τ^{IIT} is the marginal individual income tax rate to be paid in the retirement years, and S is the factor used to compute the discounted expected value of an annuity that begins paying at age a_R. Feldstein and Samwick do their calculations for 1992 under the assumption that benefit rules (a_R, PIA/AIME, τ^{IIT}) will never change from the values dictated by current law. In fact, rules did change between 1976 and 1992 and probably will change again. Rather than making guesses about what each age group in 1976 knew or expected about the policy and mortality parameters, I use Feldstein and Samwick's parameters. I use their highest discount rate of 6%/yr to reflect (1) expectations of the benefit cuts in one form or another that so many experts and laymen have been forecasting and (2) the associated political risk premium.

Feldstein and Samwick also perform calculations for married men whose wives do not have substantial lifetime earnings. I do not review their calculations here, but assume that roughly half of all men earning below the cap expect to be in this category when they retire and therefore take a simple average of the formula above and the increment to discounted expected benefits enjoyed by this second category of men.

In summary, my "PIA endogenous" calculations for men earning less than the social security cap average the increment to discounted expected benefits implied by Feldstein and Samwick's (1992) columns 2 and 4 of their Table 1. This increment is subtracted from the 1976 payroll tax rate.[51]

A.2 WORKERS AGED 62–71

The same PIA effects act as a subsidy to work for those 62 and over who have not elected to begin receiving their social security benefits. I extend

51. Feldstein and Samwick (1992) report the increment in 5-year intervals, which I interpolate geometrically.

Feldstein and Samwick's calculations by adjusting their annuity factor $S(a_R, a, r)$ for 6% interest as age nears 65 and, after age 65, for mortality as indicated below. Those working after age 61 are also still liable for payroll taxes.

In addition to the payroll taxes, two major provisions of social security have discouraged work for pay by the elderly: earnings limits and delayed retirement credits. The earnings limit is the maximum earnings an elderly worker can earn without losing some of his old-age benefits. The limit does not apply to those aged 70 or higher. For earnings above this amount ($2760 per year in 1976 for those aged 62–71), old-age benefits are reduced (employment status does not and did not affect Medicare eligibility, although since 1980 employment status can affect the health insurance premiums paid by an elderly person (U.S. House Ways and Means Committee, 1996, Section 3, p. 223)). The rate of reduction during the 1970s was $1 for each $2 over the limit.

For those *workers* aged 62–71 who are receiving social security, I compute the additional tax on work as 0 if he earns below the limit and 0.5 if he earns at or above it. It is important to note that these men are typically part-time workers earning a lower *hourly* pretax wage than they would have earned if they had worked full time, so the use of their pretax wage in my calculations in Section 3.2 already includes some of the disincentive effects of social security.

For workers aged 72+, no additional tax or PIA subsidy is computed, although the low hourly earnings of these men is to some degree a consequence of social security's encouraging them to work fewer hours in their sixties.

If a worker elects not to receive any old-age (or disability) benefits, he is credited with some additional old-age benefits when he later retires. These credits, called *delayed retirement credits* (DRC), are however small enough (1%) that most people aged 65–71 were effectively penalized for delaying retirement.[52] A person aged 65–71 of average health planning to retire in 1976 lost over 90% of a year's benefits by delaying retirement an additional year. However, only if K is small is this provision a *marginal* tax on work.

For someone who first retired in 1976, it is straightforward to compute the amount of these benefits, the implicit tax on nonretirement for that

52. Hurd and Boskin (1984) and others argue that the decision to retire at age 65 is not distorted, because the delayed retirement credits are "actuarially fair." This may have been true at the time of writing their paper for a person of average mortality risk, but not before 1979 when the DRC was only 1% and before 1975 when there was no DRC (U.S. House Ways and Means Committee, 1996, Section 1). The DRC for those aged 62–64 was 8.4%, which I treat as actuarially fair.

year and (assuming K is small) the implicit marginal tax on work. However, too little is known about the earnings histories of CPS men who have not yet retired to compute exactly the benefits they would enjoy if they had retired in 1976. If these potential benefits were known, the implicit tax on work (in addition to the IIT, payroll taxes, and the accumulation of SS wealth considered previously) could be computed as

$$[0.92 + (a - 62)0.005]\,\frac{b_i}{e_i}$$

where b_i/e_i is person i's ratio of potential benefits to actual earnings. The factor in brackets is greater than zero and a function of age, because the delayed retirement credits were actuarially unfair.[53] I estimate the ratio b_i/e_i as 0.4 for single men, which is a fairly typical replacement rate reported by the U.S. House Ways and Means Committee (1996, Table 1–14) for cohorts retired by 1976. The value 0.6 is used for married men over 65, under the assumption that their retirement also entitles their spouses to benefits equal to half of the amount of their own benefits.

Appendix B. Two Models of Reporting Errors

Two of my four corrections for CPS hours reporting bias rely on a model of reporting errors. The first is a *compression model* (B-1) where each worker reports hours that are closer to a stereotypical number $\tilde{n}$ than are his actual hours; the second is a *liars model* (B-2) where some apparently random workers choose to report the stereotypical number $\tilde{n}$:

$$\ln \hat{n} - \ln \tilde{n} = \theta(\ln n - \ln \tilde{n}), \qquad n > 0, \tag{B-1}$$

$$\hat{n} = \begin{cases} n & \text{w.p. } 1 - \theta, \\ \tilde{n} & \text{w.p. } \theta, \end{cases} \tag{B-2}$$

where n is true hours and $\hat{n}$ is reported hours.

Remember that, by definition, aggregate labor supply is the product of average hours conditional on positive hours and the fraction working positive hours. In order to derive the effect on aggregate labor supply of reporting bias among those reporting positive hours during some time interval, we need to model both aggregate labor supply and the fraction working positive hours (the *employment rate*). My life-cycle model pro-

53. Details of my derivation of the factor in brackets—which equals $1 - 0.01S(a_R,a,r)$—are available upon request.

vides a quite tractable model of aggregate labor, but, unless K is known, not a model of the employment rate. Even if K were known, my model of the employment rate is not particularly tractable. So, as an approximation, I assume a log-linear model for true age-group hours averaged across those with positive hours:

$$\ln n_t^a = \eta \ln \tilde{w}_t^a + u_t^a, \qquad n_t^a > 0, \tag{B-3}$$

with n_t^a the arithmetic mean of true hours among those with positive hours, and $\tilde{w}_t^a$ the age-group *geometric*-average after-tax market value of time. n_t^a is assumed to be uncorrelated across age groups with $\tilde{w}_t^a$ and with the error term in equation (10). Total age-group hours N_t^a are still determined according to equation (10). It is assumed that measured average age-group hourly earnings $\hat{w}_t^a$ differ from true average age-group hourly earnings $\tilde{w}_t^a$ according to

$$\hat{w}_t^a = \tilde{w}_t^a n_t^a / \hat{n}_t^a. \tag{B-4}$$

Up to an aggregation bias term, the compression model implies equation (B-1) for average reported age-group hours conditional on positive hours ($\hat{n}_t^a$) as a function of average true age-group hours conditional on positive hours (n_t^a). Given the compression factor θ, the reported hours elasticity with respect to the measured wage $\hat{\eta}$, and the R^2 of the reported-hours equation, we can infer the true hours elasticity η from equations (10), (B-1), (B-3), and (B-4):[54]

$$\eta = \left(\frac{\theta}{\hat{\eta}} \frac{\theta - (1-\theta)\hat{\eta}}{\theta - (1-\theta)\hat{\eta}\,(1/R^2)} - (1-\theta) \right)^{-1},$$

while the elasticity of true hours with respect to the reported wage is $\hat{\eta}/\theta$.

Given also the reported total-hours elasticity with respect to the measured wage $\hat{\sigma}$, we can infer the true total-hours elasticity σ:

$$\sigma = \eta + (\hat{\sigma} - \hat{\eta}) \frac{\eta\theta}{[1 + (1-\theta)\eta]\hat{\eta} \dfrac{\theta - (1-\theta)\hat{\eta}\,(1/R^2)}{\theta - (1-\theta)\hat{\eta}} - (1-\theta)\eta\hat{\eta}}$$

54. All of the estimated parameters are to be understood as probability limits.

The elasticity of true total hours with respect to the measured wage is $\hat{\sigma} - \hat{\eta} + \hat{\eta}/\theta$. Similar calculations (not shown here) can be made for the liars model.

Both the compression and the liars model have a parameter θ which can be calibrated by comparing reported hours in the CPS and the time diaries. For example, the micro-level standard deviation of log hours is twice as large in the diaries as in the CPS and (appropriately correcting for the different sample sizes) the standard deviation across groups aged 25–55 is almost three times as large in the diaries. This suggests a compression parameter of $\theta = 0.4$ or $\theta = 0.5$. I use a conservative $\theta = 0.7$ for the compression-model calculations reported in the text.

It is beyond the scope of this paper to build and formally test various models of reporting errors, but I do comment on their "realism." Four facts come out of an elementary comparison of the distribution of positive "hours last week" (or "usual hours last year") reported by men aged 25–55 from the March 1977 CPS with positive hours worked in the synthetic week reported by men aged 25–55 from the 1975–76 time diaries. First, close to 50% of CPS respondents report exactly 40 hours, and 75% report working exactly 52 weeks in 1976, as compared to 19% of time-diary synthetic weeks between 38.0 and 42.0 hours. When either CPS measure of weekly hours is multiplied by weeks worked in 1976, about 40% report exactly 40 hours as 1976 average weekly hours. Second, the standard deviation of log reported weekly hours is twice as large in the time diary as in the CPS. Third, the fraction of CPS workers reporting exactly 40 hours is very slightly increasing in age until age 63, after which it is about half of what it is for younger workers, although the age-group variance of log reported weekly hours increases steadily with age in both the CPS and diary studies. Fourth, the central tendency of the distribution of weekly diary hours is roughly 40 hours, with a majority of synthetic weeks between 30 and 50 hours.

The first fact is more consistent with the liars model, since the compression model does not literally predict that many would report exactly 40 hours, although the compression model might be amended to account for the first fact by introducing some rounding to the nearest integer. Both models are consistent with the second fact. The third fact suggests that the liars model is more appropriate up to age 63 and the compression model after age 63, since the compression model predicts that a widening of the distribution of true hours would decrease the fraction of CPS respondents reporting hours within a given distance of 40. However, the fourth fact is more easily derived from the compression model, since, in contrast to the liars model, the stereotype is related to the distribution of misreported hours even for those not reporting the stereotype.

Notice that the liars model also justifies the procedure used to produce results in the last two rows of Table 2 if we assume in addition that all of those reporting 40 are liars. Actually, the procedure is justified even if liars are not random with respect to their wage, as long as they are random with respect to other determinants of hours.

REFERENCES

Abowd, J. M., and D. Card. (1989). On the covariance structure of earnings and hours changes. *Econometrica* 57(2, March):411–445.

Aiyagari, S. R., L. J. Christiano, M. Eichenbaum. (1992). The output, employment, and interest rate effects of government consumption. *Journal of Monetary Economics* 30(1, October):73–86.

Alogoskoufis, G. S. (1987a). Aggregate employment and intertemporal substitution in the U.K. *Economic Journal* 97(June):403–415.

———. (1987b). On intertemporal substitution and aggregate labor supply. *Journal of Political Economy* 95(5, October):938–960.

Altonji, J. G. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy* 94(3, Part 2, June):S176–S215.

Ashenfelter, O. (1984). Macroeconomic analyses and microeconomic analyses of labor supply. In *Carnegie-Rochester Conference Series on Public Policy*, K. Brunner and A. H. Meltzer (eds.). Vol. 21, Autumn, pp. 117–156.

Auerbach, A. J., and L. J. Kotlikoff. (1987). *Dynamic Fiscal Policy.* Cambridge, U.K.: Cambridge University Press.

Barro, R. J. (1981). Output effects of government purchases. *Journal of Political Economy* 89(6, December):1086–1121.

———. (1987). *Macroeconomics.* New York: John Wiley and Sons, 1987.

———, and C. Sahasakul. (1983). Measuring the average marginal tax rate from the individual income tax. *Journal of Business* 56(October):419–452.

Bartel, A., and P. Taubman. (1979). Health and labor market success: The role of various diseases. *Review of Economics and Statistics* 61(1, February):1–8.

Baxter, M., and R. G. King. (1993). Fiscal policy in general equilibrium. *American Economic Review* 83(3, June):315–334.

Becker, G. S., and C. B. Mulligan. (1997). The endogenous determination of time preference. *Quaterly Journal of Economics*, 112(3, August):729–758.

———, and K. M. Murphy. (1988). A theory of rational addiction. *Journal of Political Economy* 96(4, August):675–700.

Ben-Porath, Y., (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75(4, August):352–365.

Bils, M. J. (1985). Real wages over the business cycle: Evidence from panel data. *Journal of Political Economy* 93(August):666–689.

Birnbaum, J. (1997). Washington's Power 25. *Fortune* 136(11, December):8.

Burtless, G. (1986). Social security, unanticipated benefit increases, and the timing of retirement. *Review of Economic Studies* 53(5, October):781–805.

Card, D. (1994). Intertemporal labor supply: An assessment. In *Advances in Econometrics: Sixth World Congress*, Volume II, C. A. Sims (ed.). Econometric Society Monograph No. 24. Cambridge: Cambridge University Press, pp. 49–80.

Coleman, T. S. (1984). Essays on aggregate labor market business cycle fluctuations. University of Chicago. PhD Dissertation. December.

Costa, D. L. (1998). *The Evolution of Retirement*. Chicago: Univeristy of Chicago Press.
Diamond, P. A., and J. A. Mirlees. (1978). A model of social insurance with variable retirement. *Journal of Public Economics* 10(3, December):295–336.
Feldstein, M., and A. Samwick. (1992). Social security rules and marginal tax rates. *National Tax Journal* 45(1, March):1–22.
Fraker, T., R. Moffitt, and D. Wolf. (1985). Effective tax rates and guarantees in the AFDC program, 1967–82. *Journal of Human Resources* 20(Spring):251–263.
Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.
Ghez, G. R., and G. S. Becker. (1975). *The Allocation of Time and Goods over the Life Cycle*. New York: Columbia University Press for NBER.
Gibbons, R., and K. J. Murphy. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy* 100(3, June):468–505.
Gruber, J., and D. Wise. (1997). Social security programs and retirement around the world. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 6134. August.
Hall, R. E. (1978). Stochastic implications of the life cycle–permanent income hypothesis: Theory and evidence. *Journal of Political Economy* 86(6, December):971–987.
———. (1980). Labor supply and aggregate fluctuations. *Journal of Monetary Economics* 12(6, Supplement, Spring):7–33.
———. (1988). Intertemporal substitution in consumption. *Journal of Political Economy* 96(2, April):339–357.
Hansen, G. D. (1985). Indivisible labor and the business cycle. *Journal of Monetary Economics* 16(3, November):309–327.
Hausman, J. A. (1985). Taxes and labor supply. In *Handbook of Public Economics*, A. J. Auerbach and M. Feldstein (eds.). Amsterdam: North-Holland.
———, and D. A. Wise. (1985). Social security, health status, and retirement. In *Pensions, Labor, and Individual Choice*, D. A. Wise (ed.). Chicago: University of Chicago Press, pp. 159–181.
Heckman, J. J. (1974). Life cycle consumption and labor supply: An explanation of the relationship between income and consumption over the life cycle. *American Economic Review* 64(1, March):188–194.
———. (1975). Estimates of a human capital production function embedded in a life cycle model of labor supply. In *Household Production and Consumption*, N. E. Terleckyj, (ed.). New York: Columbia University Press, pp. 227–258.
———. (1993). What has been learned about labor supply in the past twenty years? *American Economic Review* 83(2, May):116–121.
———, L. Lochner, and C. Taber. (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1(1).
Hicks, J. R., (1939). *Value and Capital*. Oxford: Clarendon Press.
Horvath, F. W. (1982). Forgotten unemployment: Recall bias in retrospective data. *Monthly Labor Review* 105(March):40–43.
Hurd, M. D., and M. J. Boskin. (1984). The effect of social security on retirement in the early 1970s. *Quaterly Journal of Economics* 99(4, November):767–790.
Judd, K. L. (1987). The welfare cost of factor taxation in a perfect-foresight model. *Journal of Political Economy* 95(4, August):675–709.

Juster, F. T. (1986). Response errors in the measurement of time use. *Journal of the American Statistical Association* 81(394, June):390–402.
———, and F. P. Stafford. (1991). The allocation of time: Empirical findings, behavioral models, and problems of measurement. *Journal of Economic Literature* 29(June):471–522.
Krueger, A. B., and J. S. Pischke. (1992). The effect of social security on labor supply: A cohort analysis of the notch generation. *Journal of Labor Economics* 10(4, October):412–437.
Kydland, F., and E. C. Prescott. (1982). Time to build and aggregate fluctuations. *Econometrica* 50(6, November):1345–1370.
Lucas, R. E., Jr. (1970). Capacity, overtime, and empirical production functions. *American Economic Review* 60(2, May):23–27.
———. (1972). Expectations and the neutrality of money. *Journal of Economic Theory* 4(2, April):103–124.
———, and L. A. Rapping. (1969). Real wages, employment, and inflation. *Journal of Political Economy* 77 (5, September).
MaCurdy, T. E. (1981). An empirical model of labor supply in a life cycle setting. *Journal of Political Economy* 88(6, December):1059–1085.
Mankiw, N. G. (1989). Real business cycles: A new Keynesian perspective. *Journal of Economic Literature* 3(3, Summer):79–90.
Mincer, J. (1974). *Schooling, Experience, and Earnings.* New York: Columbia University Press.
———. (1997). The production of human capital and the life cycle of earnings: Variations on a theme. *Journal of Labor Economics* 15(1, Part 2, January):S26–S47.
Modigliani, F., and A. Sterling. (1983). Determinants of private saving with special reference to the role of social security—cross-country tests. In *The Determinants of National Saving and Wealth,* F. Modigliani and R. Hemming (eds.). London: MacMillan, pp. 24–55.
Mulligan, C. B. (1995). The intertemporal substitution of work—what does the evidence say? University of Chicago, Population Research Center. Discussion Paper 95-11. June.
———. (1997). Employment after AFDC. University of Chicago. Working Paper. November.
———. (1998). Microfoundations and macro implications of indivisible labor. Federal Reserve Bank of Minneapolis Discussion Paper 126, July.
Myers, R. J. (1993). *Social Security,* 4th ed. Philadelphia: University of Pennsylvania Press.
Pencavel, J. (1986). Labor supply of men: A survey. In *Handbook of Labor Economics,* O. Ashenfelter and R. Layard (eds.). Amsterdam: North Holland.
Plosser, C. I. (1989). Understanding real business cycles. *Journal of Economic Literature* 3(3, Summer):51–77.
Rogerson, R., and P. Rupert. (1991). New estimates of intertemporal substitution: The effect of corner solutions for year-round workers. *Journal of Monetary Economics* 27(2, April):255–269.
Rosen, S. (1972). Learning and experience in the labor market. *Journal of Human Resources* 7(3, Summer):326–342.
Rosenzweig, M. R. (1995). Welfare, marital prospects and nonmarital childbearing. Paper presented at a Symposium on the Economic Analysis of Social Behavior. Fraser Institute, Chicago. December 1.

Ruhm, C. J. (1990). Bridge Jobs and Partial Retirement. *Journal of Labor Economics* 8(4, October):482–501.
Shea, J. (1995). Myopia, liquidity constraints, and aggregate consumption: A simple test. *Journal of Money, Credit, and Banking* 27(3, August):798–805.
Skinner, J. (1987). A superior measure of consumption from the panel study of income dynamics. *Economics Letters* 23(2):213–216.
Smith, J. (1977). Family labor supply over the life cycle. *Explorations in Economic Research* 4(2, Spring):205–276.
Stafford, F. P., and G. J. Duncan. (1985). The use of time and technology by households in the United States. In *Time, Goods, & Well-Being*, F. T. Juster and F. P. Stafford (eds.). Ann Arbor, MI: Survey Research Center Institute for Social Research, University of Michigan.
Sweet, J. A. (1973). *Women in the Labor Force*. London: Seminar Press.
United States Congress, House Committee on Ways and Means. (1996). *Overview of Entitlement Programs, 1996 Green Book*. Washington, DC: U.S. Government Printing Office.

Comment

ROBERT E. HALL
Hoover Institution and Department of Economics, Stanford University; and NBER

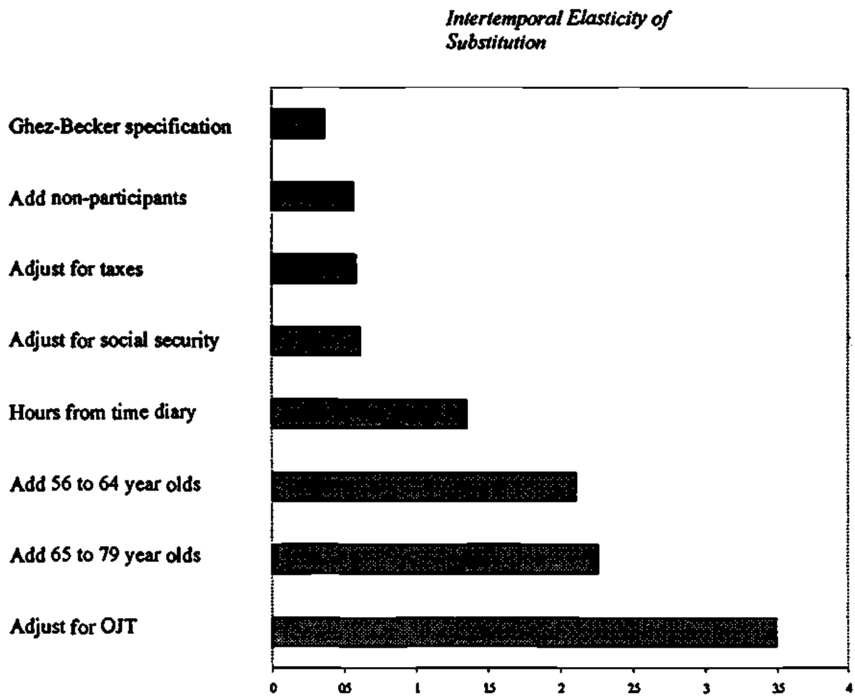
When I first read this paper, I wavered a bit from my historical position that the intertemporal elasticity of substitution is fairly low. But my wavering did not last long. The paper has not moved me much in the direction of elasticities over, say, 0.4. Still, I admire the brave attempt.

Figure 1 shows the Mulligan staircase—the dramatic increase achieved by correcting what seem to be flaws in the earlier literature deriving from Ghez and Becker's first insight that age differences reveal pure substitution effects. But the paper fails to come to grips with the basic weakness of the Ghez-Becker approach—its reliance on a strong and implausible identifying condition. That condition is that all age effects are linear in labor supply. Estimates of the intertemporal elasticity of substitution are found from comparing the departures from linear growth in hours with departures from linear growth in wage rates. Under the assumption that preferences can be related in a flexible way to age, the elasticity is not identified.

To see this, consider a general specification of the disamenity of working $L(\tau)$ hours at time τ:

$$\min \int z(\tau)\,\frac{L(\tau)^{1+1/\sigma}}{1+1/\sigma}\,d\tau$$

Figure 1



subject to

$$\int e^{-r\tau}w(\tau)L(\tau)\,d\tau = \text{present value of consumption less wealth.}$$

Here $z(\tau)$ is a weight applied to work in time τ, σ is the intertemporal elasticity of substitution, r is the interest rate, and $w(\tau)$ is the hourly wage rate. Labor supply satisfies the first-order condition

$$L(\tau) = \left(\frac{\lambda e^{-r\tau}w(\tau)}{z(\tau)}\right)^{\sigma},$$

where λ is the Lagrange multiplier associated with the wealth constraint. In logs,

$$\log L(\tau) = \sigma[\log \lambda - r\tau + \log w(\tau) - \log z(\tau)].$$

In general, if σ and $z(\tau)$ are unknown parameters, they are not identified. Ghez and Becker made the strong identifying assumption

$$\log z(\tau) = \alpha + \beta\tau.$$

It is hard to see why this is compelling. In particular, it seems likely that the disamenity of work rises sharply after age 60. Not only does this violate the identifying assumption of linearity, but it coincides with the most important variation in the data that pins down σ. A plausible view of retirement—in direct conflict with the identifying assumption—is that the disamenity of work rises in the years after age 60.

The other big problem—one Mulligan flags repeatedly—is the lack of identification resulting from sample selection issues. This is not an arcane econometric issue, and the people who raise it should not be dismissed as perfectionists who stand in the way of practical research.

Sample selection becomes particularly acute when older people are included in the sample. Ghez and Becker's restriction to preretirement age groups can be defended as an attempt to limit the sample selection problem.

Consider the following model: With probability $\pi(\tau)$, a worker suffers a major decline in productivity because of a disabling medical condition such as stroke or heart disease. The worker's wage drops below his reservation level, and he withdraws from the labor force. Over the same range, the wages earned by those workers who do not suffer the disability decline slowly with age. Consider the example illustrated by Figure 2. Mulligan's procedure will attribute the substantial declines in labor supply caused by the disability process to the small changes in wages of participants. In the example, Mulligan's regression is

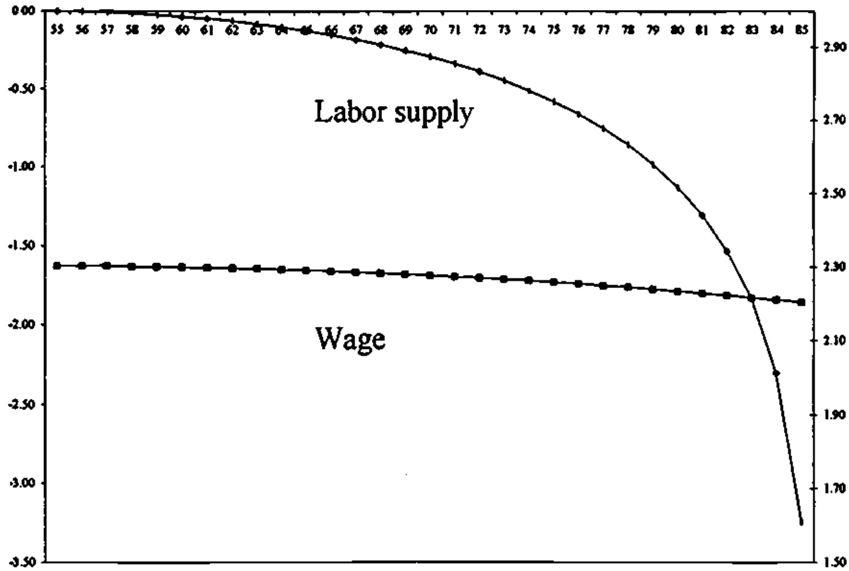
$$\log N_\tau = -107 + 44.6\log w_\tau + 0.0$$

So the estimated value of the elasticity is ridiculously high, 44.6. Large biases can result from sample selection.

These two issues of identification are much more important in Mulligan's work than in earlier work because of his extension into low-participation, high-disability groups.

I like the work with AFDC recipients much better. There is no serious identification problem either from general age-related preferences or from sample selection. But there is little surprise value in the AFDC results. The values of the intertemporal elasticity of substitution in Mulligan's Table 5 are in line with earlier work on female labor supply and are well below the surprising values of over 3 that Mulligan gets for men.

Figure 2



The stated motivation for the research is macro interest in intertemporal substitution in labor supply, and, in any case, this is the *Macro Annual.*

I've taken a pretty close look at the macro evidence in the light of Mulligan's equation (1) and get reasonably sharp estimates of around 0.3, driven both by predictable variation in the real interest rate and in wages. So, if Mulligan is right about micro cohort data, there is a problem reconciling his findings with macro behavior. There are enough reasons to question the accuracy of the macro findings so that the reconciliation could wind up favoring the micro view. But I'm far from convinced after reading this paper, or the ingenious companion coming out in the *Journal of Political Economy.*

Is the value of the intertemporal elasticity of substitution important for macro? John Campbell's careful work suggests that no value of σ delivers a reasonable story about employment fluctuations. One of the reasons may be that the standard dynamic stochastic general equilibrium (DSGE) model has so much intertemporal substitution in produced goods—through the timing of investment—that it makes little use of intertemporal substitution in labor supply. People don't work harder when goods are needed, because goods can be obtained by deferring investment.

In any case, Mulligan is pursuing an old issue at this point. The current crop of DSGE models delivers employment fluctuations in a new and more satisfactory way than intertemporal substitution in labor supply. These models consider a third use of time—job serach—and treat recessions as periods when there are an unusually large number of people searching. This seems more promising than the intertemporal substitution model, where recessions are times when people are substituting toward more leisure.

Comment

Jörn-Steffen Pischke
Massachusetts Institute of Technology

1. Introduction

Since the active research on intertemporal labor supply during the 1970s and 1980s, things have been mostly quiet in this area. Various reviewers have concluded that the intertemporal elasticity of substitution for hours is small for men, probably in the range from zero to one, and likely below 0.5. Many of us have come to think that this assessment is unlikely to change dramatically. When a central question has been laid to rest, it is certainly useful when someone comes along every now and then and stirs up some dust. Casey Mulligan does a remarkable job of stirring in this paper, reporting elasticities as high as 6.5. This is particularly noteworthy in that he focuses on variation in hours and wages over the life cycle, an area where large elasticities have been especially hard to find. Once the dust has settled, however, I suspect we will be back much closer to the previous range. I will organize my comments around the two empirical exercises in the paper: the estimates with synthetic-cohort data and the termination of AFDC benefits.

2. Synthetic-Cohort Data

The intertemporal substitution hypothesis says that expected hours should respond to expected, or evolutionary, changes in wages. One reason for wages to move systematically, and clearly well known among workers, is the tendency for wages to rise with age. This low-frequency variation in hours should be of particular interest to macroeconomists: if people are willing to substitute leisure over the phases of their life cycle, they are likely to be willing to substitute over business-cycle frequencies as well. Mulligan presents regressions of the logarithm of hours on the logarithm of wages by age group, an exercise which has been carried out many times before. The paper discusses six potential adjustments to the standard exercise. The main contribution is his attempt to put actual data to use in estimating the effects of these adjustments. The six adjustments are (1) including older workers in the estimation, (2) including nonemployment as an hours choice, (3) using data from a time-diary study as a potentially more accurate source of hours information, (4) incorporating on-the-job training in the calculation of the hourly wage, (5) allowing for taxes, and (6) adjusting for composition bias. All of these are potentially important adjustments, and I found it instructive to see what difference they make. To facilitate my discussion of these six adjustments, I have created a data set of mean hours and wages by age group from the 1980 Census comparable to Mulligan's CPS sample. Quantitatively, the most important adjustments are the first four, so I will discuss them in some detail.

2.1 INCLUDING OLDER WORKERS

Both wages and hours vary relatively little for workers between the ages of 35 and 55. Thus, in standard analyses of life-cycle data, most of the variation in hours and wages comes from the difference between prime-age and young workers. Since hours and wages differ for older workers as well, this seems potentially a good reason for exploiting this additional variation in the data. In practice, I am rather skeptical that the hours choices of older workers teach us much about the intertemporal substitution elasticity. Mulligan tries to address two potential problems for workers past the retirement age: the fact that social security changes the incentives workers face, and the problem that the older workers for whom we observe a wage are a highly selected group. Mulligan uses data from the 1980 Census by assigning older workers the wages of younger full-time workers in the same occupation and schooling group and then comparing the imputed wages of those retiring with those who continue working. This calculation of the composition bias adjusts for selective retirement between groups but not for possible biases arising from within group selection.

The fact that the social security system changes the after-tax wage is incorporated in the tax calculations in the paper. But Mulligan acknowledges that it is difficult to collapse the multitude of constraints and implicit taxes into a single tax rate relevant for the cohort averages. More importantly, the tax calculation does not incorporate the fact that workers are first eligible to receive social security benefits at age 62 and Medicare at age 65. This omission is perfectly consistent with the model, because benefit availability should not matter for the hours decisions of a "life-cycling" worker, since benefits are part of life-cycle wealth. But social security and Medicare wealth are hard to borrow against to finance leisure earlier in life, and prudent consumers tend to be reluctant to make use of the borrowing opportunities that are available. This constraint is relevant, because many individuals do not own substantial assets other than durable goods even at ages close to retirement. The median financial net worth of households with a male head age 55 to 61 in the 1983 Survey of Consumer Finances is $5,000. Including housing wealth, the median net worth is $94,000. I believe that some of the spikes in retirement rates at ages 62 and 65 are explained by the presence of the social security system. Costa's (1995) work is a case in point. She finds that the availability of army pensions for Union veterans after the Civil War substantially changed the retirement behavior of those who qualified for these pensions, suggesting that the presence of old-age benefits matters for retirement decisions and that the effect does not operate through wages.

Mulligan discusses another major difference between the old and the young, which is the fact that the young are healthier. This is important to the degree that health affects tastes for leisure directly. He finds little empirical evidence for a direct effect of health on hours choices. He also cites historical evidence to support his claim that health may not have much of a direct effect on labor-supply behavior. The fact that older men seemed to be in much worse health around the turn of the century but worked more than the elderly do now is hardly convincing evidence that health does not influence the taste for leisure. Americans are far richer now and therefore may find it far easier to finance leisure at times when it is particularly valuable. The social security disability and old-age pension programs make this easy even for individuals who have not provided for these circumstances on their own.

I am also unconvinced by the evidence presented by Mulligan on the effect of health on hours choices. The Census also asks questions about disabilities which limit the household member's ability to work or which prevent him from working. These disability questions are available for all respondents, and therefore they do not have to be imputed for the age groups below age 51 as with data used by Mulligan. In the first three rows of Table 1 I report the estimated elasticities from Mulligan's paper (column 1) and my reestimates with the Census (column 2) for the various age ranges he looks at. These estimates use hours for all men, not just those working, since much of the variation in hours for older men comes from the employment margin. I will have more to say on the distinction between workers and nonworkers below. The point estimate

Table 1 ESTIMATES OF THE INTERTEMPORAL ELASTICITY OF SUBSTITUTION FOR VARIOUS AGE GROUPS AND SPECIFICATIONS

			Census estimates		
Row	*Age Group*	*Mulligan's estimates (1)*	*No Additional Controls (2)*	*Controlling Disability Limits Work (3)*	*Controlling Disability Prevents Work (4)*
			Geometric-Average Wage		
1	25–55	0.57	0.59	0.49	0.46
			(0.02)	(0.03)	(0.03)
2	24–64	1.41	1.11	0.30	0.18
			(0.11)	(0.20)	(0.12)
3	24–79	2.86[a]	2.91	3.67	1.73
			(0.18)	(0.35)	(0.45)
			Arithmetic-Average Wage		
4	25–55		0.59	0.50	0.46
			(0.03)	(0.05)	(0.04)
5	24–64		1.11	−0.02	0.02
			(0.15)	(0.20)	(0.13)
6	24–79		3.60	2.04	−0.47
			(0.47)	(0.94)	(0.56)

Each entry represents the coefficient from a separate regression of the log of mean annual hours for an age cell on the mean of the log of the hourly wage for the age cell (in rows 1–3) or on the log of the mean of the hourly wage for the age cell (in rows 4–6). All regressions also include a constant and a linear term in age. Regressions in column (3) also control for the fraction of the age group reporting a disability which limits the kind or amount of work a person can do. Regressions in column (4) also control for the fraction of the age group reporting a disability which prevents a person from working on a job. All cell means are computed from the 1980 Census of Population 5% PUMS. Standard errors are reported in parentheses.

[a]This estimate is not reported in the published version of Mulligan's paper and is taken from a previous draft.

for the age group 25–55 is almost exactly the same, while our estimates differ a bit for the samples including older men.

If I include the fraction of an age group who report a disability which either limits or prevents work (columns 3 and 4 of Table 1), some of the estimated elasticities fall substantially. Unlike Mulligan, I find no significant intertemporal substitution effects for the age group 24–64 once health status of the age group is controlled for. For the age group 24–79, the estimated elasticities remain substantial, but the estimates differ by a factor of 2 depending on the exact definition of the disability variable. This is despite the fact that the raw correlation of the two disability measures in the sample is 0.99! This makes me quite skeptical whether we are able to control adequately for the effect of health in these regressions.

My skepticism is heightened by the fact that these regressions seem rather sensitive to changes in the functional form used. When the arithmetic average of the wage (i.e. the log of the average wage) is used as a regressor instead of the geometric average (i.e. the average of the log wages), I find substantially lower estimates for the elasticity once older groups are included. This result is again in contrast to Mulligan's estimates in his Table 3. The difference between the two wage measures results from the fact that the geometric-average wage declines much more strongly for workers over age 60, thus producing an age–wage profile which tracks the age–hours profile much more closely when older workers are included. This means that there is less need for the health indicators to explain the behavior of hours for the elderly. While the specification using the geometric-average wage may be more sensible, I find it more comforting that these functional form issues are not of importance for the age group below 55.

Mulligan concludes that retirement might be a life event particularly suitable for the study of intertemporal substitution. I find the complications introduced by using older workers overwhelming. Since the results are very sensitive to important specification issues, I feel that it is much safer to rely on the age group below 55. I will focus the rest of my comments on them.

2.2 INCLUDING NONEMPLOYMENT

Not working at all during the year is a valid choice of hours, and studies of intertemporal substitution have often neglected this margin because we do not observe the wages of nonworkers. But the male employment rate is clearly hump-shaped over the life cycle. In 1980, it was 93% at age 25, peaked around 95% for men in their thirties, and drops to 86% at age 55. Thus, the employment rate varies most for workers in their late forties and fifties, a period when wages are rather stable. The combination implies that including nonworkers raises the estimated substitution elasticity substantially because a little wage variation has to account for a lot of hours variation. But I doubt that this is the full story. An important reason why men do not work during their prime-age years (apart from unemployment, which rarely lasts an entire year) is, again, health. The fraction of workers reporting a disability which prevents them from working rises from 1.4% at age 25 to 10% at age 55. In fact, this fraction has a correlation with the employment rate of −0.95. Rows 1 and 2 of Table 2 compare the estimates of the intertemporal substitution elasticity for working men and for all men including nonworkers, controlling for

Table 2 ESTIMATES OF THE INTERTEMPORAL ELASTICITY OF SUBSTITUTION FOR THE AGE GROUP 25–55 FOR VARIOUS SPECIFICATIONS

			Census Estimates		
Row	*Specification*	*Mulligan's Estimates (1)*	*No Additional Controls (2)*	*Controlling Disability Limits Work (3)*	*Controlling Disability Prevents Work (4)*
1	Employed workers' hours and wages	0.37	0.36 (0.02)	0.38 (0.03)	0.37 (0.03)
2	All men's hours, employed workers' wages	0.57	0.59 (0.03)	0.50 (0.05)	0.46 (0.04)
3	As in row 2, but using CPS ORG usual weekly hours		0.56 (0.02)	0.52 (0.05)	0.49 (0.05)
4	As in row 2, but using CPS ORG weekly hours last week		0.61 (0.03)	0.59 (0.05)	0.55 (0.05)
5	As in row 4, but using CPS ORG hourly earnings		0.79 (0.04)	0.62 (0.05)	0.61 (0.05)
6	As in row 4, but using CPS ORG hourly earnings adj. for overtime		0.83 (0.05)	0.64 (0.04)	0.61 (0.05)
7	As in row 6, but wages adj. for learning by doing using Shaw est.		0.44 (0.02)	0.38 (0.02)	0.38 (0.03)

Each entry represents the coefficient from a separate regression of the log of mean annual hours for an age cell on the log of the mean hourly wage for the age cell. All regressions also include a constant and a linear term in age. Regressions in column (3) also control for the fraction of the age group reporting a disability which limits the kind or amount of work a person can do. Regressions in column (4) also control for the fraction of the age group reporting a disability which prevents a person from working on a job. All cell means are computed from the 1980 Census of Population 5% PUMS. Standard errors are reported in parentheses.

the fractions reporting a disability.[1] These health variables do not matter for men below age 55 who are working (row 1), while they reduce the estimates somewhat once we use hours for all men.

2.3 MISREPORTING OF HOURS AND WAGES

Answers to "usual weekly hours" questions in the Census Bureau surveys are well known to produce spikes at 40 hours, particularly for men. It seems plausible to me that some rounding by respondents masks actual hours fluctuations over the life cycle. I also agree with the idea of including nonwork investment activities, like job search in the measured

1. The estimates in Table 2 use the arithmetic average of the wage. I switch to this variable here because this makes some of the later adjustments I wish to make much easier.

hours. Hence, the time-diary data which Mulligan uses should in principle allow better inferences on the intertemporal substitution elasticity.

Unfortunately, the available time-diary sample is small and therefore produces noisy means for the hours measure. This is very visible in Mulligan's Figure 1. A salient feature in these data among the group 25–55 years old are two outliers with low hours of about 1200 at age 54 and about 900 at age 55. The observations for the next few ages are higher again and lie between 1500 and 2000 hours. The data imply that the typical man at age 55 who is employed at all worked less than 20 hours a week. I find this implausible, given that the same age group reports *more* than 40 hours in the CPS and Census. I doubt that the average worker in this group spends more than 20 hours a week at the doctor or having coffee breaks. I stress this issue because it is these two observations which more than double Mulligan's estimate of the intertemporal elasticity of substitution. When I replace annual hours for 54-year-olds by 1200 and for 55-year-olds by 900 in the Census data, my estimates change from 0.59 to 1.32 for the specification using hours for all men including nonworkers. Comparing this with Mulligan's numbers of 0.57 and 1.36 makes clear that these two data points drive the result. Moreover, the *t*-statistic on the substitution elasticity falls from 23 to 4 when I make this change, indicating that OLS agrees with me that these two data points are very different from the rest of the sample.

Realizing that sampling error is important in the diary data and may affect the estimates, Mulligan uses three other methods to adjust the CPS data for possible reporting bias. He assumes that respondents may report hours between the norm and the truth (the compression model) or that some respondents report the norm rather than the truth (the liars model). Both models are plausible, but their use requires either knowledge of or an assumption about the degree of misreporting. Mulligan gauges this quantity by comparing the standard deviation of hours by age group in the CPS and in the diary data for the ages 25–55. Of course, we know that for those ages the two outliers for 54- and 55-year-olds in the diary data will have a big effect on the estimated standard deviation. Therefore, these exercises do not address my concern with the diary data either. The final method to adjust the CPS data is to discard everybody reporting 40 hours. This idea relies on the strong assumption that every report of 40 hours is erroneous. If this is incorrect, and some men actually work 40 hours a week, even when their wages change, then this adjustment will lead to an overestimate of the intertemporal substitution elasticity. Since there are important institutional reasons which imply that 40-hour responses may be roughly correct, we have reason to doubt these estimates.

Coffee breaks apart, why do weekly hours fluctuate for male workers? We know that part-time work is rare for men before they reach retirement age. For salaried workers, hours may vary, but this variation should have no particular effect on pay during the current period. A potential source in variation in weekly hours that remains is overtime by hourly workers, even though Mulligan's indivisible-labor model ignores this possibility if a work session is a day or longer. The questions in the Census or CPS annual demographic supplement are ill suited to pick up these variations. So I went to the 1979 merged Outgoing Rotation Group (ORG) files of the CPS and recorded data on "hours worked last week." These data come from throughout the year, and recall bias for this question should be much less important than in the surveys asking about the previous year. Thus, these data are likely to give a more accurate picture of deviations of hours from their usual level, although I agree that they are inferior to the continuous recording done in diary studies. I constructed annual hours by multiplying the CPS hours last week for the age group with their weeks worked from the Census. I also did the same construction with the "usual weekly hours" variable from the CPS. This latter variable should give an answer very similar to the annual hours from the Census. The result is shown in row 3 of Table 2, and a comparison with row 2 reveals that this is in fact true for the estimated elasticity, although usual hours from the CPS tend to be somewhat lower than usual hours from the Census. Using hours last week raises the estimated elasticity a bit, but not dramatically (row 4). Like the diary hours, using hours last week produces an annual hours measure which is below the Census measure for older workers, but there is not much difference for younger workers. In addition, the differences are not large: about 50 hours a year for workers aged 55. Thus, I believe that the adjustment due to the diary hours goes in the right direction, but I am not convinced of the magnitudes shown by Mulligan.

If hours are measured incorrectly, the measurement of the wage is also affected. Mulligan tries to correct the wage measure by using the diary hours to compute the hourly wage and instruments this measure with the wage constructed from the CPS alone. Since the first stage of this regression does not fit very well, the whopping elasticity estimate of above 6 is not very informative. The sampling distribution of the just-identified IV estimator has fat tails when the correlation of the instrument with the endogenous regressor is low, so that it easily produces crazy estimates.

Combining information from the Census and the CPS allows us to exploit the superior wage measure from the CPS. I calculated hourly wages as earnings last week divided by usual weekly hours for salaried workers and used the reported hourly wage for hourly workers. At least for hourly workers, this wage measure should be much more accurate than dividing annual earnings by annual hours, even those collected from a diary study. Row 5 in Table 2 shows that using this CPS wage measure raises the estimated elasticity by 10 to 30%. Once we are considering hours fluctuations due to overtime, we should also adjust wages for this. An hourly worker working more than 40 hours a week has to be paid time and a half for the overtime hours according to the Fair Labor Standards Act. For a work session where a worker is in the overtime range, this becomes the relevant wage for the decision about hours. I assume that hourly workers report the straight-time wage in the CPS and label every hourly worker reporting more than 40 hours last week as working overtime (thus neglecting multiple job holding). The wage of these workers is 1.5 times the reported wage. This adjustment makes little difference, as can be seen in row 6 of Table 2.

2.4 ON-THE-JOB TRAINING

Not all hours at work are actually spent in productive activities. Some time may be spent accumulating more human capital. This means that the wage for an hour of actual production is higher than the hourly wage rate reported by a worker. Since accumulation of human capital tends to be concentrated during the early years of a worker's life, making this adjustment flattens the wage profile, and therefore increases the estimated elasticity of substitution. Mulligan's Table 1 shows that the elasticity estimate rises from 0.57 to 0.61 when training hours from the time-diary data are used, or to 0.76 when the estimated hours from Heckman, Lochner, and Taber (1998) are used. I am actually surprised that either adjustment makes so little difference.

A number of assumptions go into this adjustment. We should only deduct hours of work which are not compensated by the employer. Thus, hours spent in firm-specific training, a large part of which may be paid for by the employer, should not necessarily be deducted. The assumptions in the paper with respect to the issue are actually relatively conservative, since most on-the-job training seems to be general (see Loewenstein and Spletzer, 1997). However, there is some empirical evidence suggesting that employers also seem to pay for a good part of general training (see, e.g., Barron, Berger, and Black, 1997). Acemoglu and Pischke (1999) provide a theoretical rationale for this finding: If labor markets are not competitive, and general training raises the rents which employers may capture from the workers, then employers are willing to invest in general skills. As a consequence, I suspect that Mulligan's adjustment is an upper bound for the effect of general training, because employers may well pay their workers even for most hours spent in general training.

Mulligan works with the standard training model of Gary Becker and Jacob Mincer, in which learning only takes place when the worker does not engage in productive activities. An alternative, and probably complementary, approach is the learning-by-doing model, which implies that it is an additional hour actually spent at work which makes the worker more productive. This also means that the relevant value of an hour of work has to be adjusted upwards when skill accumulation takes place. The relevant price is not just the current wage but also the discounted effect of the additional hour on all future wages, weighted appropriately by the hours worked in the future. We would expect that this has the same effect of making the correct wage profile flatter and therefore raising the estimated elasticity of substitution. The implications of this model for intertemporal substitution have been studied by Shaw (1989). However, she allows the effect of an hour of work today on wages next year to depend on the current wage. She finds that a high current wage enhances learning by doing, so that her estimates actually imply a steeper and more concave profile for the true value of an hour of work, because most learning by doing takes place during the high-wage years when individuals are in their late thirties and forties. Using her estimates of the human-capital production function, I imputed the corrected value of an hour of work using the cohort means from the Census and CPS. This is a crude way of doing things. It would be much preferable to do this calculation with the micro data and aggregate up, because the relationship between the wage next period and today's hours is rather nonlinear. This can only be done with a panel, so I ignore Jensen's inequality and feed average hours and wages into this function anyway. The results are displayed in row 7 of Table 2. The intertemporal elasticity of substitution falls by over a third. I do not want to defend these estimates as a better adjustment of the effects of training. I rather view them as an illustration of what can happen when some different estimates on human-capital formation are used in place of the ones Mulligan focuses on. I think we have too little empirical knowledge about the form of the human-capital production function and the relative importance of different channels of on-the-job training, so that it remains rather unclear whether adjusting for training actually raises or lowers the elasticity of substitution.

Career concerns or rat races among young workers generate behavior very similar to the learning-by-doing model. If more hours mean higher output, and employers use output to make inferences about the ability of workers (to set future wages), then an additional hour worked today will result in a payoff in the future. Since career concerns become less important with age, hours choices are more distorted for the young. Unfortunately, there seems to be no simple way of adjusting for the resulting distortion in the hours. This is because in equilibrium employers take the behavior of the workers into account in setting wages. They reward inferred ability, not the fact that associates in law firms and assistant professors have tried to jam the signal on ability by overly hard work. This means that, without a lot of structure, observed wages and hours by themselves will contain no information on how important career concerns are. But career concerns likely imply a higher intertemporal elasticity of substitution than we will typically estimate.

2.5 OTHER ADJUSTMENTS AND SUMMARY

The remaining adjustments are empirically much less important, at least in the group of workers below retirement age. One is for taxes. Reported wages are before taxes. As an individual's earnings rise with age, he or she gets pushed into higher tax brackets. So the after-tax wage profile will be flatter, once more generating a (slightly) higher elasticity.

Since we included nonworkers in the hours calculation, we have to address the potential composition bias resulting from the fact that we do not observe wages for these workers. This bias may differ by age. I like the way Mulligan uses panel data to get at the composition bias. This is again most important for older workers, because employment rates at young ages vary less dramatically. I have followed an alternative strategy myself, which is to assume that nonworkers are individuals with low potential earnings. Consequently, I assigned them the 10th-percentile wage within their schooling, race, and age cell. Like Mulligan's adjustment, this approach also produced lower elasticities, but the difference was equally minor.

Overall, I agree much more with the ideas underlying the specific adjustments made to the life-cycle estimates by Mulligan than with some of the empirical estimates he obtains. I find the results from the adjustments which make the biggest difference, using older workers and using diary data, the most dubious. I agree that many of the adjustments will raise the estimated elasticity. But based on those where we have reliable data, the elasticity seems to be more like 0.6 than 0.3, not out of the ballpark of previous estimates. All this said, it still remains an open question whether the hours choices of individuals over their life cycle really reflect intertemporal substitution.

3. *The Termination of AFDC Benefits*

The second empirical exercise in the paper is very different from the first. It examines the change in labor-supply behavior of women when their AFDC benefits end because the youngest child turns 18. AFDC rules impose a substantial tax rate up to 100% on the earnings of benefit recipients, thus making work much less attractive before the youngest child is age 18. In order to estimate the elasticity of intertemporal substitution implied by this change in effective wages, Mulligan compares the hours of women who received AFDC benefits when their youngest child was 17, at the times when the child was 15 and 19. For example, in the most restricted sample, average hours go up by 37 log points. Wages have changed by $-\ln(1 - \text{tax rate})$, or -0.51 when the tax rate is 0.4. Dividing these two numbers, you get the substitution elasticity of 0.73 in the bottom right-hand corner of Mulligan's Table 5. Some mothers in the sample will have received AFDC when their child was 15, and some will not. This means that the change in the implicit tax rates differs for these women. Mulligan assumes an average tax rate for the whole group. It might seem preferable to use the micro-level heterogeneity in these tax rates to estimate the effect of intertemporal substitution. But these tax rates depend on labor-supply behavior, and Mulligan's method filters out the variation which is solely due to the AFDC rules. This is effectively an application of instrumental variables.

What is unfortunate is that we are not given more information about the average tax rate, because the estimated elasticity depends fairly strongly on the assumed rate. This could be done in principle by calculating the actual benefit reduction rate faced by the individuals in the micro data. As a first approximation, I would neglect complications such as the costs of obtaining earnings disregards in this exercise. Nevertheless, the PSID probably does not have enough information to do this, because surveys are only done once a year. The Survey of Income and Program Participation (SIPP) collects monthly data on employment, income, and participation in government programs and might be more suited for such an analysis. It would also allow the researcher to incorporate participation in other government transfer programs which are tied to the presence of a child in the household, primarily housing benefits. Burtless (1990) suggests that the effective marginal tax rate resulting from the combination of programs might well be as high as 0.8 or 0.9, much above the base value of 0.5 assumed by Mulligan. I therefore suspect that the more appropriate calculations in Mulligan's Table 5 are the ones with the higher tax rates, producing the lower elasticities.

Uncertainty about the estimated elasticity also results from the small sample sizes in the PSID (unfortunately, I suspect that pooling all the SIPP panels would yield just about as many, or as few, observations). For the employment rates in Mulligan's Table 4 it is possible to calculate standard errors from the information in the table. For the most restrictive sample in column 5, the employment rate when the child is 15 rises from 0.35 with a standard error of 0.07 to 0.43 with a standard error of 0.06, not a significant change. I imagine that the magnitudes of standard errors on the hours would look similar. It would therefore be useful to have these results corroborated in other samples before we draw strong conclusions.

Even if we accept the finding as a fact that women raise their annual hours by 37% when their AFDC benefits end, there is still the question of whether this is really a response to a change in the value of work, and therefore due to intertemporal substitution. Looking at consumption changes to corroborate this interpretation is clever. Mulligan finds that women reduce their food and housing expenditures by 22 cents for each dollar of AFDC income they lose, close to Gruber's (1996) finding of 30 cents based on within-state changes in benefit levels. The benefit changes Gruber analyzed should produce wealth effects, since they were presumably unanticipated by recipients. Thus, the consumption response in Mulligan's sample seems rather large.

But alternative explanations to forward-looking behavior are consistent with a small consumption response as well. Other sources of income, such as help from family and friends, may replace the previous AFDC income and therefore help smooth consumption. The child itself may start going to work between age 15 and 19 and contribute to the income of the family. Some mothers start work or raise their hours, as we have seen, and their earnings help keep their consumption up. But this behavior may be purely myopic and have nothing to do with intertemporal substitution. Another potential test of a forward-looking versus a myopic explanation would be to compare the consumption and labor-supply responses of mothers in states with different benefit levels. The benefit level should not matter for the intertemporal substitution story (only the tax rate matters, which is set federally). But women in a high-benefit state like California may be in for more of a shock when their benefits run out than their compatriots in a low-benefit state like Texas. Unfortunately, I cannot think of a data set large enough to generate decent sample sizes for this by state. But until I see some further evidence, I also remain skeptical that what we see in the behavior of welfare mothers is intertemporal substitution.

REFERENCES

Acemoglu, D., and J.-S. Pischke (1999). The structure of wages and investment in general training. *Journal of Political Economy* 107(3), forthcoming.
Barron, J. M., M. C. Berger, and D. A. Black (1997). *On-the-Job Training*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
Burtless, G. (1990). The economist's lament: Public assistance in America. *Journal of Economic Perspectives* 4(1):57–78.
Costa, D. (1995). Pensions and retirement: Evidence from Union army veterans. *Quarterly Journal of Economics* 110(2):297–319.
Gruber, J. (1996). Cash welfare as a consumption smoothing mechanism for single mothers. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 5738.
Heckman, J. J., L. Lochner, and C. Taber (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1(1):forthcoming.
Loewenstein, M. A., and J. R. Spletzer (1997). General and specific training: Evidence and implications. Bureau of Labor Statistics. Manuscript.
Shaw, K. L. (1989). Life-cycle labor supply with human capital accumulation. *International Economic Review* 30(2):431–456.

Discussion

John Shea suggested that the exclusion of family size and composition from the analysis could be a significant omission. He also pointed out that the countervailing income and substitution effects of wage changes on labor supply weaken the *a priori* case for a high elasticity of labor supply. He mentioned evidence from studies of cab drivers, whose behavior apparently reflects a substitution effect but also an income effect, as reflected in a weekly income target. Mulligan replied that his model can give new insights into the interaction of substitution and income effects, noting for example that, in his model, an increase in the real wage will lead individuals to increase the number of work sessions but reduce average session length.

David Laibson raised the issue of liquidity constraints, the existence of which may confound efforts to estimate the elasticity of intertemporal substitution. For example, labor supply peaks in midlife, which is also the time when family consumption needs are highest.

John Cochrane pointed out possible selection problems in the use of the sample of elderly workers. For example, the observed decline in hours and wages after age 65 could reflect a situation in which only lower-productivity individuals continue to work, as they cannot afford to retire.

Responding to a comment by Pischke, Mulligan argued that the way that human-capital accumulation is financed isn't critical; what is important is that the rate of accumulation is higher for younger workers, which should be interpreted as a higher implicit wage for those workers. More generally, Mulligan stressed that it was not the intention of his paper to consider all possible explanations for observed lifetime patterns of hours and wages. Rather, the idea was to see how far we can get by extending earlier research that takes the life-cycle model seriously.