1 Introduction

Research on intertemporal labor supply decisions is very active, partially because of the availability of better data that track individuals over a long time and partially because of improvements in computational methods and resources, making estimation of complicated dynamic models feasible. Some fundamental issues are:

- The theory generates hypotheses and parameter restrictions that can and need to be tested using data. One question is if the proto-typical model can explain the rich set of stylized facts about labor supply over the lifecycle that have been established over the last few decades and that seem to hold across time and countries (see for example Card (1994) and the paper by Bick, Lagakos and Spitz-Oener (forthcoming in the AER) which we discussed briefly in the first class. Most importantly, average hours worked per individual increase strongly between the ages of 20 and 30, remain roughly constant until the age of 55 and drop sharply afterward. Similar facts hold for labor market participation rates. At the same time, life-cycle profiles of real wages are concave, with large increases early in the life-cycle and a slight decrease before retirement.

- These empirical regularities are qualitatively remarkably similar across countries, though the timing of labor market entry and retirement vary substantially.

- In particular, while labor supply shows nearly identical behavior among individuals in the stable part of the life-cycle (approx. age 35-55) in most industrialized countries, with almost constant hours of work (conditional on working), gradual though slow wage growth, and stable labor force participation rates, there are larger differences along the extensive margin between countries. The age of labor market entry and exit vary dramatically, possibly because differences in the education system and the retirement benefit system. Labor force participation rates of women differ across countries, too.

- Individual and aggregate labor supply vary considerably over time. In particular, business cycle fluctuations correlate strongly with hours worked, both on the individual and the economy-wide level. Can a dynamic neo-classical model of labor supply reproduce such patterns?

- A dynamic model of labor supply can be applied to various dimensions of policy analysis. The general question is how workers adjust their behavior to policy changes and other changes in the economy. As an example, many believe that taxes on labor income lead to strong distortions in the economy. Some macro-economists, among them Nobel-Price Laureate Edward Prescott, argue that higher long-run unemployment rates and lower hours of work observed in European countries relative to the US and Canada can be explained (or are caused) by the relatively higher taxes on labor income in the former countries.
• Retirement is a particularly strong manifestation of intertemporal decision making: Employees work a lot early in the life-cycle to accumulate wealth that can be used to finance their retirement. Consequently, one needs to have an intertemporal model of labor supply to study the effects of policies that change the retirement system.

The static model of labor supply describes a world in which workers make a one-shot decision about how much to consume and how much to work. Time does not play a role. There is no "past" or "future", and thus the model is not suited for the study of dynamics of labor supply over the life-cycle. There is no concept of a "career" in the static model, since it does not "track" a worker's behavior over time. But should we really care about these short-comings of the static labor supply model? What is the gain of formulating a dynamic model which is much more difficult to solve than the static model? From the perspective of "basic research", the dynamic model tries to explain dynamic patterns of labor supply behavior. It does so by introducing only few additional parameters. Consequently, the dynamic model generates more theoretical predictions about labor supply than the static model, and it thereby provides additional restrictions on data that can be tested by an econometrician. Furthermore, because the model generates labor supply dynamics for the same individual over the life-cycle, the econometrician needs to rely on panel data (or at least quasi-panels) that follow individuals over time. This is in sharp contrast to the static model that can be estimated using cross-sectional data. Clearly, one can identify more parameters from panel data than from cross-sectional data. For example, individual fixed effects that capture unobservable skill components of workers cannot be identified from cross-sectional data, but can be estimated from panel data. Dynamic labor supply models "generate" econometric panel data models, and therefore provide a means for identifying many more structural parameters and components of unobserved heterogeneity than static models that are estimated from cross-sectional data.

From a policy perspective, there are also large gains from relying on intertemporal labor supply models. As an example, consider the labor supply elasticity \( \eta_{h,w} \) computed from the static labor supply model. This elasticity predicts by how much workers adjust their hours of work when the wage rate changes. In particular, it predicts by how much workers change their labor supply behavior when a proportional tax \( \tau \) on wages is introduced, which essentially reduces the wage received by the employee from \( w \) to \( w(1 - \tau) \). As mentioned in the introduction, there are numerous economists who believe that this elasticity is large. Now suppose one collects cross-sectional data on wages and hours, runs a regression \( \log h_i = \alpha_0 + \alpha_1 \log w_i + \varepsilon_i \), where the regression coefficient of \( \alpha_1 \) is an estimate of the static labor supply elasticity, and one finds that this estimate is small. Should we then conclude that labor supply is inelastic? There are many examples for why such a conclusion would be premature. For example, suppose that individuals choose their labor supply by calculating both, present and future costs and benefits. In this case, their decision is dynamic, not static. Also assume that future retirement benefits depend on how much individuals have worked over...
the life-cycle: In many industrial countries, Individuals who work a lot are entitled to higher retirement benefits. In this case, there is an additional incentive to work, even if wages are low. Hence, the estimate of \( \alpha_1 \) will be small not because the labor supply elasticity is low, but because individuals anticipate that by reducing hours of work by a lot reduces future retirement benefits. To measure the true labor supply elasticity one thus needs to hold constant the effect of labor supply behavior on future benefits. The example thus introduces an intertemporal dimension to a worker’s decision problem which is absent from the static model. In this case, one only reaches at an unbiased estimate of labor supply elasticities - and therefore at sound policy conclusions - when one relies on a dynamic model.

2 A Model of Intertemporal Labor Supply

The intertemporal labor supply model differs from the static model in at least three dimensions. First, utility does not only depend on present labor supply and consumption, but on the whole sequence of labor supply and consumption over the life-cycle. We write utility as \( U(C_0, C_1, ..., C_T, l_0, l_1, ..., l_T) \) where \( T \) is the last period in a workers’ life-cycle. Second, there are potential sources of uncertainty about future returns, benefits and costs, and thus individuals need to form expectations about the future. Hence, they maximize expected rather than deterministic utility. Third, the budget constraint is dynamic in the sense that individuals can shift resources between periods by using savings or other insurance mechanisms. As a consequence, today’s decisions influence future choice sets.

The specifics of the proto-typical dynamic labor supply model are as follows: Individuals need to choose sequences of consumption and hours of leisure for each period of a life-cycle, starting in period 0 and going to period \( T \). We follow the literature and assume that utility is time-separable, i.e. the utility of life-cycle sequences of consumption and leisure equals the discounted sum of period-\( t \) utilities of contemporaneous consumption and leisure:

\[
U(C_0, C_1, ..., C_T, l_0, l_1, ..., l_T) = \sum_{t=0}^{T} \beta^t * U(C_t, l_t),
\]

where \( \beta \) is a discount factor that lies in \((0, 1)\).

The workers’ budget set is dynamic in the sense that resources can be shifted across periods by accumulating debt or savings. One thus needs to specify the financial instruments the worker can use to transfer her earnings over time. We again follow the literature and assume that the worker can freely borrow or save at a risk-free interest rate of \( r \). For notational simplicity we assume that this interest rate remains constant over time, although it is straightforward to allow for time-varying returns. The intertemporal budget contraint is then given by
\[ A_{t+1} + C_t = w_t \cdot h_t + (1 + r) \cdot A_t \]  

where \( A_t \) is savings (or debt) carried into period \( t \). The budget constraint equates period \( t \) expenditures - savings carried into period \( t+1 \) and contemporaneous consumption - with period \( t \) income - labor earnings plus principle and interest on the current asset holdings.

It is important to note that there is a fundamental market incompleteness embedded in the constraint: Given that the only asset available is risk-free with an interest rate \( r \) we rule out the existence of a complete set of Arrow-Debreu securities. In other words, the set of financial instruments available is restricted to one risk-free asset. The household thus cannot write state-contingent claims on future occurrences. This rules out insurance contracts against future bad shocks such as unemployment or an accident. The only way an individual can insure herself against such shocks is by accumulating savings invested in the risk-free asset.

We also assume that there is uncertainty about the future, which can be driven by stochastic shocks to preferences (which would enter the utility function) or to wage rates (which would enter the budget constraint). At this point we follow the majority of the literature by assuming that preferences are non-stochastic, but that wages follow some stochastic process described by the conditional probability distribution \( \Gamma(w_t) \) which we write as \( w_{t+1} \sim \Gamma_t(w_t) \). This stochastic process needs to be estimated from the data. Note that I index \( \Gamma \) by \( t \) which means that the transition density of wages can change from period to period. This means that the stochastic process is allowed to be non-stationary, an assumption that can be maintained only within finite-horizon models.\(^1\) The worker maximizes expected rather than deterministic utility subject to the intertemporal budget constraint by choosing sequences of consumption and leisure and using current wages \( w_t \) to form expectations about the future:

\[
\max_{\{C_t, l_t\}_{t=0}^T} E \sum_{t=0}^T \beta^t \cdot U(C_t, l_t) 
\]

subject to:

\[
A_{t+1} + C_t = w_t \cdot h_t + (1 + r) \cdot A_t
\]

\[
A_0 = 0
\]

\[
l_t = H - h_t
\]

\[
w_{t+1} \sim \Gamma_t(w_t)
\]

\[
C_t, l_t, h_t \geq 0.
\]

This is a difficult maximization problem. One can solve it by using variational techniques (Lagrangians or Hamiltonians) or by using the more flexible method labeled as Dynamic Programming. At this

---

\(^1\)See the notes on Dynamic Programming for a discussion of stationarity in infinite horizon models.
point you should refresh your memory of Dynamic Programming since this is the method we will be using frequently. You may find the notes posted on the course website useful. We will also cover some numerical aspects of Dynamic Programming in a later section.

The state variables in period \( t \), i.e. the variables that are pre-determined at the beginning of the period but influence the individuals’ decision and whose evolution over time can be "controlled" by the individual, are \((A_t, w_t)\), and the controls are \((C_t, l_t, h_t)\). The period-\( t \) Bellman-equation is given by

\[
V_t(A_t, w_t) = \max_{C_t, l_t, h_t} \{U(C_t, l_t) + \beta * E[V_{t+1}(A_{t+1}, w_{t+1})|A_t, w_t, C_t, l_t]\} 
\tag{4}
\]

\( s.t. \ A_{t+1} + C_t = w_t * h_t + (1 + r) * A_t; \ A_0 = 0; \ l_t = H - h_t \)

\( w_{t+1} \sim \Gamma_t(w_t) \)

\( C_t, l_t, h_t \geq 0 \)

Note that the maximization is over the values \((C_t, l_t, h_t)\) taking as given the state variables \((A_t, w_t)\) rather than a sequence \((C_t, l_t, h_t)_{t=0}^T\). This is the difference between solving (3) recursively by Dynamic Programming or as a maximization over entire sequences using variational techniques. First assume that at the optimum the constraints \(C_t, l_t, h_t \geq 0\) are not binding. Then the first-order condition for the maximization problem on the RHS of the Bellman-equation with respect to \(C_t\) is given by

\[
\frac{\partial U(C_t^*, l_t^*)}{\partial C_t} = \beta * E \left[ \frac{\partial V_{t+1}(A_{t+1}, w_{t+1})}{\partial A_{t+1}}|A_t, w_t, C_t^*, l_t^* \right] 
\tag{5}
\]

and the Envelope Theorem gives

\[
\frac{\partial V_t(A_t, w_t)}{\partial A_t} = (1 + r) * \left( \frac{\partial U(C_t^*, l_t^*)}{\partial C_t} \right). 
\tag{6}
\]

Combining the two equations we get

\[
\frac{\partial U(C_t^*, l_t^*)}{\partial C_t} = \beta * (1 + r) * E \left[ \frac{\partial U(C_{t+1}^*, l_{t+1}^*)}{\partial C_{t+1}}|A_t, w_t, C_t^*, l_t^* \right] 
\tag{7}
\]

This equation is called **Consumption Euler Equation** and plays a central role in the literature that studies consumption dynamics over the life-cycle. It equates the expected marginal rate of substitution between consumption in two subsequent periods with the discounted gross return on savings \(\beta * (1 + r)\).

The Euler Equation describes the optimal evolution of consumption over time and imposes a restriction on the growth rate of expected marginal utilities, reflecting the consumption-smoothing motive: Individuals use savings to smooth life-cycle trajectories of consumption.

Next consider the optimal choice of leisure in period \( t \). The first-order condition is given by

\[
\frac{\partial U(C_t^*, l_t^*)}{\partial l_t} = \beta * w_t * E \left[ \frac{\partial V_{t+1}(A_{t+1}, w_{t+1})}{\partial A_{t+1}}|A_t, w_t, C_t^*, l_t^* \right] 
\Rightarrow \frac{\partial U(C_t^*, l_t^*)}{\partial l_t} = w_t * \frac{\partial U(C_t^*, l_t^*)}{\partial C_t} 
\tag{8}
\]
This is a remarkable result: It is the same condition for the optimal choice of leisure and consumption like in the static model! Hence, individuals form expectations about the future evolution of earnings when choosing optimal sequences of consumption and hours of work, but within each period the static marginal rate of substitution between consumption and leisure is always equal to the static wage. In other words, the within period marginal rate of substitution between consumption and hours of work does not depend on any variables that are associated with the intertemporal dimension of the model. Individuals adjust their labor supply period-by-period to smooth consumption.

Finally, we need to address the problem of corner solutions. Like in the static labor supply model it is customary to consider preferences that rule out corner solutions for consumption. Thus, we assume that the first-order condition for consumption always holds with equality. For hours worked, however, we need to take into account the possibility that the household chooses not to work at all. In this case, leisure hits the upper boundary of $H$, and the first-order condition for leisure does not hold with equality. Instead, the household cannot increase $l_t$ any further to decrease its marginal utility until it equals its marginal cost, and as a consequence we have $\frac{\partial U_t(C_t, l_t)}{\partial l_t} > w_t \cdot \frac{\partial U_t(C_t, l_t)}{\partial C_t}$ if $l_t = H$.

Without any further restrictions on preferences and the sources and shape of future uncertainty one cannot get any further than this. In the next sections we will address the issue of translating these abstract equations into an econometric model that can be confronted with the data.

### 3 Estimation of Frisch Labor-Supply Equations

The earlier literature attempting to test the model empirically focusses on deriving estimable dynamic labor supply equations from the first-order conditions. The focus is on getting reliable estimates of intertemporal labor supply elasticities. However, it is not immediately obvious what kind of income we want to define elasticities for. Do we want to estimate labor supply elasticities with respect to life-cycle earnings, or intertemporal, but predictable wage movements, or wage shocks that are unpredictable? In this section we discuss one seminal paper addressing these issues. This paper formulates an econometric model of intertemporal labor supply from the Euler Equation and estimates various types of labor supply elasticities. The main labor supply equation is a log-linear equation which is commonly referred to as "Frisch Labor Supply Equation". Much of the literature focusses on wage elasticities with respect to year-to-year movements in predictable wage movements, holding constant life-cycle marginal utility of wealth. This type of labor supply elasticity is labelled "Frisch Labor Supply Elasticity".

#### 3.1 MaCurdy (JPE, 1981)

MaCurdy makes two main assumptions to reach at Frisch labor supply equations:
1. People have perfect foresight, i.e. there is no uncertainty about the future.

2. Utility is not only time-separable, but also separable between consumption and leisure. In particular,
\[ U(C_t, l_t) = Y_{1,t} * C_t^{\omega_1} - Y_{2,t} * h_t^{\omega_2}, \]
where \( h_t \) is labor supplied, and \( \omega_1 \in (0,1), \omega_2 > 1 \).

3. Individuals always work at least some time, so that the FOC hold with equality.

Obviously, these assumptions are strict, but they turn out to be very useful to obtain a simple and transparent intertemporal labor supply equation. As it turns out, modelling uncertainty about the future is an extremely difficult task, and it is still at the center of major academic debates in contemporary labor-and macro-economics. We will return to this issue later. The second assumption implies that marginal utility of consumption does not depend on labor supply, and vice versa. It is central to obtaining closed-form labor supply equations. The third equation avoids the need to specify selection equations, an issue we return to in the section about Dynamic Discrete Choice models.

We shall note that Dynamic Programming was not used widely at the time of MacCurdy’s study. He instead solves the deterministic model using a dynamic version of the Lagrangian. The Lagrangian multiplier for the period-\( t \) budget constraint measures the marginal utility of wealth from period \( t \) onward - much like in static models of consumer behavior where the multiplier is equal to the marginal utility of income. Denote this multiplier by \( \lambda_t \) and note that at the optimum the marginal utility of wealth is given by \( \frac{\partial V_t(A_t, w_t)}{\partial A_t} \) since the value function summarizes all the discounted sum of future utilities along the optimal path of consumption and leisure choices. In other words we have
\[
\lambda_t = \frac{\partial V_t(A_t, w_t)}{\partial A_t} = (1 + r) \left( \frac{\partial U(C_t, l_t)}{\partial C_t} \right)
\]
\[
= (1 + r) \beta \left( \frac{\partial V_{t+1}(A_{t+1}, w_{t+1})}{\partial A_{t+1}} \right) \quad \text{(given assumption 1, there is no expectation operator)}
\]
\[
= (1 + r) \beta \lambda_{t+1}.
\]

Iterating backward we get
\[
\lambda_t = \lambda_{t-1} \left( \frac{1}{(1 + r)} \right) = \lambda_{t-2} \left[ \frac{1}{(1 + r)} \right]^2 = \ldots = \lambda_0 \left[ \frac{1}{(1 + r)} \right]^t
\]
where \( \lambda_0 \) is the marginal value of life-cycle wealth at the optimum. Note that this is an endogenous variable which needs to be solved for if one wants to obtain fully structural labor supply equations. MacCurdy expresses discounting in terms of a discount rate \( \rho \) defined by \( \beta = \frac{1}{1+\rho} \). A higher discount rate means that individuals are less patient: \( \beta \) is lower and future earnings are discounted at a higher rate. With this notation we get, \( \lambda_t = \lambda_0 \left( \frac{1+\rho}{1+\rho} \right)^t \). Using the fact that for small \( \rho, r \) we have \( \ln(1+\rho) \approx \rho, \ln(1+r) \approx r \) we obtain:
\[
\ln \lambda_{t+1} = \ln \lambda_0 + (t + 1) * (\rho - r).
\]  
(11)

Note that the relationship \(\lambda_t = (1 + r) * \beta * \lambda_{t+1}\) which gives us this equation is just a different way of expressing the Euler equation, so we have already utilized equation (7).

The first-order condition for leisure (8) yields (where we now compute the FOC with respect to labor supply, not leisure!):

\[
\begin{align*}
\hat{Y}_{2,t} * \omega_2 * h_{t}^{\omega_2-1} & = w_t * \beta * \lambda_{t+1} \\
\Rightarrow \ln h_t & = \left(\frac{1}{\omega_2 - 1}\right) * (\ln w_t + \ln \beta + \ln \lambda_{t+1} - \ln \hat{Y}_{2,t} - \ln \omega_2).
\end{align*}
\]  
(12)

Plugging in our expression for \(\ln \lambda_{t+1}\) we get

\[
\ln h_t = \left(\frac{1}{\omega_2 - 1}\right) * (\ln w_t + \ln \beta + \ln \lambda_0 + (t + 1) * (\rho - r) - \ln \hat{Y}_{2,t} - \ln \omega_2).
\]  
(13)

Conditional on age (i.e. the index \(t\)) equation (13) is a deterministic relationship between labor supply, wage rates, the marginal utility of life-cycle wealth, age, and the structural parameters of the model. Conditional on the right-hand side variables and parameters, the model predicts that every individual of age \(t\) supplies the same amount of labor. In reality, this is not true since there is a lot of unexplained variation in the data. Therefore, to translate equation (13) into an econometric model we need to introduce observed and unobserved heterogeneity among agents. This means that we need to index variables by \(i\) (the individual) as well, thus clarifying that the econometric model associated with the Frisch labor supply equation is an econometric panel-data model. MaCurdy assumes that \(\ln \lambda_0\) can be treated like a fixed effect: for individual \(i\): It is unobserved, constant, enters the Frisch-labor supply equation in each period of a life-cycle in the same way, and varies across individuals. As a consequence, MaCurdy defines a composite fixed effect \(F_i = \delta * (\ln \lambda_0 + \ln \beta - \sigma_i - \ln \omega_2)\), where \(\delta = \frac{1}{\omega_2 - 1}\). Finally, letting \(u_{it} = \delta * u_{it}'\) yields the econometric model of intertemporal labor supply:

\[
\begin{align*}
\ln h_{it} & = F_i + \delta * \ln w_{it} + (t_i + 1) * \delta * (\rho - r) + u_{it}.
\end{align*}
\]  
(14)

This is MaCurdy’s main equation, and it can be estimated by using a fixed-effects panel data estimation procedure, i.e. by regressing log-labor-supply on log-wages, a lead in age, and a set of individual fixed
effects. Alternatively, one can generate a first-difference version of (14), which yields

$$\ln h_{it} - \ln h_{it-1} = \delta * (\rho - r) + \delta * (\ln w_{it} - \ln w_{it-1}) + \varepsilon_{it}$$  \hspace{1cm} (15)$$

where $\varepsilon_{it} = u_{it} - u_{it-1}$. This is simply a regression of changes in log-hours worked on changes in log-wages and can be estimated by OLS.\(^2\) To perform this task, MaCurdy collects Panel data from the Panel Study of Income Dynamics (PSID) from 1967 to 1976. The main parameter of interest is $\delta$, the intertemporal labor supply elasticity or Frisch-elasticity.

**A Note on Identification** When formulating and estimating econometric models, one of the central questions is whether and which parameters can be identified. If a parameter is non-identified it cannot be determined or known perfectly even if the sample size goes to infinity. Let $G(\theta, Z)$ be the estimation equation used to determine the model parameters $\theta$, given data matrices $Z$. We say that $\theta$ is identified if $G(\theta, Z)$ has a unique solution in $\theta$. For example, in Maximum Likelihood estimation $G(\theta, Z)$ is the set of first-order conditions for the likelihood, and in GMM-estimation $G(\theta, Z)$ is the set of moment conditions.

In OLS-estimation, we have a vector of data points for the dependent variable, say $y$, and a matrix of data points for the independent variables, say $X$, so that we can set $Z = (X, y)$. Under the assumption that the regression error is conditionally independent, given $X$, the estimation equation using population moments is given by $\theta = [E(X'X)]^{-1} * E(X'y)$. It has a unique solution if the rank-condition is satisfied, i.e. if $E(X'X)$ has full rank. In this case, $\theta$ is identified because it can be uniquely pinned down using information contained in the sample. Hence, the conditional independence assumption together with the rank condition are sufficient to identify $\theta$.

Now let’s go back to equation (15). Which parameters are identified? Define $\alpha = \delta * (\rho - r)$. As noted above, equation (15) can be estimated by running a regression of changes in log-hours on changes in log-wages. The intercept provides an estimate of $\alpha$, while the coefficient on $(\ln w_{it} - \ln w_{it})$ provides an estimate of $\delta$. Hence, if the rank condition is satisfied, which in this case corresponds to the case in which there is enough variation in $(\ln w_{it} - \ln w_{it})$, then $\alpha$ and $\delta$ are identified. This is good news since the parameter of interest is $\delta$, the Frisch labor supply elasticity. However, the structural parameters $\rho$ and $r$ are not identified. This is because we only have an estimate of $\alpha$, but an infinite number of combinations of $\rho$ and $r$ satisfy the equation $\alpha = \delta * (\rho - r)$, given $\delta$. Unless we have some other information about $\rho$ or $r$, we can only identify the difference $(\rho - r) = \frac{\alpha}{\delta}$.

How should we interpret $\delta$? It is the elasticity of intertemporal labor supply with respect to the wage

\(^2\)As you probably have learned in an econometrics course, using the first-difference version has the advantage that the fixed effects are eliminated and therefore do not need to be estimated. However, because of the differencing one needs to “throw away” the first observation for each individual, introducing a statistical inefficiency.
rate, holding the marginal utility of wealth \( \lambda_{t+1} \) constant. The last part of the sentence is essential: How can we vary wages over the life-cycle by holding \( \lambda_{t+1} \) constant for each \( t \), which in turn implies that life-cycle wealth is held constant? This is only possible when one redistributes labor income across periods in such a way that a wage reduction in one period is compensated by a wage increase in another period. Consequently, \( \delta \) measures how labor supply varies across periods as a response to anticipated and pre-determined variation in wages across periods. It provides the answer to the question whether individuals work more in periods in which they earn higher wages, conditional on anticipating these wage increases.

To gain some more intuition for what \( \delta \) measures, suppose that at labor market entry, i.e. for \( t = 0 \), the worker knows that she will earn constant wage rates over the life-cycle. She also knows that in some period \( \tau > 0 \), the government will increase tax rates on wages, thus decreasing received wage rates from period \( \tau \) onwards. Then \( \delta \) determines the extent to which individuals allocate their labor supply over the life-cycle, anticipating the introduction of the tax. Since \( \delta \) is essentially a type of substitution effect it must be positive. Hence, the model predicts that individuals reduce their labor supply upon introduction of the anticipated tax.

Unfortunately the econometric model (14) does not help recovering all model parameters that are necessary to calculate behavioral adjustments of wage changes. It only helps calculating labor supply responses with respect to wage changes along a pre-determined wage trajectory. It cannot tell us the differences in labor supply across individuals who have different wage trajectories, because any such differences cause income and wealth effects, i.e. changes in \( \lambda_0 \).

MaCurdy thus specifies a functional relationship between \( F_i \) (which contains \( \lambda_0 \)) and exogenous variables that "shift" \( F_i \) and therefore generate wealth effects. In the subsequent literature, this part of the paper has not gained so much attention. My guess is that this has three major reasons: First, such an approach relies on very strict exogeneity and parametric assumptions. Second, the approach heavily relies on the assumption that individuals have perfect foresight. However, many changes in labor supply are due to unanticipated changes which cause shifts in \( \lambda_0 \). Note that this includes unanticipated changes in government policies that generate intertemporal substitution effects and wealth effects. Third, shifts in \( F_i \) generate income effects which reduce labor supply. Hence, \( \delta \), which is a substitution effect with the unambiguous prediction that labor supply increases, is an upper bound of the labor supply response with respect to wage changes. Consequently, if this parameter is small, the usefulness of the intertemporal model is cast into doubt, irrespective of the size of the wealth effect, which reduces intertemporal labor adjustments even further. But as discussed above, one can recover the parameter \( \delta \) by estimating (14) without the need to know how \( F_i \) is determined.

The estimates of \( \delta \) MaCurdy obtains lie in the range of .10 to .23, with t-statistics between .80 and 2.42. The literature has interpreted this as evidence for small and only marginally significant intertemporal labor elasticities. Given that these estimates are cleaned of wealth effects, which are contained within the
$F_t$, the total behavioral adjustment of labor supply to changes in wage rates will be even smaller.

3.2 Limitations

The result in MaCurdy’s study is generally viewed as evidence that intertemporal elasticities of labor supply are small. Many researchers thus question the usefulness of the model. Some conclusions for applied work are that the strong volatility of hours worked over the business cycle are not explained by voluntary behavioral adjustments of workers, and that policies that reduce wage rates do only have minor impacts on intertemporal labor supply. But such conclusions are premature given the limitations of the study. Here I list a few of them that motivate the work we discuss in the subsequent sections.

- The econometric models above estimate the impact of wage variation over the life-cycle on labor supply decisions, holding constant the marginal utility of wealth. However, it is not clear what drives this variation in wages, thus also questioning if the $E[\lambda_{t+1}]$ are truly held constant. Are changes in wages permanent or transitory? Are they anticipated or not? Suppose for example that an individual is "hit" by a large unanticipated shock to wages. An example is a significant increase in wages because unions have gained unexpected successes in the bargaining procedures. In this case, there is a large change in wages, causing an update in expectations about future wealth and hence $E[\lambda_{t+1}]$. Since the wealth effect goes in the opposite direction as the intertemporal substitution effect, the labor supply adjustment will be smaller than $\delta$. However, the econometric models estimated above cannot control for such unexpected changes in $E[\lambda_{t+1}]$ and thus "pack" them into the estimates of $\delta$. In other words, $\delta$ is potentially biased downwards since the approaches above do not control well for unexpected wealth effects.

- The models abstract from human capital investment decisions. Imai and Keane (2004) show that this can lead to severe downward biases. To see why, suppose that labor supply is linked to human capital accumulation, e.g. because of experience effects or learning-on-the-job. Young people, who are the group of individuals earning the lowest wages, have the largest incentive to accumulate human capital, because they have almost the entire working part of the life-cycle ahead of them. Hence, they have an extra incentive to work even when wages are low. In contrast, older people who are close to retirement have very little incentive for human capital investments since there is not so much time left to reap the benefits. Thus, even though older people earn the highest wages, they do not face an extra incentive to work. In sum, there is a negative correlation between the increasing life-cycle profile of wages and the incentive to work because of human capital effects. When not controlling for human capital accumulation, one interprets the flat life-cycle profile of hours worked together with the increasing wage rates as evidence for small intertemporal substitution effects. In a quite
complicated paper, Imai and Keane (2004) control for human capital effects and find labor supply elasticities as high as 1.7.

- The models abstract from mobility across jobs that might come with wage changes but not with hours changes. Some employers might be able to provide higher wages and better benefits to employees without asking for longer work hours because they use labor more efficiently. Clearly, workers want to eventually work in such firms. In the model presented above the type of mobility discussed here shows up as a small correlation between wage changes and changes in hours worked.

- Behavioral adjustments in hours worked may take place at the extensive rather than the intensive margin of labor supply, implying that we need to incorporate the discrete decision whether to work or not. This introduces additional numerical and econometric challenges since a model with both margins of labor supply are a model with both, discrete and continuous decision variables. Furthermore, for a zero-one decision variable there is not even a definition of a worker-level elasticity, implying that one needs to aggregate to get a proper measure of labor supply adjustments at the extensive margin. A growing literature in macroeconomics argues that in the aggregate, much of the reaction in labor supply is driven by adjustments along the extensive margins, but so far no paper has estimated a micro-level structural model that incorporates both margins. For a state-of-the-art paper, see Erosa et al. (2015, ReStud) that attempts to fill this gap. They formulate a life-cycle stochastic dynamic programming model of labor supply at the intensive and the extensive margin and calibrate it to micro data on labor supply and wage dynamics. Their model delivers a small Frisch-elasticity, but a large elasticity of labor force participation rates, just as observed in the data.

- **MaCurdy actually never truly solves the intertemporal labor supply model.** In MaCurdy, labor supply depends on $\lambda_t$ which is endogenous and depends on wages, wealth and model parameters. But to solve the model one needs to find solutions for consumption and labor supply that only depend on the exogenous variables of the model at period $t$. To draw a parallel to a model you are probably very familiar with, consider the micro-textbook static model of household choices between two consumption goods. In that model, the Lagrangian is given by $U(C_1, C_2) + \lambda * (M - p_1C_1 - p_2C_2)$, where $\lambda$ is the Lagrange-multiplier, $M$ is monetary income, and $p_1, p_2$ are goods prices. You then use the FOC for the two consumption goods to derive the condition that the marginal rate of substitution is equal to relative prices, and you use the FOC for $\lambda$ to solve consumption in terms of goods prices, income, and preference parameters. As an analogy, MaCurdy only uses the FOC for one good and the condition for $\lambda$ to characterize consumption and leisure, keeping one equation unused. MaCurdy essentially estimates Euler-Equations which are conditions on marginal rates of substitution while a fully structural approach would need to go one step further to solve for the policy functions which provide dynamic structural labor supply and consumption demand equations.
The major reason for MaCurdy not solving the full model is probably the fact that there is no closed form solution of the model under all but the strictest assumptions on model parameters, preferences and uncertainty. Consequently, one needs to solve the model numerically, something which has become possible only with the advent of powerful computing resources and numerical Dynamic Programming techniques. We turn to these approaches below as they are now dominating the literature and can be predicted to become even more prevalent.

4 Experimental Evidence: Fehr and Goette (2007)

As stated above, one of the limitations of MaCurdy's empirical analysis is the problematic assumption that wage variation used to identify \( \delta \) in equation (14) keeps the \( \lambda_t \) constant. This assumption is violated if for example there are unexpected permanent shocks to worker productivity, e.g. because of a firm closure or a promotion. In this case, the wage change observed in the data is not only associated with an intertemporal substitution effect, but also with a wealth effect. Equation (14) does not control for such wealth effects. An "extreme" approach to get around this problem is to conduct an experiment in which the econometrician does not only have a set of control variables, but in fact "controls" the transitory wage fluctuations altogether. However, such an experiment needs to make sure that the controlled wage variation indeed only generates an anticipated wage change with a negligible wealth effect. This is the approach taken by Fehr and Goette. In their work, they focus on a very narrow labor market - the one for bicycle messengers in Zurich, Switzerland - and provide an anticipated transitory wage hike to a random subset of these messengers. This is a great experiment because the experiment is designed in a way that ties the empirical results directly to the model to be tested (here: the intertemporal model of labor supply).

As a consequence of this design, the parameter of interest can be interpreted as the structural Frisch-labor supply elasticity.

Bicycle messengers in Zurich can choose to sign up for 5-hours shifts during which they perform the task of delivering mail across the city by bike, which yields a revenue of \( r_{it} \) for worker \( i \) in shift \( t \). They are employed by a firm, and the firm keeps a fraction \( (1 - w) \) of the revenue. Thus, the "wage rate" for the messengers is \( w \), which for the firm considered in the paper is .39 for men and .44 for women.

The experimental setup is roughly described as follows: In August of the year the experiment was conducted, the researchers (NOT the firm) announce to a subset of workers that they will receive a temporary wage hike of 10 percentage points for the month of September. This wage hike is financed by a research fund, NOT by the firm. Consequently, male messengers are paid a wage rate of .49, and female messengers are paid a wage rate of .54 during the month of September. The theory of intertemporal labor supply predicts that such an anticipated wage hike will reduce labor supply before and after September, and that
it will increase labor supply for the month of September. Because the wage hike is truly transitory and small relative to life-cycle earnings, the labor supply reduction before and after September can be expected to be small. Furthermore, being announced and thus expected, it is "evolutory".

It is important to highlight that the wage hike is paid only to a random subset of messengers. To understand why, consider the regression estimated by Fehr and Goette:

\[
\ln h_{it} = X_{it}/\beta + F_i + \delta * T_{it} + u_{it}
\]  

(16)

where \( X_{it} \) is a set of controls, e.g. weather conditions, \( F_i \) is a messenger fixed effect that controls for messenger-specific productivity and taste differences, and \( T_{it} \) is a dummy variable equal to 1 if messenger \( i \) gets paid the wage hike in period \( t \). As you know from econometrics courses, to get consistent estimates of \( \delta \) - the intertemporal labor supply elasticity - the assumption of conditional independence has to hold: 

\[
E[u_{it}|X_{it}, F_i, T_{it}] = 0
\]

But it is hard to justify this assumption if the wage hike is paid to every messenger as there might be factors that are unobserved by the econometrician, that change between August and September, and that affect labor supply of every messenger. One example is a worsening of traffic conditions because of new road constructions, or an improvement of traffic conditions because the holiday season comes to an end. In this case, the wage hike is correlated with such unobserved factors, therefore violating the conditional independence assumption. The problem is that if every cyclist gets paid the wage increase, then \( T_{it} \) varies only across time, but not across messengers, i.e. \( T_{it} = T_t \), and the time-series variation in \( T_t \) is potentially correlated with the error term.

But because this is a randomized experiment, there is a solution: The wage hike is paid only to a randomly selected subgroup of messengers, referred to as the treatment group. In this way, the messengers who do not get the payment serve as a control group. Since it is chosen by a random number generator who of the messengers is part of the treatment group, \( T_{it} \) is randomly selected. Hence, \( T_{it} \) varies both, across individuals and time, and this variation is uncorrelated with any observed or unobserved variable by construction. You might find it helpful to think of medical trials, in which the effectiveness of a new medication is tested, as an analogy to the experiment discussed here. In medical trials, the medication to be tested is given to a random subset of the test persons (the treatment group), and a placebo is given to the rest (the control group). Only if there are significant differences in health outcomes between the two groups can one conclude that the medication has a significant effect. Likewise, in the bike messenger experiment, the researchers provide a wage increase to a random subset of the "test persons".

Fehr and Goette use a regression framework to estimate the intertemporal labor supply elasticity. The crucial difference to the common OLS-framework and its applications you know from standard econometrics courses is that Fehr and Goette use experimental data. Most of the literature that evaluates policies, including the paper by Eissa and Liebman (1996), uses non-experimental data instead. For example, many researchers want to estimate the returns to certain government policies, such as publicly funded workplace
training programmes, or the returns to certain individual choices, such as the returns to education. But these kind of programmes or choices often depend on variables that are unobserved to the econometrician. As a consequence, the observable variable capturing programme-participation or education is correlated with the error term and the conditional independence assumption is violated. Conducting an experiment avoids this problem if randomization of the treatment is properly executed. What makes the experiment by Fehr and Goette particularly interesting is that they design it such that the "causal effect" of treatment can be interpreted as a structural labor supply elasticity. The crucial ingredient here is that the wage hike is transitory and announced - exactly the kind of wage variation the defines the Frisch-labor supply elasticity.

The first step in the analysis of experimental data is to convince the reader that randomization has worked, indeed. To see how randomization could fail, suppose for example that Fehr and Goette offer the wage increase to a randomly selected subgroup of bicycle messengers at the firm in question, but some cyclists decide not to accept the wage increase. If the cyclists who decline the offer are systematically different from those who accept it, $T_i$ is not truly exogenous anymore, and again there might be correlation with the error term. Another example in which randomization might fail if the cyclists who get the wage increase make side-payments to those who do not get it, e.g. because of social pressure. Fehr and Goette show a table in which they compare certain observables, e.g. age, hours work before the announcement of the experiment, etc. of those in the treatment group with those in the control group. If randomization works, then there should not be any significant differences between the two groups prior to the experiment.

The authors show that randomization has worked. They then run the regression (16) for different outcome variables, not only for log-labor supply. They find the following: First, the intertemporal elasticity of revenue generated by the cyclist is very large. Second, so is the elasticity of the number of 5-hours shifts worked per cyclist. Third, the elasticity of revenue per shift, which is taken as a proxy for effort, is actually negative. In other words, cyclists in the treatment group do work significantly more hours during the month in which the wage increase is paid to them, but at the same time the effort per shift decreases for these cyclists. In total, the first effect is stronger so that the elasticity of total revenue generated by each cyclist in the month of the experiment is positive.

Fehr and Goette spend a lot of time on investigating why effort decreases in the treatment group. Their conclusion is that one needs non-standard preferences to explain the results. My take on this is that there are several other explanations that do not require one to turn to non-mainstream behavioralist explanations, e.g. decreasing marginal productivity of cyclists (more 5 hours shifts mean that cyclists are potentially more exhausted and cannot perform as well during a shift) or social pressure that dictates them not to earn too much more than those in the control group. At the same time, the finding that the cyclists increase the number of shifts worked by a large amount points towards substantial intertemporal labor supply elasticities - something which contradicts the results in MaCurdy. Of course we should keep
in mind that these results do not have external validity to a nationwide economy.