1
The Labor Wedge

Throughout this book, I study the interaction of optimizing households and firms in a closed economy. I begin in this chapter by developing a competitive, representative-agent version of the model. The chapter has two objectives. First, I introduce much of the notation that I rely on throughout the book. Because of this, I include details in this chapter that are not really necessary for the second, more substantive objective: I use the model to measure and analyze the behavior of the *labor wedge*, the wedge between the marginal rate of substitution of consumption for leisure and the marginal product of labor. I confirm the well-known result that the labor wedge tends to rise during recessions, so the economy behaves as if there is a countercyclical tax on labor. The remainder of the book explores whether extending the model to incorporate labor market search frictions can explain the behavior of the labor wedge.

I start the chapter by laying out the essential features of the model: optimizing households, optimizing firms, a government that sets taxes and spending, and equilibrium conditions that link the various agents. In section 1.2, I use pieces of the model to derive a static equation that relates hours worked, the consumption–output ratio, and the labor wedge. Section 1.3 discusses how I measure the first two concepts and uses these measures to calculate the implied behavior of the labor wedge in the United States. I establish the main substantive result: that the labor wedge rose strongly during every recession since 1970. I show the robustness of my results to alternative specifications of preferences in section 1.4 and discuss the possibility that the results are driven by preference shocks in section 1.5. I finish the chapter with a brief discussion in section 1.6 on the empirical relationship between the fluctuations in hours, which I analyze here, and fluctuations in employment and unemployment, which are the main topic of subsequent chapters.

1.1 A Representative-Agent Model

I denote time by $t = 0, 1, 2, \ldots$ and the state of the economy at time t by s_t. Let $s^t = \{s_0, s_1, \ldots, s_t\}$ denote the history of the economy and $\Pi(s^t)$

denote the time-0 belief about the probability of observing an arbitrary history s^t through time t. Exogenous variables like aggregate productivity, government spending, and distortionary tax rates may depend on the history s^t. At date 0, there is an initial capital stock $k_0 \equiv k(s^0)$ and an initial stock of government debt $b_0 \equiv b(s^0)$. The capital stock is owned by firms, while households hold the debt and own the firms.

Households

A representative household is infinitely lived and has preferences over history-s^t consumption $c(s^t)$ and history-s^t hours of work $h(s^t)$. To start, I assume that preferences are ordered by the utility function

$$\sum_{t=0}^{\infty} \sum_{s^t} \beta^t \Pi(s^t) \left(\log c(s^t) - \frac{\gamma \varepsilon}{1+\varepsilon} h(s^t)^{(1+\varepsilon)/\varepsilon} \right), \tag{1.1}$$

where $\beta \in (0,1)$ is the discount factor, $\gamma > 0$ measures the disutility of working, and, as I show below, $\varepsilon > 0$ is the Frisch (constant marginal utility of wealth) elasticity of labor supply.

This formulation implies that preferences are additively separable over time and across states of the world. It also implies that preferences are consistent with balanced growth—doubling a household's initial assets and its income in every state of the world doubles its consumption but does not affect its labor supply. This is consistent with the absence of a secular trend in hours worked per household, at least in the United States (Aguiar and Hurst 2007; Ramey and Francis 2009). I maintain both of these assumptions throughout this book. The formulation also imposes that the marginal utility of consumption is independent of the worker's leisure. This restriction is more questionable and so I relax it in section 1.4 below.

The household chooses a sequence for consumption and hours of work to maximize utility subject to a single lifetime budget constraint,

$$a_0 = \sum_{t=0}^{\infty} \sum_{s^t} q_0(s^t)(c(s^t) - (1 - \tau(s^t))w(s^t)h(s^t) - T(s^t)). \tag{1.2}$$

The household has initial assets $a_0 = a(s^0)$. In addition, $\tau(s^t)$ is the labor income tax rate, $w(s^t)$ is the hourly wage rate, and $T(s^t)$ is a lump-sum transfer in history s^t, all denominated in contemporaneous units of consumption.[1] Thus $c - (1-\tau)wh - T$ represents consumption in excess of after-tax labor income and transfers, which is discounted back to time 0

[1] One can easily extend the model to include a consumption tax. Then $\tau(s^t)$ measures the total tax wedge: the cost to an employer of providing its worker with one unit of the consumption good.

according to the intertemporal price $q_0(s^t)$. That is, $q_0(s^t)$ represents the cost in history s^0 of purchasing one unit of consumption in history s^t, denominated in units of history-s^0 consumption. Put differently, $q_0(s^t)$ is the history-s^0 price of an Arrow–Debreu security that pays one unit of consumption in history s^t and nothing otherwise. Equation (1.2) states that the household's net purchase of Arrow–Debreu securities in history s^0 must be equal to its initial assets a_0.

It will be useful to define the assets of the household, following history s^t, as

$$a(s^t) = \sum_{t'=t}^{\infty} \sum_{s^{t'}|s^t} q_t(s^{t'})(c(s^{t'}) - (1 - \tau(s^{t'}))w(s^{t'})h(s^{t'}) - T(s^{t'})),$$

where the notation $s^{t'}|s^t$ indicates that the summation is taken over histories $s^{t'}$ that are continuation histories of s^t, i.e., $s^{t'} \equiv \{s^t, s_{t+1}, s_{t+2}, \dots, s_{t'}\}$ for some states $\{s_{t+1}, s_{t+2}, \dots, s_{t'}\}$. Then $q_t(s^{t'})$ is the price of a unit of consumption in history $s^{t'} = \{s^t, s_{t+1}, s_{t+2}, \dots, s_{t'}\}$ paid in units of history-s^t consumption. The absence of arbitrage opportunities requires that $q_0(s^t)q_t(s^{t+1}) = q_0(s^{t+1})$ for all s^t and for all $s^{t+1} \equiv \{s^t, s_{t+1}\}$. Equivalently, the lifetime budget constraint implies a sequence of intertemporal budget constraints,

$$a(s^t) + (1 - \tau(s^t))w(s^t)h(s^t) + T(s^t) = c(s^t) + \sum_{s^{t+1}|s^t} q_t(s^{t+1})a(s^{t+1}), \tag{1.3}$$

so assets plus labor income plus transfers in history s^t is equal to consumption plus purchases of assets in continuation histories s^{t+1}.

Firms

The representative firm owns the capital stock $k_0 = k(s^0)$ and has access to a Cobb–Douglas production function, producing gross output $z(s^t)k(s^t)^\alpha h^d(s^t)^{1-\alpha}$ in history s^t, where $z(s^t)$ is history-contingent total factor productivity,[2] $k(s^t)$ is its capital stock, $h^d(s^t)$ is the labor it demands, and $\alpha \in [0, 1)$ is the capital share of income. A fraction δ of the capital depreciates in production each period, while at the end of period t, the firm purchases any capital that it plans to employ in period $t + 1$. That is, history-$s^{t+1} \equiv \{s^t, s_{t+1}\}$ capital $k(s^{t+1})$ is purchased in history s^t and so must be measurable with respect to s^t. The present value

[2] Although I do not place explicit restrictions on the productivity process, I do assume that a worker's expected utility is finite so her optimization problem is well-behaved. This is ensured if productivity is bounded but is true under substantially weaker conditions, if productivity does not grow too fast.

of the firm's profits is then given by

$$J(s^0, k_0) = \sum_{t=0}^{\infty} \sum_{s^t} q_0(s^t)(z(s^t)k(s^t)^{\alpha} h^d(s^t)^{1-\alpha} + (1-\delta)k(s^t) - k(s^{t+1}) - w(s^t)h^d(s^t)). \quad (1.4)$$

Note that this expression presumes that the firm does not pay any taxes. I do this for notational simplicity alone. In particular, any payroll taxes are rolled into the labor income tax rate τ. The firm chooses the sequences $h^d(s^t)$ and $k(s^{t+1})$ to maximize J.

I can also write the value of the firm's profits from history s^t on as

$$J(s^t, k(s^t)) = \sum_{t'=t}^{\infty} \sum_{s^{t'}|s^t} q_t(s^{t'})(z(s^{t'})k(s^{t'})^{\alpha} h^d(s^{t'})^{1-\alpha} + (1-\delta)k(s^{t'}) - k(s^{t'+1}) - w(s^{t'})h^d(s^{t'})).$$

This implies the recursive equation

$$J(s^t, k(s^t)) = z(s^t)k(s^t)^{\alpha} h^d(s^t)^{1-\alpha} + (1-\delta)k(s^t) - k(s^{t+1}) - w(s^t)h^d(s^t) + \sum_{s^{t+1}|s^t} q_t(s^{t+1})J(s^{t+1}, k(s^{t+1})). \quad (1.5)$$

The value of a firm that starts history s^t with capital $k(s^t)$ comes from current production $z(s^t)k(s^t)^{\alpha}h^d(s^t)^{1-\alpha}$ minus the cost of investment $k(s^{t+1}) - (1-\delta)k(s^t)$ minus labor costs $w(s^t)h^d(s^t)$ plus the value of starting the following period in history $s^{t+1} \equiv \{s^t, s_{t+1}\}$ with $k(\{s^{t+1}\})$ units of capital.

Government

A government sets the path of taxes, transfers, and government debt to fund some spending $g(s^t)$. I assume government spending is wasteful or at least is separable from consumption and leisure in preferences. The government faces a budget constraint in any history s^t,

$$b(s^t) = \sum_{t'=t}^{\infty} \sum_{s^{t'}|s^t} q_t(s^{t'})(\tau(s^{t'})w(s^{t'})h(s^{t'}) - g(s^{t'}) - T(s^{t'})), \quad (1.6)$$

so debt $b(s^t)$ is equal to the present value of future tax receipts in excess of spending and lump-sum transfers. Again, this is equivalent to a sequence of budget constraints of the form

$$b(s^t) + g(s^t) + T(s^t) = \tau(s^t)w(s^t)h(s^t) + \sum_{s^{t+1}|s^t} q_t(s^{t+1})b(s^{t+1}), \quad (1.7)$$

so initial debt plus current spending and transfers is equal to current tax revenue plus new debt issues.

Market Clearing

There are three markets in this economy: the labor market, the capital market, and the goods market. All of them must clear in equilibrium. Labor market clearing dictates that labor supply equals labor demand in all histories, $h(s^t) = h^d(s^t)$. Capital market clearing dictates that household assets are equal to firms' valuation plus government debt, $a(s^t) = J(s^t, k(s^t)) + b(s^t)$. Goods market clearing dictates that output plus undepreciated capital is equal to consumption plus government spending plus next period's capital stock:

$$z(s^t)k(s^t)^\alpha h^d(s^t)^{1-\alpha} + (1-\delta)k(s^t) = c(s^t) + g(s^t) + k(s^{t+1}).$$

One can confirm that goods market clearing is implied by the household budget constraint (equation (1.3)), the firm's value function (equation (1.5)), the government budget constraint (equation (1.7)), and capital and labor market clearing. This is an application of Walras's law.

Equilibrium

Given arbitrary paths for government spending $g(s^t)$, taxes $\tau(s^t)$, and government debt $b(s^t)$, an equilibrium consists of paths for consumption $c(s^t)$, labor supply $h(s^t)$, labor demand $h^d(s^t)$, capital $k(s^t)$, assets $a(s^t)$, transfers $T(s^t)$, intertemporal prices $q_0(s^t)$, and the wage rate $w(s^t)$ such that:

- $\{c(s^t)\}$, $\{h(s^t)\}$, and $\{a(s^t)\}$ solve the household's utility-maximization problem, maximizing equation (1.1) subject to the budget constraint (1.2) given $\{q(s^t)\}$, $\{w(s^t)\}$, $\{\tau(s^t)\}$, and $\{T(s^t)\}$;
- $\{h^d(s^t)\}$ and $\{k(s^t)\}$ maximize firms' profits in (1.4) given $\{q_0(s^t)\}$ and $\{w(s^t)\}$;
- the government budget is balanced, so equation (1.6) holds; and
- the labor, capital, and goods markets clear.

1.2 Deriving the Labor Wedge

To see the implications of this model for the labor wedge, I focus on a subset of the equilibrium conditions. First, consider the household's choice of history-s^t consumption and labor supply. These must satisfy the first-order conditions

$$\beta^t \Pi(s^t) \frac{1}{c(s^t)} = \lambda q_0(s^t) \tag{1.8}$$

and

$$\beta^t \Pi(s^t)\gamma h(s^t)^{1/\varepsilon} = \lambda q_0(s^t)(1-\tau(s^t))w(s^t), \tag{1.9}$$

where λ is the Lagrange multiplier on the budget constraint: equation (1.2). Note from the second equation that a one percent increase in the after-tax wage $(1-\tau)w$ raises labor supply h by ε percent, holding fixed the Lagrange multiplier λ and the intertemporal price $q_0(s^t)$. Thus ε is the Frisch elasticity of labor supply, a key parameter in this chapter.

In any history with positive probability, $\Pi(s^t) > 0$, we can eliminate $\lambda q_0(s^t)/\beta^t \Pi(s^t)$ between these equations and solve for the wage:

$$w(s^t) = \frac{\gamma c(s^t) h(s^t)^{1/\varepsilon}}{1-\tau(s^t)}. \tag{1.10}$$

This states that the wage is equal to the tax-adjusted marginal rate of substitution (MRS) between consumption and leisure.

Next turn to the firm's choice of history-s^t labor demand. From equation (1.4), the necessary first-order condition is

$$w(s^t) = (1-\alpha)\frac{y(s^t)}{h^d(s^t)}, \tag{1.11}$$

where $y(s^t) = z(s^t)k(s^t)^{\alpha} h^d(s^t)^{1-\alpha}$ is the firm's gross output. Equation (1.11) states that the wage is equal to the marginal product of labor (MPL).

Eliminate the state-contingent wage between equations (1.10) and (1.11) and impose labor market clearing, $h^d(s^t) = h(s^t)$. Solving for $\tau(s^t)$ gives

$$\tau(s^t) = 1 - \frac{\gamma}{1-\alpha}\frac{c(s^t)}{y(s^t)} h(s^t)^{(1+\varepsilon)/\varepsilon}. \tag{1.12}$$

This static equation explains how the tax rate τ affects the consumption–output ratio c/y and hours worked h.

It is worth stressing that this relationship holds even though productivity, government spending, and distortionary taxes may be time varying or stochastic. Expectations of these changes are all captured by the current consumption–output ratio. For example, if productivity is currently below trend, the consumption–output ratio will be high and, to the extent that labor supply is elastic, labor supply will be low. An increase in government spending without a corresponding change in contemporaneous taxes will tend to reduce the consumption–output ratio and raise hours worked in an offsetting manner.

Prescott (2004) uses a version of equation (1.12) to examine the effect of tax variation over time and across countries on labor supply. More precisely, he uses a slightly different functional form for preferences, with period utility function $\log c + \gamma \log(100 - h)$, where 100 represents the available amount of time per week. He then calibrates γ to match the average number of hours worked across a broad set of countries, $\bar{h} \approx 20$. With this functional form, the Frisch elasticity of labor supply is $100/h - 1$, or about 4 on average. My choice of functional forms brings the issue of the elasticity of labor supply to the forefront of the discussion.

In addition, I focus on a different implication of this equation. Under the hypothesis that business cycle fluctuations are not primarily due to changes in taxes and transfers, I interpret cyclical variation in $\tau(s^t)$ as the labor wedge: the wedge between the MRS and the MPL. More precisely, I measure hours and the consumption–output ratio at quarterly frequencies. By making appropriate assumptions about the disutility of working γ, the capital share α, and the Frisch elasticity of labor supply ε, I back out the labor wedge from equation (1.12). This approach builds upon a substantial body of research, including Parkin (1988), Rotemberg and Woodford (1991, 1999), Hall (1997), Mulligan (2002), and Chari et al. (2007).

1.3 Measurement

To measure the labor wedge for the United States using equation (1.12), I need time series of the consumption–output ratio and hours worked, as well as values for the parameters ε, γ, and α. Nominal consumption and output data are available at quarterly frequencies from the National Income and Product Accounts (see table 1.1.5 therein). Output is gross domestic product, while consumption is personal consumption expenditures on nondurable goods and services.

I focus on the most comprehensive available series on hours: Prescott et al.'s (2008) measure of total hours worked relative to the noninstitutional population with ages between 16 and 64, which has been available quarterly since 1959.[3] This series is based primarily on data originally collected as part of the Current Population Survey (CPS), a monthly survey of households that is used to construct the unemployment rate. The total number of hours worked is equal to the product of the number of civilians at work and the average hours worked by a person at work,[4] plus the number of military personnel, who are assumed to work for

[3] The authors have recently extended their data set back to 1947. The results that I report here are, if anything, stronger in this longer sample.

[4] See www.bls.gov/cps/, series LNU02005053 and LNU02005054, for the monthly data since June 1976. Prescott et al. (2008) obtained earlier data from table A-24 of the Bureau of Labor Statistics publication *Employment and Earnings*.

forty hours per week. The population is equal to the civilian noninstitutional population with ages between 16 and 64 plus the number of military personnel.

I compare the results with those based on a measure of hours paid per adult in the civilian noninstitutional population from the Current Employment Statistics (CES), a monthly survey of business establishments.[5] The main drawback of this survey is that data on hours are unavailable for the government (and military) sector as well as for farm workers, proprietors, unpaid family workers, and supervisors. In addition, the CES measures hours paid rather than hours worked, and thus includes vacation time, sick days, and so on. This series has been available since 1964.

I also use an unpublished series for hours worked constructed by the Bureau of Labor Statistics (BLS) as part of the Major Sector Productivity and Costs program.[6] I again divide this series by the adult noninstitutional population. Although this measure relies primarily on data from the CES, it also uses data from the CPS to estimate the hours worked by workers who are not covered by the establishment survey. It also adjusts the CES data to convert hours paid into hours worked. In principle, the coverage of this series should be similar to that of Prescott et al. (2008).

Rather than take a stand on the Frisch labor supply elasticity ε, I consider a range of possible values and report four of them: $\varepsilon = 0.5, 1, 4,$ and ∞. The lowest value is toward the upper range of elasticities for prime-age mens' hours that many microeconomists consider plausible (see Blundell and MaCurdy 1999). The value $\varepsilon = 4$ is in line with the elasticities that macroeconomists frequently use in representative-agent business cycle and growth models. For each value of the elasticity, I set the ratio $\gamma/(1-\alpha)$ so that the average labor wedge, measured using equation (1.12), is 0.4 from 1959 to 2007, consistent with the tax wedge that Prescott (2004) reports. The results are similar if the average labor wedge is 0.3 or 0.5.

Figure 1.1 shows the implied behavior of the labor wedge using the CPS measure of hours. Two patterns stand out. First, there has been a trend decline in the labor wedge since around 1980. Arguably this reflects underlying movements in labor and consumption taxes. Second, the labor wedge is countercyclical. I indicate National Bureau of Economic Research (NBER) recession dates with gray bands. Regardless of

[5] See www.bls.gov/ces/, series CES0500000034.

[6] See www.bls.gov/lpc/, series PRS84006033, for the business sector. The series I use augments this with estimates of hours worked in the government sector. I am grateful to Simona Cociuba for providing me with this data and for clarifying the relationship between the different series for hours.

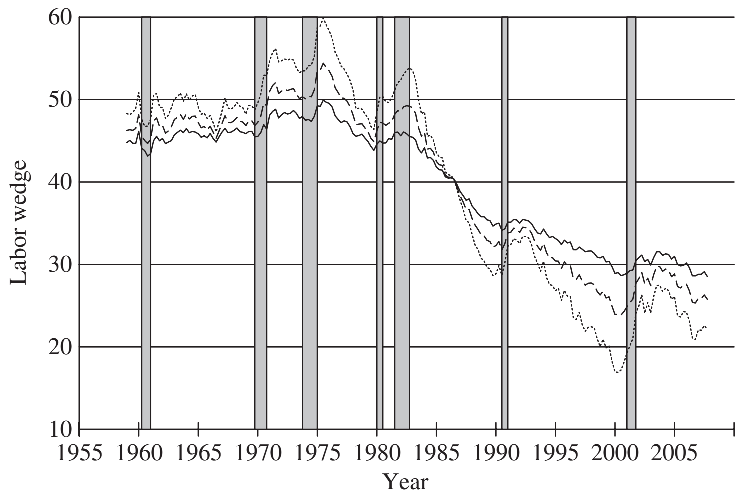

Figure 1.1. The U.S. labor wedge from equation (1.12) using the CPS measure of hours. The solid line shows $\varepsilon = 4$, the dashed line shows $\varepsilon = 1$, and the dotted line shows $\varepsilon = \frac{1}{2}$. In each case, I fix the remaining parameters to ensure that the average labor wedge is 0.40. The gray bands show NBER recession dates.

the elasticity of labor supply, the labor wedge rose during every recession except the first, with more pronounced fluctuations when labor supply is less elastic. This effect does not disappear even when labor supply is infinitely elastic.[7]

To emphasize this pattern, figure 1.2 shows the difference between $\log \tau$ and its trend, where I measure the trend using a Hodrick–Prescott (HP) filter with a standard smoothing parameter: 1,600 for quarterly data. It is easy to see a sharp increase in the labor wedge during every recession except the one in 1960. The magnitude of the implied cycles depends on the elasticity of labor supply. For example, with $\varepsilon = 1$, the period around the 1990 recession is associated with a ten percent increase in the labor wedge relative to trend, while with $\varepsilon = 4$, the increase was almost six percent.[8] Higher values of the labor supply elasticity only slightly dampen the implied fluctuations—even if the Frisch elasticity is infinite, the labor wedge rose by four percent during this time period. Conversely, smaller values of the Frisch elasticity amplify fluctuations in the labor wedge.

[7] Observe that the elasticity of labor supply enters equation (1.12) as $(1 + \varepsilon)/\varepsilon$. This means that an elasticity of 4 and an infinite elasticity have nearly the same effect on the labor wedge.

[8] With $\varepsilon = 1$, the labor wedge was five percent below trend in the third quarter of 1990 and rose to five percent above trend by the second quarter of 1992. With $\varepsilon = 4$, it rose from three percent below trend to three percent above trend during the same period.

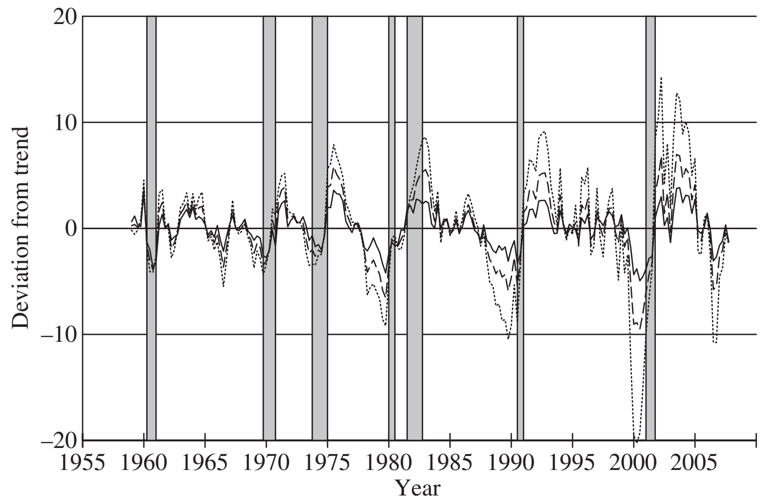

Figure 1.2. Deviation of the labor wedge from log trend, measured as an HP filter with parameter 1,600, using the CPS measure of hours. The solid line shows $\varepsilon = 4$, the dashed line shows $\varepsilon = 1$, and the dotted line shows $\varepsilon = \frac{1}{2}$. In each case, I fix the remaining parameters to ensure that the average labor wedge is 0.40. The gray bands show NBER recession dates.

Figure 1.3 displays the same findings slightly differently, depicting the annual growth rate of the labor wedge. Again, the labor wedge grew during every recession except the first one, in 1960. The magnitude of fluctuations in the growth of the labor wedge depends on the elasticity of labor supply.

Table 1.1 summarizes these results. The first row in the upper part of the table shows the standard deviation of the detrended consumption–output ratio, detrended hours, and the detrended labor wedge for four different labor supply elasticities. When the elasticity is small, the labor wedge is four times more volatile than hours and five times more volatile than the consumption–output ratio, while the relative volatilities of the labor wedge and hours are similar when the elasticity is large. The remaining entries show the contemporaneous correlation between the labor wedge (for different elasticities) and the consumption–output ratio and hours. The correlation with the consumption–output ratio disappears when the elasticity is high enough, but the labor wedge is strongly negatively correlated with hours, regardless of the elasticity of labor supply. The bottom part of the table shows the analogous results for the annual growth rate of the labor wedge, the consumption–output ratio, and hours. They are quantitatively very similar. It looks as if hours growth is negative when the labor income tax rate is rising, regardless of the elasticity of labor supply.

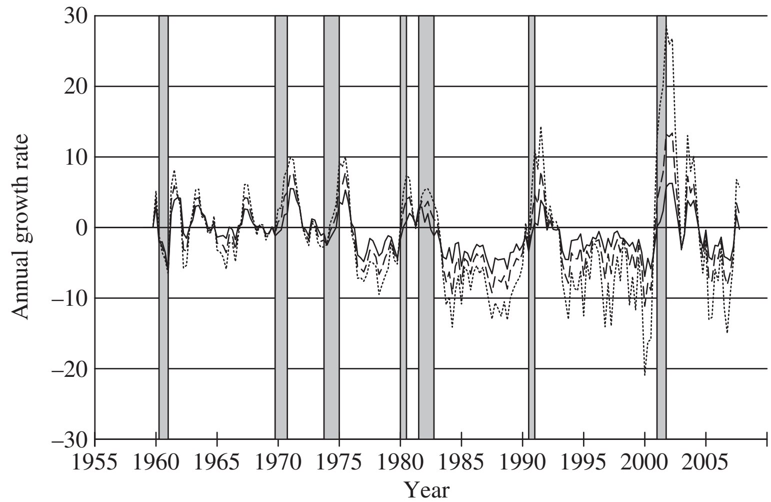

Figure 1.3. Annual growth rate of the labor wedge using the CPS measure of hours. The solid line shows $\varepsilon = 4$, the dashed line shows $\varepsilon = 1$, and the dotted line shows $\varepsilon = \frac{1}{2}$. In each case, I fix the remaining parameters to ensure that the average labor wedge is 0.40. The gray bands show NBER recession dates.

Table 1.1. Comovement of hours, the consumption–output ratio, and the labor wedge τ for four different values of the labor supply elasticity ε using the CPS measure of hours.

(a) All series detrended with HP filter with parameter 1,600

Standard deviations

c/y	h	$\varepsilon = 0.5$	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = \infty$
0.010	0.013	0.055	0.031	0.018	0.014

Correlation matrix

	$\varepsilon = 0.5$	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = \infty$
c/y	0.338	0.278	0.049	−0.131
h	−0.795	−0.835	−0.745	−0.628

(b) Annual growth rates

Standard deviations

c/y	h	$\varepsilon = 0.5$	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = \infty$
0.015	0.018	0.079	0.045	0.027	0.022

Correlation matrix

	$\varepsilon = 0.5$	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = \infty$
c/y	0.256	0.163	−0.088	−0.260
h	−0.803	−0.835	−0.733	−0.617

Table 1.2. Comovement of the labor wedge and hours with labor supply elasticity $\varepsilon = 1$. The first column uses the hour series from Prescott et al. (2008) from 1959 to 2007, constructed primarily from the CPS. The second uses hours data from the CES from 1964 to 2007. The third uses unpublished hours data constructed by the BLS to measure labor productivity from 1959 to the first quarter of 2006.

	CPS	CES	BLS
All series detrended with HP filter with parameter 1,600			
Standard deviation h	0.013	0.018	0.014
Standard deviation τ	0.031	0.049	0.035
Correlation (h, τ)	−0.835	−0.868	−0.883
Annual growth rates			
Standard deviation h	0.018	0.026	0.020
Standard deviation τ	0.045	0.069	0.051
Correlation (h, τ)	−0.835	−0.872	−0.879

There are several different ways to understand these results. Viewed through the lens of a model with a competitive labor market, fluctuations in the labor income tax rate drive fluctuations in hours. If in reality the tax rate is constant, the model underpredicts fluctuations in hours worked at business-cycle frequencies, given the observed time path of the consumption–output ratio. This is an old critique of competitive models of the labor market: such models can only explain part of the cyclical fluctuations in hours worked, particularly when labor is supplied relatively inelastically.

Table 1.2 shows that the main conclusions hold with the alternative measures of hours. I assume that the elasticity of labor supply is $\varepsilon = 1$ and examine how alternative measures of hours affect the behavior of the labor wedge. These other measures raise the volatility of the labor wedge, and by more than they raise the volatility of hours. In addition, the correlation between the two series is, if anything, increased. I find similar results with other values of the labor supply elasticity and so conclude that this result is robust to the exact measure of hours. The competitive model cannot explain all of the movement in hours relative to the consumption–output ratio if the labor income tax rate is constant.

One possible explanation for this pattern is that labor tax rates are in fact countercyclical. This hypothesis has some supporters. For example, in a recent paper, Mertens and Ravn (2008) measure tax shocks using the Romer and Romer (2007) narrative analysis of tax policy. They conclude that tax shocks account for eighteen percent of the variance of output at business cycle frequencies. Perhaps most provocatively, they find that the 1982 recession was caused by workers' anticipation of future tax cuts. Of course, I have shown that expectations of future tax cuts may affect both the consumption–output ratio and the hours worked, but not the labor wedge, so such behavior cannot easily explain the patterns in the data. In any case, most economists seem to be skeptical that tax movements alone can explain the observed variation in the labor wedge.

1.4 Alternative Specification of Preferences

A second possible explanation for the behavior of the labor wedge is that in this model either the MRS or the MPL is misspecified. The specification of the MPL depends only on the assumption of a Cobb–Douglas aggregate production function. Macroeconomists are justifiably reluctant to abandon this assumption because it ensures that the capital and labor shares of national income, as well as the interest rate, are constant, consistent with the Kaldor (1957) growth facts.

The specification of household preferences is also tightly constrained by long-run restrictions.[9] Maintain the assumption that preferences are separable across time and states of the world, but relax the assumption of additive separability between consumption and leisure. To be consistent with balanced growth—the absence of a long-run trend in hours—and a constant Frisch elasticity, preferences over consumption and leisure must be ordered by the utility function

$$\sum_{t=0}^{\infty}\sum_{s^t}\beta^t \Pi(s^t)\frac{c(s^t)^{1-\sigma}(1+(\sigma-1)(\gamma\varepsilon/(1+\varepsilon))h(s^t)^{(1+\varepsilon)/\varepsilon})^{\sigma}-1}{1-\sigma}. \tag{1.13}$$

As before, $\gamma > 0$ parameterizes the disutility of work and $\varepsilon > 0$ is the Frisch labor supply elasticity. The new parameter $\sigma > 0$ determines the substitutability between consumption and leisure. The limit as $\sigma \to 1$

[9] Hall (2009) does not impose balanced-growth preferences, but instead assumes that permanently doubling wages reduces hours worked by fifteen percent, ceteris paribus. Thus income effects outweigh substitution effects in labor supply. He also allows for measurement error in consumption, hours, employment, and productivity. Finally, he allows for rigid wages, as I discuss in section 4.4. These departures jointly lead him to conclude that the labor wedge is in fact relatively small and acyclic. I confirm that such a substantial departure from balanced-growth preferences reduces the correlation between the measured labor wedge and hours worked, but I find that it actually increases the volatility of the measured labor wedge. Moreover, according to the model without balanced-growth preferences, a sharp increase in the labor wedge, roughly from 0 to 0.5, kept hours worked from rising between 1959 and 1973. It is unclear what can explain this secular increase in the labor wedge, except for model misspecification.

nests the separable case in equation (1.1). The case where $\sigma > 1$ is of particular interest, since this implies that the marginal utility of consumption is higher when households work more, consistent with standard models of time allocation (Becker 1965). In any case, utility is increasing and concave in consumption and decreasing and concave in hours of work.

With these preferences, the first-order conditions for consumption and hours are

$$\beta^t \Pi(s^t) c(s^t)^{-\sigma} \left(1 + (\sigma - 1)\frac{\gamma\varepsilon}{1+\varepsilon} h(s^t)^{(1+\varepsilon)/\varepsilon}\right)^{\sigma} = \lambda q_0(s^t),$$

$$\beta^t \Pi(s^t) c(s^t)^{1-\sigma} \left(1 + (\sigma - 1)\frac{\gamma\varepsilon}{1+\varepsilon} h(s^t)^{(1+\varepsilon)/\varepsilon}\right)^{\sigma-1} \sigma\gamma h(s^t)^{1/\varepsilon} = \lambda q_0(s^t)(1 - \tau(s^t))w(s^t).$$

The Frisch demand system expresses consumption $c(s^t)$ and hours $h(s^t)$ as functions of the Lagrange multiplier λ, the intertemporal price $q_0(s^t)$, and the wage rate $w(s^t)$. Eliminating $c(s^t)$ between the first-order conditions gives

$$\sigma\gamma h(s^t)^{1/\varepsilon} = \left(\frac{\lambda q_0(s^t)}{\beta^t \Pi(s^t)}\right)^{1/\sigma} (1 - \tau(s^t))w(s^t),$$

so the Frisch elasticity of labor supply—the elasticity of hours with respect to the wage holding fixed the intertemporal price and the Lagrange multiplier—is in fact ε. Moreover, eliminate $\lambda q_0(s^t)/\beta^t \Pi(s^t)$ between the first-order conditions to get

$$w(s^t) = \frac{\sigma\gamma c(s^t) h(s^t)^{1/\varepsilon}}{(1 - \tau(s^t))(1 + (\sigma - 1)(\gamma\varepsilon/(1+\varepsilon))h(s^t)^{(1+\varepsilon)/\varepsilon})}$$

whenever $\Pi(s^t) > 0$. Eliminate the wage using the firm's first-order condition, equation (1.11), and market clearing, $h(s^t) = h^d(s^t)$. This gives

$$\tau(s^t) = 1 - \frac{\gamma\sigma(c(s^t)/y(s^t))h(s^t)^{(1+\varepsilon)/\varepsilon}}{(1-\alpha)(1 + (\sigma - 1)(\gamma\varepsilon/(1+\varepsilon))h(s^t)^{(1+\varepsilon)/\varepsilon})}. \tag{1.14}$$

This is a modest generalization of equation (1.12). Once again, one needs to know only the consumption–output ratio, hours worked, and the value of four parameters in order to compute the labor wedge.

To understand the quantitative implications of this expression, I use the hours series from Prescott et al. (2008). I fix the labor share at the conventional value of $1 - \alpha = \frac{2}{3}$ and, for different values of the substitutability parameter σ and the elasticity of labor supply ε, I choose the disutility of work parameter γ to ensure an average labor wedge of 0.40.

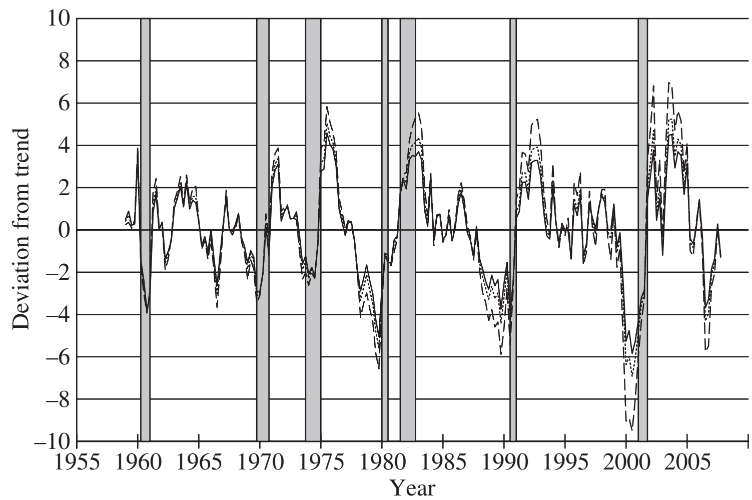

Figure 1.4. The deviation of the labor wedge from log trend, measured as an HP filter with parameter 1,600, using equation (1.14) and the CPS measure of hours. The dashed line shows $\sigma = 1$, the dotted line shows $\sigma = 2$, and the solid line shows $\sigma = 4$. In each case I set $\varepsilon = 1$ and $\alpha = \frac{1}{3}$ and fix the remaining parameter γ to ensure that the average labor wedge is 0.40. The gray bands show NBER recession dates.

Figure 1.4 shows the time series behavior of the labor wedge with the Frisch elasticity fixed at 1. The dashed line corresponds to the limit as σ converges to 1, the additively separable case that I analyzed before, while the remaining two lines show $\sigma = 2$ and $\sigma = 4$. Raising the substitutability between consumption and leisure modestly reduces the magnitude of fluctuations in the labor wedge but does not qualitatively change the results. I do not show the results with a higher elasticity of labor supply, but they too are similar.

Additionally, the microeconomic behavior of the model is unreasonable when σ is much larger than 1. Consider the following thought experiment. Two workers normally work and consume the same amount. One year, however, one is unable to work, for example because of an idiosyncratic shock to the disutility of work γ, while the other continues to work the average number of hours that I observe in the data. With complete markets, the two workers want to keep their marginal utility of consumption λ equal through this episode. How much lower is the consumption of the unemployed worker, c_{u}, than the consumption of the employed worker, c_{e}? The first-order condition for consumption implies that

$$\frac{c_{\mathrm{e}}}{c_{\mathrm{u}}} = 1 + (\sigma - 1)\frac{\gamma\varepsilon}{1+\varepsilon}h^{(1+\varepsilon)/\varepsilon},$$

where h is the number of hours worked by the employed worker. With $\sigma = 1$, consumption is equal for the two workers. With $\varepsilon = 1$ and $\sigma = 1.5$, consumption falls by 13 log points for the unemployed. With $\sigma = 2$, consumption should fall by 20 log points, and with $\sigma = 4$ it should fall by 31 log points.

Are these numbers reasonable? Aguiar and Hurst (2005) provide some guidance. They find that food consumption expenditures drop by about seventeen percent at retirement, accompanied by a fifty-three percent increase in the time spent on food production. This is consistent with values of σ near 2. Much stronger substitutability is inconsistent with the numbers in Aguiar and Hurst (2005) and does not resonate introspectively. In any case, even with infinite substitutability between consumption and leisure, $\sigma \to \infty$, the results do not change appreciably. With the labor supply elasticity fixed at $\varepsilon = 1$, the standard deviation of the annual growth rate of the labor wedge is more than twice the corresponding number for hours and the correlation between the two series is -0.76.

1.5 Preference Shocks

A third theoretical possibility is that the representative agent's disutility of work, γ, is stochastic. This modifies equation (1.12) to read

$$\tau(s^t) = 1 - \frac{\gamma(s^t)}{1-\alpha}\frac{c(s^t)}{y(s^t)}h(s^t)^{(1+\varepsilon)/\varepsilon},$$

where $\gamma(s^t)$ is the history-contingent disutility of work. An economist who ignored variation in the disutility of work would then falsely conclude that there are fluctuations in the labor wedge, even if labor taxes are constant in the data and the model is otherwise correct.

Many recent quantitative macroeconomic models allow for such a preference shock. An unobserved demand shock plays an important role in explaining aggregate fluctuations in Rotemberg and Woodford (1997). They interpret the shock as a combination of a preference shock and a shock to government spending.[10] Erceg et al. (2000) and Smets and Wouters (2003) also have a quantitatively important preference shock in their models of monetary policy. More recently, Galí and Rabanal (2004) find that a preference shock explains fifty-seven percent of the variance

[10] Unlike preference shocks, government spending can be measured and is not strongly correlated with the labor wedge. In any case, the model I have developed here allows for government spending shocks, but these do not affect the labor wedge equation.

of output and seventy percent of the variance of hours in their estimated dynamic stochastic general equilibrium model.[11]

A closely related theoretical possibility is that workers have time-varying market power in labor supply. This is often formalized by assuming that each household is the monopoly supplier of a heterogeneous type of labor and sets the wage to maximize its utility. Recessions are periods when different types of labor are poor substitutes, so households are better able to exploit their market power, reducing hours to drive up wages. A number of recent papers (including Smets and Wouters (2003, 2007)) have emphasized time-varying wage markups as an important source of business cycle shocks. In Smets and Wouters (2007), the wage markup shock accounts for a fifth of the variance in output and over half the variance in inflation at a ten-quarter horizon. Galí et al. (2007) also find an important role for markup fluctuations, but reach a different conclusion. They argue that some other, unspecified, primitive shock causes countercyclical fluctuations in markups, which in turn generates a countercyclical labor wedge.

Like many economists, I have a strong prior belief that the changes in the disutility of labor and changes in wage markups do not drive business cycle fluctuations.[12] Although households may differ in their disutility of work and the disutility may change over time for some households, one would expect those movements to average out in a large economy. Still, the empirical success of preference and markup shocks is revealing. They work because, viewed through the lens of a market-clearing model, recessions look like times when workers choose to supply less labor

[11] Galí and Rabanal (2004, p. 271) write that the preference shock can be "interpret[ed] more broadly as a (real) demand shock."

[12] For example, Modigliani (1977, p. 6) writes

> Sargent (1976) has attempted to remedy this fatal flaw by hypothesizing that the persistent and large fluctuations in unemployment reflect merely corresponding swings in the natural rate itself. In other words, what happened to the United States in the 1930's was a severe attack of contagious laziness! I can only say that, despite Sargent's ingenuity, neither I nor, I expect, most others at least of the nonmonetarists' persuasion are quite ready yet to turn over the field of economic fluctuations to the social psychologist.

Mankiw (1989, p. 82) writes

> Alternatively, one could explain the observed pattern without a procyclical real wage by positing that tastes for consumption relative to leisure vary over time. Recessions are then periods of "chronic laziness." As far as I know, no one has seriously proposed this explanation of the business cycle.

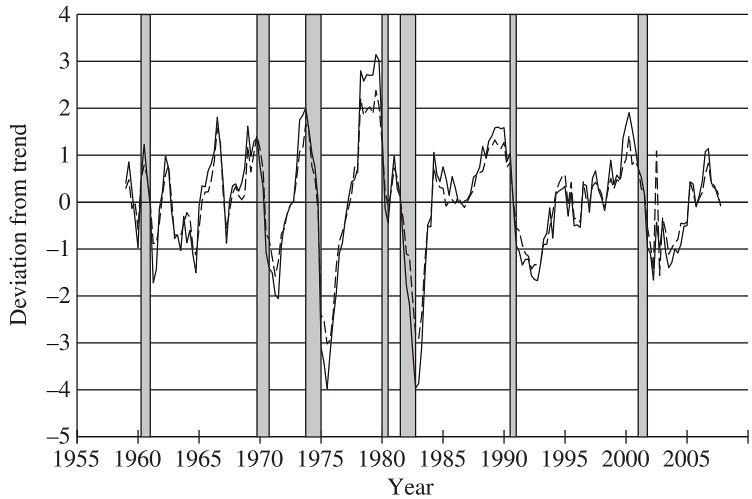

Figure 1.5. Deviation of per capita hours and the e-pop ratio from log trend, measured as an HP filter with parameter 1,600. The solid line shows the deviation of hours from log trend and the dashed line shows the deviation of employment. The gray bands show NBER recession dates.

than is predicted by the model and expansions look like times when they choose to supply more labor.

1.6 From Hours to Unemployment

In the remainder of the book, I explore whether search-and-matching models, based on Pissarides (1985) and Mortensen and Pissarides (1994), can help to explain the behavior of the labor wedge. In doing so, I switch from a focus on the behavior of hours worked to a focus on the behavior of employment and unemployment. This is because search costs introduce a nonconvexity into households' decision problems that emphasizes the binary decision of whether to work, rather than the continuous decision of how many hours to work each week. The data indicate that this focus is, for the most part, appropriate. Figure 1.5 shows that the correlation between the detrended employment–population (e-pop) ratio and the detrended hours per adult is 0.97 and detrended hours are only slightly more volatile than detrended employment, with a relative standard deviation of 1.3.[13] In words, most business cycle frequency fluctuations in hours are accounted for by fluctuations in

[13] I compute hours per member of the noninstitutional population with ages between 16 and 64 using the measure of hours in Prescott et al. (2008). I use their measure of employment as well, and divide by the same population measure.

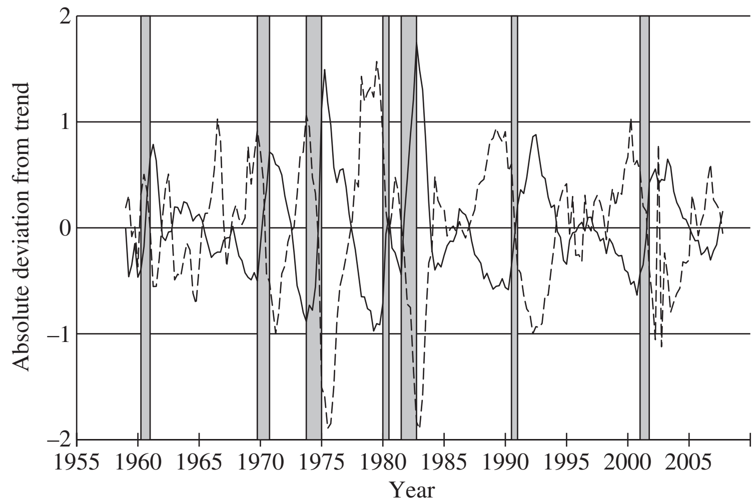

Figure 1.6. Deviation of the e-pop and u-pop ratios from trend, measured as an HP filter with parameter 1,600. The dashed line shows the deviation of the e-pop ratio from trend and the solid line shows the deviation of the u-pop ratio. The gray bands show NBER recession dates.

employment, rather than by fluctuations in the number of hours per employee.

As in many search models, the book also focuses on the margin between employment and unemployment, abstracting away from entry and exit from the labor force. Again, this is empirically reasonable at business cycle frequencies. Figure 1.6 shows that the correlation between the absolute deviation of the e-pop ratio from trend and the absolute deviation of the unemployment–population (u-pop) ratio from trend is −0.90, with a relative standard deviation of 1.4.[14] When employment falls below trend, most of the workers show up as unemployed, rather than dropping out of the labor force.

[14] I use the standard BLS measure of unemployment, based on the CPS.