Chapter 12

Labor supply

Richard Rogerson and Giovanni L. Violante

12.1 Introduction

Hours worked are a fundamental input in the aggregate production function and, consequently, a crucial determinant of an economy's output per capita. Macroeconomists conceptualize the total hours worked at a given point in time as the equilibrium outcome where demand meets supply in the labor market. In this chapter, we examine the factors influencing labor supply.

The macroeconomic study of labor supply is motivated by a set of stylized facts regarding the variation in hours worked: across countries, over time both in the long run at the low frequency of economic growth and in the short run at the business cycle frequency, throughout an individual's life cycle, and across demographic groups, such as men and women, young and old, and skilled and unskilled workers. Understanding these variations hinges on three key elements. First, accurately identifying the incentives and disincentives that influence individuals' decisions to work. Second, determining the sensitivity of individuals to these incentives, measured by labor supply elasticities. Third, appropriately aggregating these elasticities to derive the macroeconomic patterns of hours worked. Each of these steps involves subtle considerations and is the focus of an extensive body of research.

The chapter is organized as follows. Section 12.2 summarizes some key facts about variation in hours of work. Section 12.3 presents a benchmark static model of labor supply and introduces several key elasticity concepts that are relevant for understanding how labor supply responds to changes in the economic environment. It also describes several enrichments of the static model. Section 12.4 extends our results derived for the benchmark static model of Section 12.3 to a dynamic stochastic life-cycle model. There, we also discuss the empirical literature that estimates labor supply elasticities with micro data, including a discussion of how optimization frictions can affect estimation. Section 12.5 derives the restrictions that balanced growth imposes on preferences over consumption and hours worked. Section 12.6 discusses how macroeconomists have used representative agent models to tackle variation in hours worked across space and over time. First, we explain how differences in the level of taxes across countries can account for the large observed differences in hours worked. Second, we illustrate the role of the Frisch elasticity in fluctuations of hours worked over the business cycle. The final two sections of the chapter describe extensions that have important implications for labor supply elasticities and their measurement. Section 12.7 generalizes the dynamic model of Section 12.4 to allow for human capital accumulation, thereby introducing dynamic returns to current labor supply. Section 12.8 introduces models that explicitly consider labor supply along both the extensive margin (employment) and intensive margin (hours when employed). We highlight how these additional features have important implications for labor supply elasticities.[1]

A key message that we want the reader to take away from this chapter is that a relatively simple and parsimonious model of labor supply can help us understand many of the patterns found in the data on aggregate labor market outcomes. For this reason it is important for macroeconomic models to explicitly incorporate labor supply. Importantly, this message should not be interpreted to imply that all labor market outcomes reflect desired labor supply by workers. Modern models of unemployment stress the possibility that frictions in the labor market create a wedge between labor market outcomes and desired labor supply. Frictional labor markets are discussed in Chapter 20.

12.2 Facts about hours of work

In this section we present some basic facts about hours of work, focusing on differences in hours of work in several contexts. The first subsection documents differences across countries, and decomposes these differences along the intensive (hours per worker) and extensive (fraction of people working) margins. The second subsection examines differences across time, and in particular secular trends and business cycle fluctuations. The third subsection documents differences across demographic groups within the U.S.

12.2.1 Differences across countries

One of the most basic statistics for macroeconomists is aggregate hours of work normalized by some measure of population.

Table 12.1: Hours of work per person aged 15+ relative to the U.S. 2015-2019

<0.75	**(.75,.85)**	**(.85,.95)**	**>.95**
Italy (0.69)	Finland (0.77)	UK (0.85)	Canada (0.96)
France (0.70)	Austria (0.79)	Sweden (0.90)	Australia (0.98)
Belgium (0.72)	Norway (0.80)	Ireland (0.91)	U.S. (1.00)
Greece (0.73)	Netherlands (0.82)	Japan (0.91)	New Zealand (1.07)
Denmark (0.74)	Portugal (0.85)	Switzerland (0.93)	Korea (1.12)
Germany (0.74)			
Spain (0.75)			

[1]To the student who wants to deepen their understanding of labor supply, we recommend the classic handbook chapter by Blundell and MaCurdy (1999) and the more recent survey articles by Keane (2011), Keane and Rogerson (2012, 2015), and Rogerson (2024).

Table 12.2: Labor supply along the intensive and extensive margin 2015-2019

Panel A: Extensive Margin: Employment to Population (%)				
<50	**(50,55)**	**(55,60)**	**(60,65)**	**>65**
Greece (40.9)	Belgium (50.0)	Ireland (57.7)	U.S. (60.1)	Switzerland (65.1)
Italy (44.1)	France (50.5)	Austria (57.9)	UK (60.4)	New Zealand (66.9)
Spain (48.6)	Portugal (52.3)	Denmark (58.3)	Korea (60.7)	Sweden (67.6)
	Finland (54.4)	Germany (58.8)	Australia (61.7)	
		Japan (59.0)	Canada (61.8)	
			Netherlands (62.3)	

Panel B: Intensive Margin: Annual Hours Worked per Employed Person				
<1500	**(1500,1650)**	**(1650,1750)**	**(1750,1850)**	**> 1850**
Germany (1388)	Austria (1502)	Spain (1693)	New Zealand (1761)	Greece (1941)
Denmark (1395)	France (1516)	Japan (1693)	U.S. (1825)	Korea (2026)
Norway (1423)	UK (1535)	Canada (1699)		
Netherlands (1435)	Finland (1549)	Italy (1718)		
Sweden (1466)	Switzerland (1563)	Ireland (1720)		
	Belgium (1577)	Portugal (1736)		
		Australia (1737)		

Table 12.1, adapted from Rogerson (2024), compares aggregate hours of work per person aged 15 and older across 23 OECD economies for the period 2015-2019. To facilitate comparisons all values are reported relative to the U.S. The value for the U.S. is 1096 hours per year. These differences in aggregate hours per person reflect differences both in the hours per person employed (the *intensive* margin) and the fraction of the population that is employed (the *extensive* margin). Table 12.2 presents differences along each of the two margins for the same countries and time period as Table 12.1. As a reference point for interpreting the values in Panel B, we note that an individual who works 40 hours a week for 50 weeks will have 2000 annual hours.

12.2.2 Differences over time

Macroeconomists are also very interested in the time series changes in aggregate hours of work. Figure 12.1 plots the log of aggregate hours per person aged 15 and older for the G7 economies over the period 1950-2019. Two properties are worth noting. First, most of the countries experience a very large decline in aggregate hours per person over this period, with several countries experiencing declines of roughly 30 percent. The U.S. is somewhat of an exception, with hours in 2019 only marginally lower than in 1950 and with no statistically significant trend. Second, the large differences that we documented in Table 12.1 for the period 2015-2019 are not a constant feature of the data. Although the U.S. has one of the highest values for aggregate hours per person in 2019, it actually has the lowest value among these countries in 1950.

It is also of interest to look at the separate time series for the intensive and extensive margins. Figure 12.2 shows these series for the G7 economies over the same time period.

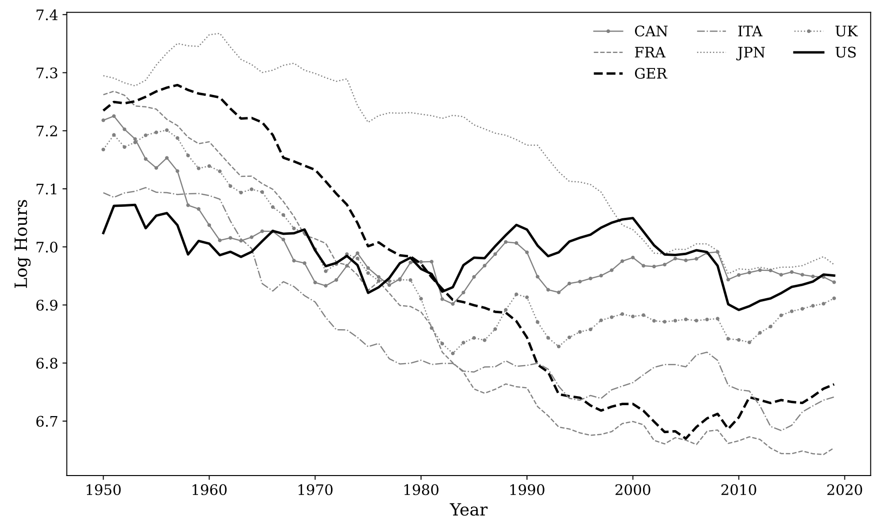

Figure 12.1: Log average annual hours worked per person for the G7 countries from 1950-2019

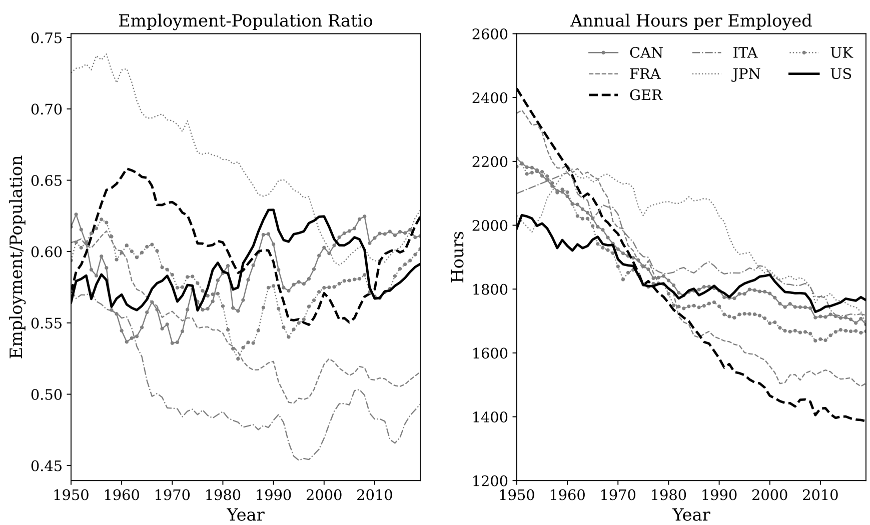

Figure 12.2: Intensive and extensive margins of labor supply for G7 countries from 1950 to 2019.

Notes: Left panel: employment to population ratio. Right panel: annual hours per employed person.

There are some notable differences between these two figures. Regarding the intensive margin, we see that all countries experience relatively large secular declines, though there is a large range in the magnitude of the decline. Germany experiences a reduction of roughly 40 percent, while the U.S. experiences a decline of roughly 10 percent. The country level experiences for changes along the extensive margin are much more varied. Some countries witness a significant secular decline, while others exhibit relatively little secular change, and still others feature a secular increase.

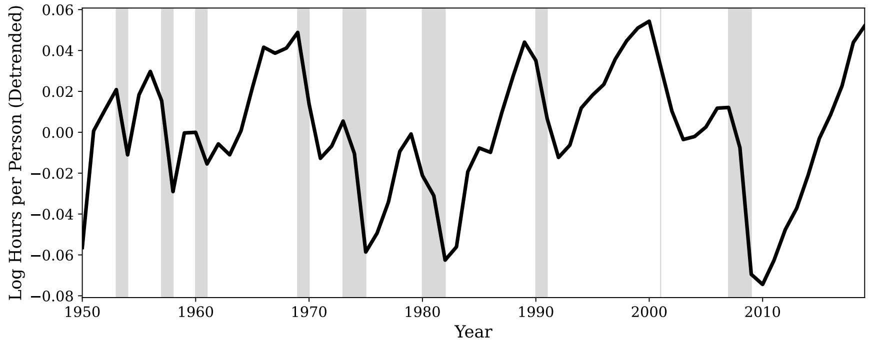

Figure 12.3: Business cycle fluctuations worked per person for the U.S. from 1950-2019

.

Notes: Calculated as the residual of a regression of log average annual hours worked per person on a quartic polynomial in time.

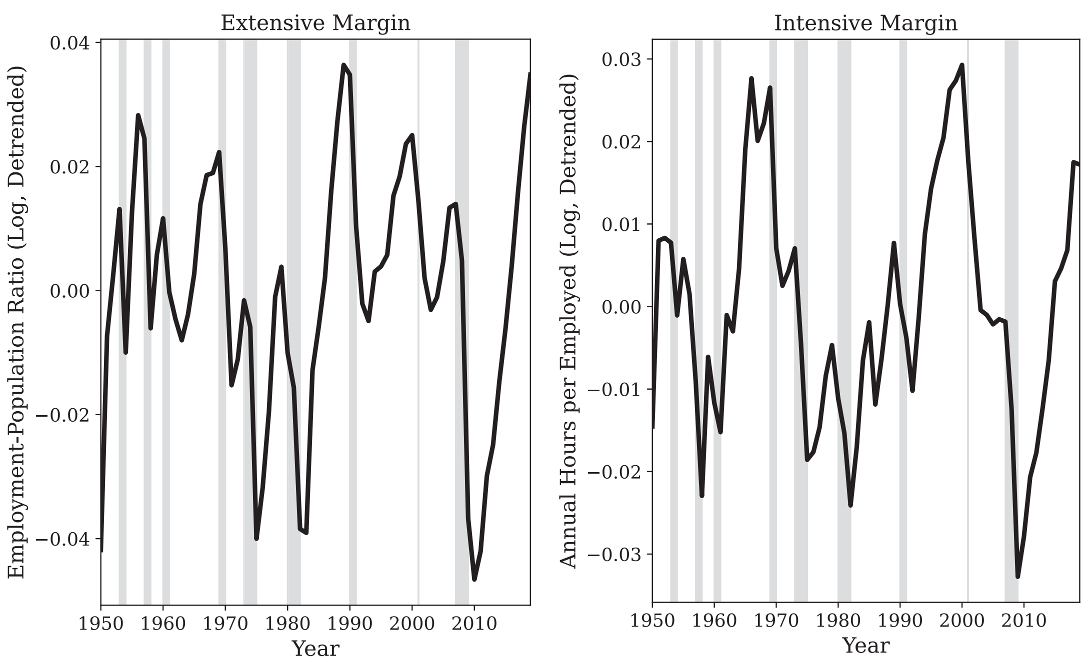

Figure 12.4: Business cycle fluctuations in labor supply on the intensive (left panel) and extensive (right panel) margins for the U.S. from 1950-2019.

Notes: The business cycle variation in the intensive margin of labor supply is calculated as the residual of a regression of log annual hours worked per employed on a quartic polynomial in time. The business cycle variation in the extensive margin of labor supply is calculated as the residual of a regression of log employment to population ratio on a quartic polynomial in time.

When examining time series for hours worked, macroeconomists are often interested in distinguishing between trend behavior and deviations from trend, typically associated with business cycle fluctuations.

There are various methods used to perform this decomposition. Here we simply use a quartic polynomial in time to identify the trend and define the business cycle to be deviations from this trend. Figure 12.3 displays the business cycle component of aggregate hours for the U.S. economy, with grey shaded areas indicating NBER recession dates. This figure displays the well-known fact that aggregate hours display large fluctuations over the business cycle, with the movements from peak to trough exceeding ten percentage points in many instances. One might also be interested in the nature of business cycle fluctuations along the intensive and extensive margins separately. The two panels of Figure 12.4 show this for the U.S. economy. The left panel displays fluctuations along the extensive margin while the right panel shows fluctuations along the intensive margin. Both exhibit relatively large fluctuations, though the movements along the extensive margin tend to be somewhat larger.

12.2.3 Differences across demographic groups

To this point we have focused on purely aggregate measures of labor supply. Another important feature of the data is that hours of work vary quite dramatically across demographic groups within a country. To illustrate this we rely on data from the American Time Use Survey for the year 2023. This survey produces data on average time devoted to work per day broken down by various demographic characteristics. Table 12.3 shows how time devoted to work varies by gender and age group.

Table 12.3: Hours of work per day

Age	All	Male	Female
15 − 19	1.27	1.24	1.29
20 − 24	3.67	3.81	3.52
25 − 34	5.02	5.90	4.13
35 − 44	5.04	5.79	4.29
45 − 54	4.79	5.45	4.15
55 − 64	4.02	4.73	3.35
65 − 74	1.42	1.88	1.02
75+	0.37	0.55	0.23
Total	3.56	4.17	2.98

For both males and females we see a very pronounced hump shape for time devoted to work over the life cycle, with hours peaking during the 25-54 age range and falling off quite substantially at younger and older ages. Although both males and females display the same hump-shaped pattern, the life cycle profile for female time devoted to work is substantially lower than the profile for males, with the lone exception of work among 15-19 year olds. Gender differences have also changed dramatically over time. Figure 12.5 shows the evolution of the employment to population ratio for males and females in the U.S. since

1950. At the same time that the male employment to population ratio has declined by more than ten percentage points the female employment to population ratio has increased by roughly 20 percentage points. These large changes in gender employment differences are a pervasive feature of the data.

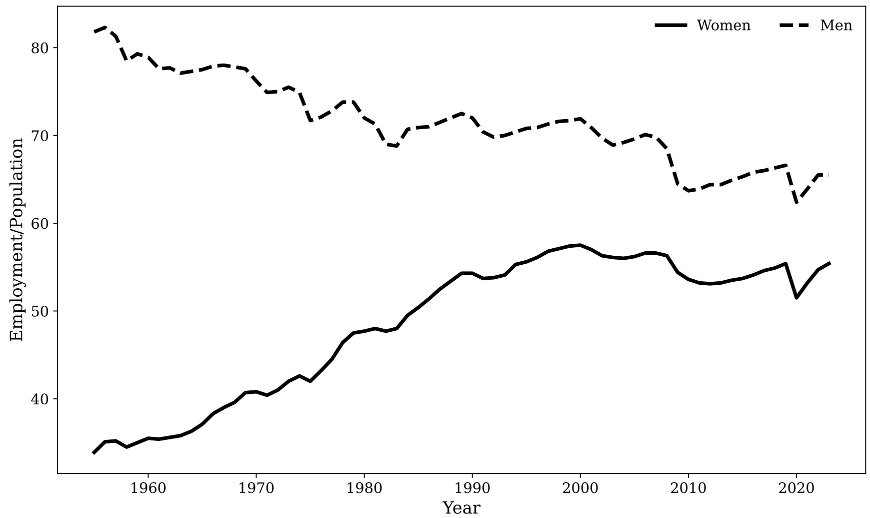

Figure 12.5: Employment to population for the U.S. by gender from 1950-2019.

In addition to large differences in hours of work by age and gender, there are also large differences by educational attainment. Table 12.4 shows average time devoted to work per day by educational attainment for five educational attainment groups: less than high school (<HS), exactly high school (HS), some college but not a college degree (SC), a four year college degree but nothing further (C) and more than a college degree (>C).

Table 12.4: Hours of work per day by education

<HS	HS	SC	C	>C
2.89	3.39	3.90	4.08	4.17

This table shows that average time devoted to work increases substantially with educational attainment, with college educated workers devoting roughly 20% more time to work relative to high school educated workers.

12.2.4 Summary

In this section we have documented that there are large differences in hours of work across countries, across time, and across demographic groups. Understanding the forces that account for these differences is an important goal of economics research. One key question in this context concerns the extent to which these differences in hours of work reflect different choices by individuals, and if so, what factors account for these different choices. Answering this question requires that one think about the forces that shape individual labor supply decisions. The goal of this chapter is to introduce the student to the basic economics of labor supply.

12.3 The theory of labor supply: static models

Much of this chapter is devoted to understanding the forces that shape labor supply in dynamic stochastic settings. But because the same forces that shape labor supply in simpler settings will also play an important role in dynamic stochastic environments, it is useful to begin our analysis of labor supply by considering a simple textbook model of labor supply in a static and deterministic setting. This setting is sufficient to introduce several elasticity concepts that are important in understanding how labor supply responds to various changes in the economic environment.

12.3.1 A benchmark model of labor supply

In this section we study a standard textbook model of individual labor supply and examine some of the forces that shape optimal labor supply. We consider an individual with preferences over consumption (c) and hours worked (ℓ). While the discussion in this section can be generalized to a large class of preferences, to facilitate exposition we focus on a special case which is commonly used in the macroeconomics literature. Specifically, we assume that preferences are separable between consumption and hours of work:

$$U(c, \ell) = u(c) - v(\ell),$$

with the functions $u(c)$ and $v(\ell)$ of the following form:

$$u(c) = \frac{1}{1-\sigma} c^{1-\sigma} \text{ and } v(\ell) = \frac{\psi}{1+1/\gamma} \ell^{1+1/\gamma},$$

where ψ, σ, and γ are all positive constants. As we show later in this chapter, this relatively parsimonious specification of preferences leads to a theory of labor supply that helps us understand many features documented in the previous section.

The price of consumption is normalized to one, the individual faces a wage rate of w, and has non-labor income equal to I. The individual thus solves the following problem:

$$\max_{c,\ell} U(c, \ell)$$

subject to

$$c = w\ell + I,\ c \geq 0, \text{ and } \ell \geq 0.$$

Because the marginal disutility of work is equal to zero when $\ell = 0$, the individual will always choose $\ell > 0$. Substituting the budget equation into the utility function one obtains the first-order condition:

$$\frac{\psi \ell^{1/\gamma}}{(w\ell + I)^{-\sigma}} = w. \tag{12.1}$$

This equation has the interpretation that the individual chooses hours of work ℓ so that the marginal rate of substitution between consumption and working time is equal to the level of the real wage w, i.e., the price of time relative to consumption.

Individuals differ both in the wage rate that they face and their level of non-labor income. It is of interest to ask what our model implies for how these differences will affect labor supply. Economists typically find it useful to summarize these effects as elasticities. If we write the solution to the optimal labor supply problem as $\ell(w, I)$, then the wage and income elasticities, denoted by $\varepsilon^M_{\ell,w}$ and $\varepsilon_{\ell,I}$ can be written as:

$$\varepsilon^M_{\ell,w} = \frac{w}{\ell}\ell_1(w, I)$$

and

$$\varepsilon_{\ell,I} = \frac{I}{\ell}\ell_2(w, I),$$

where $\ell_i(w, I)$, $i = 1, 2$ is the partial derivative of $\ell(w, I)$ function with respect to the ith argument. Using the implicit function theorem and equation (12.1) and carrying out some simple algebra yields the following two expressions:

$$\varepsilon_{\ell,I} = -\frac{\sigma}{1/\gamma + \sigma(1 - s_I)}s_I \tag{12.2}$$

and

$$\varepsilon^M_{\ell,w} = \frac{1 - \sigma(1 - s_I)}{1/\gamma + \sigma(1 - s_I)}, \tag{12.3}$$

where $s_I = I/(w\ell + I)$ is the share of non-labor income in total income.

Note that there is an intimate connection between curvature parameters in the utility function and these two labor supply elasticities. This implies that knowledge of labor supply elasticities provide information about the parameters of the utility function. We begin by discussing the income elasticity. Given that both σ and γ are positive and $s_I \leq 1$, we see that the income elasticity is always negative if s_I is positive.[2] The intuition for this result is straightforward. Higher non-labor income implies a higher level of consumption for any given level of hours. Because the marginal utility of consumption is decreasing, higher consumption decreases the marginal benefit of working and thus causes the individual to reduce hours of work.

Next we turn to the wage elasticity. One of the most fundamental questions in the context of labor supply is how does the optimal choice of ℓ respond to an increase in w? Looking at equation (12.3), one sees that the sign of the denominator is necessarily positive, but that the sign of the numerator could be positive or negative depending upon the values of σ and s_I. The reason for this ambiguous prediction is that an increase in w has two opposing effects on the optimal choice of ℓ. On the one hand, an increase in w increases the reward to additional work at the margin, thereby creating an incentive for additional work. This is what economists call the **substitution effect**. On the other hand, holding hours

[2] If $s_I = 0$ then the income elasticity is equal to zero. But importantly, this does not imply that the effect of income on hours is zero. The same algebraic procedure that one uses to solve for the income elasticity implies that $\ell_2(w, I) = -\sigma(1 - s_I)/[1/\gamma + \sigma(1 - s_I)]$. In particular, when s_I equals zero we see that $\ell_2(w, I) = -\sigma/[1/\gamma + \sigma]$, which is necessarily negative.

fixed, an increase in w raises the consumption of the individual, which makes additional consumption less valuable at the margin, thereby creating an incentive for less work. This is what economists call the **income effect**.

It is useful to focus on the special case in which non-labor income is equal to zero. In this case we have:

$$\varepsilon^M_{\ell,w} = \frac{1-\sigma}{1/\gamma + \sigma}.$$

It follows that the wage elasticity is strictly positive if σ is less than one, strictly negative if σ is greater than one and exactly equal to zero if σ is equal to one. Put somewhat differently, when $\sigma < 1$ the substitution effect dominates the income effect and when $\sigma > 1$ the income effect dominates the substitution effect. When $\sigma = 1$ changes in the wage have no effect on optimal labor supply. Importantly, when $\sigma = 1$ neither the income nor the substitution effect is equal to zero; rather, they are exactly offsetting. Specifically, the magnitude of each effect is increasing in the value of γ.

The elasticity of hours with respect to the wage that we have calculated is referred to as the **Marshallian elasticity**, and explains why we have included a superscript M. Our preceding discussion suggests that this elasticity can be thought of as the sum of two pieces, one reflecting the substitution effect and the other reflecting the income effect. Holding hours fixed, the additional income associated with a marginal increase in w is equal to level of hours worked, so that the size of the income effect is given by $\ell \cdot \ell_2(w, I)$. Given that the total effect on hours is given by $\ell_1(w, I)$, we can compute the magnitude of the substitution effect as the difference between the total effect of the wage change and the income effect associated with the wage change:

$$\text{substitution effect} = \ell_1(w, I) - \ell \cdot \ell_2(w, I).$$

To express this in terms of elasticities we multiply both sides by w/ℓ and rearrange to obtain:

$$\varepsilon^H_{\ell,w} = \varepsilon^M_{\ell,w} - \frac{1-s_I}{s_I}\varepsilon_{\ell,I},$$

where $\varepsilon^H_{\ell,w}$ is what economists refer to as the **Hicksian elasticity**. This elasticity reflects the change in hours worked in response to a change in the wage rate holding utility constant. Substituting from our earlier expressions we obtain:

$$\varepsilon^H_{\ell,w} = \frac{1}{1/\gamma + \sigma(1-s_I)}. \tag{12.4}$$

A key result is that the Hicksian elasticity is always positive. Moreover, it is always larger than the Marshallian elasticity.

It is perhaps useful to provide a concrete example to illustrate the usefulness of the different elasticities that we have introduced. An important policy question in which labor supply plays a key role is the effects of tax and transfer programs. To pursue this we introduce two policy parameters into the individual labor supply problem that we have been studying: a proportional tax on labor income at rate τ and a lump-sum transfer payment denoted by T. The individual's labor supply problem now becomes:

$$\max_{c,\ell}\; U(c, \ell)$$

subject to

$$c = (1-\tau)w\ell + I = (1-\tau)w\ell + y + T,\ c \geq 0, \text{ and } \ell \geq 0,$$

where y is non-labor income apart from the government transfer payment T. If we are solely interested in the effect of changes in the proportional tax rate on hours of work, we see that a change in $(1-\tau)$ is equivalent to a decrease in w, so that the effect of interest is given by the Marshallian elasticity. If we are solely interested in the effect of a change in transfer payments on hours of work, we see that a change in T is equivalent to a change in I, so that the effect of interest is given by the income elasticity. Changing tax rates and transfer payments individually have implications for the government budget. If one wants to study changes in tax and transfer policies that are budget neutral then one could focus on the case in which the transfer payment is equal to the amount of revenue raised by the tax, i.e., $T = \tau w\ell$. In this case there is no income effect associated with the change in τ and the appropriate elasticity for assessing the impact on hours is the Hicksian elasticity. In a highly cited paper, Prescott (2004) essentially used this framework to study the effect of differences in tax and transfer systems between the U.S. and several Western European countries on hours of work. He found that the effects were large. We examine this in more detail in Section 12.6.1.

12.3.2 Richer versions of the static labor supply model

In this section we consider three extensions to illustrate additional channels that have also been shown to play an important role in shaping labor supply.

Home production

In our previous analysis, all of the work that individuals did occurred in the market at the going wage of w, and all of their consumption was purchased in the market. In reality, individuals engage in many work activities outside of the market and there are several components of consumption that individuals can produce for themselves without purchasing them in the market. Some prominent examples include cooking and cleaning services, yard work, and child care. Economists use the term home production to describe the time that individuals devote to these activities. Classic references include Becker (1965) and Gronau (1977). We now generalize our previous analysis to allow for home production.

To do this we introduce the notion of the home production function and consider two alternative types of time spent working: ℓ_m will be time spent working in the market and ℓ_h will be time spent on home production. As before, time spent working in the market will generate income according to the market wage rate w. Time spent working at home will now generate home production, which we capture by the production function $y_h = A_h\ell_h$, where A_h captures the productivity of time spent in home production.[3] We now assume that the total consumption of an individual is given by the composite of market consumption denoted

[3]More generally, one could generalize this to allow for diminishing marginal product by writing $y_h = A_h f(\ell_h)$, where the function f is strictly concave. One could also add capital as an additional input. We discuss the significance of this generalization at the end of the subsection.

by c_m and consumption of home production, denoted by c_h:

$$c = g(c_m, c_h).$$

Our analysis will follow much of the literature and assume that the function g is given by a constant elasticity of substitution function:

$$g(c_m, c_h) = [\theta c_m^\eta + (1-\theta)c_h^\eta]^{\frac{1}{\eta}}$$

with $0 < \eta < 1$, indicating that home and market consumption are substitutes. We assume that the disutility of work depends only on the sum of market work and home production time, though one could generalize this.

The individual now solves the following optimization problem:

$$\max_{\ell_m, \ell_h} u\left(g(\ell_m w + I, A\ell_h)\right) - v(\ell_m + \ell_h)$$

subject to

$$\ell_m \geq 0 \text{ and } \ell_h \geq 0,$$

where the functions u and v are as before, though for much of our analysis the specific functional forms for these two functions will not play any role. From this maximization problem we obtain the following two first-order conditions:

$$\begin{aligned} \ell_m &: \quad u' g_1 w = v' \\ \ell_h &: \quad u' g_2 A = v', \end{aligned}$$

where g_i, $i = 1, 2$, is the partial derivative of the g function with respect to the ith argument. Dividing the two first-order conditions by each other gives:

$$\frac{g_2}{g_1} = \frac{w}{A}. \tag{12.5}$$

This equation has an intuitive economic interpretation: it states that the marginal rate of substitution between home and market consumption is equal to the marginal rate of transformation between home and market consumption. For our purposes we want to focus on one particular implication. Specifically, assuming g is a CES aggregator and focusing on the special case in which $I = 0$ we obtain the following expression for the left hand side of equation (12.5):

$$\frac{g_2}{g_1} = \frac{1-\theta}{\theta}\left[\frac{c_h}{c_m}\right]^{\eta-1} = \frac{1-\theta}{\theta}\left[\frac{A\ell_h}{w\ell_m}\right]^{\eta-1}.$$

Substituting this into equation (12.5) and rearranging terms yields:

$$\frac{\ell_h}{\ell_m} = \left[\frac{1-\theta}{\theta}\right]^{\frac{1}{\eta-1}} \left[\frac{w}{A}\right]^{\frac{\eta}{\eta-1}}. \tag{12.6}$$

If $0 < \eta < 1$ then the exponent on w/A is negative and we have that the ratio of home to market work is negatively related to the ratio of the market wage to the level of home productivity.

The key message from this extension is that technical change that affects the productivity of time spent in home production can have important effects on labor supply to the market sector. Greenwood, Seshadri, and Yorukoglu (2005) used a generalization of this model to argue that technological changes in home production played a major role in the rise of female labor force participation observed in the first part of the 20th century.[4]

Household labor supply

The second extension considers labor supply from a household perspective when we explicitly consider households that consist of multiple individuals. The motivation for this is straightforward: the vast majority of total hours of work in modern economies is accounted for by individuals who live in households with multiple members. The key point we make here is that when looking at labor supply at the individual level, it is not only the market wage of that individual that matters, but also the market wages of other individuals in their household. Here we make this point in the context of the most commonly studied situation, that of two member households.

Studying optimal labor supply in multi-member households necessarily raises the question of how to specify the objectives and interactions of individuals within the household. Do they behave strategically with regard to each other? Do they behave cooperatively? The literature on family labor supply has considered different possibilities, but here we will focus on a common and particularly tractable specification, in which we assume that all family members share a common objective function. In the literature this is referred to as the unitary household model.[5] It eliminates any potential for strategic considerations between household members since there is no disagreement over which outcomes are best. While a two-person household model is a natural setting in which to also consider home production, we will focus on the basic model without home production in order to minimize notation.

For ease of exposition we will refer to our two members as male and female, using subscripts m and f respectively. The household allocation will consist of total consumption and hours of work for each member. Once again we assume a separable utility function:

$$U(c, \ell_f, \ell_m) = u(c) - v(\ell_f) - v(\ell_m),$$

where for simplicity we have assumed that disutility from working takes the same form for both members. The household now faces the following budget equation:

$$c = w_f \ell_f + w_m \ell_m + I,$$

where w_j is the wage rate that member j faces, and as before, I is non-labor income of the household. Another effect is present in this model beyond the income and substitution

[4]Their argument was that the invention of home capital goods like the washing machine decreased the set of tasks for which labor was useful in home production. This served to increase the average product of labor in home production at the same time that it decreased the marginal product of labor in home production. Capturing these effects requires extending our analysis to explicitly include capital in home production. But in such an extension, the key implication of the theory is that a decrease in the marginal product of time in home production holding wages constant will lead to a decrease in home production time relative to market work.

[5]See Bourguignon and Chiappori (1992) for an early discussion and analysis of models that explicitly model households as composed of distinct individuals.

effects that we have previously noted. Specifically, the labor supply of each member depends not only upon non-labor income and their own wage, but also on the wage of the other household member. To see this, we derive the first-order conditions for the two choices of labor supply:

$$\begin{aligned} \ell_f &: \; u'(w_f\ell_f + w_m\ell_m + I)w_f = v'(\ell_f) \\ \ell_m &: \; u'(w_f\ell_f + w_m\ell_m + I)w_m = v'(\ell_m). \end{aligned}$$

Looking at the first equation, it is apparent that higher labor earnings of the male member act like an increase in non-labor income and thus will lead to lower labor supply of the female member. And similarly, higher labor earnings for the female member will decrease the labor supply of the male member. This indicates that the two labor supplies are jointly determined. Dividing the two FOCs by each other gives:

$$\frac{v'(\ell_f)}{v'(\ell_m)} = \frac{w_f}{w_m}.$$

This implies that relative wages are positively correlated with relative hours. Our functional form for v implies the following relationship between relative hours and relative wages:

$$\frac{\ell_f}{\ell_m} = \left[\frac{w_f}{w_m}\right]^{\gamma}.$$

Previously we showed how the value of the curvature parameter γ affected the response of labor supply of a single individual to changes in their own wage rate. In this setting this parameter also dictates the extent to which the household reallocates work across household members in response to changes in relative wages. An immediate implication is that the model has a force that can lead to specialization within the household, with the higher wage individual engaging in much more market work. If we were to generalize this setting to explicitly include home production as an activity, we would obtain the prediction that holding all else constant, differences in market wages are a force that leads to specialization in home and market work within the household. This force predicts that an exogenous reduction in the gender wage gap will lead to an increase in the labor supply of married females relative to married males.[6]

This extension also provides insight into a phenomenon known as the added worker effect. This describes the tendency for a secondary earner within a family to increase their labor supply when the primary earner experiences an unemployment spell. While our labor supply model does not account for unemployment, it does explain why an exogenous reduction in ℓ_m will imply an increase in ℓ_w.

The productivity of leisure activities

In our discussion thus far, we have specified preferences in terms of the disutility of work, so that leisure has been somewhat in the background. In this subsection we consider an

[6] See, for example the analysis in Attanasio, Low, and Sánchez-Marcos (2008).

alternative specification that places more emphasis on leisure. Previously we defined utility over consumption and hours and wrote:

$$U(c, \ell) = u(c) - v(\ell).$$

In this section we will instead define utility over consumption and leisure. Leisure will be defined as the difference between total discretionary time and time spent working. Denoting leisure by l, and letting $\bar{\ell}$ be the total amount of discretionary time, we have that:

$$l = \bar{\ell} - \ell$$

where as before, ℓ is time devoted to market work. Again assuming preferences that are separable, we will now write utility as:

$$\tilde{U}(c, l) = u(c) + \tilde{v}(l),$$

where $\tilde{v}$ is now a strictly concave function. Analogous to our earlier analysis, a commonly used functional form in this case is:

$$\tilde{v}(l) = \frac{1}{1 - 1/\gamma} l^{1-1/\gamma},$$

where $\gamma > 0$.

While the two specifications are very similar, there are some differences worth noting. First, in the previous specification, we noted that the solution for ℓ was always interior, independently of the value for non-labor income I. This followed from the fact that the marginal disutility of work was equal to zero when $\ell = 0$. This is no longer the case. For example, if non-labor income is sufficiently high, it is possible that $\ell = 0$ is optimal. Second, the previous specification implied that the elasticity of utility with respect to work, i.e., $\ell v'(\ell)/v(\ell)$ was constant. This is no longer true, as the elasticity of the utility with respect to time spent working is now given by $-\ell \tilde{v}'(\bar{\ell} - \ell)/v(\bar{\ell} - \ell)$. This value approaches minus infinity as ℓ approaches $\bar{\ell}$ and leisure tends to zero.

Starting from the specification in which we write utility as $\tilde{U}$, we now consider a simple extension to illustrate the point made by Aguiar, Bils, Hurst, and Charles (2021) regarding how changes in technology associated with leisure activities will directly impact labor supply choices. To do this we dig deeper into the specification of how leisure produces utility. In particular, we assume that there are J different leisure activities, which we index by j. An individual allocates their total leisure time l across the J different leisure activities, and we denote by l_j the amount of leisure time devoted to leisure activity j. In the simplest specification, total utility from leisure is then written as:

$$\tilde{v}\left(\sum_j \frac{(A_j l_j)^{1-\delta}}{1-\delta}\right),$$

where the A_j can be interpreted as the productivity of leisure time in leisure activity j. In this setting, increases in some or all of the A_j will mechanically act like an increase in the value of ψ in our original specification of utility. At this level, one might think that this is purely a game of words about whether we label the change in ψ a change in preferences versus a change in the productivity of leisure. The key contribution of Aguiar et al. (2021) was to develop a structure that allows us to examine this more deeply. A full analysis is beyond the scope of what we will cover here, but we can provide a few details. By examining the evolution of trend changes in the allocation of leisure relative to how the allocation of leisure time changes during a large downturn in economic activity, the authors were able to assess the extent to which the trend changes reflect changes in the A_j. One of their key findings is that there was a large increase in the productivity of leisure time allocated to computer and video games, and that this in turn led to a large increase in the total amount of time devoted to leisure time by young males.

12.3.3 Dynamic labor supply: a first look

Our analysis of the basic static labor supply problem emphasized wage and income elasticities as two important forces that shape the response of labor supply to changes in the economic environment. In dynamic settings there is another elasticity that is important for understanding labor supply responses. This will feature prominently in later parts of this chapter, but we first introduce it here by extending our previous analysis to a two period setting.

We now assume that an individual has preferences over consumption and labor supply given by:

$$U(c_1, c_2, \ell_1, \ell_2) = \sum_{t=1}^{2} \beta^{t-1} \left[\frac{1}{1-\sigma} c_t^{1-\sigma} - \frac{\psi}{1+1/\gamma} \ell_t^{1+1/\gamma} \right],$$

where the parameters σ, γ, and ψ are as before and $\beta \in (0, 1)$ is the discount factor. The individual faces a real wage of w_1 in the first period and w_2 in the second period. Importantly, we also assume that the individual can borrow and/or save at the interest rate r, and thus faces a lifetime budget constraint given by:

$$c_1 + \frac{c_2}{1+r} = w_1 \ell_1 + \frac{w_2 \ell_2}{1+r} + I,$$

where I is now initial non-labor wealth.[7]

If $\beta = 1/(1+r)$ and $w_1 = w_2$ the optimal choice of ℓ is the same in both periods and is essentially the same as analyzed in the previous subsection, modulo normalizing non-labor income appropriately. Our interest in this subsection is in the situation in which there are differences in wages in the two periods. To pursue this we derive first-order conditions for the individual lifetime maximization problem. Letting λ be the Lagrange multiplier on the lifetime budget constraint, and focusing on the case in which $\beta = 1/(1+r)$ we obtain the

[7] If the individual is not able to move purchasing power across time by either saving or borrowing then the individual will effectively solve two static problems of the form that we previously studied.

following first-order conditions:

$$\begin{aligned} c_1 &: c_1^{-\sigma} = \lambda \\ c_2 &: c_2^{-\sigma} = \lambda \\ \ell_1 &: \psi \ell_1^{1/\gamma} = \lambda w_1 \\ \ell_2 &: \psi \ell_2^{1/\gamma} = \lambda w_2. \end{aligned}$$

Note that the value of the Lagrange multiplier λ is the marginal utility of consumption. Dividing the last two equations by each other one obtains:

$$\left[\frac{\ell_1}{\ell_2}\right]^{1/\gamma} = \frac{w_1}{w_2},$$

or taking logs,

$$\log \ell_1 - \log \ell_2 = \gamma \left[\log w_1 - \log w_2\right].$$

Recalling that $\gamma > 0$, this expression indicates that hours and wage rates are always positively correlated intertemporally. Loosely speaking, because purchasing power can be transferred across time, the individual should work relatively more when wages are relatively high. The strength of this effect is dictated solely by the preference parameter γ, and for this reason γ is referred to as the intertemporal elasticity of substitution for labor supply. This elasticity, which holds the marginal utility of consumption fixed, is also known as the **Frisch elasticity**. This elasticity plays a key role in understanding the role of labor supply in the context of business cycle fluctuations. This equation has played a key role in work that estimates the Frisch elasticity using panel data on individuals.[8]

12.4 Dynamic models of labor supply: theory and estimation

We now present a fully dynamic stochastic life-cycle model, and generalize our derivations for the three labor supply elasticities (Marshallian, Hicksian, and Frisch) in Section 12.3. The key challenge, relative to the static model, is that non-labor income is now endogenous because it contains savings, a choice variable for the household. To overcome this challenge, we use the so-called "two-stage budgeting" approach (see Blundell and MaCurdy, 1999 and Keane, 2011). This method splits the dynamic optimization problem of the household in two stages. In the first stage, the household allocates its lifetime resources intertemporally, by choosing savings in each period. Once savings are chosen, non-labor income is given, and the second-stage problem, where the household chooses hours worked, is analogous to the static model

Consider an individual i who lives for T periods (where $T = \infty$ is a special case), has discount factor $\beta > 0$, and derives utility from consumption c_{it} and disutility from hours

[8] For future reference we note that combining the equation for consumption and labor in period t one obtains the equation $\log \ell_t = (\gamma/\psi) + \gamma \log w_t + \gamma(1-\sigma) \log c_t$, which can also be used to provide an estimate of γ if one has data on hours, wages and consumption.

worked ℓ_{it}. We normalize the time endowment to 1. Let w_{it} be the hourly wage of the individual at date t. Wages fluctuate stochastically and are a source of uncertainty for individuals. The individual can trade a risk-free asset a_{it} with constant gross rate of return $R = 1 + r$, subject to a borrowing limit $\underline{a}_i$. The government levies a proportional tax τ on labor income and pays a lump-sum transfer $\mathcal{T}$.

For a given initial wealth a_{i0}, the optimization problem of the individual is

$$\max_{\{c_{it},\ell_{it}\}_{t=0}^{T}} \mathbb{E}_0 \sum_{t=0}^{T} \beta^t U\left(c_{it}, \ell_{it}\right) \tag{12.7}$$

subject to

$$c_{it} + a_{i,t+1} = Ra_{it} + (1-\tau)\, w_{it}\ell_{it} + \mathcal{T},$$

$$a_{i,t+1} \geq -\underline{a}_i, \quad c_{it} \geq 0, \text{ and } 0 \leq \ell_{it} \leq 1.$$

In addition, we require the No-Ponzi condition $a_{i,T+1} \geq 0$ which, in the infinite horizon version of the model, becomes $\lim_{T\to\infty} a_{i,T}/R^T \geq 0$, as explained in Section 4.3.1.

For much of this section, it is useful to specialize to the same preferences used in Section 12.3:

$$U\left(c_{it}, \ell_{it}\right) = \frac{c_{it}^{1-\sigma} - 1}{1-\sigma} - \psi \frac{\ell_{it}^{1+1/\gamma}}{1+1/\gamma} \tag{12.8}$$

with $\sigma \geq 0$ and $\gamma \geq 0$. Recall that $1/\sigma$ is the intertemporal elasticity of substitution for consumption.

12.4.1 Derivations of labor supply elasticities

We start by defining non-labor income net of saving as

$$I_{it} = ra_{it} + \mathcal{T} - \left(a_{i,t+1} - a_{it}\right). \tag{12.9}$$

This definition is convenient because it allows us to rewrite the time t budget constraint as

$$c_{it} = (1-\tau)\, w_{it}\ell_{it} + I_{it}, \tag{12.10}$$

which is isomorphic to the formulation in the static model of Section 12.3. This is the cornerstone of the two-state budgeting method. To fully understand this approach, normalize the time endowment to 1, and rewrite the budget constraint as

$$c_{it} + (1-\tau)\, w_{it}\left(1 - \ell_{it}\right) = Y_{it},$$

where Y_{it} is "full income", i.e. the maximum possible income earned by the individual at time t taking $a_{i,t+1}$ as given, or

$$Y_{it} = (1-\tau)\, w_{it} + R_t a_{it} + \mathcal{T} - a_{i,t+1}. \tag{12.11}$$

The two-state budgeting approach exploits the idea that one can split the dynamic optimization problem of the household (12.7) in two steps. First, the household decides how to optimally allocate its life-time full income period by period. This amounts to choosing a sequence $\{a_{i,t+1}\}_{t=0}^{T}$ based on the Euler equation, the intertemporal optimality condition that describes the trade-off between consuming today and consuming next period.[9] Next, given $a_{i,t+1}$, the individual chooses (c_{it}, ℓ_{it}) at every t based on their intratemporal first-order condition and the budget constraint. Thus, in what follows, we focus on a generic time t and take $a_{i,t+1}$ as given.

Marshallian Elasticity. We start by deriving the expression for the Marshallian elasticity $\varepsilon^{M}_{\ell,w}$. Recall from the discussion of Section 12.3 that this is the total elasticity of hours worked ℓ_{it} to a change in the wage w_{it}. By total, we mean that it incorporates both the substitution and the income effect. Namely, because in calculating it we do not compensate the change in the wage with a corresponding change in non-labor income of the opposite sign, i.e. we keep I_{it} fixed, this elasticity is also called "uncompensated."

Start by differentiating the budget constraint (12.10) with respect to w_{it}, c_{it} and ℓ_{it}. Rearranging terms, we obtain:

$$\frac{d\log c_{it}}{d\log w_{it}} = \left(1 - s^{I}_{it}\right) + \left(1 - s^{I}_{it}\right)\frac{d\log \ell_{it}}{d\log w_{it}}, \tag{12.12}$$

where $s^{I}_{it} = I_{it}/c_{it} < 1$ is the non-labor income share of consumption expenditures. The intratemporal first-order condition with respect to ℓ_{it} yields

$$c_{it}^{-\sigma} w_{it}\left(1-\tau\right) = \psi \ell_{it}^{\frac{1}{\gamma}}. \tag{12.13}$$

Differentiating (12.13) with respect to w_{it}, c_{it} and ℓ_{it}, we arrive at:

$$\sigma\frac{d\log c_{it}}{d\log w_{it}} + \frac{1}{\gamma}\frac{d\log \ell_{it}}{d\log w_{it}} = 1. \tag{12.14}$$

Combining (12.12) and (12.14), we obtain the final expression for the Marshallian elasticity

$$\varepsilon^{M}_{\ell,w} \equiv \frac{d\log \ell_{it}}{d\log w_{it}} = \frac{1-\sigma\left(1-s^{I}_{it}\right)}{1/\gamma + \sigma\left(1-s^{I}_{it}\right)} \tag{12.15}$$

which, modulo the different definition of s^{I}_{it} adapted to the dynamic model, is the same as equation (12.3) in the static model of Section 12.3. Once again, note that $\varepsilon^{M}_{\ell,w}$ can be positive or negative, depending on whether the substitution effect is larger or smaller than the income effect.

Hicksian elasticity. We now turn to the Hicksian or "compensated" elasticity. This elasticity measures how hours worked optimally respond to a wage change that is compensated by an equal change in non-labor income of the opposite sign. This compensation offsets the income effect from the wage change, thus the Hicksian elasticity only captures the substitution effect that leads individuals to allocate more time to work (less to leisure) when the wage (the price of leisure) goes up.

Let $\ell_{it}\left(w_{it}, I_{it}\right)$ denote optimal hours worked expressed as a function of wage and net non-labor income from the first-order condition (12.13). Differentiating this function, we obtain:

$$d\ell_{it}\left(w_{it}, I_{it}\right) = \frac{\partial \ell_{it}\left(w_{it}, I_{it}\right)}{\partial w_{it}}dw_{it} + \frac{\partial \ell_{it}\left(w_{it}, I_{it}\right)}{\partial I_{it}}dI_{it},$$

[9] We derived and discussed the Euler equation for consumption in Chapter 4.

and rearranging, we arrive at:

$$\frac{d\ell_{it}\left(w_{it},I_{it}\right)}{dw_{it}}\frac{w_{it}}{\ell_{it}}=\frac{\partial\ell_{it}\left(w_{it},I_{it}\right)}{\partial w_{it}}\frac{w_{it}}{\ell_{it}}+\frac{\partial\ell_{it}\left(w_{it},I_{it}\right)}{\partial I_{it}}\frac{I_{it}}{\ell_{it}}\left(\frac{dI_{it}}{dw_{it}}\frac{w_{it}}{I_{it}}\right).$$

The term on the left-hand side is the Marshallian elasticity. The first term on the right-hand side can be interpreted as the Hicksian elasticity (i.e., the pure substitution effect) as long as the second term exactly compensates the individual with a change in non-labor income equal to the labor income shift, or as long as $dI_{it}/dw_{it}=(1-\tau)\,\ell_{it}$. Thus, expressing the equation above in terms of elasticities, we arrive at the so-called Slutsky equation:

$$\varepsilon^{H}_{\ell,w}=\varepsilon^{M}_{\ell,w}-\varepsilon_{\ell,I}\left(\frac{1-s^{I}_{it}}{s^{I}_{it}}\right), \tag{12.16}$$

which is the dynamic counterpart of equation (12.4) in the static model of Section 12.3.

To obtain the elasticity of hours worked to net non-labor income, differentiate (12.13) with respect to ℓ_{it} and I_{it}. Simple algebra yields

$$\varepsilon_{\ell,I}\equiv\frac{d\log\ell_{it}}{d\log I_{it}}=\frac{-\sigma s_{it}}{1/\gamma+\sigma\left(1-s_{it}\right)}$$

which, substituted into (12.16) , delivers the Hicksian labor supply elasticity

$$\varepsilon^{H}_{\ell,w}=\frac{1}{1/\gamma+\sigma\left(1-s^{I}_{it}\right)}. \tag{12.17}$$

Once again, note that the Hicksian is always positive. In addition, $\varepsilon^{H}_{\ell,w}\geq\varepsilon^{M}_{\ell,w}$ where equality holds in the absence of income effects, i.e. when period utility is linear in consumption, or $\sigma\to 0$.

The fact that the expressions for Marshallian and Hicksian elasticities are the same in the static and dynamic model suggests that these are fundamentally static concepts, as both hold constant the intertemporal allocation of resources. These concepts are, therefore, especially useful to analyze shifts in wages or in taxes that are perceived as being permanent by households. Examples of the former are a permanent, or very persistent, increase in individual labor productivity, a promotion associated with a pay raise, or a move to a better-paid job. Examples of the latter are income tax reforms that are expected to last indefinitely. In the absence of binding borrowing constraints, following this sort of wage increases (or a tax cuts), the household would spend roughly all the extra income every period and there would be no, or little, change in saving patterns.[10] If, however, we want to analyze temporary wage changes (e.g., a seasonal bonus) or income tax changes (e.g., tax reforms with expiring provisions) that lead to a shift in saving patterns, we need a different concept of elasticity. This motivates us to introduce the concept of the Frisch elasticity.

Frisch elasticity. The Frisch elasticity of labor supply describes the response of hours worked to a change in the wage keeping the marginal utility of wealth constant. It is the

[10] There might be, however, a change in non-labor income if, for example, some of the additional tax reveues are redistributed as lump-sum transfers to households. As in the static model, this makes the Hicksian the relevant elasticity.

relevant concept to analyze how either a transitory or an anticipated wage change affects hours. Small enough temporary wage fluctuations have only a negligible effect on lifetime wealth, and anticipated wage changes convey no new information. Thus neither one impacts the marginal utility of wealth. As a result, the Frisch elasticity is especially useful to assess the implications of wage fluctuations over the business cycle or of temporary tax changes.

It is easy to derive an expression for the Frisch elasticity for generic period utility $U(c_{it}, \ell_{it})$ without specializing to a particular functional form for now. Let the marginal utility of wealth for individual i at date t be λ_{it}. The first-order conditions with respect to consumption and hours worked for the problem specified previously in (12.7) are (U_c is the marginal utility of consumption, U_ℓ is the marginal disutility of working, and U_{ij}, $i, j = c, \ell$ represents a second derivative)

$$U_c = \lambda_{it} \tag{12.18}$$

and

$$-U_\ell = \lambda_{it}(1-\tau) w_{it}. \tag{12.19}$$

Differentiating (12.19), while keeping λ_{it} constant, yields

$$-\ell_{it} U_{\ell\ell} \frac{d\ell_{it}}{\ell_{it}} - U_{\ell c} dc_{it} = \lambda_{it}(1-\tau) w_{it} \frac{dw_{it}}{w_{it}}.$$

Using (12.19) again to substitute out $\lambda_{it}(1-\tau) w_{it}$ and rearranging leads to:

$$\ell_{it} U_{\ell\ell} \frac{d\ell_{it}}{\ell_{it}} + \ell_{it} U_{\ell c} \frac{dc_{it}}{d\ell_{it}} \frac{d\ell_{it}}{\ell_{it}} = U_\ell \frac{dw_{it}}{w_{it}}. \tag{12.20}$$

Differentiating (12.18) gives

$$\frac{dc_{it}}{d\ell_{it}} = \frac{-U_{c\ell}}{U_{cc}}$$

which, substituted into (12.20), gives the general expression for the Frisch elasticity:

$$\varepsilon^F_{\ell,w} = \frac{U_\ell}{\ell_{it} U_{\ell\ell} - \ell_{it}(U^2_{\ell c}/U_{cc})}. \tag{12.21}$$

It is useful to derive the Frisch elasticity for some particular functional forms of the period utility function. For (12.8), it is easy to see from (12.21) that $\varepsilon^F_{\ell,w} = \gamma$. It is now clear why (12.8) is such a common preference specification in macroeconomics: its two curvature parameters fully control two key elasticities, the intertemporal elasticity of substitution $(1/\sigma)$ and the Frisch elasticity (γ). Comparing (12.15), (12.17) and (12.21) for the separable utility (12.8) we conclude that

$$\varepsilon^F_{\ell,w} \geq \varepsilon^H_{\ell,w} \geq \varepsilon^M_{\ell,w}$$

with equality holding when $\sigma \to \infty$. In this case, utility becomes quasi-linear and the income effect vanishes.

Consider now the **GHH utility function** introduced by Greenwood, Hercowitz, and Huffman (1988)

$$U(c_{it}, \ell_{it}) = \frac{\left(c_{it} - \psi \frac{\ell_{it}^{1+1/\gamma}}{1+1/\gamma}\right)^{1-\sigma} - 1}{1-\sigma}. \tag{12.22}$$

Applying (12.21) yields again that $\varepsilon^F_{\ell,w} = \gamma$. Another key property of these preferences is that the income effect is zero. Thus, for GHH utility, $\varepsilon^M_{\ell,w} = \varepsilon^H_{\ell,w} = \varepsilon^F_{\ell,w} = \gamma$.

We conclude by noting that not all utility functions imply a constant Frisch elasticity. For example, for the specifications

$$U(c_{it}, \ell_{it}) = \frac{\left(c_{it}^{1-\psi}(1-\ell_{it})^{\psi}\right)^{1-\sigma} - 1}{1-\sigma} \quad \text{and} \quad U(c_{it}, \ell_{it}) = \frac{c_{it}^{1-\sigma} - 1}{1-\sigma} + \psi \frac{(1-\ell_{it})^{1-\gamma}}{1-\gamma}$$

it is the Frisch elasticity of leisure that is constant, while the Frisch elasticity of hours now depends on hours worked ℓ_{it}. For the first specification, we have

$$\varepsilon^F_{\ell,w} = \left(\frac{1-\ell_{it}}{\ell_{it}}\right)\left(\frac{1-(1-\psi)(1-\sigma)}{\sigma}\right)$$

and for the second

$$\varepsilon^F_{\ell,w} = \left(\frac{1-\ell_{it}}{\ell_{it}}\right)\frac{1}{\gamma}.$$

Thus, in both cases, individuals who work more hours are less responsive to temporary wage changes.

Expression (12.21) was derived only under the assumption that individuals optimize intratemporally. Nothing was explicitly assumed about intertemporal saving behavior, nor about asset market structure. One could be, therefore, tempted to conclude that, as long as equation (12.19) holds, observing the response of hours worked to a transitory wage change identifies γ correctly. This deduction is, however, incorrect. For example, for a constrained individual even a temporary increase in labor income affects the marginal utility of wealth and that, in turn, affects optimal labor supply.

To further appreciate this point, and illustrate the intertemporal nature of this elasticity, note that abstracting from uncertainty and in the absence of borrowing constraints, under separable utility (12.8), the Euler equation for problem (12.7) yields

$$c_{it}^{-\sigma} = \beta R c_{i,t+1}^{-\sigma}. \tag{12.23}$$

Substituting the intratemporal first-order condition (12.13) into (12.23) we obtain

$$\frac{d\log(\ell_{i,t+1}/\ell_{it})}{d\log(w_{i,t+1}/w_{it})} = \gamma.$$

Thus, as shown for the 2-period model in Section 12.3, the Frisch elasticity dictates the response of relative hours worked across two periods with respect to changes in relative wages. If taxes were time varying, γ would also describe the elasticity of relative hours to relative changes in taxes between the two periods. The larger is γ the more individuals will shift hours worked in response to temporary wage or tax changes. Importantly, this derivation hinges on the fact that equation (12.23) holds with equality, i.e. liquidity constraints do not bind, and there is no uncertainty and thus no precautionary motive. In the next section, we discuss how the presence of these two factors complicate the estimation of γ.

12.4.2 Estimation of Frisch elasticity from micro data

In the separable utility specification, commonly used in macroeconomics, the two key elasticity parameters are σ and γ, so historically there has been great interest in estimating both of them. The parameter σ governs both the intertemporal elasticity of substitution for consumption and risk aversion. Under the former interpretation, $1/\sigma$ measures the sensitivity of consumption growth to the interest rate. Under the second one, it measures, for example, how households allocate their portfolio between risky and safe assets, for a given equity premium, or how large the equity premium can be, for a given volatility of aggregate consumption (see Chapter 7). Thus, the estimation of σ has been mostly the domain of the consumption and asset pricing literatures.[11] For this reason, here we focus on γ and present a brief overview of the empirical challenges underlying the estimation of the Frisch elasticity from micro data, and how the macro-labor literature has evolved over time to tackle these challenges.

In what follows, we assume that households solve (12.7) and have rational expectations, i.e. they know the stochastic process underlying fluctuations in w_{it}. In addition, we assume that utility takes the separable functional form in (12.8), therefore the parameter to be estimated is γ. We discuss two complementary approaches to the measurement of γ, a structural approach and an experimental one.

Structural approach

Taking logs of the two optimality conditions (12.18) and (12.19), and using the notation $const_i$ for individual specific constant terms, we obtain:

$$\log c_{it} = -\frac{1}{\sigma} \log \lambda_{it} \tag{12.24}$$

and

$$\log \ell_{it} = const_i + \gamma \log w_{it} + \gamma \log \lambda_{it}. \tag{12.25}$$

The minimum requirement to obtain an empirical estimate of γ from equation (12.25) is the availability of longitudinal micro data on individual hourly wages and hours worked. The key challenge is that λ_{it}, the marginal utility of consumption for individual i, is unobservable. Suppose we naively try to estimate equation (12.25) by OLS, possibly using individual fixed effects to absorb the individual-specific constant term. This approach would be problematic because λ_{it}, which enters the residual, is likely to be correlated with the wage when the latter has a persistent component that affects wealth and consumption. In this case, $\text{Cov}(\log w_{it}, \log \lambda_{it}) < 0$, which induces a downward bias in the estimator $\hat{\gamma}$. The same logic applies if the estimation is done in first differences. The empirical literature has tackled this problem in three ways, by using (i) instrumental variables, (ii) data on expected wage growth, and (iii) data on consumption expenditures.

Instrumental variables. The most natural way to address this potential bias in OLS is through instrumental variables. Express (12.25) in first differences between time $t-1$ and

[11] As we discuss later in this chapter, conditional on assuming that utility is separable between consumption and hours, macroeconomists have often focused on the case in which $\sigma = 1$ since it is the only value consistent with a balanced growth path solution that features constant hours.

time t

$$\Delta \log \ell_{it} = \gamma \Delta \log w_{it} + \gamma \Delta \log \lambda_{it}. \tag{12.26}$$

To substitute out the Lagrange multiplier note that, from household optimization, the optimal saving decision yields the following Euler equation expressed in logarithms

$$\log \left(\lambda_{it} - \phi_{it}\right) = const_i + \log \mathbb{E}_t \left[\lambda_{i,t+1}\right], \tag{12.27}$$

where ϕ_{it} is the the multiplier on the borrowing constraint. Assuming that λ_{it} is conditionally log-normally distributed, then

$$\log \mathbb{E}_t \left[\lambda_{i,t+1}\right] = \mathbb{E}_t \left[\log \lambda_{i,t+1}\right] + \frac{1}{2} \operatorname{Var}_t \left(\log \lambda_{i,t+1}\right), \tag{12.28}$$

where Var_t is the conditional variance. Under rational expectations, we can substitute (12.28) into (12.27) and write (12.27) as

$$\log \left(\lambda_{it} - \phi_{it}\right) = const_i + \log \lambda_{i,t+1} - \xi_{i,t+1} + \frac{1}{2} \operatorname{Var}_t \left(\log \lambda_{i,t+1}\right),$$

where $\xi_{i,t+1}$ is the prediction error for the marginal utility of consumption which is orthogonal to all information available to worker i at time t. Using a first-order approximation of the left-hand-side around $\phi^* = 0$, and recognizing that $\operatorname{Var}_t \left(\log \lambda_{i,t+1}\right) = \operatorname{Var}_t \left(\xi_{i,t+1}\right)$, we obtain

$$\log \lambda_{it} - \frac{\phi_{it}}{\lambda_{it}} = const_i + \log \lambda_{i,t+1} - \xi_{i,t+1} + \frac{1}{2} \operatorname{Var}_t \left(\xi_{i,t+1}\right).$$

This equation can be used to substitute out the Lagrange multiplier λ_{it} in (12.26) and obtain

$$\Delta \log \ell_{it} = \gamma \Delta \log w_{it} + \gamma \left[\xi_{it} - \frac{\phi_{i,t-1}}{\lambda_{i,t-1}} - \frac{1}{2} \operatorname{Var}_{t-1} \left(\xi_{it}\right)\right]. \tag{12.29}$$

Once again, one would expect the residual ξ_{it} to be negatively correlated with $\Delta \log w_{it}$: positive news about wage growth, unless very transitory, would translate into an unexpected growth in consumption, and hence a decline in ξ_{it}.

MaCurdy (1981) abstracts from potentially binding liquidity constraints and uninsurable uncertainty, i.e. assumes $\phi_{i,t-1} = 0$ and $\operatorname{Var}_{t-1} \left(\xi_{it}\right) = 0$, and uses lagged wage growth $\Delta \log w_{i,t-1}$ as an instrument to deal with the correlation between ξ_{it} and $\Delta \log w_{it}$ in equation (12.29). In practice, this is a relatively ineffective strategy for two reasons. First, in surveys hourly wages are typically measured as earnings divided by hours worked, but hours are notoriously measured with error. By itself, this yields a downward bias in the estimate of γ. This is the first reason why lagged wage growth is a weak instrument. The second one is that if the statistical process for wages is very persistent, then the autocorrelation of wage growth with past wage growth is very small –for example, in the limiting case in which wages follow a random walk this autocorrelation is exactly zero.

Wage growth expectations. One can always write

$$\Delta \log w_{it} = \mathbb{E}_{t-1} \left[\Delta \log w_{it}\right] + \varepsilon_{it} \tag{12.30}$$

where, under rational expectations, the shock ε_{it} is orthogonal to predicted wage growth $\mathbb{E}_{t-1}[\Delta \log w_{it}]$. Substituting (12.30) into (12.26) , we obtain

$$\Delta \log \ell_{it} = \gamma \mathbb{E}_{t-1}[\Delta \log w_{it}] + \gamma (\varepsilon_{it} + \Delta \log \lambda_{it}). \tag{12.31}$$

Under the permanent-income hypothesis –i.e., once again abstracting from liquidity constraints– consumption growth between $t-1$ and t, and therefore $\Delta \log \lambda_{it}$, is only affected by the news accruing over that period, i.e. by ε_{it}, but not by anything which is already incorporated in the information set at time $t-1$. As a result, with an external measure of expected wage growth one could estimate γ consistently. Pistaferri (2003) estimates equation (12.31) directly using Italian survey data which contain information about individual subjective earnings expectations, and obtains an estimate of $\gamma = 0.7$.

Domeij and Floden (2006) point out that both the wage growth expectations approach and the IV strategy fail if there are binding borrowing constraints. Consider first equation (12.31) . Individuals with very steep predictable wage growth are more likely to be constrained because they would like to consume more out of their future income stream, which induces a negative correlation between $\mathbb{E}_{t-1}[\Delta \log w_{it}]$ and the growth in the marginal utility of consumption $\Delta \log \lambda_{it}$. This argument keeps holding in equation (12.29) if one uses $\Delta \log w_{i,t-1}$ as an instrument for $\Delta \log w_{it}$. In addition, there is another form of bias coming from constrained agents. In the absence of liquidity constraints, households choose to work hard in periods with temporarily high wages and to enjoy more leisure when wages are low. However, when borrowing constrained agents receive transitory negative income shocks, in order to avoid cutting consumption drastically they will increase labor supply in response to falling wages. This force is another source of downward bias in the estimation of the Frisch elasticity.

Low (2005) advances a related critique. Even in the absence of binding constraints, uninsurable income risk leads to a precautionary saving motive, as shown in Chapter 11. One way in which workers can increase their precautionary saving is to work more and use the additional earnings to increase saving, a sort of precautionary labor supply motive. From equation (12.29) it is clear that if this motive is strong (i.e., $\text{Var}_{t-1}(\xi_{it})$ is large) when wage growth is high, precautionary labor supply is yet another source of downward omitted variable bias. This would be the case, for example, if young workers who typically face steep wage growth also hold low amounts of wealth.

Consumption data. If consumption expenditures are directly observable in the data, one can exploit equation (12.24) to use consumption in place of λ_{it}. Combining (12.24) and (12.26) and expressing the resulting equation in first differences yields

$$\Delta \log \ell_{it} = \gamma \Delta \log w_{it} - \gamma \sigma \Delta \log c_{it}.$$

Altonji (1986b) implements this approach for the U.S. using data on the Panel Study of Income Dynamics (PSID), a panel dataset with information on earnings, hours worked, and expenditures on some selected consumption goods (e.g., food, clothing, and shelter). The advantage of this approach is that it is based on the envelope condition which always holds with equality, so it is immune to binding borrowing constraint. Its key limitation is that consumption expenditures in surveys are measured with error and, as a result, can have a weak correlation with the true marginal utility of consumption. A second problem is that food is often used as a proxy for total consumption because it is better measured and more commonly present in surveys. However, it only represents a relatively small share of the budget for most households.

Experimental and quasi-experimental evidence

In light of the challenges in estimating the Frisch elasticity through structural approaches, the literature has explored, in parallel, RCTs or quasi-exogenous empirical variation that can shed light on this key parameter. One example of the former approach is Fehr and Goette (2007). They conducted a randomized field experiment at a bicycle messenger service in Zurich, Switzerland. The bicycle messengers are paid solely on commission, i.e. they retain a share of the revenues they generate. The authors implemented an exogenous and transitory increase of 25 percent in the rate of commission for a random subgroup of workers. Since the wage was increased only during four weeks, its impact on the workers' lifetime wealth is negligible. By comparing the treated group with the control group, the authors estimated an intertemporal elasticity of labor supply which exceeds 1. Bianchi, Gudmundsson, and Zoega (2001) is an example of quasi-exogenous variation from the macro data. This paper exploits a tax-reform in Iceland resulting in a year, 1987, free of labor income taxes. This tax-free year created a strong incentive for intertemporal substitution of work, but a minimal income effect because the reform applied only to one specific year, and thus did only imply a small change in life-time resources. As a result, the tax-free year offers a rare natural experiment suitable to estimate the Frisch elasticity. The authors estimate an average Frisch elasticity around 0.4.

Optimization frictions

Optimization frictions occur when households encounter costs associated with updating previously optimal decisions in light of the new environment. Examples are physical costs of adjusting the choice variable, costs of acquiring new information, or behavioral biases such as the so called "status-quo" bias. Chetty (2012) pointed out that these frictions can pose a challenge to estimate labor supply elasticities from microdata.

To illustrate this point, we consider a static model of a household with GHH preferences. This preference specification, albeit non-standard, allows us to abstract from income effects— which simplifies the derivations— and to focus on the role of optimization frictions. Consider the static optimization problem of a household:

$$U\left(c^{*},\ell^{*}\right)=\max_{c,h}\ \log\left(c-\psi\frac{\ell^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}\right)-\Delta\mathbb{I}_{\{\ell\neq\ell^{*}\}} \tag{12.32}$$

subject to

$$c=w\ell.$$

Recall that under these preferences Frisch, Marshallian, and Hicksian elasticities are all the same and equal to γ. Note the fixed re-optimization cost $\Delta>0$, expressed in utility terms, which kicks in whenever the choice of hours worked differs from its current level which we assume to be at the optimum, given the current wage w. It is easy to see that the optimal choices of hours worked ℓ^* and consumption c^* when the wage equals w are:

$$\ell^* = \left(\frac{w}{\psi}\right)^\gamma \tag{12.33}$$

and

$$c^* = w\ell^*,$$

and utility evaluated at the optimum is

$$U(c^*, \ell^*) = -\log(1+\gamma) - \gamma \log \psi + (1+\gamma)\log w,$$

which is increasing in the wage and decreasing in the disutility of labor ψ, as expected.

Consider now the effect of an increase in the wage from w to $w(1+\varepsilon)$ on utility and, for the moment, ignore re-optimization costs, i.e. assume the usual frictionless optimization environment. This utility gain can be decomposed into the direct effect of the wage change, holding ℓ fixed at the old choice ℓ^*, plus the effect induced by the behavioral response of reoptimizing labor supply, keeping consumption fixed at the new choice:

$$U(c_\varepsilon^*, \ell_\varepsilon^*) - U(c^*, \ell^*) = [U(c_\varepsilon^*, \ell^*) - U(c^*, \ell^*)] + [U(c_\varepsilon^*, \ell_\varepsilon^*) - U(c_\varepsilon^*, \ell^*)],$$

where the notation $(c_\varepsilon^*, \ell_\varepsilon^*)$ denotes the updated choices after the wage increase, and (c^*, ℓ^*) denotes the old choice before the wage changed.

From (12.32) evaluated at $(c_\varepsilon^*, \ell^*)$ and (c^*, ℓ^*) one obtains that

$$U(c_\varepsilon^*, \ell^*) - U(c^*, \ell^*) = \log(1+\varepsilon(1+\gamma)),$$

and note that this gain is not costly to achieve, so households will always update their consumption decision, conditional on their choice of hours worked. From (12.32) evaluated at $(c_\varepsilon^*, \ell_\varepsilon^*)$ and $(c_\varepsilon^*, \ell^*)$ one obtains that

$$U(c_\varepsilon^*, \ell_\varepsilon^*) - U(c_\varepsilon^*, \ell^*) = (1+\gamma)\log(1+\varepsilon) - \log(1+\varepsilon(1+\gamma)). \tag{12.34}$$

This is the term we are interested in because it represents the utility gain from re-optimizing the choice of hours, which is costly. This term is different from zero only up to second order. To see this, consider a second order Taylor expansion of both terms on the right-hand side of (12.34) around $\varepsilon = 0$ (or $\ell_\varepsilon^* = \ell^*$):

$$\begin{aligned}(1+\gamma)\log(1+\varepsilon) &\simeq (1+\gamma)\varepsilon - \frac{1}{2}(1+\gamma)\varepsilon^2 \\ \log(1+\varepsilon(1+\gamma)) &\simeq (1+\gamma)\varepsilon - \frac{1}{2}(1+\gamma)^2\varepsilon^2\end{aligned}$$

and note that up to a first order these two expressions are the same, so full optimization of hours worked only leads to second order utility gains approximately of size

$$U(c_\varepsilon^*, \ell_\varepsilon^*) - U(c_\varepsilon^*, \ell^*) \simeq \frac{1}{2}\gamma(1+\gamma)\varepsilon^2. \tag{12.35}$$

Now, we introduce optimization frictions through the fixed utility loss Δ. In order for the individual to choose to change hours in order to respond to the wage change, Δ must be smaller than (12.35) , or

$$\varepsilon > \varepsilon_{\min} = \left(\frac{2\Delta}{\gamma\,(1+\gamma)}\right)^{\frac{1}{2}}.$$

Note that the higher the labor supply elasticity, the lower this threshold is because, when γ is very large, even small changes in hours lead to large changes in utility, as apparent from (12.34). Thus, in the presence of fixed cost of re-optimization, hours might not be responsive to small changes in wages, which creates a downward bias in the estimation of the structural parameter γ. However, hours would respond with elasticity γ to large enough changes in wages. Specifically, in the presence of optimization frictions of size Δ, it descends from (12.33) that we would only observe changes in log hours worked (approximately) larger than

$$\Delta \log \ell^{*} > \gamma\varepsilon_{\min} = \left(\frac{2\Delta\gamma}{1+\gamma}\right)^{\frac{1}{2}}.$$

Taking stock, there might be several frictions that induce agents to deviate from the optimal choices predicted by standard frictionless economic models. When these frictions are salient, microeconometric studies of labor supply can be uninformative, unless they analyze episodes where wage changes are large enough to overcome these frictions. We conclude by noting that this is a more general insight that applies beyond labor supply.

12.5 Labor supply and balanced growth

As illustrated in Chapter 13, developed economies tend to grow at a constant rate, and this growth path is "balanced,"meaning that consumption, investment, and output grow at the same rate. Hours worked, however, cannot keep growing at a constant rate because, inevitably, at some point they would hit the time endowment and stop growing, violating balanced growth. Figure 12.1 shows that in the U.S. hours worked per person have remained relatively stable over time. King, Plosser, and Rebelo (1988) showed that this property of long-run labor supply puts important restrictions on preferences.

Consider the social planner formulation of the neoclassical growth model where output is produced by a Cobb-Douglas production function with constant capital share α (another property of balanced growth):

$$\max_{\{C_t, L_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U\left(C_t, 1 - L_t\right) \tag{12.36}$$

subject to

$$C_t + K_{t+1} = Z_t^{1-\alpha} K_t^{\alpha} L_t^{1-\alpha} + (1-\delta)\, K_t$$

and

$$K_0 \text{ given.}$$

Let Z_t denote labor-augmenting technological change which we assume to grow at a constant rate g, i.e. $Z_{t+1} = Z_t\,(1+g)$ for all t. Along a balanced growth path, consumption, investment, capital and output all grow at rate g and hours are constant. What do these features imply for U? In Appendix 12.A we prove that the period utility function U is consistent with balanced growth—in particular with the observation that hours worked are constant as the economy grows at a constant rate—if and only if

$$U(C, 1-L) = \begin{cases} \dfrac{C^{1-\sigma}-1}{1-\sigma} v\,(1-L) & \text{if } \sigma \neq 1 \\ \log\,(C) + v\,(1-L) & \text{if } \sigma = 1. \end{cases} \tag{12.37}$$

In a static setting with only labor income, these specifications have the implication that a higher wage rate has no effect on desired hours of work, i.e., income and substitution effects exactly cancel out. This is what allows hours worked to remain constant along the balanced growth path in an economy with positive productivity growth, which in turn implies growth in real wages over time.

Figure 12.1 also shows that, for many countries, hours worked have slowly diminished over the post-war period. Thus, the postwar U.S. experience, over which hours have shown no net decrease and which is the main argument for the use of "balanced-growth preferences," eq. (12.37), is to some extent a striking exception more than a representative feature of modern economies. In the next section we discuss factors that can account for the variation in hours worked across countries and over time.

12.6 Labor supply across countries and over time

Two classic issues in the study of labor supply are that (i) there exist large differences in hours worked per capita across countries, as shown in Table 12.1, and (ii) hours worked vary widely over time at the business cycle frequency, as shown in Figure 12.4. This section serves to illustrate how representative household models have been used to shed light on the role of labor supply in each of these contexts.

A key message from this section is that the optimal labor supply implied by a representative agent model with a simple and parsimonious specification of preferences–a period utility function of the form

$$U(C_t, L_t) = \log(C_t) - \frac{\psi}{1+1/\gamma} L_t^{1+1/\gamma}$$

accounts for an important part of variation in aggregate hours of work, both across countries and over time. Put somewhat differently, labor supply is relevant for understanding aggregate labor market outcomes. While quantitatively important, this parsimonious model of labor supply does not capture all aspects of the data, indicating that some of the variation found in the data reflects factors other than labor supply.

12.6.1 Labor supply and taxation across countries

A variety of factors can explain cross-country differentials in hours worked. A chief candidate is the level of taxes and transfers because they strongly influence work incentives. Prescott (2004) shows that households work less in countries with high tax rates on labor, and develops a simple theoretical framework to assess this hypothesis quantitatively.

Consider the problem of an infinitely-lived representative household who faces tax rates τ_c and τ_ℓ on consumption and labor income, and receives $\mathcal{T}_t$ as lump-sum transfers:

$$\max_{\{C_t, L_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \log C_t - \psi \frac{L_t^{1+1/\gamma}}{1+1/\gamma} \tag{12.38}$$

subject to

$$(1+\tau_c) C_t + K_{t+1} = (1-\tau_\ell) w_t L_t + r_t K_t + (1-\delta) K_t + \mathcal{T}_t.$$

A representative firm produces the final good with a Cobb-Douglas technology so the resource constraint for this economy is

$$C_t + G_t + K_{t+1} - (1-\delta) K_t = Y_t = Z_t^{1-\alpha} K_t^{\alpha} L_t^{1-\alpha},$$

where G_t is government consumption. The government uses tax revenue to finance spending on government consumption G_t and the lump sum transfer $\mathcal{T}_t$, subject to a period-by-period balanced budget. Note that the specification of preferences assumes that households do not value government consumption.[12]

Prescott assumes Z_t grows at a constant rate and $G_t = gY_t$ and focuses on the level of hours worked along the balanced growth path. Starting from the first-order condition for the optimal intratemporal allocation of consumption one can derive the following expression for hours along the balanced growth path:

$$L^* = \left[\frac{(1-\tau)(1-\alpha)}{\chi\psi}\right]^{\frac{\gamma}{1+\gamma}}, \tag{12.39}$$

where

$$\tau = \frac{\tau_\ell + \tau_c}{1+\tau_c}$$

is the effective tax return to labor income and $\chi \equiv C_t/Y_t$ is the (constant) consumption-output ratio along the balanced growth path.

Prescott takes the view that the capital share α, the labor disutility shifter ψ, and the Frisch elasticity γ are common across countries. Under this assumption, the only two terms in (12.39) that can explain cross-country differences in hours worked are τ and χ.[13]

As a concrete example, consider the U.S. and Italy. Prescott reports that in the mid-1990s, the average American aged 15-64 worked 25.9 hours per week, while the average Italian worked only 16.5 hours per week. The ratio between the two is a staggering 1.57. Prescott reports that, for the mid-1990s, τ equals 0.40 in the U.S. and 0.64 in Italy, and χ is 0.81 in the U.S. and 0.69 in Italy. If we set $\gamma = 1$, i.e. a unitary aggregate Frisch elasticity, we obtain a ratio of hours worked between the U.S. and Italy predicted by the theory of 1.42

[12] The results of the analysis are unaffected if one alternatively assumes that individuals value G_t but it enters utility separably with respect to C_t and L_t.

[13] Variation in χ in the data can come from variation in g or variation in the ratio of investment to output. The model as specified does not generate variation in the investment to output ratio. As a practical matter, variation in investment to output is relatively unimportant in this context. We note that variation in capital income tax rates is one possible source of variation in the investment to output ratio.

compared to 1.57 in the data. Calculations for a number of other countries confirm that this simple theory can account for much of the cross-country differences in labor supply.

Recall our earlier discussion in which we argued that the elasticity that matters for permanent changes in taxes is either the Marshallian when the tax revenue is not rebated, or the Hicksian when it is rebated. Prescott set $\sigma = 1$ based on balanced growth path considerations. Conditional on fixing the value of σ, both of these elasticities are determined by the value of γ, which is why the key elasticity parameter in equation (12.38) is γ.

The previous discussion focused on differences in hours worked across a sample of advanced economies. Bick, Fuchs-Schündeln, and Lagakos (2018) studied differences in hours across a broader sample of countries and found a systematic negative relationship between hours of work and the level of development as measured by GDP per capita. While it is true that less developed economies tend to have smaller tax and transfer systems than advanced economies, Bick et al. (2018) argue for an alternative explanation: the reason that hours of work are higher in poorer economies is because income effects are larger than substitution effects.

As discussed earlier, a commonly used specification of the period utility function that displays perfectly offsetting income and substitution effects is the following:

$$U(C_t, L_t) = \log(C_t) - \frac{\psi}{1+1/\gamma} L_t^{1+1/\gamma}.$$

There are two modifications to this specification that are commonly used to capture the possibility that income effects are larger than substitution effects. The first introduces a subsistence consumption term $\bar{C}$, resulting in a period utility function given by:

$$U(C_t, L_t) = \log(C_t - \bar{C}) - \frac{\psi}{1+1/\gamma} L_t^{1+1/\gamma}.$$

With this specification, individuals need to work enough hours in order to reach consumption of at least $\bar{C}$.[14] It follows that very low wages will lead individuals to choose to work long hours in order to reach consumption of at least $\bar{C}$. A key feature of this specification is that the magnitude of the income effect relative to the substitution effect decreases with the level of consumption, with the two effects being perfectly offsetting in the limit as consumption becomes very large.

A second specification that allows for the income effect to be larger than the substitution effect is:

$$U(C_t, L_t) = \frac{1}{1-\sigma} C_t^{1-\sigma} - \frac{\psi}{1+1/\gamma} L_t^{1+1/\gamma},$$

where $\sigma > 1$. Unlike the previous specification with a subsistence consumption term, a key feature of this specification is that the income effect is always larger than the substitution effect. Boppart and Krusell (2020) show that this specification gives rise to a balanced growth path in which hours of work decrease at a constant rate. They argue that such a balanced

[14] Our discussion here implicitly assumes that it is feasible for the individual to achieve consumption greater than $\bar{C}$. More generally, one would need to modify this specification to define utility when consumption is less than $\bar{C}$.

growth path offers a better description of long run changes in hours worked among current advanced economies. Ohanian, Raffo, and Rogerson (2008) offer an alternative interpretation of this time series evidence. They argue that changes in tax rates over time can account for much of trend behavior of hours among OECD economies after 1960. Whether income effects dominate substitution effects, and if so, whether this effect diminishes as consumption grows, remains an open issue.

12.6.2 Labor supply and business cycle fluctuations

When discussing the Frisch elasticity, we have stated that this elasticity is especially relevant to understand the role of labor supply in accounting for the large fluctuations of hours worked over the business cycle. Figure 12.4 shows that hours worked are very volatile over the cycle and are pro-cyclical. To illustrate the role of the Frisch elasticity for aggregate fluctuations in hours, it is useful to return to the household problem in (12.38), allowing now productivity Z_t to be stochastic and to be the driver of economic fluctuations. Equation (12.39) which determines optimal hours worked still holds, but now the consumption-output ratio varies over time. Expressed in logarithms, equation (12.39) yields

$$\log L_t = const - \frac{\gamma}{1+\gamma} \log\left(\frac{C_t}{Y_t}\right), \tag{12.40}$$

where *const* is a constant term. Expectations of changes in productivity are all captured by the current consumption-output ratio. Suppose productivity is random and is currently above trend, i.e. the economy is in a boom, but expected to revert towards the trend, for example because it follows an AR(1) process. The permanent income hypothesis suggests that consumption should increase less than income, as some of this increase in income is saved (see Chapter 11). Then, the consumption-output ratio will be low, and labor supply will be above its steady-state value. The sensitivity of aggregate hours worked L_t to fluctuations in aggregate productivity, and hence in the consumption-output ratio, is mediated by the value of the aggregate Frisch elasticity γ.

If the Frisch elasticity is infinite (as implied by the indivisible labor model of Rogerson (1988) and Hansen (1985) discussed later in this chapter) then the ratio $\gamma/(1+\gamma)$ equals one and log hours move one for one with changes in $\log(C_t/Y_t)$. But notably, even a Frisch elasticity of 1, a value consistent with the empirical evidence as we will argue later in Section 12.8, produces a response which is already half the size of the infinite-elasticity case. There is broad agreement that the Frisch elasticity is central to the business cycle behavior of aggregate labor supply.

We note that, in absence of capital and in a closed economy, $C_t = Y_t$ at every t, thus these preferences consistent with balanced growth imply that hours worked are invariant to any productivity change, including temporary ones. Since the procyclicality of aggregate hours is a salient stylized fact of developed economies, it is important for macroeconomic models that aim to be consistent with this key feature of the data to include capital and savings decisions.

Before concluding we should point out that, when taken to the data, equation (12.40) yields, typically, a poor approximation of aggregate hours dynamics at the business cycle frequency. In particular, hours decrease more than what would be predicted by movements in the consumption-output ratio during recessions. The time-varying term missing from equation (12.40) is called the *labor wedge* and acts like a countercyclical labor income tax in the intratemporal first-order condition of the representative agent. Chapter 14 discusses a number of possible interpretations of the labor wedge.[15] The most plausible explanation is that the neoclassical labor supply model which underlies (12.40) abstracts from labor market frictions, bilateral monopoly power in wage setting, and unemployment, all factors that can affect that relationship. Chapter 20 is devoted to these topics.

12.7 Dynamic returns to labor supply

The analyses of labor supply described so far all assume wages evolve exogenously over the life cycle. There is a long tradition in macroeconomics that takes a more plausible view that wages increase when individuals accumulate work experience or human capital. Imai and Keane (2004) show that explicitly incorporating human capital accumulation into a life-cycle model has sharp implications for the measurement of labor supply elasticities.

To focus on human capital, we abstract from borrowing constraints in the life-cycle problem (12.7), but we add dynamic returns to work. The individual problem becomes:

$$\max_{\{c_{it},\ell_{it}\}_{t=0}^{T}} \mathbb{E}_t \sum_{t=0}^{T} \beta_i^t \left[\frac{c_{it}^{1-\sigma}}{1-\sigma} - \psi \frac{\ell_{it}^{1+1/\gamma}}{1+1/\gamma} \right]$$

subject to

$$c_{it} + a_{i,t+1} = Ra_{it} + (1-\tau)\, w_{it}\ell_{it} + \mathcal{T}$$

and

$$w_{it} = \left(1 + \kappa \sum_{j=0}^{t-1} \ell_{ij}\right) w_{i0}.$$

The second constraint is the new feature of the model. It states that the hourly wage at time t depends on all past hours worked from $j=0$ to $t-1$. Thus, in this model, the return to an hour of work is not just the current wage as in (12.7). It consists of the current wage plus the expected present value of increased earnings in all future periods because working an additional hour today raises the wage permanently from today onward. Imai and Keane refer to this second component of the total return to work as the human capital term.

The first-order condition with respect to hours worked at date t is given by

$$\psi \ell_{it}^{\frac{1}{\gamma}} = \lambda_{it}\,(1-\tau) \left[w_{it} + \underbrace{\kappa w_{i0} \sum_{j=1}^{T-t} R^{-j}\mathbb{E}_t\left[\ell_{i,t+j}\right]}_{\text{human capital}} \right], \tag{12.41}$$

where we have used the Euler equation to simplify the right-hand side of (12.41). This optimality condition illustrates that the human capital term (the second term in the right-hand

[15] Shimer (2009) provides a critical overview of such interpretations.

side) creates a wedge between the current wage and the return to work. As a consequence, γ no longer captures the Frisch elasticity. The parameter γ is the elasticity of hours worked to the total return to work, whereas the Frisch elasticity measures the impact of an anticipated or transitory change in w_{it} on ℓ_{it}, i.e., keeping the marginal utility of wealth constant.

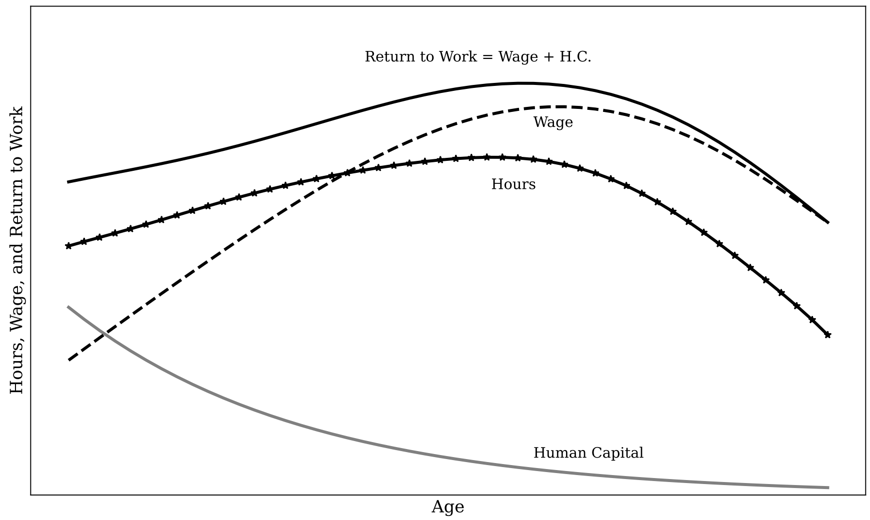

Figure 12.6: Life cycle profile of hours, wages, the human capital term, and the return to work in the Imai-Keane model

To understand why the presence of dynamic returns to work can affect the estimation of γ and of the intertemporal labor supply elasticity, consider Figure 12.6 adapted from Keane and Rogerson (2015). This figure plots the life cycle paths of hours, wages, the human capital term, and the return to work. Note that the human capital term is high early on, but it declines with age. The reason is that the return to working more in order to increase future wages is larger when the individual is young, i.e. t is small, as clear from the sum in equation (12.41) because the worker can enjoy the higher wages for longer. Ignoring these dynamic returns and naively estimating γ from the relation between current wages and current hours would lead one to underestimate its true value because, over the first half of the life-cycle, hours are relatively flat while wages grow steeply. Instead, γ should be gauged from the joint dynamics of hours and the return to work, whose path is much flatter because the profiles of hourly wages and human capital tend to offset each other.

Turning to the Frisch elasticity, it is clear that the larger is the gap between the return to work and the current wage, i.e., the human capital term, the smaller the Frisch elasticity will be, since the optimal decision on how much to work at time t depends both on current and dynamic returns. Increasing the current return w_{it} may not have a large impact on ℓ_{it} if the dynamic returns are large. In line with this logic, Imai and Keane estimate a Frisch elasticity that increases with age from 0.4 to 2.0.

Another interesting implication of the human capital model is that it can alter the ranking of labor supply elasticities. Recall that in Section 12.4 we showed that in the benchmark dynamic model, Frisch > Hicksian > Marshallian. This ordering is no longer necessarily true here. To understand this point, compare a transitory tax increase that only affects today's after-tax earnings and a permanent tax increase that affects all future disposable earnings.

Obviously, the income effect which pushes in the direction of working longer hours is more powerful for a permanent tax hike. However, the substitution effect will also be stronger. The reason is that a permanent tax increase reduces the total return to work —both the current wage and the human capital term. Under some parameterizations, the substitution effect can dominate the income effect. Thus, although human capital dampens labor supply responses to transitory tax changes (and this effect is especially strong for young workers), it magnifies the impact of permanent changes.

12.8 Labor supply along the intensive and extensive margins

In Section 12.2, we highlighted that variation in aggregate hours of work are accounted for by variation along both the extensive (number of people employed) and intensive (hours per person employed) margins. The models presented in the previous sections of this chapter have the feature that all of the variation in aggregate hours occurs along the intensive margin, with no consideration of the extensive margin. A large body of work has shown that explicitly modeling both margins of adjustment is important for understanding labor supply responses. In this section we study models that incorporate intensive and extensive margins and illustrate some of their implications. One of the key messages is that labor supply elasticities estimated using individual panel data may not be informative about aggregate labor supply elasticities, i.e., how aggregate hours of work respond to changes in the average real wage.

12.8.1 A static model with indivisible labor

We begin by studying a static model that forces all adjustment to be along the extensive margin and contrasting its properties with those of a model in which all adjustment takes place at the intensive margin. The intensive margin economy will essentially be a general equilibrium version of the benchmark static model introduced in Section 12.3. In particular, we assume a unit mass of identical individuals, each with preferences of the form:

$$U(c,\ell) = u(c) - \psi v(\ell)$$

with u and v taking on the same functional forms as in Section 12.2:

$$u(c) = \frac{1}{1-\sigma}c^{1-\sigma}, \quad v(\ell) = \frac{1}{1+1/\gamma}\ell^{1+1/\gamma}.$$

There is a constant returns to scale aggregate production function that uses only labor:

$$Y = AL,$$

where L is total labor input. Individuals in this economy are allowed to choose any value for their hours, but for reasons that will become clear we will normalize the total discretionary time of each individual to equal one, so that ℓ must lie between 0 and 1.

The extensive margin only economy is identical with one exception: instead of allowing each individual to freely choose their hours of work, we now limit individuals to two choices: either $\ell = 0$ or $\ell = \bar{\ell}$, where $\bar{\ell}$ is a positive constant lying strictly between 0 and 1. This restriction can be interpreted as capturing the fact that many jobs come with a standard workweek, so that the key choice that an individual makes is whether to work at the standard workweek or to not work. Because labor can only be supplied in the fixed quantity $\bar{\ell}$, this model is referred to as the indivisible labor model; the economy in which individuals can freely choose ℓ will be referred to as the divisible labor model.

As a first step, we examine what a social planner would do in each of these economies. In the divisible labor model, it is easy to show that the Social Planner will give all individuals the same allocation of consumption and hours. These values are the solution to:

$$\max_{c,\ell} \; U(c,\ell)$$

subject to

$$c = A\ell,\; c \geq 0, \text{ and } 0 \leq \ell \leq 1.$$

In the indivisible labor economy, the social planner's problem is slightly more complicated. By assumption, the social planner can only give all individuals the same allocation of consumption and hours if either all individuals work $\bar{\ell}$ or all individuals do not work at all. But in general, the social planner may want to have only a fraction of all individuals work. Denote this fraction by e. Conditional on having a fraction e of individuals employed, the social planner has to decide how to allocate the resulting output of $Ae\bar{\ell}$ among all of the individuals in the economy. Because preferences are separable and the function u is strictly concave, it is optimal for the social planner to spread consumption equally across all individuals, independently of whether they are chosen to work. An alternative to having the social planner treat identical individuals asymmetrically is for the social planner to offer all individuals the same random allocation, namely, that they work with probability e and receive the same consumption c independently of the realization of the randomness. Formulated this way, the social planner in the indivisible labor economy will solve:

$$\max_{e,c} \; eU(c,\bar{\ell}) + (1-e)U(c,0)$$

subject to

$$c = A\bar{\ell}e,\; 0 \leq e \leq 1, \text{ and } c \geq 0.$$

Note that we are implicitly appealing to the law of large numbers; given that there is a unit mass of individuals, if each has a probability e of working then we know that a fraction e of workers will be selected to work.

Next we examine the objective function of the social planner in the indivisible labor economy in more detail. Recalling that U is assumed to be separable we have:

$$\begin{aligned} eU(c,\bar{\ell}) + (1-e)U(c,0) &= u(c) - ev(\bar{\ell}) - (1-e)v(0) \\ &= u(c) - e[v(\bar{\ell}) - v(0)] - v(0). \end{aligned}$$

The term $v(0)$ is just a constant, and adding a constant to an objective function does not affect the optimal choices, so that the objective function is equivalent to:

$$u(c) - \tilde{\psi}e$$

where $\tilde{\psi} = v(\bar{\ell}) - v(0)$. Compare this to the objective function of the social planner in the divisible labor economy, which is:

$$u(c) - \frac{\psi}{1 + 1/\gamma} \ell^{1+1/\gamma}.$$

Recalling our discussion of the Frisch elasticity in the simple two period example earlier in this chapter, we see that whereas the divisible labor economy has an aggregate Frisch elasticity equal to γ, the indivisible labor economy has an infinite aggregate Frisch elasticity. The indivisible labor economy highlights the need to distinguish between individual and aggregate elasticities. In the divisible labor economy, all individuals work the same amount and movements in aggregate labor supply are exactly mirrored in movements in individual labor supply. But in the indivisible labor economy, a given individual either works or does not work, so individual labor supply does not vary smoothly. Nonetheless, aggregate labor supply does vary smoothly with changes in the fraction of individuals that work.

In summary, this analysis of indivisible labor offers two key takeaways. First, it creates a disconnect between the properties of aggregate labor supply and individual labor supply. Second, it explains why aggregate labor supply might have a very large Frisch elasticity independently of the Frisch elasticity that each individual has.

12.8.2 Decentralizing the social planner's allocations

Our discussion so far has focused on the allocations that would be chosen by a social planner. In the divisible labor economy it is straightforward to show that the social planner allocation would be implemented if we studied a competitive equilibrium. In the indivisible labor economy we argued that a social planner would use randomization. It is not immediately obvious how to implement such allocations as competitive equilibria. Following Hansen (1985) and Rogerson (1988), we briefly describe how this can be done if we allow individuals to sell employment probabilities. If w is the wage per unit of work, then offering to work $\bar{\ell}$ hours with probability e will result in labor earnings of $e\bar{\ell}w$. The firm demands a certain number of total work hours, which must be equal to the total supply of labor in equilibrium. But appealing to the law of large numbers, this is not an issue since aggregate labor supply will be deterministic even if each individual offers a random amount of hours.

While the above discussion shows how one can formally define a market structure such that the social planner's allocation is implemented as a competitive equilibrium, one might argue that this market structure with employment lotteries does not seem to be a good description of the market structure we observe in real-world economies. Ljungqvist and Sargent (2006) have shown that the social planner's allocation can be implemented as a competitive equilibrium with a more compelling market structure. Their argument requires explicit consideration of an economy with many time periods. To make the argument precise we will focus on an economy in continuous time that runs from time 0 to time T. At each instant of time this economy resembles the static indivisible labor economy that we described above, and we assume that all parameters are constant over time. For simplicity, we assume that individuals evaluate lifetime utility as the integral of utility over time without any discounting.

If the social planner's optimal allocation in the static economy was for everyone to work with probability e^* and receive consumption c^*, then it follows that an optimal allocation in the dynamic economy will be to choose these values at each point in time. The key insight of Ljungqvist and Sargent is that from a lifetime perspective, choosing to work with probability e at each instant of time is equivalent to deterministically choosing to work a fraction e of time. If an individual chooses to deterministically work in a fraction e of the time periods, there will be some instants in which they have labor income equal to $w\bar{\ell}$ and other instants in which they have labor income equal to zero. But if they have access to markets that allow for borrowing and saving, and the interest rate is equal to zero, they can achieve a smooth profile for consumption equal to $e\bar{\ell}w$ in each period. Because we assumed that individuals do not discount future utility, it turns out that the competitive equilibrium interest rate will be equal to zero. Additionally, if we normalize the price of consumption to equal unity at each instant, the equilibrium wage rate will be constant and equal to A. In the competitive equilibrium each individual solves the following problem (a continuous-time formulation is adopted so that we can avoid the integer constraint):

$$\max \int_0^T U(c(t), e(t)\bar{\ell})dt$$

subject to

$$\int_0^T c(t)dt = \int_0^T e(t)\bar{\ell}Adt,\ e(t) \in \{0,1\},\text{ and } c(t) \geq 0.$$

The solution to this problem is such that the individual wants to choose the $e(t)$ such that $\int_0^T e(t)dt = e^*$ and $c(t) = c^*$ but is indifferent about at which instant they work. That is, they care about how much they work over their lifetime but are indifferent about the timing of that work. Because this is true for all individuals it is possible to have a fraction e^* work in each period, so that the social planner's allocation can be implemented in competitive equilibrium without having any markets for random supply of labor.

12.8.3 Allowing for heterogeneity

Our previous analysis argued that an indivisible labor economy with identical individuals gives rise to an aggregate Frisch elasticity that is infinite if the aggregate employment rate is interior.[16] In this subsection we argue that this result hinges on the assumption of identical individuals. This is intuitive. In an economy in which everyone is identical, their reservation wage, i.e., the wage at which they are indifferent between working and not working, is the same for all individuals. In such a situation a small change in the wage around the reservation wage can move everyone from wanting to work to not wanting to work and vice versa. In contrast, if there is heterogeneity in reservation wages then a small change in the wage rate will only change the decisions of a small group of people. We now develop this idea more formally in one specific context.

[16] If the social planner chooses $e^* = 1$ then by continuity a small change in the wage at any instant holding all other wages constant will not result in any change in e^*, implying an aggregate Frisch elasticity equal to zero.

For simplicity we return to a static economy. We assume that individuals are identical ex ante, but are subject to idiosyncratic shocks to their disutility of working parameter ψ. In particular, each individual will receive an iid draw from a distribution with cdf $F(\psi)$ and density $f(\psi)$. The aggregate production function is the same as before.

We again focus on the social planner's problem for this economy. Because all individuals are the same ex ante, we assume that the social planner maximizes the equal weighted integral of expected utility across individuals. The social planner will choose individual allocations contingent on the realizations of disutility of working, and can be written as two functions, $e(\psi)$ and $c(\psi)$, which represent the work probability and consumption allocation respectively for an individual with realization ψ. With separable utility, the social planner will continue to allocate the same consumption to all individuals independently of their disutility of working. For the employment decision, the most efficient way to have a fraction e of all individuals work is to choose the fraction e of individuals with the lowest draw for the disutility of work and to have them work with probability one, while the other individuals work with probability zero. This amounts to choosing a reservation disutility level ψ^* with the property that $F(\psi^*) = e$. Every choice of e is associated with a unique choice of ψ^*, which we will write as $\psi^*(e)$. It then follows that the total disutility associated with an employment rate of e, which we denote by $\tilde{v}(e)$ is given by:

$$\tilde{v}(e) = \int_0^{\psi^*(e)} \psi f(\psi) d\psi$$

and the social planner's problem can be written as:

$$\max \; u(c) - \tilde{v}(e)$$

subject to

$$c = Ae\bar{\ell}, \; 0 \leq e \leq 1, \text{ and } c \geq 0.$$

The key point is that the function $\tilde{v}$ is no longer linear; as e increases the social planner will be choosing workers with progressively higher values of ψ to work. The aggregate Frisch elasticity will now depend on the local properties of the $\tilde{v}$ function, which will in turn depend on the local properties of the density function f. This clearly illustrates that the aggregate Frisch elasticity depends critically on the distribution of heterogeneity in the population.

The previous discussion considered a static model, but one can easily generalize it to a dynamic economy by assuming that individuals receive iid draws of ψ at each instant from the distribution $F(\psi)$. As in the previous subsection, the social planner's solution can be implemented in a competitive equilibrium without any insurance markets if we allow complete markets for borrowing and lending. Our discussion has focused on heterogeneity in preferences, but similar effects are obtained if one considers heterogeneity in productivity.

A key question is to assess the implications of indivisible labor when allowing for empirically relevant sources of heterogeneity. In particular, what does such a model imply for the values of individual and aggregate labor supply elasticities. One of the first quantitative analyses of this question was undertaken by Chang and Kim (2006).[17] They consider an aggregate model populated by infinitely-lived two-member households in which labor supply

[17] Additional issues were explored in Chang and Kim (2007) and An, Chang, and Kim (2009).

is indivisible, individuals are subject to idiosyncratic productivity shocks and markets for credit and insurance are incomplete. Consistent with the unitary household model that we described in Section 12.3.2, households consist of a male and a female, and the household has preferences given by:

$$\sum_{t=0}^{\infty} \beta^t \left[2 \log \left(\frac{c_t}{2} \right) - \psi_m \frac{\ell_{mt}^{1+1/\gamma}}{1+1/\gamma} - \psi_f \frac{\ell_{ft}^{1+1/\gamma}}{1+1/\gamma} \right],$$

where c_t is household consumption, and ℓ_{mt} and ℓ_{ft} are hours worked by the male and female household member. Each individual can only supply 0 or $\bar{\ell}$ units of labor in any period. Individual productivity, denoted by z_t, is stochastic and follows the stochastic process:

$$\log z_{jt+1} = \rho_j \log z_{jt} + \varepsilon_{jt+1},\ j = m, f. \tag{12.42}$$

The process is gender specific but is the same for all individuals of a given gender. The realizations of the ε_{jt} are iid across individuals and over time. A worker with productivity z_t has labor earnings $w_t z_t \bar{\ell}$ if working, where w_t is the wage per efficiency unit of labor in period t.

The production side of the economy is the same as in the standard growth model. There is a Cobb Douglas aggregate production function that uses capital and efficiency units of labor, output can be used as either investment or consumption, and capital depreciates at a constant rate δ.

Chang and Kim study a competitive equilibrium assuming the following market structure. Each period there are markets for capital and labor services as well as output, but the market for labor services does not allow workers to sell employment probabilities. Following the work of Aiyagari (1994) and Huggett (1993) discussed in Section 11.4, markets for credit and insurance are incomplete. Specifically, there are no markets for insurance against idiosyncratic productivity shocks, and households are able to save and borrow in a credit market, subject to an exogenous borrowing limit.

Chang and Kim calibrate this model assuming that a period is equal to one quarter and show it matches the cross-sectional heterogeneity in earnings and wealth found in the data reasonably well, with the exception that it is not able to capture the extreme right tail of the wealth distribution. The model implies that individuals move between employment and nonemployment as their productivity and asset holdings evolve. Chang and Kim did not assess the extent to which the transitions in the model match those found in the data, but subsequent related work has shown this to be the case.[18]

They proceed to study the properties of individual and aggregate labor supply in their calibrated model. First, they simulate histories lasting 120 quarters for a sample of households in the steady state, aggregate the observations to produce a panel data set at annual frequency and then run the following panel regression using individuals with positive hours in each year:

$$\log \ell_{it} = \gamma(\log w_{it} - \log c_{it}) + \varepsilon_{it}. \tag{12.43}$$

[18] Bils, Chang, and Kim (2012) examine this in a slightly more general model. Krusell, Mukoyama, Rogerson, and Şahin (2017) further extend this analysis by including search frictions and considering movements between employment, unemployment and out of the labor force, though they consider single individual households.

Recall that in Section 12.3 we noted that this equation can be derived in the context of a two period setting if the function u is given by log. Their estimation exercise yields estimates for γ of 0.41 and 0.78 for males and females respectively. The key finding is that a standard labor supply regression using individual data generated by their model yields a relatively small estimate of the Frisch elasticity for men, and a larger and moderate estimate for women.

Their second exercise follows in the spirit of the literature on real business cycles. In particular, they now assume an aggregate technology shock that follows an AR(1) process and simulate the economy to produce aggregate time series data for hours, consumption and wages, and then run the same regression as above but now using aggregate time series data. This produces an estimate for γ of 1.08. Notably, this estimate is higher than both of the estimates obtained when using individual data.

Lastly, they consider a representative household model with preferences of the form:

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

where ℓ_t is now allowed to take on any value in the interval $[0,1]$. Assuming the same process for aggregate technology shocks, they solve for the value of $\tilde{\gamma}$ that generates fluctuations in aggregate hours that are the same as in the heterogeneous agent economy with indivisible labor, and find a value for $\tilde{\gamma}$ of approximately 2. This value is roughly five times as large as the Frisch elasticity estimated from individual data on males.

Their analysis has two key messages. First, allowing for empirically reasonable individual heterogeneity and risk-sharing in a model with indivisible labor dramatically lowers the implied aggregate Frisch elasticity, reducing it from infinity to around 2. But second, the model still implies a large disconnect between the labor supply elasticity obtained using standard methods on microdata for continuously employed individuals and the aggregate elasticity.

12.8.4 Models with intensive and extensive adjustment

So far in this section we have focused on models in which all adjustment in aggregate hours occurs either along the intensive margin or along the extensive margin. As emphasized in the first section of this chapter, data shows that there is important adjustment along both margins. In this subsection we develop some models that feature adjustment along both margins. The key feature that we use to generate this outcome is to assume a convex mapping from the hours supplied by an individual to the efficiency units of labor associated with these hours. There are two specifications of this form that are commonly found in the literature. One is to assume that there is a fixed time cost associated with working:

$$\ell^e = \max(0, \ell - \ell_f), \tag{12.44}$$

where ℓ is the time devoted to work, ℓ^e denotes the efficiency units of labor and $\ell_f > 0$ is the fixed time cost associated with working. The key distinction is that ℓ is what enters the utility function of the individual while ℓ^e is what enters the production function. The fixed cost ℓ_f can represent some combination of the time associated with getting to and from work and the time required to get set up once at work. The presence of fixed costs makes it suboptimal for workers to choose to work low numbers of hours and creates an incentive to concentrate hours in order to minimize the total fixed costs.

The second specification assumes a smooth convex mapping from hours of work to efficiency units of labor supplied:

$$\ell^e = \ell^\theta,$$

where $\theta > 1$ ensures that this relationship is convex, and $\theta = 1$ represents the standard textbook case in which time devoted to work is the same as the efficiency units of work. This specification is consistent with a wage penalty for part-time work and a wage bonus for overtime work. Similar to the fixed-cost specification, when $\theta > 1$ there is an incentive for an individual to concentrate their hours of work, since for example, an individual who works 25 hours in each of two periods will earn higher income if they instead chose to work 50 hours in one of the periods and 0 hours in the other period.

Because both of these specifications incentivize individuals to concentrate hours of work and avoid periods in which they work relatively few hours, they both have a force that captures the essence of the indivisible labor assumption. In fact, if we adopt either of these specifications in an economy with identical workers, then for sufficiently large values of ℓ_f or θ the solution to the social planner's problem will look like the outcome in an indivisible labor economy, with one group of individuals working $\bar{\ell}$ units and the remaining individuals all working 0 hours. But importantly, the value of $\bar{\ell}$ will be endogenously determined, depending upon the primitives of the model, and in particular the value of ℓ_f or θ.

When individuals are heterogeneous, these specifications lead to labor allocations in which variation in aggregate hours reflects changes both in the fraction of individuals employed and hours per employed person. Rogerson and Wallenius (2009) illustrated this in the context of a life cycle model to examine the effects of tax and transfer programs. They used the fixed time cost of work specification and introduced heterogeneity by assuming that productivity varied in a deterministic manner over the life cycle. In particular, assume that time is continuous, an individual lives from $t = 0$ to $t = T$, and that productivity varies deterministically over the life cycle according to the function $z(t)$. The individual maximizes lifetime utility, which is given by (again, a continuous-time formulation is employed to avoid the integer constraint):

$$\int_0^T [u(c(t)) - v(\ell(t))]dt,$$

where for simplicity we assume no discounting. As above, the individual faces the mapping from hours of work $\ell(t)$ to efficiency units $\ell^e(t)$ as specified in equation (12.44). The wage per one unit of $z(t)\ell^e(t)$ is constant over time and equal to w, and the individual is allowed to borrow and save at the interest rate of 0 subject to the constraint that they do not die with any debt. That is, they face a lifetime budget constraint given by:

$$\int_0^T c(t)dt = w\int_0^T z(t)\max(0, \ell(t) - \ell_f)dt.$$

To make things concrete, assume that $z(t)$ is continuous, has a single peak, and that $z(0) = z(T) = 0$. The optimal profile for life cycle hours in this model has the following form. There is a threshold value z^* for individual productivity such that when productivity is below z^* the individual chooses $\ell = 0$. When productivity is greater than z the individual chooses to work positive hours, and hours are an increasing function of productivity. Thus, at the individual level, lifetime labor supply features adjustment along both the intensive and extensive margins.

Rogerson and Wallenius use this framework to quantitatively assess its implications for some labor supply elasticities. For their quantitative work they adopt $u(c) = \log c$ and $v(\ell) = \psi\ell^{1+1/\gamma}/(1+1/\gamma)$, and assume that life-cycle productivity $z(t)$ is piecewise linear. They consider values of γ ranging from .10 to 2.00 and in each case calibrate the model to match some key properties of life cycle labor supply.

Two key results emerge. First, they assess the response of aggregate hours to a permanent tax and transfer policy that includes a balanced budget constraint and find that the response is roughly independent of the value of γ. Specifically, increasing the tax rate from 30% to 50% decreases aggregate hours by roughly 20% for all values of γ in the range of 0.10 to 2.00. Recall from our previous discussion that this policy exercise reflects the value of Hicksian elasticity of labor supply.

Second, although γ has virtually no effect on the change in aggregate hours, it does have a quantitatively important effect on the relative importance of adjustment along the intensive and extensive margins. For example, when $\gamma = 2.00$, the intensive margin accounts for over 60% of the total decrease in hours, while when $\gamma = .10$ this value is less than 5%. If a researcher used the benchmark model from Section 12.2 to interpret steady state differences in aggregate hours worked across two Rogerson-Wallenius economies with the same low value of γ but differing scales of their tax and transfer system, they would infer a value of γ that is more than an order of magnitude larger than the true underlying value of γ. This happens because the change in aggregate hours includes responses on both the extensive and intensive margins and the implied value of γ must proxy for adjustment along both margins. The greater is the adjustment along the extensive margin, the greater the disconnect between the true value of γ and the value of γ that the intensive margin only model will infer from the aggregate data. A key message from the analysis of Chang and Kim (2006) continues to hold in this setting that features adjustment along both intensive and extensive margins: there is a large disconnect between the parameters that characterize aggregate and individual labor supply. In particular, labor supply elasticities estimated on micro panel data using workers with positive hours are not particularly informative for predicting the aggregate effects of permanent changes in taxes. Moreover, the aggregate elasticity is large.

Because the Rogerson and Wallenius (2009) model is somewhat stylized, it is of interest to consider robustness to allowing for richer and more realistic empirical specifications. Erosa, Fuster, and Kambourov (2016) go quite far in assessing this. Specifically, they extend the Rogerson-Wallenius model along many dimensions in order to match a wide variety of features of wages and hours worked for males between the ages of 25 and 61. Their analysis allows for multiple sources of heterogeneity (both idiosyncratic shocks as in Chang and Kim (2006) and permanent productivity differences), multiple nonconvexities (fixed utility costs of working in addition to nonconvex earnings), time aggregation, and measurement error in wages. While these features do matter for the empirical properties of the model and its ability to replicate the salient features of the data, the conclusions are broadly similar. They find that the aggregate labor supply elasticity to a temporary unanticipated wage change is 1.75.