Work and Leisure

Jón Steinsson*
University of California, Berkeley

January 26, 2024

One of the most important economic issues we all grapple with in life is how much to work. This issue has many dimensions. Some of these dimensions we grapple with as individuals and others we grapple with as a society. As individuals, we choose which type of job to pursue. Some jobs demand long hours. Some jobs are physically demanding. Some jobs require large amounts of education or other types of training. Some jobs have flexible hours. Some jobs have generous vacation policies. Some jobs have generous paternity leave. Some jobs accrue generous retirement benefits. We each have different preferences, abilities, and life circumstances, and these factors lead us to make different choices regarding what type of career to pursue.

As individuals, we also decide when to retire; we decide whether to work full-time or part-time; we decide whether to take time off work to raise our children or to enjoy ourselves; and we decide whether to work at all (if our economic circumstances allow such a choice). Typically, we face a trade-off that largely boils down to the notion that the more time and effort we devote to working the higher our income will be.

We grapple with many of these same issues as a society. In most countries, various laws and regulations influence people's choices regarding work. Many countries have public pension systems that influence the incentives people face to retire. In some cases, countries even have a mandatory retirement age for certain jobs. Many countries also have public policies that govern working hours, paternity

*I would like to thank Gerard Martin-Escofet, Ethan McClure and Venance Riblier for excellent research assistance. I would like to thank Janice Eberly, Emi Nakamura, and Moritz Schularick for valuable conversations and Emmanuel Saez, Michael Sigenthaler, and Jósef Sigurdsson for providing me with estimates from their studies. First posted in November 2023.

leave, and vacation policies. In addition, collective bargaining agreements between unions and employers often shape the incentives workers have regarding, working hours, paternity leave, and vacation policies.

There is a simplistic view that is quite prevalent in popular discourse that more work is always better. Politicians place immense emphasis on job creation. 'Jobs! jobs! jobs!' seems to be a constant rallying cry in politics. This is understandable; the risk of having no job at all is an extremely serious and frightening risk for those of us who need to provide for ourselves and our families. But once one gets beyond the basic notion of there being enough jobs, it is not clear that more work is always better. After all, for many – perhaps most – a job is a means to an end, not an end in itself.

People work to earn income. But the time they spend working takes away from the time they have to enjoy the fruits of their labor. This implies that ideally there is a balance between work and leisure. This chapter explores this balance. We begin by deriving a condition for optimal labor supply by workers. We then use this condition to explore several issues. First, we consider how the balance between work and leisure has changed as people have become more affluent over the last 200 years. Second, we analyze the Covid recession using the concepts we have developed. Third, we ask why Europeans work so much less than Americans. And finally, we explore how the ideas developed in this chapter differ from the highly influential analysis of similar issues by Karl Marx.

1 The Trade-Off Between Work and Leisure

How much people work over their lifetime depends both on how much they want to work and on whether they are able to find the amount and type of work that they would like. Finding enough work of the type that matches a person's skills can be difficult. Many people end up being unemployed, underemployed, or employed in jobs not well suited for their skills for significant amounts of time. Despite the importance of these problems, we will largely abstract from them in this chapter so that we can focus exclusively on gaining a deeper understanding of the determinants of how much people want to work.

Our goal is to present a stylized description – a model – of people's choices regarding how much to work. Our starting point is to assume that people act in a purposeful manner when making decisions about how much to work. This raises the question of how we should model purposeful behavior by people? The most basic model of this type in economics is the rational choice model, which assumes that: 1) people face a set of options and need to make a choice among these options; 2) they have preferences over these options (i.e., they can rank these options from the least preferred to the most preferred); 3) they choose the option that they most prefer.

Textbooks on microeconomics typically discuss the foundations of the rational choice model in quite a bit of detail – e.g., what properties should a person's preference ordering have to be viewed as rational. Readers are referred to Varian (2014) and Kreps (1990) for more thorough treatments of these foundations. The assumptions of the rational choice model are clearly strong assumptions. (Can we really articulate the entire set of options we face, let alone rank them from least preferred to most preferred?) But it yields a relatively simple and (arguably) quite insightful framework.

A celebrated result in microeconomic theory is that under certain conditions people's preferences can be represented by a function. This function maps each option into a number, which is called the utility that option yields. The function is called the *utility function*. Purposeful behavior then amounts to maximizing a person's utility function over the set of option the person can choose among.

1.1 The Household's Utility Function

Here, we focus on people's choice regarding how much to work, and the implied trade-off between work, leisure, and consumption. To this end, consider a household with the following utility function

$$U(C, H),$$

where C denotes consumption and H denotes hours worked. This utility function is quite "broad brush" in that we are ignoring many details, such as what types of jobs people choose and which goods they choose to purchase with their income. We are also ignoring issues having to do with how work and consumption is divided up within the household. We do this to simplify the analysis, which allows us to focus more clearly on certain aspects of the household's problem which it is important to understand before other complications are added.

To further simplify our analysis, we assume that the utility function $U(C, H)$

takes the form

$$U(C) - V(H). \tag{1}$$

Here, $U(C)$ denotes utility from consumption of goods and services and $V(H)$ denotes disutility from supplying labor. In equation (1), utility from consumption and disutility from work enter the overall utility function in an "additively separable" manner – i.e., the two components are simply added together to get overall utility. Additive separability implies that the extra utility derived from an extra unit of consumption does not depend on the amount the household is working. More formally, the marginal utility of consumption does not depend on hours worked. This is violated if, for example, working more leads the household to buy more fancy clothes, spend more on meals, or buy more gas because of more commuting to and from work. By assuming that the utility function is additively separable, we are abstracting from these effects, for simplicity.

In the equations above, H denotes hours worked per household. The aggregate amount of labor supplied is then $L = NH$, where N is the size of the population. (Strictly speaking the number of households.) It is mathematically convenient to assume that there are a continuum of households in the economy, i.e., one household per point on an interval of the real line. This means that there are an infinite number of households and each household is infinitesimally small relative to the size of the economy. In this case, N refers to the "size" of the population, i.e., the length of the interval.

But in what units should we measure the size of the population? In fact, it does not matter in what units we measure the population. This means that we can choose whichever units we want for N. The choice of the units to measure N in is called a normalization. A particularly convenient normalization is to set $N = 1$. With this unit size, there is a household at each point on the interval from 0 to 1 and the total size of the population is 1. This normalization amounts to choosing the unit we use to coincide with the size of the population.

1.2 Properties of the Utility Function

What properties should the household's utility function have? First, consider the function $U(C)$. This function is meant to represent the utility households derive from consumption of goods and services. Suppose for illustrative purposes that the household consumes only coconuts. Since coconuts are a "good" thing (in the household's view), consuming more coconuts yields more utility. In other words,

$U(C)$ is increasing in C, or equivalently the first derivative of $U(C)$ is positive: $dU(C)/dC > 0$. To save on notation, we use a prime to denote a derivative: $U'(C) = dU(C)/dC$. Using this notation, we have that $U'(C) > 0$.

What about the second derivative of $U(C)$? Consider first a household that is consuming very few coconuts. For this household, an extra coconut is very valuable – i.e., yields substantial utility. Contrast this with a household that is already consuming a huge number of coconuts. This household is, in all likelihood, already rather satiated in coconuts and an extra coconut is only of modest value – i.e., yields a modest amount of extra utility. This comparison suggests that it is reasonable to assume that coconuts (and consumption of goods and services more generally) have diminishing marginal value to households: an extra coconut is most valuable when households have few coconuts and the marginal value of an extra coconut falls as the household acquires more coconuts. We capture this logic mathematically by assuming that the second derivative of the function $U(C)$ is negative: $U''(C) = d^2U(C)/dC^2 < 0$. A utility function that has this property is said to exhibit diminishing marginal utility.

We assume that the function $V(H)$ is also upward-sloping – i.e. has a positive first derivative. Mathematically, we write this as $V'(H) = dV(H)/dH > 0$. Recall that there is a negative sign in front of $V(H)$ in equation (1). This implies that $V(H)$ represents the household's *dis*utility (the negative of utility). By making $V(H)$ upward-sloping, we are assuming that the household dislikes working.

Is it obvious that $V(H)$ should be upward sloping? Might it not be the case that people enjoy their jobs? Many people are certainly quite miserable when they do not have a job. Furthermore, many people desire that their life have meaning and some find this meaning (partly) in the work they do. These considerations suggest that work may yield positive utility rather than negative utility for some people. But two points are important to keep in mind in this context. First, some of the unhappiness people suffer from when they can't find work is due to the loss of income and therefore the loss of consumption that joblessness entails. Second, even those for whom being employed is an important source of meaning, typically do not want to work every waking hour. In other words, even for these people, there comes a point when working another hour yields disutility rather than utility. People's jobs may therefore yield positive utility in aggregate, but they likely yield disutility *on the margin*. It is the utility work yields on the margin that is important for our purposes. For simplicity, we therefore abstract from the fact that $V(H)$ may be negatively sloped in certain intervals and focus on the fact that it is positively sloped in the interval that is relevant for our analysis.

Finally, we consider the second derivative of $V(H)$. Unlike $U(C)$, we assume that $V(H)$ has a positive second derivative: $V''(H) = d^2V(H)/dH^2 > 0$. To see why, consider a construction worker that usually works 40-50 hours per week but occasionally works much more when the job they are working on is behind schedule and an important deadline is looming. For such a worker, working an extra few hours in a normal 40 hour week is likely not very costly – i.e., yields modest disutility. Contrast this with working a few extra hours during a week when the worker has already worked 90 hours (e.g., six 15 hour days in a row). The 91st and 92nd hours in a work week are likely to be much more costly to the worker than the 41st and 42nd hours. (I speak from experience on this point.) Mathematically, this logic can be represented by the function $V(H)$ having a positive second derivative, which implies that the first derivative is increasing. We refer to this property as increasing marginal disutility of labor.

An alternative specification of the utility function, which is sometimes more convenient is

$$U(C) + V(1 - H). \tag{2}$$

In this formulation, the function $V(\cdot)$ denotes not the disutility from labor but rather the utility from leisure. Here, we have normalized the total amount of time available to the household to 1. The household works H hours. The remaining $1 - H$ hours are then leisure time. The argument of $V(\cdot)$ is therefore leisure rather than labor in equation (2). And since leisure yields utility rather than disutility (at the margin), $V(1 - H)$ enters the overall utility function with a positive sign (and has a negative second derivative).

1.3 The Household's Budget Constraint

The household maximizes its utility function – equation (1) – subject to a *budget constraint*. We consider a setting where the household's budget constraint is simply

$$C = wH + T, \tag{3}$$

where w denotes the hourly wage the household receives for working, and T denotes all other sources of income, including any transfer the household may receive from the government. The wage w is measured in consumption units. So, for example, if the consumption good is coconuts, the wage is in units of coconuts per hour. The additional income is also measured in consumption units (e.g., coconuts).

The structure of the budget constraint is that the left-hand-side represents all sources of spending (only consumption in our case), while the right-hand-side represents all sources of income. The constraint encodes the simple notion that the household cannot spend more than it earns. This implies that it cannot choose a high value of C without also choosing a high value of H. For simplicity, we are abstracting from savings (introduced in chapter XX) and taxes (introduced in section 4 of this chapter). Strictly speaking, the household could spend less than it earns. But as long as $U'(C) > 0$, this is not optimal. So, we ignore this possibility.

1.4 The Household's Maximization Problem

The household's problem is to maximize its utility function – equation (1) – subject to its budget constraint – equation (3). The household's problem is thus a constrained optimization problem. This problem is simple to solve. There are actually several ways we could go about this. Perhaps the simplist approach is to start by plugging the budget constraint into the utility function to eliminate C. This yields

$$U(wH + T) - V(H). \tag{4}$$

This step transforms the problem from a problem where the household seeks to optimally choose two variables (C and H) subject to one constraint into a problem where the household seeks to choose one variable (H) in an unconstrained manner (with the constraint now embedded in the objective function).

The next step is to differentiate equation (4) with respect to H and set the resulting derivative to zero. This yields

$$wU'(wH + T) - V'(H) = 0.$$

We can now use the budget constraint to rewrite this equation as

$$wU'(C) = V'(H). \tag{5}$$

This equation is a necessary condition for household optimization. In other words, it describes what must be true for households to be optimally trading off consumption and leisure. (The assumptions we have made on the shape of the utility function guarantee that the second-order sufficient conditions for household optimization also hold.)

The derivation above is straightforward from a mathematical perspective. But as with the solution to many optimization problems in economics, it can be performed in a mindless "plug-and-chug" manner that results in little insight. To understand the economics embedded in equation (5), it is important to think carefully about the terms that appear in the equation. For this purpose, it is useful to adopt a "variational" perspective.

Suppose the household is considering working H hours. To assess whether this choice is optimal, the household might contemplate working a little bit more. If the household were to work a little bit more, the marginal cost of this would be the marginal disutility of work: $V'(H)$ "utils" per hour worked. The marginal benefit would be the marginal value of the extra consumption the household could then afford. This is equal to the wage w the household earns (in units of coconuts per hour worked) times the marginal value of extra income (in units of "utils" per coconut). The marginal value is therefore $wU'(C)$ utils per hour worked.

This discussion shows that the left-hand-side of equation (5) is the marginal value of working more, while the right-hand-side of that equation is the marginal cost of working more. Equation (5), therefore, simply says that for the household to be optimizing it must set the marginal value of working more equal to its marginal cost.

If to the contrary, $wU'(C) > V'(H)$, the household could raise its utility by working more, since the marginal value of the extra income it earns from working more is (in this case) higher than the marginal disutility of the work involved. And if $wU'(C) < V'(H)$, the household could raise its utility by working less, since (in this case) the marginal disutility of working is larger than the marginal value of the resulting income. It is only when $wU'(C) = V'(H)$ that the household can neither raise its utility by working more or by working less. This is the "sweet spot" where the marginal utility of the consumption extra work allows for is exactly equal to the marginal disutility of the work involved.

A slight rearrangement of terms in equation (5) yields

$$\frac{V'(H)}{U'(C)} = w. \tag{6}$$

This way of writing the household's optimality condition has the same interpretation as before except in different units. In this case, the right-hand-side of the equation is the marginal benefit of working more in units of coconuts per hour (rather than utils per hour). The interpretation of the left-hand-side is a bit trickier, but turns out to be the marginal cost of working more in units of coconuts per hour.

To see this, notice first that $U'(C)$ is the marginal value of coconuts in units of utils per coconut. Loosely speaking, $U'(C)$ is the number of extra utils the household

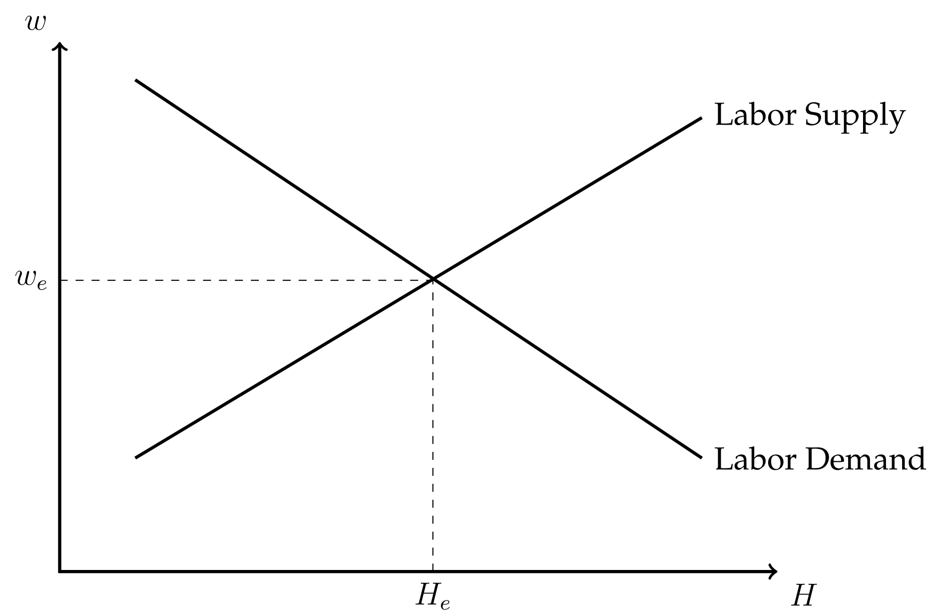

Figure 1: Labor Market Equilibrium Holding Household Consumption Fixed

get from one extra coconut (at the margin). If one extra coconut yields $U'(C)$ utils, then $1/U'(C)$ coconuts yield one util (at the margin). In other words, $1/U'(C)$ is the price of one util in coconuts (at the margin). This means that multiplying $V'(H)$ by $1/U'(C)$ transforms the marginal cost of working more from units of "utils per hour" to units of "coconuts per hour."

1.5 The Household's Labor Supply Curve

Equation (5) (or equivalently equation (6)) is the household's labor supply curve. It implicitly describes how much labor H the household supplies as a function of the wage w. We learn from equation (5) that the household's labor supply depends not only on the wage w but also on the household's level of consumption C. The fact that consumption is, in turn, partly determined by the wage and the amount the household works adds important extra complication, which we will explore in detail in the next section. For now, however, we make the simplifying assumption that consumption is fixed and unaffected by the wage w and the household's labor supply H. In this case, we can represent the combination of values of the wage w and hours worked H that satisfy equation (5) by an upward-sloping curve in (H, w) space. Figure 1 plots an illustrative curve of this type.

Why is the set of (H, w) "points" that satisfy equation (5) in Figure 1 upward sloping? The reason for this is that $V''(H) > 0$. To see this, consider first a point (H, w) that satisfies equation (5). Then consider what value of H makes equation (5) hold for a slightly larger value of w. Increasing w increases the left-hand-side of equation (5). For equation (5) to hold, the right-hand-side must increase as well. What change in H will lead to such an increase? Since $V''(H) > 0$, an increase in H increases $V'(H)$ and therefore increases the right-hand-side of equation (5). This means that the values of H that make equation (5) hold are increasing in w. In other words, the set of points that satisfy equation (5) is an upward sloping set of points. (In Figure 1, I plot this set of points as being a straight line. This is not necessarily the case and is only done for simplicity.)

In the last chapter, we derived the economy's labor demand curve from a firm's optimization problem. If we make the additional assumption that the labor market clears, we can combine these two equations to analyze equilibrium in the labor market. Analysis along these lines is often done graphically in economics. The equation describing firm demand for labor – equation XX from chapter XX – is a downward sloping curve in (H, w) space. (In chapter XX, we derived labor demand in terms of L rather than H. But recall that $L = NH$ and we can normalize $N = 1$.) The equilibrium in the labor market is then the intersection of the labor demand curve and the labor supply curve – the point (H_e, w_e) in Figure 1.

We will sometimes use such graphical analysis. But at other times we will work directly with the equations since some insights are better illustrated through direct manipulation of the equations. It is extremely important to recognize that labor demand and supply curves are simply graphical illustrations of the optimization conditions we have derived. This is the case more generally in economics: demand and supply curves are simply graphical representations of behavioral equations (often derived from household or firm optimization). We will see this equivalence again and again throughout this book.

2 Income and Substitution Effects on Labor Supply

Suppose you have a job that is paid by the hour and your boss announces that due to an upcoming important deadline your wage is temporarily double its normal level for the next week but will return to its original level in a week's time. How would this temporary change in your wage affect the amount that you decided to work over the next week? When I ask my students this question, most say rather confidently that they would work more.

Now, suppose your boss comes to you and surprises you by letting you know that they have decided to permanently double your wage. How would this permanent change in your wage affect that amount that you decide to work? When I ask my students this question, they are much less sure. Some say they would work more. Others are not sure they would change the amount that they would work. Still others say they would probably work less after the raise.

These two scenarios are meant to illustrate how an increase in one's wage has two opposing effects on one's labor supply. These effects are called the *substitution effect* and the *income effect*. An increase in one's wage makes working more lucrative and, as a consequence, leisure more expensive. (The wage is the opportunity cost of leisure.) The substitution effect captures the notion that we substitute toward lucrative activities and away from expensive activities when wages change. This type of substitution pushes us in the direction of working more as wages rise. If the substitution effect was the whole story, an increase in our wage would cause us to work more.

However, another effect of a change in our wage is that it makes us richer. Richer people consume more of most things. A good is considered "normal" if people consume more of it as they become richer. (Goods that people consume less of when they become richer are referred to as "inferior" goods.) Leisure seems to be a "normal good." (We will explore empirical evidence supporting this below.) This means that, other things equal, we choose to "consume" more leisure – i.e., work less – when we become richer. This effect is called the income effect on labor supply. If the income effect was the whole story, an increase in our wage would cause us to work less.

The degree to which a wage change is permanent affects the size of the income effect of that wage change. A permanent wage change has a much larger effect on ones life-time income than a transitory wage change. A permanent wage change, therefore, has a larger income effect than a transitory wage change. The example above of a large change that lasts only one week was designed to minimize the income effect of the wage change. Wages over one week of one's life make up a trivial portion of one's life-time income. The income effect of such a wage change is therefore likely to be very small. In this case, therefore, the substitution effect likely wins out and we work more. The case of a permanent wage change is less clear because the income effect in this case is much larger.

We can gain further insight about the relative magnitude of income and substitutions effects on labor supply with the help of the labor supply curve $wU'(C) = V'(H)$. A pure substitution effect on labor supply is the effect on labor supply of a change in the wage w when consumption C is held constant. We already worked out that this effect is positive in section 1 and plotted the labor supply curve in Figure 1 under this assumption.

The notion that we can think of C being unchanged when w changes may seem mysterious. After all, the budget constraint in our model says that $C = wH + T$. So, in our model, there is a clear link between C and w. But the simplicity of our model leads it to overstate this link. As we noted earlier, in this chapter, we are abstracting from saving. A more realistic model would introduce the passage of time: households would live for many time periods. In such a model, which we analyze in chapter XX, consumption in a particular period is not necessarily equal to contemporaneous income. Rather, households can save and borrow.

We will see in chapter XX that households may choose to save a substantial portion of any unusual transitory increase in income. This allows them to spread out the increased consumption that the transitory income allows for rather than engaging in a temporary consumption binge. A very transitory increase in income may thus result in a very minimal contemporaneous increase in consumption. Our assumption here that consumption is completely unaffected is, thus, not as unrealistic as it may seem at first. But it is clearly a simplifying assumption we make to allow us to think through a pure substitution effect.

A pure income effect on labor supply occurs when consumption C changes without any change in the wage. A simple way to think about this in our model is as a consequence of an increase in non-labor income T. An extreme example is if the household were to win the lottery. In this case, C would rise, but w would remain unchanged. The income effect on labor supply is the effect of such an increase in C on H. The labor supply curve $wU'(C) = V'(H)$ must hold both before and after this increase in C. The income effect is, therefore, the change in H needed to make $wU'(C) = V'(H)$ hold when C increases.

To find the needed change in H, we first consider how the change in C affects $U'(C)$ on the left-hand-side of the labor supply curve. The derivative of $U'(C)$ is $U''(C)$. Since $U''(C) < 0$, $U'(C)$ is decreasing in C. This means that an increase in C will lead $U'(C)$ to fall. As a consequence, the left-hand-size of $wU'(C) = V'(H)$ falls. For the equation to continue to hold, the right-hand-side must also fall. Since $V''(H) > 0$, a decrease in H will lead $V'(H)$ on the right-hand-side of the equation to fall. An increase in C that is independent of w, therefore, implies that H falls.

This logic shows formally that the income effect on labor supply is negative in our model.

Generally speaking, a change in the wage will have both an income effect and a substitution effect. We can see this mathematically by plugging the budget constraint into the labor supply curve to eliminate C and get

$$wU'(wH + T) = V'(H).$$

In this equation, w shows up in two places. It is the first w – the one furthest to the left in the equation – that results in a substitution effect and the second w – the one inside $U(\cdot)$ – that results in an income effect.

2.1 Empirical Evidence on Income and Substitution Effects

A natural question to ask at this stage is which effect is stronger? For short term changes in wages – such as a wage increase that lasts a week or a month – the income effect will be modest and it seems likely that the substitution effect will dominate. But for long-run changes in wages – such as a permanent raise or a permanent tax cut – the income effect will be larger and it is not as clear which will dominate. One way to assess this question is to consider data on wages and hours worked over long periods of time.

Figure 2 plots the evolution of real wages in the United States from 1830 to 2009. Real wages refer to wages adjusted for changes in the price level. For the 140 year period from 1830 to 1970, real wages rose steadily by about 1.5% per year. This seemingly modest growth rate, when cumulated over 140 years, added up to a staggering 800% increase in real wages. How did hours worked respond to this huge increase in wages?

Figure 3 plots weekly hours worked per worker in the United States from 1830 to 2009. In the 1830s, the average work week was extremely long by modern standards at 69 hours per week. This amounts to 11.5 hours per day, six days a week. As wages rose, the length of the work week shrank steadily. By the 1890s, it was down to 59 hours per week. By the 1920s, it was down to 50 hours per week. And by the 1970s, the average work week had fallen by almost 40% to 44 hours per week.

Figures 2 and 3 strongly suggest that in the long run the income effect on labor supply is slightly stronger than the substitution effect. This has many important implications, some of which we will explore later in the chapter. Ignoring income

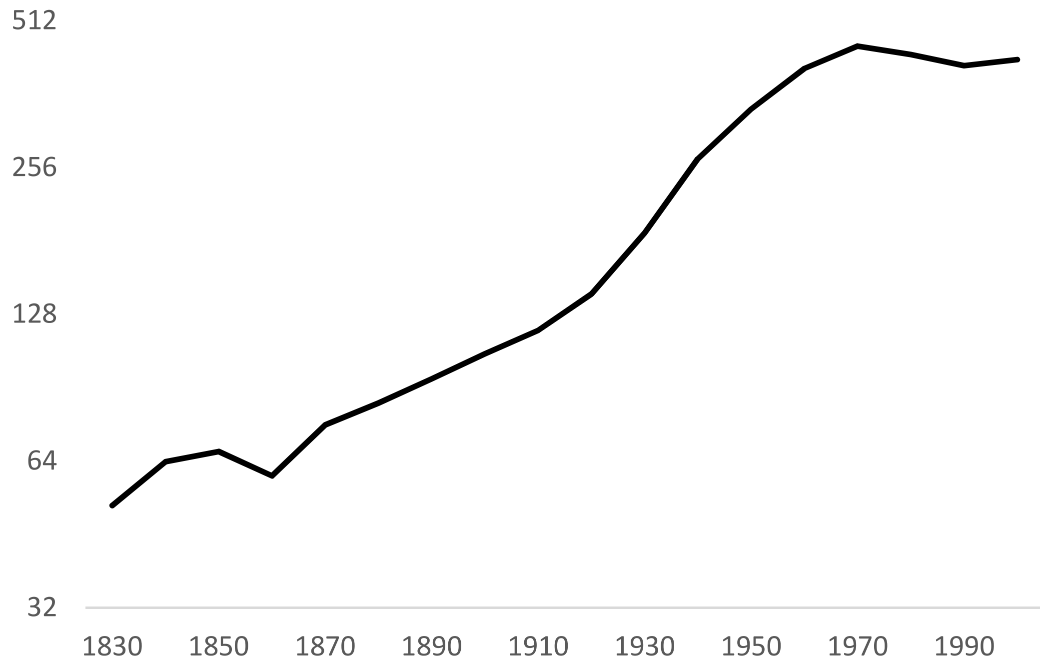

Figure 2: Real Wages in the United States

Note: This figure plots real wages in the United States from 1830 to 2009. Data for the period 1830 to 1988 are from Williamson (1995). Data from 1988 to 2020 is from the Main Economic Indicators of the OECD. The price series used to deflate nominal wages is the CPI. I plot decadal averages. That is, the data point listed as 1830 is the decadal average for 1830-1839.

effects is a common error in economic analysis. Many people seem to find substitution effects "more intuitive" than income effects and therefore implicitly or explicitly ignore income effects. An important example of this is the analysis of permanent tax cuts. It is common for people to argue that a permanent tax cut will lead people to work more. This would be the case if there were no income effects. But if income effects are larger than substitution effects, a permanent tax cut might result in people working less (depending on how the money is used, as we discuss in more detail in section 4).

While hours worked *per worker* have been on a downward trend since at least 1830, the same is not true of hours worked *per person*. Figure 4 plots hours worked per person of age 15-65 in the U.S. from 1950 to 2020. This series fluctuates quite a bit, but it does not display a downward trend. In fact, hours worked per person in 2020 are almost exactly the same as in 1950. The reason for the difference between Figures 3 and 4 is that the fraction of the population that is employed increased between 1950 and 2020. How does this affect our inference about the strength of income and substitution effects on labor supply? Does it perhaps imply that rising

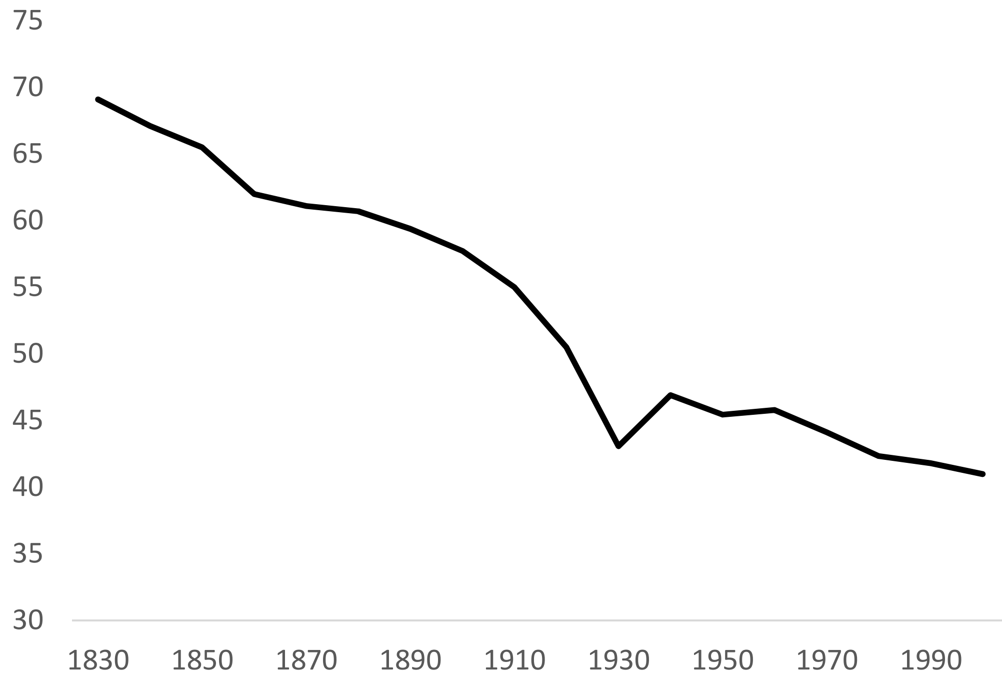

Figure 3: Weekly Hours Worked per Worker in the United States

Note: This figure plots hours worked per worker in the United States from 1830 to 2000. The data is from Boppart and Krusell (2020) who update a series from Greenwood and Vandenbroucke (2005). Greenwood and Vandenbroucke construct their series from Whaples (1990) for the period 1830 to 1880 and the *Historical Statistics of the United States: Colonial Times to 1970* and *Statistical Abstract of the United States* for the more recent period. The data points for the period 1830 to 1880 are based on the decadal censuses for these years. The data points from 1890 onward are decadal averages (i.e., the 1890 point is the decadal average for 1890 to 1899).

wages cause each worker to work less intensively but draw more workers into the labor force?

Figure 5 sheds further light on this issue. This figure plots the labor force participation rate and the employment rate for men and women separately. The labor force participation rate is the fraction of the adult population that is employed or is actively seeking employment. We say that these people are in the labor force. The employment rate is the fraction of the adult population that is employed. The difference between the labor force participation rate and the employment rate is the unemployed. Adults may be outside the labor force for a number of reasons such as being retired, in school, raising children, working in the home, enjoying leisure, unwell, or discouraged from seeking employment.

Figure 5 shows that the labor force participation rate of men has been falling steadily since 1950. In sharp contrast, the labor force participation rate for women rose substantially between 1950 and 2000. The large increase in the labor force par-

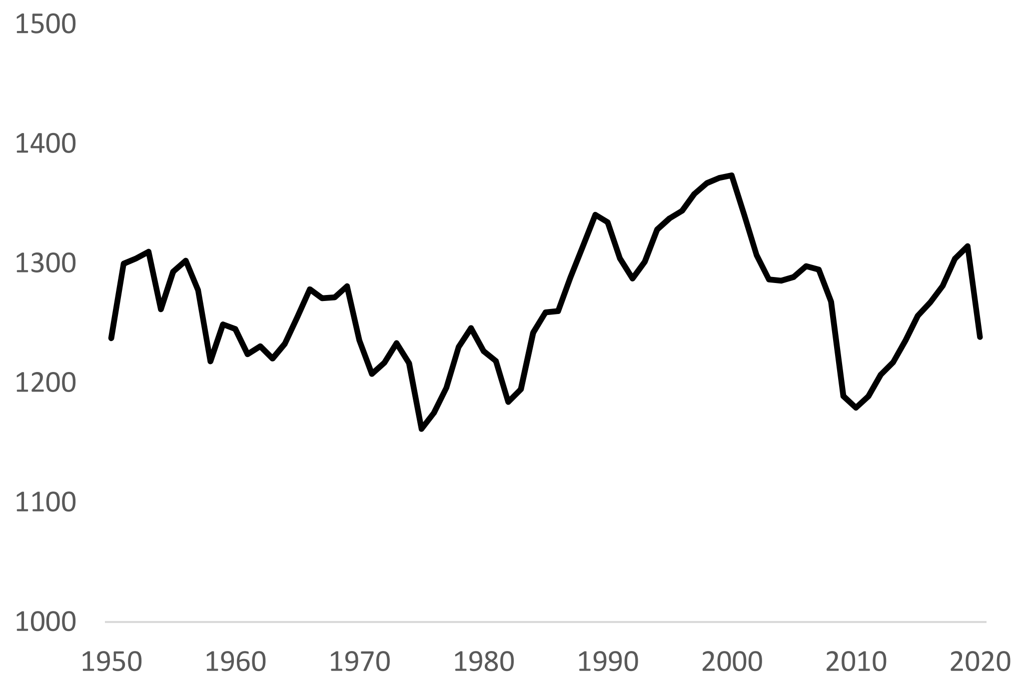

Figure 4: Annual Hours Worked per Person in the United States

Note: This figure plots annual hours worked per person of age 15-65 in the United States from 1950 to 2000. Data on total hours worked are from the Conference Board Total Economy Database. Data on the U.S. population aged 15-65 is from the OECD Employment and Labor Force Statistics database.

ticipation rate of women in the second half of the 20th century is one aspect of a more general Gender Revolution that occurred over this period. Prior to the Gender Revolution, women's employment was depressed by cultural norms and discrimination in the workplace. As attitudes towards women's work changed, many more women chose to work and the huge gap between male and female employment rates shrank. By 2000, a very substantial fraction of the gender difference in employment had disappeared. Since then the labor force participation rate and employment rate of women has been on essentially the same downward trend as the the labor force participation rate and employment rate of men.

The steady downward trend in the labor force participation rate of men from 1950 to 2020 suggests that rising incomes lead not only to shorter workweeks but also to fewer people working. As people become wealthier, they retire earlier, their children enter the labor force later, and they take more time off from being part of the labor force. However, during the second half of the 20th century, this trend was obscured when it comes to the the overall population by the Gender Revolution. For female employment, the Gender Revolution much more than counteracted rising

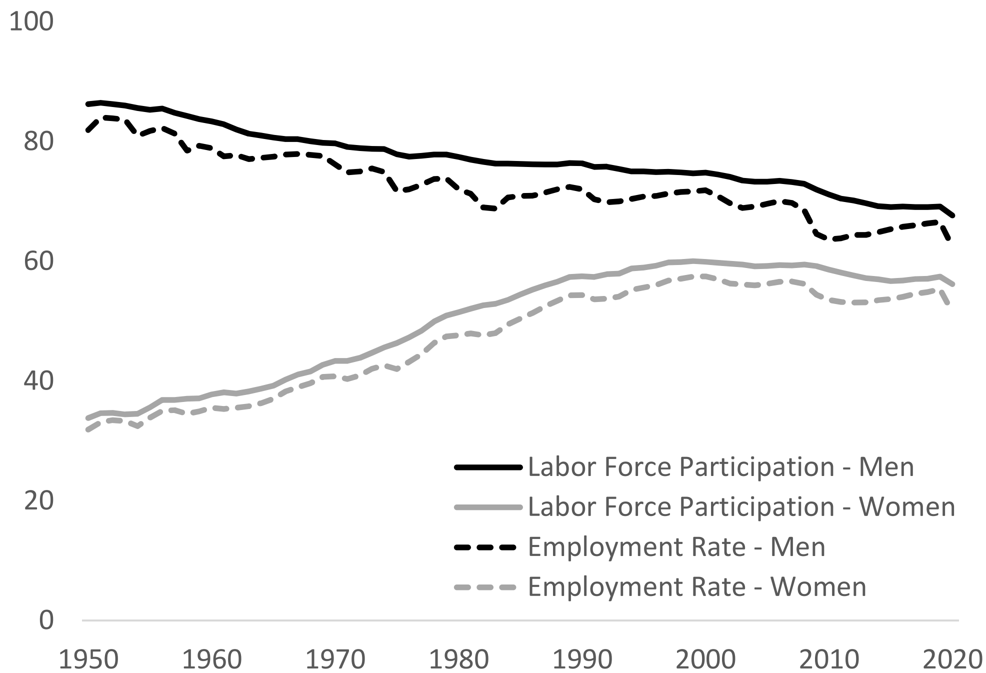

Figure 5: Employment of Men and Women in the United States

Note: This figure plots the labor force participation rate of men and women as well as the employment rate of men and women. These data are from the U.S. Bureau of Labor Statistics.

incomes over this period. As a result, the overall labor force participation rate rose from 59% to 67% between 1950 and 2000.

2.2 What we Learn about the Utility Function

The evidence on wages and hours worked presented in Figures 2-5 suggests that income effects on labor supply are slightly stronger than substitution effects in the long run. Notice that the evidence suggests a rather modest difference in the strength of these effects. Real wages have risen by roughly 800% while hours worked per worker have only fallen by a little less than 40%. Income and substitutions effects on labor supply are thus close to being equally strong in the long run. We can use this empirical result to determine more precisely than before what functional form is reasonable for us to assume for $U(C)$ and $V(H)$. Consider a household with

$$U(C) = \log C$$

$$V(H) = -\psi \log(1 - H)$$

and $T = 0$. In this case, $U'(C) = 1/C$ and $V'(H) = \psi/(1-H)$. The labor supply curve of the household then becomes

$$\frac{w}{C} = \frac{\psi}{1-H}.$$

If we plug in $C = wH$, we get

$$\frac{w}{wH} = \frac{\psi}{1-H}.$$

We can now cancel the w in the numerator against the w in the denominator on the left-hand-side of this equation and get that the labor supply curve of this household can be written as

$$\frac{1}{H} = \frac{\psi}{1-H}. \tag{7}$$

What is the effect of a change in the wage on the labor supply of this household? This may seem like a trick question since w doesn't show up in the labor supply curve. But it is not a trick question. The fact that w doesn't show up in equation (7) implies that hours worked are independent of w. An increase in w doesn't affect either the left-hand-side or the right-hand-side of equation (7). So, the change in w doesn't necessitate any change in H. The same value that led equation (7) to hold before the change in the wage leads equation (7) to hold after the change in the wage. In other words, the labor supply curve is vertical in this case.

The reason a change in the wage has no effect on labor supply when $U(C) = \log C$ and $V(H) = -\psi \log(1-H)$ is that in this case the income and substitution effect are exactly equally strong and, therefore, exactly cancel each other out. What is it exactly about these functional forms that leads the income and substitution effects to cancel out? Going through the derivation above, you will notice that it is the fact that the derivative of $\log C$ is $1/C$ that allows us to exactly cancel out the two w's on the left-hand-side of the labor supply equation. Other functional forms for $U(C)$ do not have this effect. From this we conclude that $U(C) = \log C$ or a functional form with a similar amount of curvature to $\log C$ is a reasonable function form for $U(C)$ since this function form generates an income effect on labor supply that is equally strong as the substitution effect and this is close to being true in the data.

2.3 Keynes and the Income Effect

In a famous essay titled *The Economic Possibilities of Our Grandchildren* written in 1930, the economist John Maynard Keynes speculated about the effect of steady wage increases on the amount that people would work over the next 100 years

(Keynes, 1930/1963). He said: "Suppose that a hundred years hence we are all of us, on the average, eight times better off in the economic sense than we are to-day." An eight-fold increase in income will occur over 100 years if income increases by about 2% per year over this period. For GDP, this is not very far from what has in fact occurred since 1930. The data plotted in Figure 2, however, indicate that real wages for the average worker have stagnated since 1970 and "only" increased three-fold between 1930 and 2020. The real wages of workers at the high end of the income distribution have, however, tracked increases in GDP more closely and, thus, increased roughly in line with Keynes' supposition.

Keynes then went on to argue that this spectacular increase in our income would result in our "absolute needs" being "satisfied in the sense that we prefer to devote our further energies to non-economic purposes." For this reason, Keynes thought that "the *economic problem* may be solved, or be at least within sight of solution, within a hundred years." Provocatively, Keynes then asked whether this will be a benefit. He was not at all sure. He worried that people would find it difficult to adjust to a situation of abundance and leisure and that this might even result in a general nervous breakdown:

> [T]here is no country and no people, I think, who can look forward to the age of leisure and abundance without a dread. For we have been trained too long to strive and not to enjoy. It is a fearful problem for the ordinary person, with no special talents, to occupy himself, especially if he no longer has roots in the soil or in custom or in the beloved conventions of a traditional society. To judge from the behavior and the achievements of the wealthy classes to-day in any quarter of the world, the outlook is very depressing! For these are, so to speak, our advance guard – those who are spying out the promised land for the rest of us and pitching their camp there. For they have most of them failed disastrously, so it seems to me – those who have income but no associations or duties or ties – to solve the problem which as been set them. (Keynes, 1930/1963, p. 368)

For these reasons, Keynes thought that despite their incredibly high level of income, people in the early 21st century would need to "do *some* work if he is to be contented." Keynes predicted that "[t]hree-hour shifts or a fifteen-hour week may put off the problem for a great while. For three hours a day is quite enough to satisfy the old Adam in most of us!"

Clearly, in Keynes' view, income effects on labor supply dominate substitution effects. In fact, his prediction was for an even more rapid decline in labor than we have actually seen occur. As we have seen, people do work less today than in 1930. But they work much more than Keynes predicted (even the well off, whose income has risen as much as he predicted). There are two potential reasons for this. One is that our desire for additional consumption does not in fact fall quite as rapidly as Keynes thought it would. Keynes likely underestimated people's insatiable desire for additional consumption. But another possible reason is that work may have become more enjoyable over time. A curious pattern that has emerged in recent decades is that hours worked of high income people have started to rise, while hours worked for people with low income have continued to fall. This may be because high income jobs have become more fulfilling in recent decades.

But are strong income effects the only way to explain the downward trend in hours worked in Figure 3? Actually, there is another possibility. It could be that the reason people work less today is that they have access to a better leisure technology. If the quality of the leisure activities available to people improves (making leisure more enjoyable), they should devote more time to leisure. There has clearly been a massive improvement in leisure technology over the past 200 years. Products such as the light bulb, motion picture, automobile, radio, television, airplane, internet, smartphone, video game, and social media have dramatically changed how people spend their leisure time.

It seems very likely that the advent of these products has contributed to the decline in hours worked. However, the timing of the drop in hours worked suggests that income effects have been a more important driver of the decline in work. In particular, much of the decline in hours worked occurred in the 19th century, while leisure technology arguably improved more in the 20th and 21st centuries. Also, the introduction of important leisure technologies was not obviously associated with a substantial drop in hours worked. For example, Fenton and Koenig (2023) found that the roll-out of television in the U.S. had modest effects on labor supply. But more research is needed to assess the relative role of income effects and changes in leisure technology on hours worked.

3 The Covid Recession

The causes of most recessions are complicated and hotly debated. This is not the case when it comes to the recession that started in February 2020. The cause of this recession was clearly the novel coronavirus that came to be known as Covid-19 (or Covid for short). Human-to-human spread of this novel virus was first detected in late 2019 in Wuhan China. The virus was difficult to contain partly because some carriers were asymptomatic while they were infectious. It initially caused an outbreak in the Wuhan area but then spread around the world. In late January and early February of 2020 outbreaks occurred in Italy and the Austrian ski village Ischgl and spread from there throughout Europe and to America eventually resulting in a world-wide pandemic.

At the start of the pandemic, the virus was completely novel to the human population. This meant that people had no specific antibodies to help kill the virus, which contributed to its high virality and morbidity. Little was known about how the virus spread and how severe its morbidity was. Initially, the WHO advised that the virus spread primarily through contact with contaminated surfaces. Later it was recognized that the virus, in fact, spread primarily through respiratory droplets.

In March 2020 massive outbreaks were detected in Northern Italy and New York City. Caseloads, hospitalizations, and deaths grew explosively in these areas. As the month wore on, hospitals became overwhelmed by the flood of infected people. This resulted in many infected people being turned away or not receiving proper care. On March 8th, Italian Prime Minister Giuseppe Conte took the drastic step of announcing a lockdown of much of Northern Italy. The following day he expanded the lockdown to the entire country. In the United States, a stay-at-home order was issued for the San Francisco Bay Area on March 16th followed by a state-wide order on March 19th for California. A number of other countries and U.S. states issued similar orders in the following days and weeks. By mid April, large parts of Western Europe and the United States were under lockdown. In China, Wuhan had been placed under lockdown on January 23rd.

The Covid pandemic eventually killed millions of people worldwide. Figure 6 plots cumulative Covid deaths per million inhabitants over time in six countries. The pandemic hit countries in waves. There was an initial wave in the spring of 2020, another larger wave in the fall of 2020, and several additional waves after that time. These waves were partly associated with new variants of the virus. Early on, novel mutations made the virus more contagious and more deadly. These variants were initially referenced by the country they originated in – the U.K. variant and the South African variant – but were later given official names. The delta variant, which first appeared in late 2020 and became the dominant strain in 2021 was particularly deadly. In 2022, however, the the omicron variant became dominant. This

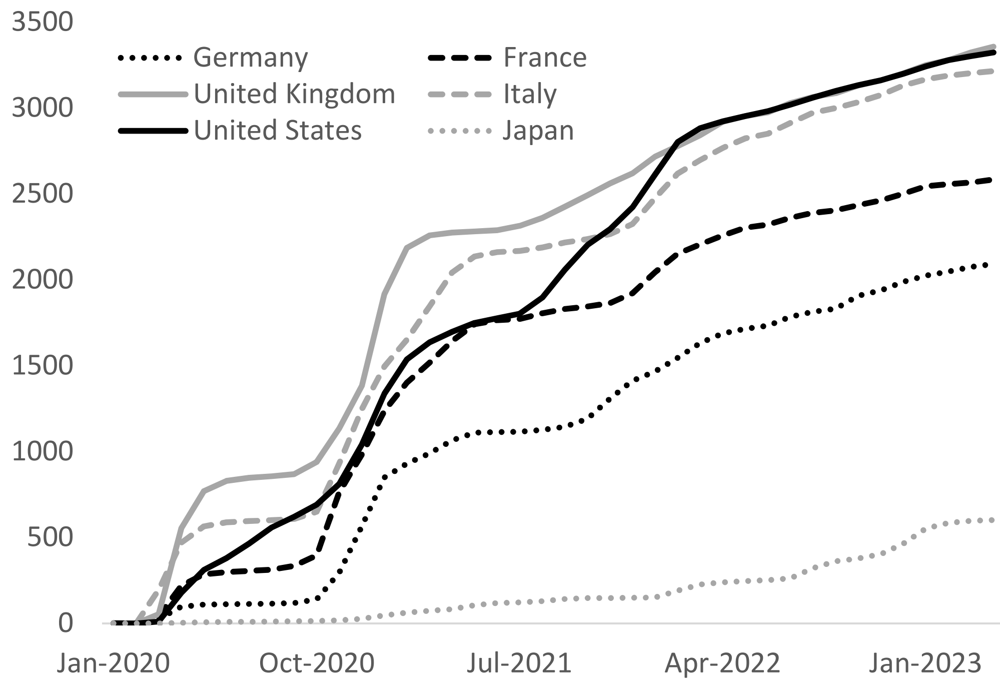

Figure 6: Cumulative Covid Deaths Per Million Inhabitants

Note: This figure plots cumulative Covid deaths per million inhabitants for several countries. The period plotted is January 2020 to April 2023. The data are from Our World in Data.

variant was similarly contagious as delta but substantially less likely to result in hospitalization and death.

Figure 6 shows that death rates in different countries were quite heterogeneous. Countries in East Asia managed to keep death rates much lower than in Europe and the United States. Japan illustrates this pattern in Figure 6. Early in the pandemic, death rates within Europe were extremely variable. The U.K. and Italy were hit very hard in the first wave, while France and especially Germany did better. Later in the pandemic death rates rose substantially in most of European countries. By the spring of 2023, the death rate in the U.K., the U.S., and Italy had reached roughly 3,300 per million. In France, it was 2,600; in Germany, it was 2,100; while in Japan, it was 600.

The onset of the Covid pandemic caused a deep recession. Figure 7 plots the unemployment rate in the United States and Germany. We can see a huge spike in unemployment in March and April 2023 in the United States. Over the two month period from February to April 2023, the unemployment rate rose from 3.5% to 14.7%. This is by far the largest increase in unemployment in the United States since the Great Depression, much larger than the increase in unemployment over the course

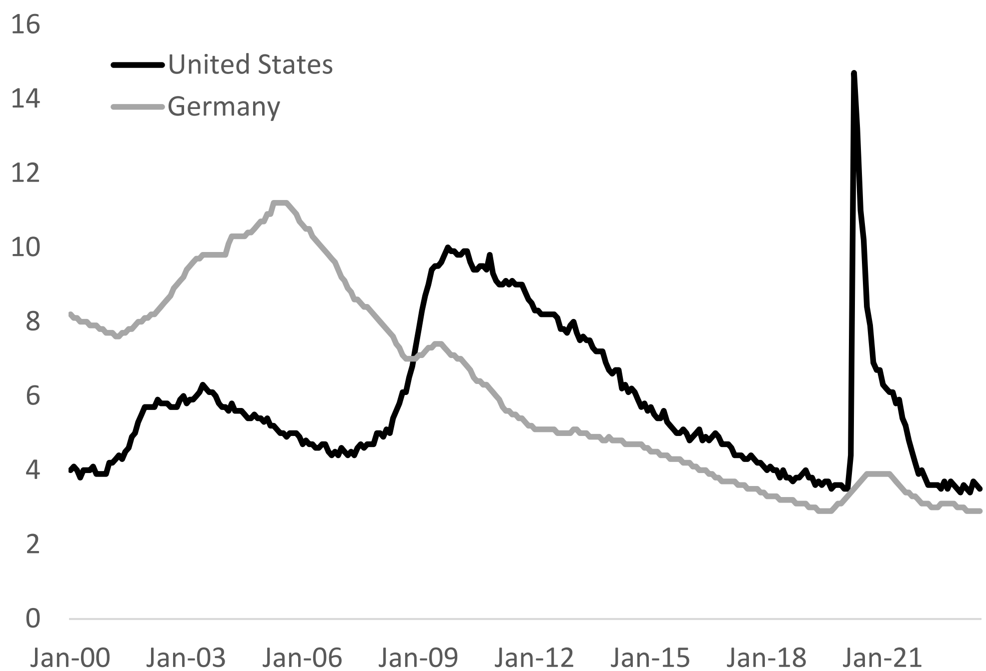

Figure 7: Unemployment in the United States and Germany

Note: This figure plots the unemployment rate in the United States and Germany for the period January 2000 to July 2023. The data are from the OECD Main Economic Database.

of the Great Recession of 2007 to 2009.

In sharp contrast to earlier recessions, the economy recovered from the Covid recession very rapidly. By October 2020, the unemployment rate in the United States had fallen back down to 6.9% and in March 2022 – when the pandemic was still ongoing – unemployment in the U.S. was down to 3.6%. After the Great Recession, it took more than seven years for unemployment to return to pre-recession levels. In the case of the Covid recession, this happened in only two years.

Germany also experienced a deep but brief recession in 2020 as a consequence of Covid. Real GDP in Germany fell by 9% in the second quarter of 2020 before rebounding strongly. In sharp contrast to the United States, the Covid recession did not result in a large spike in unemployment in Germany. This is due to the different way in which the German government responded to the crisis. We will discuss this in more detail below.

3.1 A Model of the Covid Recession

The reason why economic activity fell at the onset of Covid is simple: people wanted to limit contact with each other so as to avoid contracting the virus and avoid spreading it in case they were already infected. This led people to limit their consumption of certain types of goods and services, those that involved contact with other people, such as face-to-face shopping, going to restaurants, air travel, concerts, and sports events. It also led people not to want to go to work in cases where working involved substantial contact with other people. In addition to people's private desires to engage in social distancing, governments in many cases imposed lockdowns that required people to socially distance.

We can incorporate shifts in people's preferences into our labor market model by assuming that the household's utility function becomes

$$\psi_c U(C) - \psi_h V(H)$$

during Covid. Here $\psi_c < 1$ is a preference shifter that represents reduced utility from consumption during the pandemic due to fear of infection, and $\psi_h > 1$ is a preference shifter that represents increased disutility of labor supply due to fear of infection.

We assume as before that households maximize utility subject to a budget constraint. The household's budget constraint is the same as before – equation (3). Steps similar to those we took in section 1.4 then yield

$$\frac{\psi_h}{\psi_c}\frac{V'(H)}{U'(C)} = w$$

as a necessary condition for household optimization. We can rewrite this equation as

$$\frac{V'(H)}{U'(C)} = \psi w, \tag{8}$$

where $\psi = \psi_c/\psi_h < 1$. This equation shows that the preference shifters we have introduced influence behavior in the same way as a fall in the wage on the left-hand-side of equation (8) would influence behavior.

The Covid pandemic was expected to be relatively temporary (and turned out to have relatively temporary effects on the labor market). Recall that temporary changes in wages have limited income effects on labor supply. In this section, we therefore assume for simplicity that $U(C) = C$. This assumption implies that there are no income effects on labor supply. To see this, notice that in this case $U'(C) = 1$ and equation (8) becomes

$$V'(H) = \psi w. \tag{9}$$

Since labor supply is not a function of C, there are no income effects on labor supply.

Recall from chapter XX [production chapter], that labor demand by firms is

$$(1 - a)AL^{-a}K^{a} = w. \tag{10}$$

How does Covid affect this equation? Covid forced firms to adopt a wide variety of new measures to enhance the safety of their workers and customers. These measures reduced firm productivity, i.e., resulted in more inputs being needed to produce a unit of output. A simple way to model this is as a reduction in A in equation (10).

Given these equations, the effect of Covid on the economy can most simply be analyzed using a labor market supply-demand diagram. Figure 8 plots such a figure. Since $V'(H)$ is increasing in H, the labor supply curve – equation (9) – is upward sloping. Since L^{-a} is decreasing in L, the labor demand curve – (10) – is downward sloping in L. (Recall that $L = NH$ and we normalize the population N to one. So, $L = H$.) The solid lines in Figure 8 represent the pre-pandemic labor supply and labor demand curves. The equilibrium in the labor market prior to the pandemic is then at the intersection of these curves (point A in the figure).

When the pandemic strikes, ψ falls from is normal value of 1 (a normalization) to a lower value. How does this affect the labor supply curve? It shifts the labor supply curve back. We can see this by noticing that for any particular value of w, we need a smaller value of of H to make equation (9) hold with $\psi < 1$. The pandemic labor supply curve is the broken gray line in Figure 8.

The pandemic also results in a fall in the value of A in equation (10). This shifts the labor demand curve back. A smaller value of L is needed to make this equation hold for any given value of w when A falls. The pandemic labor demand curve is the broken black line in Figure 8.

Our model therefore predicts that the pandemic leads to a shift in the equilibrium of the economy from point A to point B in Figure 8. This shift leads to a fall in labor (and therefore output): point B is to the left of point A. The effect of the pandemic on wages is however ambiguous. I have drawn Figure 8 such that the pandemic wage is higher than the pre-pandemic wage. Whether this is the case depends on whether the labor supply curve or the labor demand curve shifts back more. Figure 8 is drawn with a larger shift in the labor supply curve.

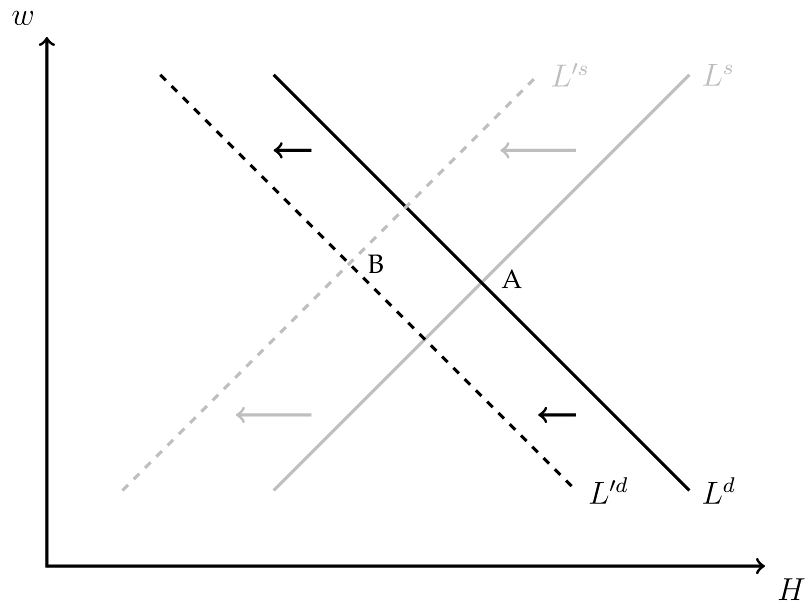

Figure 8: Labor Market Equilibrium Holding Household Consumption Fixed

3.2 Policy Responses to Covid

While it is relatively straightforward to explain why the Covid pandemic resulted in a deep recession, it is not nearly as straightforward to determine how governments should have responded to the pandemic. This was particularly difficult in the 'fog of war' at the beginning of the pandemic, when much was still unknown about its characteristics. But even after the fact, many issues remain debated and disputed.

Governments did respond in a variety of ways: lockdowns were imposed early in the pandemic, school closures were imposed for long periods in some areas, the government supported the development of vaccines, unemployment insurance was extended in a number of ways, the U.S. government sent relief checks to most adults and issued forgivable loans to smaller companies in exchange for retaining workers, it also instituted a number of loan forbearance programs, and the Federal Reserve lowered interest rates and supplied the financial system with substantial liquidity. A number of the policies implemented were novel and quite a few were very controversial. Below I discuss the main groups of policies starting with public health interventions.

3.2.1 Public Health Interventions

Early in 2020, epidemiologists issued dire warnings about the possible death toll from Covid. A particularly influential early assessment came from a group led by U.K. epidemiologist Neil Ferguson. On March 16th 2020, this group issued a report warning that absent mitigation policy or behavioral responses by the population, Covid would kill about 500 thousand people in the U.K. and 2.2 million in the U.S. over the course of the year 2020 (Ferguson et al., 2020). Ferguson and co-authors predicted that the demand for critical care hospital beds would exceed capacity by mid-April and in the "business as usual" scenario, demand for such hospital beds would rise to over 30 times capacity by summer.

To appreciate the potentially tsunami like force of an epidemic like Covid it is useful to introduce a few concepts from epidemiology. The speed with which an epidemic spreads through a population is governed by two numbers: the reproduction number and the generation time. The reproduction number R_t of an infectious disease is the average number of people an infected person infects. The reproduction number varies over time depending on mitigation policies, behavioral responses, and the rising prevalence of resistance in the population. The *basic* reproduction number R_0 is the reproduction number early in an epidemic absent mitigation, behavioral responses, and when virtually no one is resistant. Early estimates of R_0 for Covid were 2.4. An R_0 of 2.4 implies that each infected person on average infects 2.4 people. (Later estimates of R_0 for Covid were even higher (Chitwood et al., 2022).)

The generation time of an infectious disease is the average time it take from when a person is infected until they infect others. The generation time is the sum of the average incubation time and the average infection time of the virus. For Covid, early estimates of the generation time were roughly one week.

These two numbers together imply that early on Covid grew at the staggering rate of 140% per week. That meant that the number of infected doubled about every 5 days. However, at the onset of the epidemic, it was hard to tell how many people were infected. Many infected people were asymptomatic, testing was very limited, and Covid symptoms were similar to flu symptoms for many. Once an outbreak became large enough in a particular location, hospitals started getting flooded with patients making the outbreak easier to detect and salient even to non-experts. This happened in Wuhan, Milan, and New York. But by that point the rate of infection in the community was quite high and it was too late to avoid hospitals becoming overwhelmed.

Public health officials faced the challenging task of issuing warnings and imposing lockdown measures before the epidemic was clearly evident to people in the community. For such policy decisions to work well, public health officials must enjoy a high degree of trust in society. This is similar to when government officials issue evacuation orders before a major hurricane. Such orders must be issued before the storm hits, when the weather is still fine and when there is some uncertainty about the path of the storm. If an order is issued and the storm swerves off, this can undermine trust in the officials issuing the order. In the case of Covid, the living population had no prior experience with a pandemic of a similar seriousness, further increasing the degree to which the population needed to trust officials.

A crucial decision for public health officials at the beginning of a pandemic is whether to aim to suppress the disease or only to slow its spread through the population in order to mitigate its effects. The goal of suppression is to avoid most people contracting the disease. This is the strategy generally employed with Ebola and SARS, but also many diseases for which we have effective vaccines, such as measles and small pox.

For Covid a suppression strategy was difficult due to asymptomatic spread of the virus. Nonetheless, China successfully implemented a suppression strategy with its zero-Covid policy from 2020 to 2023. This policy was possible in China due to the Chinese government's ability to implement both very strict lockdowns and very invasive monitoring of the movements of its population. People's movements were tracked through an app on their mobile phones and they were required to show a green badge on this app indicating no exposure to infected people to move around.

As knowledge of Covid grew, it became clear that suppression was not a viable option in other countries. One factor was that morbidity of Covid was modest for younger people. Early estimates of the rate of hospitalization conditional on symptomatic infection for people in their 40s was 5%. This rate was even lower for younger people (less than 1% for people under 20). Early estimates of the rate of death conditional on symptomatic infection were below 1% for people in their 50s and below 0.1% for people under 40. In contrast, the death rate conditional on symptomatic infection for those over the age of 70 was high. Early estimates put this rate at 5% for people in their 70s and about 10% for people 80 and above. (All these estimates are from Ferguson et al. (2020).)

This modest morbidity for the bulk of the population limited the spontaneous behavioral response of this group. Asymptomatic spread was another important difficulty. These two factors together implied that only extremely stringent actions by the government – such as those taken in China – would result in suppression. Such actions were not considered acceptable in the case of Covid in countries other than China.

The fact that suppression was not a viable strategy (outside China) meant that most people were eventually going to contract the disease. But given this, what was the goal of public health policy? Why not simply let the disease run its course? Two important rationales were articulated for policies aimed at slowing the spread of Covid. First, unabated spread of the disease would overwhelm the healthcare system. This would results in many sick people not receiving the medical care they needed to fight off the virus and thus raise the death rate from the virus. Interventions that slowed the spread of the virus could "flatten the curve" and in this way spread the epidemic over a longer time period preventing the healthcare system from being overwhelmed at any given point in time.

Figure 9 illustrates this by plotting the evolution of infection rates in society for different assumptions about the reproductive rate. Recall that the uncontrolled reproductive rate for Covid was extremely high, perhaps as high as 3. Figure 9 shows that without any mitigation this high reproductive rate would result in infections going "through the roof" (literally off the figure). If a combination of policy and private behavioral responses lowered the reproductive rate to 2 or lower values, the infection curve is flattened out. With enough of a response, it is flattened enough that the health care system is never overwhelmed.

A second potentially important rationale for public health policy aimed at slowing the spread of the virus was to buy time for vaccine development. If a sizable fraction of the population could avoid contracting the virus before vaccines were developed, this could potentially save a large number of lives (and avoid a great number of serious but non-fatal cases). Once vaccinated, people might become immune to the disease or the vaccine would help protect against severe infection.

This is what happened during Covid. Vaccines for Covid were developed at record speed and became available in December of 2020 and early 2021 at which time about 70% of the U.S. population had managed to avoid infection (Chitwood et al., 2022). The U.S. government supported the rapid development of multiple vaccines through Operation Warp Speed. The usual lengthy process of development and trial of vaccines was accelerated by fast-tracking trials and the building of large-scale manufacturing capacity while trials were still underway. Atkeson (2023) estimates that the Covid vaccines saved roughly 750,000 lives in the United States alone through mid-2023.

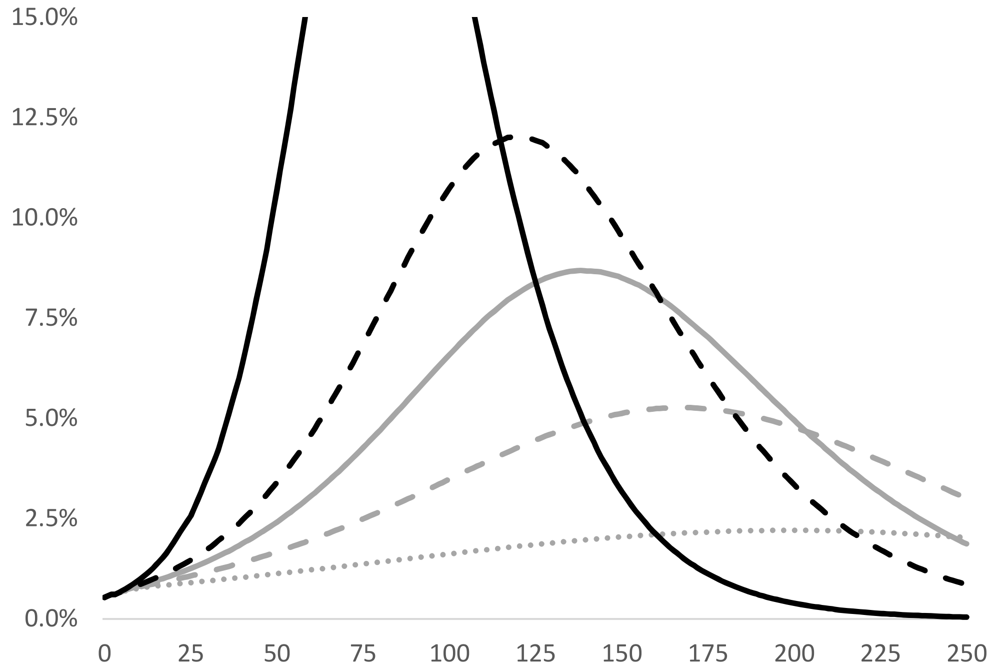

Figure 9: Dynamics of Infection Rates with Different Reproductive Rates

Note: This figure plots the dynamics of the infection rate in a society for different reproductive rates. The solid black line is for $R = 3$, the broken black line is $R = 2$, the solid gray line is $R = 1.75$, the broken gray line is $R = 1.5$, and the dotted gray line is $R = 1.25$. The horizontal axis is measured in days. These simulations are from the SEIR model of Atkeson (2020).

For the reasons discussed above, slowing the spread of an epidemic with public health interventions can potentially save a large number of lives. But such interventions also have important costs which makes them controversial. The least invasive and therefore least controversial action the government can take to slow the spread of an epidemic is to issue public health warnings regarding the epidemic. Warnings can be immensely important in that they can help the public adjust their behavior of their own accord. Some might argue that public warnings are unnecessary since the private sector will produce information on its own. But information has a public good element and is therefore likely underproduced by the private sector. For this reason, public warnings are an important element of policy in the case of epidemics as they are in the case of weather events and natural disasters.

During the Covid epidemic, public health officials in many countries went far beyond issuing warnings. To further slow the spread of the virus, they imposed quite severe restrictions on people's freedom. As we discussed above, stay-at-home orders and other forms of lockdowns were imposed. Schools were closed, in some cases for long periods. Even once the most severe lockdowns were relaxed, restrictions were placed on large gatherings, people were ordered to socially distance, and they were ordered to wear masks.

These measures had high costs. They prevented economic activity from occurring. They prevented children from attending school. They prevented people from attending religious ceremonies. They prevented families from gathering for important events. They prevented loved ones from visiting sick and dying relatives. They outlawed parties, dates, and other socializing. And the list goes on. Most fundamentally, they encroached very seriously on people's civil liberties. In addition, the measures were broad brush. They therefore prevented all manner of activity that might have high benefit but low cost in terms of increased transmission.

In the next chapter [Markets chapter], we will discuss in detail the idea (which is very prominent in economics) that allowing people to choose freely for themselves how to behave can often lead to good outcomes. This idea challenges those that advocate severe restrictions on peoples liberty to provide a justification for such restrictions. In the case of lockdowns and other social distancing measures during an epidemic, the most natural justification is externalities.

Socializing increases the probability that a person becomes infected with the disease. This cost is borne by each person (except for medical costs that are borne by society or their insurance company). But socializing also increases the probability that a person infects other people. This cost is entirely borne by others. Since each person does not bear the full cost of their own socializing, they will not reduce their socializing as much as is socially optimal. This provides a rationale for government action to enforce social distancing during a serious epidemic.

3.2.2 Fiscal Relief Payments and Forbearance

In most recessions, economic activity is inefficiently low and a major goal of government policy is to stimulate economic activity. The Covid recession was not this type of recession. During Covid, the epidemic made it efficient to socially distance and refrain from many types of economic activity. This means that the *efficient level of economic activity* fell during Covid. The reduction in economic activity during Covid was, thus, not primarily a fall in output below the efficient level, but rather a fall in the efficient level of output. The recession during Covid was, in other words, a recession by design as opposed to a pathology that needed to be fixed.

Fiscal policy during recessions is usually aimed at stimulating the economy. But since the primarly problem during Covid was not that output was below its efficient level, stimulus was not the primary rationale for fiscal policy. The primary problem during Covid was that some sectors of the economy – those involving contact-intensive activity – were hit extremely hard, while other sectors were less affected or even benefited from the shift in demand away from contact-intensive activity. This meant that some workers lost their jobs, while others were relatively unaffected.

The primary problem was thus one of insurance. People were not insured against the Covid shock. The aim of much Covid economic policy was to mimic insurance by transferring resources from those that were relatively unaffected (including people in the future) to those that were hard hit. Absent this type of policy, there was a risk that the large fall in income of those hard hit might have important knock-on effects since these people would then reduce demand and default on various commitments (e.g., rent and mortgage payments). This could potentially have led output to fall further than the already low efficient level of output.

An important challenge was how to effectively implement the goal of providing such insurance. In late March 2020, the U.S. Congress passed the CARES act which provided over $2 trillion in various forms of fiscal relief. One part of the CARES act was fiscal relief payments to individuals ($1,200 per adult and $500 per child). Another part was the Paycheck Protection Program (PPP), which provided forgivable loans to small businesses that maintained employment during the pandemic. Legislation passed later in 2020 provided a second round of fiscal relief payments to individuals in December 2020 (this time $600 per adult and child). Finally, the American Rescue Plan act of 2021 provided a third round of fiscal relief payments in March 2021 ($1,400 per adult and child). In total, roughly $800 billion were sent out as fiscal relief payments to individuals and roughly $700 billion were spent on the PPP. Together, these payments amounted to roughly 7% of U.S. GDP.

An advantage of fiscal relief payments is that they are simple to implement. A disadvantage is that they are poorly targeted. Everyone received these payments (except those at the very top of the income distribution) whether their livelihood was adversely affected by Covid or not. One way to think about these payments is as a transfer from people in the future – who must service and pay down the debt the government incurred to make these payments – to people during the pandemic. In this sense, the fiscal relief payments transferred funds from those less affected to those more affected by Covid. But a number of policies discussed below were better targeted than these fiscal relief payments.

The combination of large fiscal relief payments and depressed spending during the pandemic led American households to build up a large amount of extra

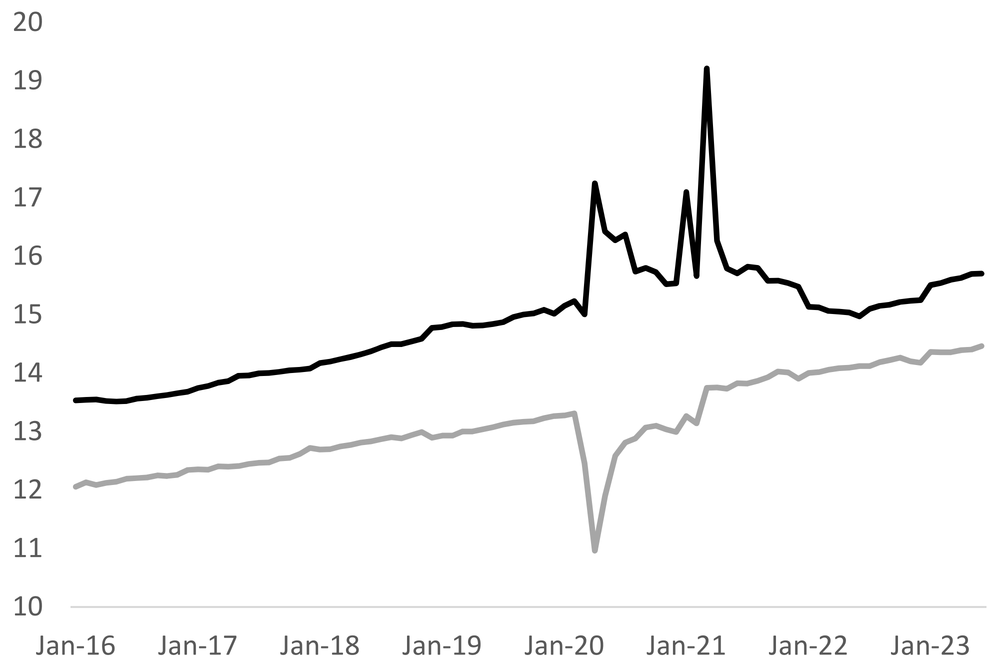

Figure 10: Disposable Income and Consumer Expenditures in the United States

Note: This figure plots real personal disposable income and personal consumer expenditures in the United States from January 2016 to June 2023. Both series are in trillions of 2012 dollars. The data are from U.S. Bureau of Economic Analysis.

savings. Figure 10 plots aggregate disposable income and consumer expenditures in the United States over the course of the pandemic. Disposable income actually rose (the three spikes are due to the three rounds of fiscal relief payments), while consumer expenditures fell sharply early on during Covid. The large increase in the difference between these two lines represents extra savings. The fact that households were flush with cash likely contributed to the speedy recovery from the Covid recession – the recovery was much more rapid than for earlier recessions – but may have also contributed to the bout of inflation that followed in 2021 through 2023.

A second form of insurance implemented during Covid was forbearance policies. The CARES act mandated that mortgage lenders provide forbearance to borrowers who claimed hardship due to Covid. Forbearance allowed most borrowers to delay mortgage payments for up to 18 months. The U.S. government provided forbearance on student loans from March 2020 to September 2023 (along with a 0% interest rate during this period). Certain states – such as California – also enacted eviction protection for renters during Covid as well as rental assistance for certain renters. Together, these policies provided substantial amounts of insurance for many households during Covid. But as with fiscal relief payments, these policies were not very well targeted since it was easy to claim hardship and such claims were not verified.

A third form of insurance enacted as a part of the CARES act was that people were allowed to withdraw up to $100,000 from their 401(k), 403(b), and IRA retirement accounts without incurring an early withdrawal penalty. This provision did not transfer money between people. Rather it allowed people to transfer their own resources from their retirement years to the Covid period. Typically, people must pay a hefty penalty for withdrawing funds early from their retirement accounts – typically a 10% penalty. The logic for waiving these penalties was that Covid was a time of particular hardship.

3.2.3 Unemployment Insurance

The United States and many other countries have long-standing unemployment insurance (UI) systems. These systems provide workers with important insurance when they loose their jobs. In most U.S. states, the regular UI system provides benefits for up to 26 weeks at a replacement rate of roughly 50% of prior pay. The system, however, has a built in feature that makes it more generous during economic downturns. This Extended Benefits program increases the maximum number of weeks of benefits workers have access to from 26 weeks to 39 week and in some cases 46 week when the unemployment rate is high and rising (i.e., at the onset of recessions). The logic for this is that jobs are hard to find at these times and workers therefore need more insurance than during normal times.

In addition to the Extended Benefits program (which has been in effect since 1970), the federal government has enacted special additional programs in every recession since 1973. In some cases, the maximum duration of benefits have become extremely long. In the aftermath of the Great Recession of 2007-2009 the maximum benefit duration reached 99 weeks.

The CARES act expanded unemployment insurance during Covid in three ways. First, it extended the maximum duration of benefits. With later additions, maximum benefit durations eventually rose to 79 weeks during Covid. Second, it expanded the scope of the UI system to former self-employed workers, contract workers, and gig workers. Finally, it provided an additional $600 per week in benefits. This last provision massively increased the average replacement rate of workers on UI. In fact, for many unemployed, the benefits they received were higher than their prior earnings. Ganong, Noel, and Vavra (2020) estimate that 76% of eligible workers received higher benefits than their prior wages and that the median replacement rate was 145%.

An important advantage of UI is that it is highly targeted. The people that receive benefits are those that have lost their jobs. The primary disadvantage of UI is that it weakens the incentives of workers to find new jobs. This disadvantage made the large expansions of UI seen in the Covid recession and the Great Recession quite controversial. Critics argue that they prolong the recession. In the summer of 2021, roughly half of U.S. states terminated additional benefit payments (which were $300 per week at that point) before these benefits were set to expire, and many of these states terminated all the Covid expansions to their UI systems. The stated goal of this early termination was to accelerate job growth.

As with many controversial policy proposals, a key difficulty is that it is hard to estimate the effect that extended UI has on unemployment. The reason for this is reverse causality. UI is extended *because* unemployment is high and rising. To estimate the effect of UI extensions on unemployment, researchers must control for the fact that the extensions occur when other factors are pushing unemployment up.

My own research on this topic uses variation in rules adopted in different U.S. states to identify the effect of UI on unemployment (Acosta et al., 2023). We find that extending the duration of UI by 13 weeks (i.e., from 26 weeks to 39 weeks) increases unemployment by about 0.3 percentage points. These are moderately sized effects (neither trivial nor huge). We also find that once the maximum duration of benefits is already quite long, further extensions have virtually no effect on unemployment. This is likely due to the fact that only a small fraction of unemployed people stay unemployed for long enough for such extensions to matter.

Ganong et al. (2023) estimate the impact of the increase in benefits levels during Covid (the extra $600 per week) on household spending and on job finding. They find that these enhanced UI benefit levels have large effects on household spending. Household spending rose by between 27 and 42 cents per dollar of extra benefits within a month. In contrast, they find that the extra benefit levels have small effects on job finding rates.

While the U.S. relied heavily on its UI system for social insurance during Covid, other countries made use of other types of policies. A particularly interesting example is the German Kurzarbeit (German for "short work") program. This program encourages firms to reduce worker hours during downturns rather than laying worker off. Under the program, workers receive 60% of their pay for hours not worked and full pay for hours worked. The German government partially funds the payments for hours not worked. The idea behind the program is to allow firms and workers to avoid costs associated with separation, hiring, and training during temporary downturns. As we saw in Figure 7, this program dramatically limited the increase in unemployment in Germany during Covid (and also during the Great Recession of 2007-2009). Several other European countries have instituted similar programs in recent years.

3.2.4 Monetary Policy

As information about the spread of Covid began to accumulate, financial markets reacted violently. Stock markets fell sharply, as did the price of other risky assets. Market liquidity deteriorated and volatility skyrocketed. These were all symptoms of a rapid and massive flight to safety among investors who feared that Covid would cause a deep recession and severe disruption in financial markets.

Figure 11 illustrates this by plotting the price of the S&P500 stock price index. The S&P500 fell by 34% between February 19th 2020 and March 23rd 2020. On that day – March 23rd 2020 – the Federal Reserve announced that it would purchase Treasury securities and agency mortgage-backed securities (MBS) "in the amounts needed to support smooth market functioning." This effectively meant the Fed stood ready to do whatever it took to maintain the smooth functioning of financial markets. The day after this announcement, the S&P500 rose by 9.4% – one of the largest single-day increases in history – and after two more days it had risen by 17.6%.

The March 23rd announcement was the most important of a series of announcements of massive support by the Federal Reserve for the flow of credit in the economy. Earlier in the month, the Fed had lowered interest rates to zero in two large steps and announce large (but not unlimited) purchases of Treasuries and agency MBS. In rapid succession, it also announced an alphabet soup of different lending programs each geared to a different segment of the the financial market. CPFF supported the commercial paper market; MMLF supported money market mutual funds, PDCF supported primary dealers; PMCCF and SMCCF supported lending to large corporation; MSLP supported lending to small and mid-sized corporation; MLF supported lending to state and local governments; and TALF supported lending to households and small businesses. The Fed also massively expanded its swap lines with foreign central banks due to heightened demand for dollars globally. Mil-

Figure 11: S&P500 Stock Price Index

Note: This figure plots he S&P500 stock price index from July 1st 2019 to December 31st 2020. The data are from the Federal Reserve Economic Database (FRED).

stein and Wessel (2021) provide a more detailed discussion of these programs. Many of them had originally been created during the Great Recession of 2007-2009. The speed with which the Fed could react during Covid owed a great deal to the experience it had gained in that earlier crisis.

The complexity of the various programs introduced by the Fed should not blind us to the relatively simple economic logic that underlay the Fed's response. The Fed was acting as a lender of last resort. For reasons that we will discuss in much greater detail in chapter XX [Money and Banking chapter], financial markets can seize up during a crisis and this can result in major economic disruptions. At such times, the central bank has a special ability to step in and lend – sometimes in massive quantities – so as to avoid financial disruption. This is exactly what the Fed did in March and April of 2020. Arguably, these actions arrested what looked to become a colossal financial crisis.

Table 1: Output Per Working-Age Person: Differences Versus the U.S.

Country	1970	1985	2000	2015
Germany	-39	-42	-27	-17
France	-32	-35	-29	-26
Italy	-51	-41	-29	-35
United Kingdom	-36	-35	-28	-25

Notes: The table reports percentage differences versus the United States in output per working-age person. Data on output and capital are from the Penn World Tables (version 10.1). Data on hours worked and the population of working age are from the OECD. Working age is defined as age 15 to 65.

4 Prosperity and Depression: The U.S. versus Europe

In a lecture titled "Prosperity and Depression," given in 2002, economist Edward Prescott described France as being depressed relative to the United States. The reason Prescott gave for describing France this way was that output per working-age person in France was about 30% lower than in the the United States. Table 1 presents estimates of output per working-age person over time for the four largest economies of Europe – Germany, France, Italy, and the United Kingdom – relative to the United States. In 2000, output per working-age person was about 30% lower than in the U.S. in all four of these countries. By Prescott's definition, they were, therefore, all depressed.

Considering the estimates in Table 1 over time, we see that these large European economies have generally been catching up to the U.S. in terms of output per working-wage person: the differences versus the U.S. were larger in 1970 than they were in 2015. But the extent to which they have been catching up has been uneven, and, apart from Germany, progress between 2000 and 2015 was modest.

Prescott's use of the term 'depressed' conjures up images of the Great Depression, when output in the United States and various other countries fell by roughly 30% over the span of a few years. The Great Depression is generally considered a major calamity during which the economic system of capitalist countries malfunctioned in a major way. Should one consider the situation in Germany, France, Italy, and the U.K. over the past few decades as being due to a similarly major malfunction of these economies?

4.1 Accounting for Differences in Output Per Working-Age Person

To gain a better understanding of what might be causing the large differences in output per working-age person in Europe versus the U.S., Prescott employed a method which is alternatively referred to as *development accounting* or *level accounting*. This method takes a simple production function as it starting point:

$$Y_{it} = A_{it}^{1-\alpha} K_{it}^{\alpha} H_{it}^{1-\alpha} \tag{11}$$

where Y_{it} denotes output in country i at time t, K_{it} denotes the capital stock in country i at time t, H_{it} denotes hours worked in country i at time t, A_{it} denotes productivity in country i at time t, and α is a parameter.

We can rewrite this production function in a form that allows for a particular decomposition of differences in income across countries. A justification for this particular decomposition is given in Prescott (2002). Since Prescott's argument involves concepts introduced later in this book, I will not present it here. But we, nevertheless, follow his analysis since the decomposition he chooses does have important advantages. We start by dividing both sides of equation (11) by Y_{it}^{α}. This yields:

$$Y_{it}^{1-\alpha} = A_{it}^{1-\alpha} \left(\frac{K_{it}}{Y_{it}}\right)^{\alpha} H_{it}^{1-\alpha}.$$

Next we divide both sides by $N_{it}^{1-\alpha}$, where N_{it} denotes the working-age population in country i at time t. This yields:

$$\left(\frac{Y_{it}}{N_{it}}\right)^{1-\alpha} = A_{it}^{1-\alpha} \left(\frac{K_{it}}{Y_{it}}\right)^{\alpha} \left(\frac{H_{it}}{N_{it}}\right)^{1-\alpha}.$$

Next we take natural logarithms on both sides, divide through by $1-\alpha$, and use lower case letters to denote per working-age population versions of corresponding upper case letters – i.e., $y_{it} = Y_{it}/N_{it}$. This yields:

$$\log y_{it} = \log A_{it} + \frac{\alpha}{1-\alpha} \log(k_{it}/y_{it}) + \log h_{it}.$$

This last equation shows that we can decompose output per working-age person into three factors: 1) productivity, 2) the capital-output ratio, and 3) hours worked per working age person. Since this holds for each country i, it holds for the difference between two countries i and j as well:

$$\underbrace{\log y_{it} - \log y_{jt}}_{\text{Output Difference}} = \underbrace{\log A_{it} - \log A_{jt}}_{\text{Productivity Factor}} + \underbrace{\frac{\alpha}{1-\alpha}\left(\log(k_{it}/y_{it}) - \log(k_{jt}/y_{jt})\right)}_{\text{Capital Factor}} + \underbrace{\log h_{it} - \log h_{jt}}_{\text{Labor Factor}}.$$

Table 2: Decomposition of Output Differences with the U.S.

	Log Difference versus United States × 100			
	GDP	Productivity	Capital	Labor
Panel A: Germany				
1970	-50	-80	10	20
1985	-55	-56	6	-5
2000	-32	-8	2	-27
2015	-19	-20	13	-12
Panel B: France				
1970	-38	-50	5	6
1985	-43	-27	6	-23
2000	-34	-1	-2	-31
2015	-31	-28	21	-23
Panel C: Italy				
1970	-71	-65	-1	-5
1985	-53	-35	0	-18
2000	-35	-21	7	-21
2015	-43	-60	32	-15
Panel D: United Kingdom				
1970	-44	-48	-2	6
1985	-43	-20	-14	-8
2000	-34	-13	-7	-13
2015	-29	-44	16	-1

Notes: The table reports log differences versus the United States multiplied by 100 for GDP per working-age person and three determinants. Data sources are the same as in Table 1.

Table 2 plots this decomposition of differences in output per working-age person for Germany, France, Italy, and the U.K. relative to the United States. The table presents this difference for four years: 1970, 1985, 2000, and 2015. The first column plots $\log y_{it} - \log y_{jt}$, while the next three plot the productivity factor, the capital factor, and the labor factor in the last equation above this paragraph. Notice that there is a subtle difference between the output differences in Table 1 and in Table 2. Table 1 reports percentage differences, while Table 2 reports log differences. These are approximately equal for small differences. (You can see this by taking a first-order Taylor series approximation of $\log x$ around x_0.) But the discrepancy is non-trivial for log differences as large as those discussed in this table.

The decompositions for the four countries in Table 2 shares certain similarities, but they also differ substantially in some cases. For concreteness, we focus on France. In 1970, output in France was 38 log points lower than in the United States. Thirty year later, in 2000, this difference was almost exactly unchanged (34 log points). So, over this 30 year period, France did not catch up to the U.S. in terms of output per working-age person almost at all.

However, when we look at the proximate causes of this large output difference, we see that they changed dramatically over this period. In 1970, the entire output gap was explained by low productivity. Actually, the French worked more than Americans in 1970, and the capital-output ratio in France was higher as well. By 2000, productivity in France had completely caught up with productivity in America. But hours worked per working age person in France had collapsed relative to the United States by 37 log points.

Comparing 1970 and 2000 for the other three countries in Table 2 yields a similar, if somewhat less extreme, story. In all cases, the gap in productivity fell dramatically between 1970 and 2000. And in all cases, hours worked per working age person fell substantially. In Germany, the fall in hours worked relative to the U.S. was even larger than in France over this period (47 log points). In Italy and the U.K. it was smaller (16 and 19 log points, respectively).

Prescott's analysis focused on France. He viewed the low level of work in France as a major malady. Below, we consider this idea in more detail. But before moving to that analysis it is interesting to consider what has happened since Prescott delivered his lecture. The year 2000 seems have been something of a higher water mark for productivity in Europe relative to the United States. By 2015, the gap in productivity had increased a great deal in all four countries. Likewise, between 2000 and 2015, hours worked per working age person in these four European countries converged with the United States, in some cases quite substantially.

4.2 Hours Worked in Europe versus the United States

It is useful to take a closer look at the evolution of hours worked in Europe versus the United States. Figure 12 plots hours worked per working-age person in the United States, France, Germany, and Italy between 1950 and 2020. Consistent with our results in Table 2, we see that Europeans worked substantially more than Amer-

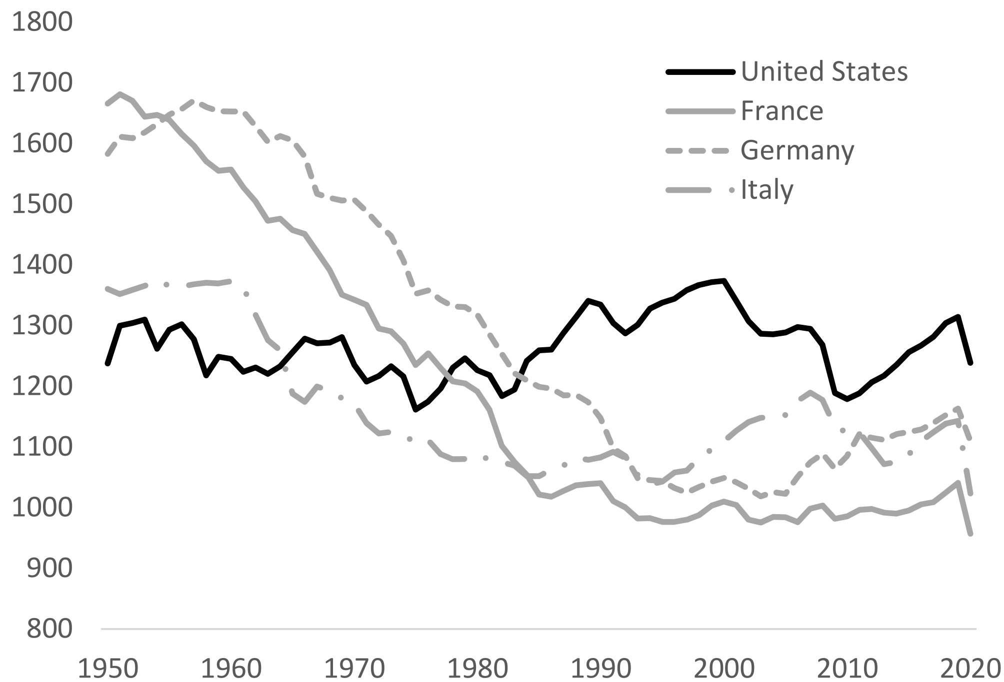

Figure 12: Annual Hours Worked per Working-Age Person

Note: This figure plots annual hours worked per person of age 15-65 in the United States, France, Germany, and Italy from 1950 to 2000. Data on total hours worked are from the Conference Board Total Economy Database. Data on the population aged 15-65 is from the OECD Employment and Labor Force Statistics database.

icans in the early post-World War II period. Germans and the French on average worked roughly 1630 hours per year in the 1950s, while Italians worked on average 1370 hours per year. Americans worked only 1280 hours per year at this time. Germans and the French, therefore, worked about 27% more than Americans at the time. This difference reflects both differences in the fraction of people working and hours per worker.

The subsequent decades saw a dramatic reversal in this pattern. Between 1960 and the mid-1980s, hours worked in Europe fell sharply, while they stayed relatively unchanged in the United States. By the 1990s, the French worked only 990 hours on average. Germans and Italians worked only slightly more at 1060 and 1070, respectively. Hours worked per person had therefore fallen by almost 40% in France, 35% in Germany, and 22% in Italy. In sharp contrast, hours worked in the United States had inched up slightly to 1330.

We can gain further insight by decomposing the difference in hours worked per working age person into several components. Let's denote hours worked per worker at time t by H_t, the number of people employed at time t by E_t, the number of people in the labor force at time t by L_t, and the population of working age at time t by N_{At}. Recall that the labor force is the number of people employed plus those that are not employed but actively seeking employment (i.e., those unemployed). With these definitions we can write:

$$\frac{H_t E_t}{N_{At}} = H_t \frac{E_t}{L_t} \frac{L_t}{N_{At}},$$

where the left-hand side is hours worked per working-age person at time t, $H_t E_t / N_{At}$, and on the right-hand side we have hours worked per worker at time t, H_t, the employment rate at time t, E_t/L_t, and the labor force participation rate at time t, L_t/N_{At}. We can now take natural logarithms of this equation to get

$$\log\left(\frac{H_t E_t}{N_{At}}\right) = \log H_t + \log\left(\frac{E_t}{L_t}\right) + \log\left(\frac{L_t}{N_{At}}\right).$$

This decomposition holds for all times, which means we can write:

$$\Delta \log\left(\frac{H_t E_t}{N_{At}}\right) = \Delta \log H_t + \Delta \log\left(\frac{E_t}{L_t}\right) + \Delta \log\left(\frac{L_t}{N_{At}}\right), \tag{12}$$

where the Δ (the Greek letter capital delta) denotes change from one period to another, i.e., $\Delta \log H_t = \log H_t - \log H_s$ where s is a different time period.

Table 3 reports estimates of the decomposition of hours worked per working-age person given by equation (12) for France and the United States. Focusing for concreteness on the difference between 1970 and 2000, we see that the vast majority of the drop in hours worked per working-age person in France resulted from a fall in hours worked per worker. This reflects factors such as a shortening of the work week and longer vacations. The employment rate did also fall non-trivially over this period. This reflects the substantial increase in unemployment in France between 1970 and 2000. But increased unemployment is nevertheless a relatively small factor in the overall picture. The labor force participation rate actually increased modestly over this period. Labor force participation of men fell, but labor force participation of women rose faster.

The contrast between France and the U.S. is quite striking. Hours worked per worker also fell in the U.S. between 1970 and 2000, but by much less than in France. The employment rate (and therefore also the unemployment rate) was almost exactly unchanged. Finally, the labor force participation rate rose much more in the U.S. than in France reflecting a much larger entry of women into the labor force in the U.S. than in France over this period.

Table 3: Decomposition of Hours Worked Per Working-Age Person

	Log Changes since 1970 × 100			
	HE/N_A	H	E/L	L/N_a
Panel A: France				
1985	-29	-19	-7	-3
2000	-28	-25	-6	3
2015	-29	-27	-9	7
Panel B: United States				
1985	5	-4	-2	11
2000	13	-4	1	16
2015	6	-7	0	13

Notes: The table reports log differences versus 1970 multiplied by 100 for hours worked per working age person (HE/N_A), hours worked per worker (H), the employment rate (E/L), and the labor force participation rate among working age people (L/N_A). Data are from the OECD.

Figure 13 decomposes the difference in employment rates between Europe and the United States by age groups for the year 2000. Interestingly, we see that employment rates in Europe were quite similar for "prime-aged" workers (those between 25 and 54). There are, however, very substantial differences in employment rates between Europe and the U.S. for young workers and for older workers.

4.3 Why Do Europeans Work So Little?

These facts raise the following question: why do Europeans work so little? When I ask my students this question, they propose a wide variety of explanations. Here are a few of the more common ideas:

- Differences in preferences and culture
- More generous social safety net (unemployment insurance, health care, etc.)
- Higher minimum wages
- Unions
- Labor market reguations (e.g., higher hiring and firing costs)
- Higher taxes

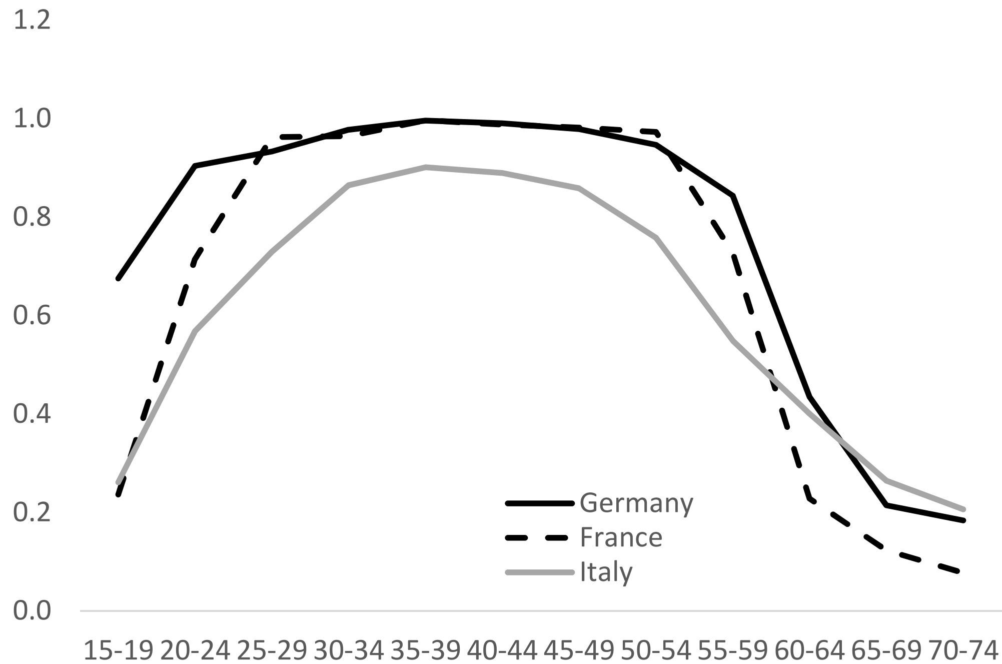

Figure 13: Employment Rates by Age Relative to the U.S.

Note: This figure plots employment rates by age groups for Germany, France, and Italy relative to the United States for the year 2000. .

There is a substantial literature that discusses each of these explanations. Given the complexity of coming up with definitive evidence on this subject, it is perhaps not surprising that there is no consensus. My own reading of the literature is that several (perhaps all) of these explanations play some role.

Let's consider preferences and culture first. One view under this heading is that Europeans – lets consider the French for concreteness – simply have a stronger preference for leisure. Relative to the past, many people in France are extremely well of financially. Why work so much? Better to enjoy the good life, take long vacations in the summer, go home early from work so as to have time to spend with the family. Perhaps this is simply a more "enlightened" or "civilized" way to live given the high standard of living France has attained than the endless toil of Americas.

Of course, this perspective is not universally held. Americans, in particular, may not agree that their way of life is less civilized than that of the French. In some cases, they may go so far as to see the French as lazy. Arguing about such differences in preferences is difficult. Some would say it is pointless.

The preferences and culture view must, however, contend with the fact that the French worked more than Americans in the 1950s and 60s. So, if you view the French as being lazy today, you must acknowledge that they have not always been lazy. If you view the French as enlightened today, you can, however, argue that their hard work 70 years ago was due to their relative poverty at that time: perhaps income effects on labor supply are stronger among the French than among Americans. I.e., perhaps the French have preferences that are closer to those that Keynes believed we all had when he wrote *The Economic Possibilities of Our Grandchildren*.

Another aspect of culture that is relevant in this context is attitudes towards female employment. Female employment has risen much faster in the United States than in Europe contributing to a faster increase in the labor force participation rate in the United States (last column of Table 3). There are likely several reasons for this difference. But cultural norms and discrimination are likely among the most important. Cultural norms regarding female employment have undergone a sea change in the United State (and parts of Northern Europe) but less so in France and Southern Europe.

Yet another preference related idea is that our preferences regarding work versus leisure may be dependent on the actions of others in a way that creates a coordination problem. Perhaps we all want to work less (e.g., take longer vacations) but only if others do this as well. Why might this be? One reason is simply that any one person's leisure may be complementary with other people's leisure: it is more fun to be on vacation when others we know are also on vacation, since in that case we can spend time with them. Also, if everyone takes a large amount of leisure, the economy will create better activities and services geared towards those that are enjoying leisure. This can increase each person's utility from leisure.

A second reason why one person's leisure may be complementary with the leisure of others is signaling. Many workers want to signal to their employers that they are "good" employees (hard working, etc.). One way to do this is to work more (take few vacation days, arrive early, leave late, etc.). But if others are doing this, each person needs to do even more to be able to send the right signal. This dynamic may lead to an inefficient "rat race" equilibrium where everyone is working really hard just because everyone else is working really hard.

Coordination problems like these can be hard to solve because they require collective action. An interesting idea – discussed, for example, in Alesina, Glaeser, and Sacerdote (2005) – is that unions in Europe may have helped solve this coordination problem. Unions can collectively bargain for more holidays and mandated vacation days. Alesina, Glaeser, and Sacerdote present estimates suggesting that about 40% of the difference in hours worked per worker (our H in Table 3) between France and the U.S. in the early 2000s was due to differences in weeks worked in the year. At the time, the French spent 7.0 weeks on holidays and vacations on average, while American spent only 3.9 weeks on holidays and vacations. France had 16 national holidays and 25 days of mandated vacation time, while the U.S. had 12 federal holidays and no mandated vacation time. Unions can also play a role in bargaining for a more generous pension system and an early retirement age. As we saw in Figure 13, some of the difference between Europe and U.S. is due to lower employment rates among those aged between 55 and 64.

But unions are not the only institutions capable of collective action. The political process is another avenue for collective action. In fact, the socialist government elected in France in 1982 played an important role in reducing hours and increasing mandatory vacation time. Another socialist government in the late 1990s passed "35 hour workweek" laws. The U.S., obviously, has a political process that could solve a coordination problem regarding work hours if this were demanded with enough force by enough voters. This has not occurred in the United States for some reason. Part of the reason may be that strong unions change the political dynamics of a country in important ways.

Another idea strongly associated with unions in Europe is the idea of "work less – work all." This is the idea that there is a limited amount of work available and this work should be spread among more workers by reducing the amount of work each worker does, e.g., by shortening the workweek. This idea was particularly prominent in the 1970s following the first OPEC oil shock, when unemployment was high in Europe and the United States. It is widely believed to have played a major role in shaping the demands of unions and left wing political parties in Europe for shorter working hours. The most prominent case being the 35 hour workweek laws in France mentioned above.

Most economists are extremely skeptical of the wisdom of "work less – work all" ideas. They do not think of work as a scarce resource. They think of labor as a scarce resource. If there is not enough work for those that want to work, this is a signal that the labor market is not working as it should. This could be for a number of reasons. One is that wages in some segments of the labor market are "too high" – meaning that they are above the marginal product of a significant fraction of workers. High minimum wages can lead to this problem. Minimum wages in Europe have indeed been much higher than in the U.S. at most times. These high minimum wages may play a role in explaining why young workers – some of which may be relatively low productivity due to their inexperience – have a hard time finding work in Europe. But high minimum wages are unlikely to explain changes in hours worked per worker (H), which as we saw in Table 3, explains the bulk of the drop in hours in France.

Another reason why the labor market may not clear is that non-wage costs of hiring workers are sufficiently high that they make it unprofitable for firms to hire workers. Such costs include payroll taxes, but also costs associated with various types of regulations (e.g., safety regulations). These costs also include the costs associated with firing a worker. Firing costs can be extremely high in some European countries. Large firing costs can deter firms from hiring workers in the first place. Large firing costs can result in particularly sever problems when the economy is hit by aggregate shocks such as the oil shocks of the 1970s or sectoral shocks that imply that it is efficient for certain sectors to shrink while others expand. In such cases, the rigidity associated with high firing costs can lead to low employment. This situation is sometimes referred to as Eurosclerosis.

A third potential reason why labor markets may not clear is if workers have weak incentives to seek employment. A generous social safety net weakens workers' incentives to seek employment. For example, unemployment insurance (UI) reduces the incentive of workers that lose their job to find new employment quickly. In the United States, maximum UI is generally 26 weeks except in recessions when it is usually extended. The benefit levels of workers on UI in the United States is also usually modest: replacement rates are usually around 50%. In many European countries, UI is both much longer and much more generous in terms of replacement rates. This was particularly the case around 2000. Since then, some countries in Europe have reformed their UI systems to make them less generous with an eye towards encouraging the unemployed to seek employment. Another example is health insurance. In the United States, health insurance is tied to employment, while in Europe health insurance is universal. The desire to have health insurance encourages Americans to seek employment, while this is not the case for Europeans.

A final potential explanation for the difference in hours worked between Europe and the U.S. that I would like to discuss is differences in taxes. Taxes are much higher in Europe than in the United States. These high taxes result in lower take-home pay which can discourage labor supply. But how large are these effects? Can they plausibly explain the large difference in hours worked between Europe and the United States? Prescott thought that they explained the entire difference. Others – e.g., Alesina, Glaeser, and Sacerdote – have argued that they explain a relatively modest share of the difference. Empirically, it is difficult to reach definitive conclusions since differences in tax rates between countries may be correlated with other factors that also affect hours worked (such as the many factors we discuss above). Given this, how can we make progress on assessing the role of taxes? The next section shows how a bit of theory can help sharpen the debate and make it more tractable.

4.4 Taxes and Labor Supply

To gain insight into the effect of taxes on labor supply, we add taxes to our model from section 1. We again consider a household with a utility function $U(C) - V(H)$. But in this case, we assume that the household's budget constraint is

$$(1+\tau_c)C = (1-\tau_l)wH + T,$$

where τ_c is a consumption tax (i.e., a sales tax or a value added tax) and τ_l is a labor income tax. (τ is the Greek letter tau.) Household maximization and manipulations analogous to those we worked through in section 1 yield the following labor supply equation:

$$\frac{V'(H)}{U'(C)} = w\frac{1-\tau_l}{1+\tau_c}.$$

This labor supply equation is identical to the one we derived in section 1 except that it is the after-tax wage $(1-\tau_l)w/(1+\tau_c)$ that shows up on the right-hand side rather than simply the wage.

Let's take a moment to make sure we understand why $(1-\tau_l)w/(1+\tau_c)$ is the after tax wage. Someone that is not thinking too carefully about this problem may think that the after tax wage is $(1-\tau_l)w$. After all, this is the wage minus the labor income tax $\tau_l w$. In other words, $(1-\tau_l)w$ is take-home pay in this economy. But what really matters is not take-home pay in "dollars" but rather how many coconuts these dollars can buy. The consumption tax raises the price of coconuts from 1 to $(1+\tau_c)$. This means that one hour of work buys not $(1-\tau_l)w$ coconuts, but rather $(1-\tau_l)w/(1+\tau_c)$ coconuts. So, the after tax wage is $(1-\tau_l)w/(1+\tau_c)$. This shows that when we think about the level of taxes in Europe and the U.S., it is important to take account of both labor income taxes and consumption taxes. Consumption taxes are much higher in Europe than in the United States.

To simplify our notation going forward we define the following overall tax wedge:

$$(1-\tau) = \left(1 - \frac{\tau_l+\tau_c}{1+\tau_c}\right) = \frac{1-\tau_l}{1+\tau_c}.$$

Written in terms of this overall tax wedge, the labor supply equation is

$$\frac{V'(H)}{U'(C)} = w(1-\tau). \tag{13}$$

Let's now consider how a change in the tax wedge affects labor supply through the lens of this labor supply equation. First, an increase in τ reduces the right-hand side of the labor supply equation. We know from section 2 that – holding C constant – this reduces labor supply. This is the substitution effect at work. But we also know that an increase in taxes may affect labor supply through an income effect. It may seem logical to think that paying taxes reduces after-tax income and therefore makes the household poorer. If this is the case, an increase in taxes will have a positive income effect on labor supply. As we saw in section 2, this type of income effect may be quite large. Perhaps even large enough to completely offset the substitution effect.

But the income effect of a tax increase turns out to be a bit more complex than just looking at after-tax income. We must also think about what the government does with the tax revenue. Here, it is useful to consider two polar cases. One polar case is that the government throws the tax revenue in the ocean. In a coconut economy, the literal interpretation of this is that the government uses the tax revenue to purchase coconuts and then throws those coconuts in the ocean. In this case, the households are clearly poorer by exactly the amount of taxes they pay and the income effects discussed in section 2 will come into play with full force.

Is this a ridiculous case to consider? Not as ridiculous as it may seem. Consider a related case, where the government raises taxes, uses the tax revenue to buy tanks, cruise missiles, and fighter jets, and then uses this equipment to fight a war on foreign soil. From the perspective of the household, this is akin to the tax revenue being thrown in the ocean. Such wars may serve vital security interests. But even if they do, the need to spend resources on defense makes us poorer. Consider the comparison between a situation where the wars are not needed (and taxes are not raised) and a situation where the wars are needed (and taxes are raised). In the second case, the households are poorer by the amount of taxes they pay.

Views about the value of non-military government spending differ sharply just as views about military spending do. Some people are quite pessimistic about the value of such spending. If government tax revenue is wasted on low value projects, households are made poorer by taxes and this has a positive income effect on labor supply.

The other polar case is one where the government simply transfers all the the tax revenue back to the households through "lump-sum" transfers. In this case, the household is not made poorer by the taxes: they get the money back through the transfer. So, there is no effect on consumption and no therefore income effect.

But if the households get the taxes they pay back through the transfer, is there then any substitution effect? Actually, there is! This has to do with the meaning of a "lump-sum" transfer. That the transfer is lump sum means that the amount each household gets as a transfer does not depend on the amount of taxes *they* pay, but rather on the total amount of taxes raised from *all* households. Since the transfer depends on total tax revenue, each household treats their own transfer as independent of their behavior: if they work more, the extra tax revenue gets shared by everyone in the economy and they get a trivial share of it back. So, the lump-sum transfer case is a case were there is a full substitution effect, but no income effect.

A related case, is one where the government spends the tax revenue on services that the households would have purchased themselves if they didn't get them from the government. Consider child care. Suppose the government taxes households and uses the money to provide child care services in the amount and of the character and quality that the households would have purchased themselves in the absence of the taxes. This is a case where the households are not made worse off by the taxes. Instead of purchasing child care in the market, they pay taxes and get that same child care from the government. A number of other government services fall into this category to a greater or lesser extent: schools, health care, security (police), etc.

Whether the government can or does provide services of the same character and quality as the private sector is hotly debated and probably varies greatly across countries and types of services. In some cases, strong arguments can be made that the government has the potential to solve market failures and provide services more efficiently than the market. Defense and insurance are candidates for this category (more on this in chapter XX [Market Efficiency chapter]). In other cases, it is likely that the government provides services less efficiently than the market.

The real world is likely somewhere in between the two polar cases discussed above. Where on that continuum the real world falls is quite important for analyzing the effect of taxes on labor supply since this determines the strength of the income effect of taxes. For the sake of the argument, we are going to follow Prescott (2002) and assume we are in the second polar case, where there is no income effect because all the taxes are spent on a combination of lump-sum transfers and services that are of equal quality and character as things the household would have purchased themselves if the government did not provide them. For concreteness, we assume that all the tax revenue is transferred back to the household lump sum.

To make quantitative predictions about the effect of taxes on labor supply we need to be more precise about the utility function. To this end, we consider the case where

$$U(C) = \log C \qquad \text{and} \qquad V(H) = \psi \frac{H^{1+\eta^{-1}}}{1+\eta^{-1}}.$$

As we discussed in section 2.2, $\log C$ is the case where a permanent change in a worker's wage has no effect on labor supply because the income and substitutions effects are equally strong. This is then also the case where, if the tax revenue were thrown in the ocean, an increase in taxes would have no effect on labor supply. But since we are assuming that the tax revenue is transferred back to households lump sum, the income effect in our case will be zero. For $V(H)$, we are assuming a power function – H to the power $1+\eta^{-1}$. Why we choose this particular form will become clear below.

With this utility function, we have that

$$U'(C) = \frac{1}{C} \qquad \text{and} \qquad V'(H) = \psi H^{\eta^{-1}}$$

and the labor supply equation becomes

$$\psi C H^{\eta^{-1}} = w(1-\tau). \tag{14}$$

In chapter XX [production chapter], we derived the following equation for labor demand when the production function of firms takes a Cobb-Douglas form:

$$w = (1-a)\frac{Y}{L}. \tag{15}$$

Combining these two equations to eliminate w yields

$$\psi C H^{\eta^{-1}} = (1-a)\frac{Y}{L}(1-\tau). \tag{16}$$

Recall from section 1 that $L = NH$ and we normalized N to 1. Also, the fact that all tax revenue is transferred lump sum back to households means that $T = \tau_l wH + \tau_c C$. Plugging this into the household's budget constraint yields $C = wH$. Recall from chapter XX [production chapter] that $wH = aY$. This means that $C = aY$. Plugging $L = H$ and $C = aY$ into equation (16) yields

$$\psi C H^{\eta^{-1}} = a(1-a)\frac{C}{H}(1-\tau).$$

Dividing through by C and rearranging then yields

$$H^{1+\eta^{-1}} = (1-\tau)\frac{a(1-a)}{\psi}.$$

Taking natural logarithms of this equation and rearranging yields

$$\log H = \frac{\eta}{\eta+1}\log(1-\tau) + \frac{\eta}{\eta+1}\log\frac{a(1-a)}{\psi}. \tag{17}$$

Now suppose that equation (17) holds in France and the United States. For simplicity, assume that η, a, and ψ are the same in these two countries (but taxes are not). We can then take equation (17) for France and subtract equation (17) for the U.S. from it. This yields

$$\log H_{\text{Fr}} - \log H_{\text{US}} = \frac{\eta}{\eta+1}[\log(1-\tau_{\text{Fr}}) - \log(1-\tau_{\text{US}})], \tag{18}$$

where the subscript Fr stands for France and the subscript US stands for the United States. This equation shows how differences in taxes translate into differences in hours worked in our model. In particular, log differences in hours worked are proportional to log differences in the tax factor $1-\tau$ between the two countries with a factor of proportionality $\eta/(\eta+1)$.

Recall that log differences are approximately equal to percentage changes when the log differences are small. This follows from the fact that a first-order Taylor series approximation of $\log x$ around $\log x_0$ is

$$\log x = \log x_0 + \frac{1}{x_0}(x-x_0) + o(x-x_0).$$

We therefore have that for small changes in the tax factor $1-\tau$, we can say that a 1% change in $1-\tau$ translates into a $\eta/(\eta+1)$ percent change in hours worked.

Clearly, η is a key parameter in assessing the potential role of taxes in explaining the difference in hours worked across countries. This parameter is called the *Frisch elasticity of labor supply*. It measured the strength of the substitution effect on labor supply. Specifically, the Frisch elasticity is the percentage change in hours worked that results from a 1% change in wages holding consumption fixed. We can see that this is η in our model by taking natural logarithms of equation (14) and rearranging:

$$\log H = \eta\log w + \eta\log(1-\tau) - \eta\log C - \eta\log\psi.$$

If w changes by 1% in this equation and the other factors on the right-hand side remain unchanged, then hours worked will increase by η percent. If η is large, the

Table 4: Tax Rates in France and the United States

	France	United States
τ_c	0.33	0.13
τ_l	0.49	0.32
Social-security tax	0.33	0.12
Marginal income tax	0.15	0.20
τ	0.62	0.40

Notes: The table reports estimates of tax rates in France and the United States. These estimates are taken from Prescott (2002).

substitution effect is strong and changes in wages (holding consumption fixed) have large effects on hours worked. If η is small, the substitution effect is small. The functional form we chose for $V(H)$ above implies that the Frisch elasticity is a constant (i.e., not a function of the wage or of hours worked). We chose that functional form for this reason.

Using some theory, we have now uncovered two specific features of the environment that matter a great deal for answering the question of how large the effect of taxes on labor supply are: 1) how is the tax revenue used (this determines the size of the income effect); 2) the Frisch elasticity of labor supply (this determines the size of the substitution effect). This is a good example of how theory can be useful. It helps us zero in on what really matters in a complex setting. Armed with these conclusions, we can focus our discussion about the plausibility of taxes explaining the difference in hours between France and the U.S. on these particular things.

4.5 Estimation of the Frisch Elasticity of Labor Supply

Prescott (2002) reports estimates of the tax wedge in France and the United States in the late 1990s. These estimates are reproduced in Table 4. We see that taxes were indeed substantially higher in France than in the United States at this time. This difference comes both from much higher consumption taxes and much higher social-security taxes. From Table 2, we also see that the difference in hours worked between France and the U.S. in 2000 was about 31 log points.

Table 5 reports the difference in hours worked that our model predicts based on the difference in tax rates reported in Table 4 for different values of the Frisch elasticity η. We calculate these predicted differences using equation (18). We see from this table that if the Frisch elasticity is between two and three, taxes can indeed explain the entire difference in hours worked. However, if the Frisch elasticity is 0.5 or smaller, taxes can explain less than half of the difference in hours even under the favorable assumption for this explanation that all tax revenue is transferred lump sum back to households.

Prescott (2002) made different assumptions about $V(H)$ than we have. But one can calculate the Frisch elasticity for his $V(H)$ and it turns out to be around three. He therefore concluded that taxes can explain the entire difference in hours worked between France and the United States. Many researchers – e.g., Alesina, Glaeser, and Sacerdote (2005) – have criticized Prescott's assumption regarding the Frisch elasticity. They argue that empirical evidence on labor supply elasticities do not support the notion that the Frisch elasticity is anywhere close to three.

Other researchers have countered by arguing that much of the existing evidence on labor supply elasticities does not capture all the ways in which hours can respond to taxes. One important distinction is between the intensive margin response – i.e., hours worked per worker – and the extensive margin response – i.e., the employment rate. Supporters of Prescott's conclusions argue that much of the existing evidence does not capture extensive margin responses partly because such responses will occur over long periods of time. For example, higher taxes will lead workers to retire earlier. But this will not happen all at once. Rather it will happen gradually. Also, higher taxes may lead young people to choose different levels of education and different occupations. Again, this may happen gradually over long periods of time.

Another important distinction is between the Frisch elasticity of labor supply and other notions of the labor supply elasticity. The Frisch elasticity is the elasticity of labor supply *holding consumption fixed*. This elasticity is particularly difficult to measure. Much of the literature that seeks to measure the labor supply elasticity uses changes in taxes as the source of variation in wages. But most tax changes are permanent and therefore potentially have substantial effects on consumption. Such tax changes will therefore not measure the Frisch elasticity but rather the Hicksian or Marshallian labor supply elasticity depending on how the tax review is used. (See Keane (2011) for a detailed discussion of the differences between these concepts.) Both the Hicksian and Marshallian elasticities are smaller than the Frisch elasticity. Small labor supply elasticities from permanent tax changes may therefore not provide solid evidence about the Frisch elasticity. However, small elasticities from permanent tax changes do provide direct measures of the (short run) effect of per-

Table 5: How Much Can Taxes Explain?

Frisch Elasticity (η)	Hours Difference
0	0
0.1	4
0.5	15
1.0	22
2.0	30
3.0	34
Actual Difference in 2000	31

Notes: The table reports the log difference in hours worked (multiplied by 100) in France and the U.S. predicted by our model using equation (18) for different values of the Frisch elasticity of labor supply η. The tax rates used are from Table 4. The bottom row of the table reports the actual difference in hours worked between France and the U.S. in 2002 from Table 2.

manent tax changes on labor supply, which is the ultimate question at hand (apart from these being short run responses and perhaps not capturing extensive margin effects).

A third complication – related to those discussed above – is that estimates constructed from small tax changes may understate the effect of the very large difference in taxes between Europe and the U.S. because of frictions that lead people not to adjust their behavior in response to small change in wages. Many jobs have fixed working hours that do not allow for changes in hours worked. Such workers can potentially find a different job with different working hours, but this can be quite costly. Even in cases where workers can take extra shifts negotiating a different work schedule with ones employer may be costly. People may also be inattentive in the short run to small changes in taxes. These "frictions" may lead people to behave in the short run as if they have a small Frisch elasticity in response to small tax changes even if their true Frisch elasticity is large. Extrapolating such estimates to larger tax changes and the long run, may then lead to an underestimate of their ultimate response to such tax changes.

To (partially) address these empirical challenges, recent empirical work has focused on tax-free years, that have occurred in several countries. In particular, Sigurdsson (2023) studies a tax-free year in Iceland and Martínez, Saez, and Siegenthaler (2021) study tax-free years in Switzerland. These settings have several advantages. First, the tax change is temporary (only lasts for one or two years), which

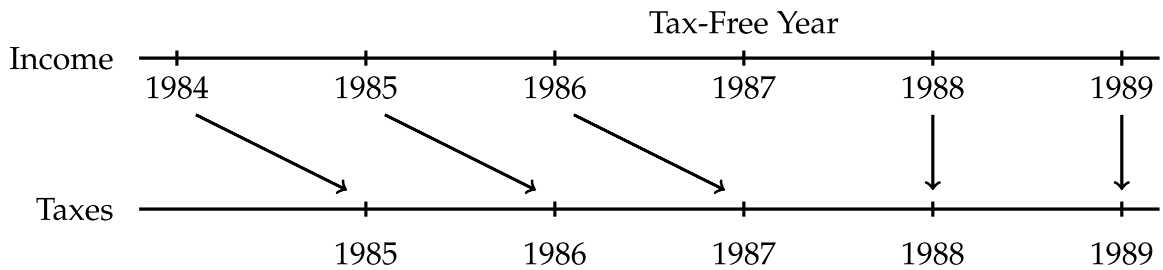

Figure 14: Iceland's 1988 Tax Reform

Note: Prior to 1988, taxes in year t were based on income in year -1. Starting in 1988, taxes in year t were based on income in year t (modern withholding system). Income in year 1987 was never taxed. This figure is based on Figure 1 in Sigurdsson (2023).

limits the size of any potential income effects. Second, the tax change is large. This makes it less likely that small frictions lead to inaction. Third, the authors are able to estimate both intensive margin and extensive margin effects.

The tax-free year in Iceland that Sigurdsson exploits occurred in 1987. The reason for this tax-free year was that Iceland adopted a modern withholding-based pay-as-you-earn income tax system at this time. Prior to 1988, income taxes in Iceland were collected with a one-year lag. This meant that taxes paid in year t were based on income earned in year $t-1$. The switch implied that income earned in 1987 was never taxed. Figure 14 depicts this visually. In 1987, Icelanders were paying taxes on income earned in 1986. In 1988, however, they started paying taxes on income earned that same year (contemporaneous withholding). To avoid everyone having to pay taxes on two year's income in the same year, taxes were never assessed on income earned in 1987. Importantly, this policy change was announced towards the end of 1986 with newspaper headlines such as "A Tax-Free Year."

How exactly can we use this tax-free year to estimate the Frisch elasticity? One approach would be to simply compare hours worked and employment in 1987 – when tax rates were zero – to hours worked and employment in surrounding years – when marginal tax rates were on the order of 40% for many workers. Figure 15 plots person-years worked in Iceland over the period 1980 to 1990. We see that labor increased steadily in the early and mid-1980s, it peaked in 1987, and then fell for the following three years. The increase in labor between 1986 and 1987 was 5.8%, the fall between 1987 and 1988 was 3.0%. But the increase between (say) 1984 and 1985 was also 3.6%. What can we conclude from these data about the Frisch elasticy?

The trouble with trying to estimate the Frisch elasticity from the data in Figure 15 is that we don't know what the counterfactual would have been absent the tax-

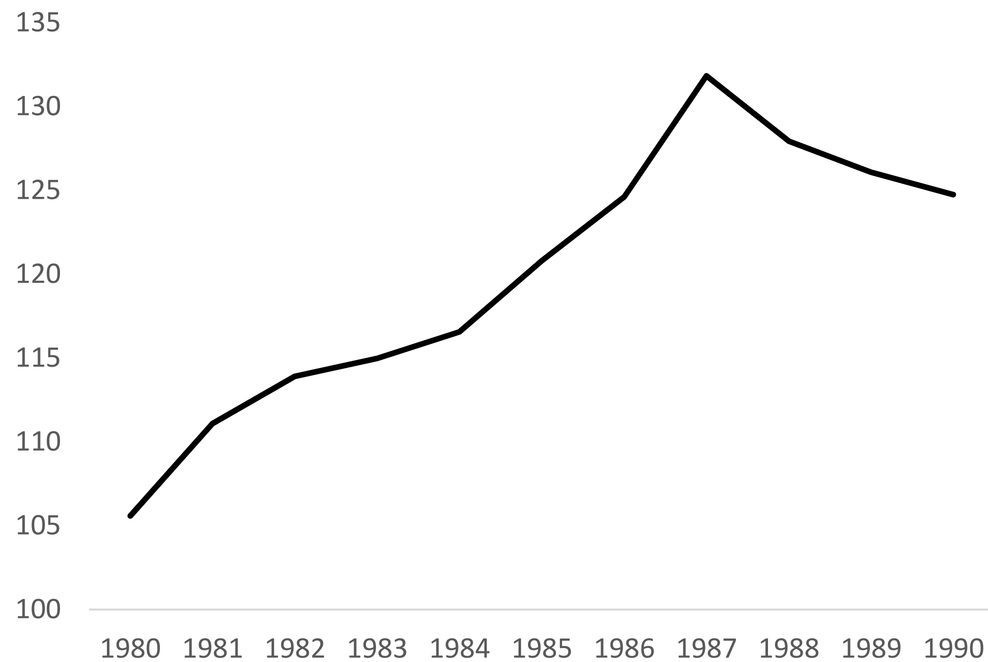

Figure 15: Aggregate Person-Years Worked in Iceland

Note: This figure plots person-years worked in Iceland between 1980 and 1990. The data is compiled from records of weeks worked. A person-year is 52 weeks worked. The vertical axis is thousands of person-years worked. The source of these data is the Icelandic Historical Statistics Database.

free year. One mundane concern is that labor in Iceland was trending upward in the 1980s. How much of the 5.8% increase between 1986 and 1987 is just due to this trend increase? This is hard to estimate since the trend may be changing over time and estimates of its slope are sensitive to what sample period is used (e.g., whether the years after 1987 are included in the calculation).

An even thornier issue is that 1987 may be special in other ways than just that it was a tax-free year. The Icelandic economy has been notoriously volatile both due to natural shocks (such as variation in fish stocks) and also due to volatile economic policy. The fish stocks around Iceland may have been unusually large in 1987 leading to an unusually large amount of fishing. Or monetary and fiscal policy may have been unusually loose during that year. A myriad of other special circumstances may have also affected the amount of work people did in Iceland in 1987. Clearly 1987 was a good year in terms of employment in Iceland and it seems plausible that the absence of income taxes played a role. But other factors may also have played a role. How do we isolate the effect of taxes?

Sigurdsson's answer to this question is to use a method called *difference-in-*

difference estimation. He notes that workers in different tax brackets faced reductions in tax rates of different sizes: those in the highest tax bracket saw their marginal tax rates fall from about 50% to zero, while those in to lowest tax bracket saw a fall from only 10% to zero. Sigurdsson's idea is to compare the response of workers in different tax brackets.

Workers in the different tax brackets were all affected by the other factors that made the tax free year unusual. For example, if monetary policy in Iceland was loose in 1987, this affected workers in all tax brackets. Therefore, if one looks at the *difference* between the behavior of workers in different tax brackets, the effect of these other factors may "cancel out."

This approach will work if – and this is not a trivial 'if' – the different groups are equally exposed to the other special things affecting the Icelandic economy in 1987. In other words, if all those other factors affect workers in the different tax brackets equally on average. In this case, one can look at the difference between the behavior of the different groups in order to isolate the effect of taxes. Looking at the difference will then "difference out" (cancel out) the effect of all other factors since both groups are equally exposed to these other factors.

What is then left is the fact that the high tax bracket group is "more exposed" to the tax change (sees a larger change in their marginal tax rate). A larger increase in labor by the high tax bracket group, relative to the lower tax bracket group, will then be interpreted as being caused by the larger fall in taxes for that group. This larger response can then be used to calculate the Frisch elasticity.

The "equal exposure to other factors" assumption discussed above is often referred to as a "parallel trends" assumption. The reason for this language is the idea that the two groups would have moved in parallel over time if not for the tax change. Any non-parallel movement in 1987 is then attributed to the tax change. Clearly, the parallel trends assumption is a non-trivial assumption that may not hold. But looking at whether the groups move in parallel in prior years is one way to build confidence in this assumption.

This method is called difference-in-difference estimation because there are two differences: over time and across groups. Implementation of this method often involves a few additional issues (such as "controls") that I ignore here for simplicity. For a more thorough introduction to difference-in-difference methods, I recommend chapter 5 of *Mastering Metrics* by Josh Angrist and Steve Pischke (Angrist and Pischke, 2015).

Figure 16 plots the difference in labor earning between workers in high versus

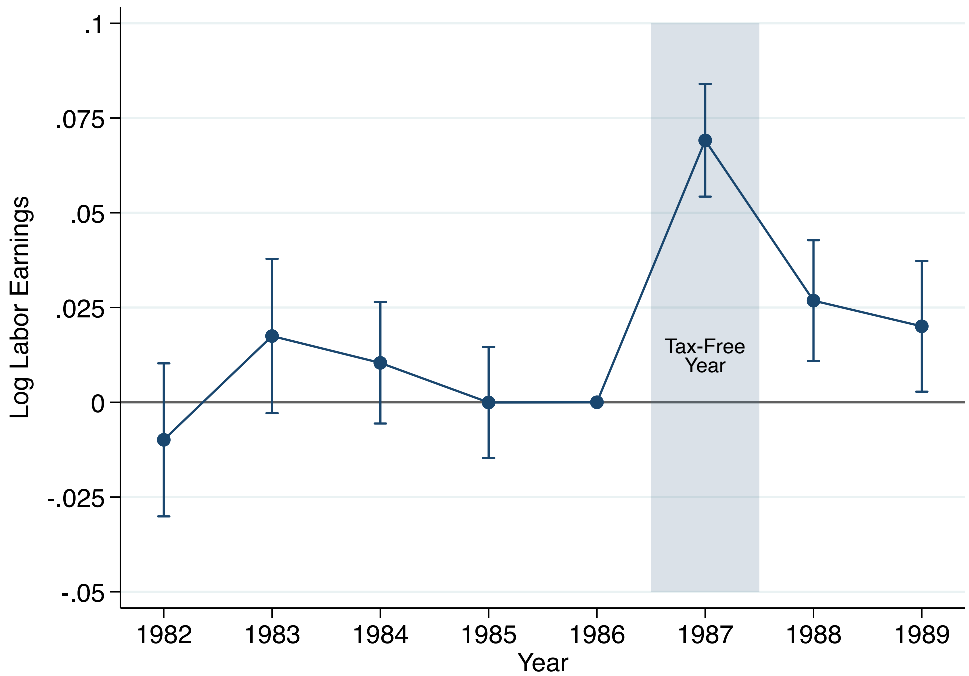

Figure 16: Difference in Labor Earnings: High versus Low Tax Bracket

Note: This figure plots the difference in labor earnings between workers in high and low tax brackets. This difference is normalized to zero in 1986. Workers are assigned to tax brackets based on income in the prior year. The difference in calculated conditional on several controls discussed in Sigurdsson (2023). This figure is the same as Figure 4(a) in Sigurdsson (2023).

low tax brackets over time. The difference in normalized to zero in 1986. Clearly, the difference in labor earnings across these groups spiked upwards substantially (by about 7%) in 1987. This is saying that workers in high tax brackets increased their labor earning about 7% more than workers in the lowest tax bracket in 1987. Sigurdsson's interpretation of this fact is that the larger drop in marginal tax rates faced by workers in the high tax brackets led them to change their labor supply by more than the workers in the lowest tax bracket.

Sigurdsson goes on to convert this fact and others like it into estimates of the Frisch elasticity. His conclusion is that workers in Iceland displayed a Frisch elasticity of about 0.4 on the intensive margin (hours worked among the employed) and 0.1 on the extensive margin (number of workers employed). The extensive margin response is heavily concentrated among young workers (aged 18 to 25). Also, it is based on a somewhat different method. Adding the intensive and extensive margin responses, Sigurdsson's results suggest an overall Frisch elasticity of about 0.5.

The other important feature of Figure 16 is the *absence* of a significant effect in years prior to 1987. The vertical bars around the estimates for each of the years are

95% confidence intervals for these estimates. The fact that all of these contain zero prior to 1987 indicates that we cannot reject statistically that the effect is zero in any of these prior years. In this sense, we can say that the high and low tax bracket groups followed "parallel trends" prior to 1987. The parallel trend is somewhat close to being violated in 1983. This was a year of extremely high inflation (above 80%) and a sharp recession in Iceland. It was therefore arguably a severe test for the parallel trends assumption.

Let's next consider evidence from tax free years in Switzerland. Martínez, Saez, and Siegenthaler (2021) study this case and come to quite a different conclusion than Sigurdsson. They estimate a Frisch elasticity that is close to zero – below 0.05. This is the case even though the methods they use are in many ways quite similar to the methods used by Sigurdsson. Martínez, Saez, and Siegenthaler employ a difference-in-difference methodology just as Sigurdsson does. However, the comparison is not between people in different tax brackets, but rather between people in different Swiss cantons. To understand why, we must discuss some details of the Swiss tax reform.

The reason for the tax free years in Switzerland was the same as in Iceland: Switzerland was moving to a modern withholding-based pay-as-you-earn system for income tax. But the reform was more complicated in several dimensions. First, the old system was biennial in many Swiss cantons. This meant that tax returns were filed every two years and taxes were paid for two years on income earned the prior two years. For these cantons, the tax reform lead to two tax free years, rather than one. Second, the tax reform was implemented in a staggered fashion across cantons. Some cantons transitioned to the new system in 1999 – implying that 1997 and 1998 were tax free, others transitioned in 2001 – implying that 1999 and 2000 were tax free, yet others transitioned in 2003 – implying that 2001 and 2002 were tax free.

The staggered implementation of the tax reform made it possible for Martínez, Saez, and Siegenthaler to use cantons that implement the reforms late as controls for those that implement the reform early and vice versa. Figure 17 illustrates this well. It plots average wage earning from 1990 to 2010 for three groups of Swiss cantons: those that had a tax holiday in 1997-1998, those that had a tax holiday in 1999-2000, and those that had a tax holiday in 2001-2002. The years of the tax holidays are highlighted in the figure with blank markers.

If the tax holiday had a substantial effect on wage earning, we should see something unusual occur in the figure when it comes to the tax holiday observations.

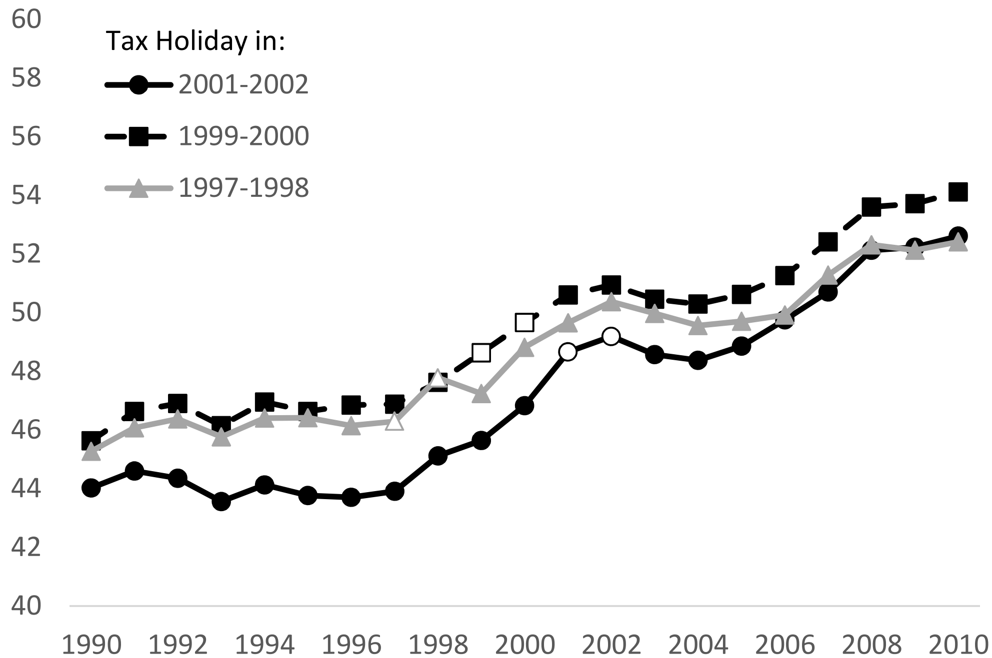

Figure 17: Average Wage Earnings in Swiss Cantons

Note: This figure plots average wages earnings in Swiss cantons which faced tax holidays in different years. The tax holiday years are listed in the legend and the marker for the line is blank in these years. This figure is a simplified version of Figure 5 panel B of Martínez et al. (2021).

"Unusual relative to what?" you may ask. Unusual relative to what is occurring in those same years to the other regions that don't have a tax holiday. Generally, the three groups share relatively similar movements over time. In other words, they move roughly in parallel. This implies that the "parallel trends" assumption holds reasonably well for these three groups. But what about in the years with tax holidays? Consider the gray line. Does this line move differently in 1997 and 1998 than the other two lines. Perhaps a little in 1998. But almost not at all. The same is true of the other two lines during the tax holidays they experience.

This result is quite remarkable. We are simply plotting the raw data. No fancy statistical methods needed. And we can very clearly see that almost nothing happens in the tax free years. Now, it is useful to get a sense for what one might have expected to happen if one believed that the Frisch elasticity was (say) 1. Since the cut in the marginal tax rate was around 20 percentage points, a Frisch elasticity of 1 would imply that wage earning should rise by 20 percent. This is about 10 thousand Swiss francs. In this case, we should therefore see a large spike in each series during the tax holiday years. In fact, however, we see almost no unusual movement. This implies that the Frisch elasticity is very close to zero. Martínez, Saez, and Siegenthaler use the size of the modest increases in Figure 17 to estimate an intensive margin Frisch elasticity of 0.026. They present similar evidence for the employment rate and use this to estimate an extensive margin elasticity of zero.

So, what should we make of this? We have evidence from two papers that give rather different answers. Actually, this is not unusual. No study is perfect and every study is special in some way when it comes to the setting being studied. A well done empirical study – as both of these are in my mind – should move readers' priors about the object being studied. But few studies will settle an issue. In this case, both studies yield estimates of the Frisch elasticity that are much lower than the estimate Prescott needed to assume to be able to conclude that taxes could explain all of the difference in hours worked between Europe and the United States. In this sense, both studies provide evidence contradicting Prescott's argument.

One possible reason for the difference is that it reflects differences in the flexibility of the labor markets in Iceland versus Switzerland. Iceland's labor market is (arguably) quite a bit more flexible than the Swiss labor market. Perhaps this helps explain the larger response in Sigurdsson's study. If this is true, it is possible that the Frisch elasticity relevant for comparing long-run differences in tax rates is even larger than Sigurdsson's estimate, since workers are presumably more able to overcome frictions in the long run. On the other hand, the countries we are trying to explain the behavior of – France, Germany, Italy, and other continental European countries – have rigid labor markets more like Switzerland than Iceland. But this is all quite speculative. The hard evidence we have from these two papers suggests modest Frisch elasticities.

5 Marx and the Mainstream

Karl Marx is arguably the most influential labor economist of all time. His ideas continue to resonate strongly with many people. When I assign readings by Marx in my macroeconomics class, many students react enthusiastically to what they read. Yet, Marx's ideas differ markedly from the ideas we have explored up until this point in this book. Given Marx's persistent appeal, these differences are worth exploring in some detail.

Marx's ideas evolved over his lifetime. I will focus on Marx's thought as it appears in a relatively early essay titled *Wage-Labor and Capital* (Marx, 1849/1978). This essay was published in 1849. It was based on a series of lectures Marx delivered in Brussels in December 1847. Notice that Marx and Engels wrote the *Communist Manifesto* at almost exactly the same time. It was published in 1848. *Wage-Labor and Capital* therefore represents Marx's thinking about capitalism at the time he and Engels wrote the *Communist Manifesto*. When Engels re-published *Wage-Labor and Capital* in 1891, he commented that it was "approximately as Marx would have written in 1891." So, arguably, it is a good representation of his ideas. In addition, *Wage-Labor and Capital* is a relatively short and highly accessible piece of Marx's writing (in contrast to his magnum opus *Capital*).

A central argument in *Wage-Labor and Capital* is that increases in productivity lower wages. Marx writes: "the productive power of labor is raised, above all, by a *greater division of labour*." (Emphasis is in the original in all quotes from Marx.) In other words, "division of labour" is Marx's preferred term for labor productivity. He then argues:

> "the greater *division of labour* enables *one* worker to do the work of five, ten or twenty; it therefore multiplies competition among the workers fivefold, tenfold and twentyfold. ... as the *division of labour* increases, labour is *simplified*. The special skill of the worker becomes worthless. He becomes transformed into a simple, monotonous productive force that does not have to use intense bodily or intellectual faculties. His labour becomes a labour that anyone can perform. Hence, competitors crowd upon him on all sides ... *Therefore, as labour becomes more unsatisfying, more repulsive, competition increases and wages decrease.*"

After some further elaboration, Marx concludes:

> Let us sum up: *The more productive capital grows, the more the division of labour and the application of machinery expands. The more the division of labour and the application of machinery expands, the more competition among the workers expands and the more their wages contract.*

How do these ideas compare to the model of the labor market we have developed thus far in this book? In fact, they are diametrically opposed to the implication of the model we have developed. In that model, the marginal product of labor is equal to $(1-a)AL^{-a}K^{a}$, where A is productivity, L is the quantity of labor, K is the quantity capital, and a is a parameter. An increase in productivity – increase in A – therefore, raises the marginal product of labor. If the labor market is competitive –

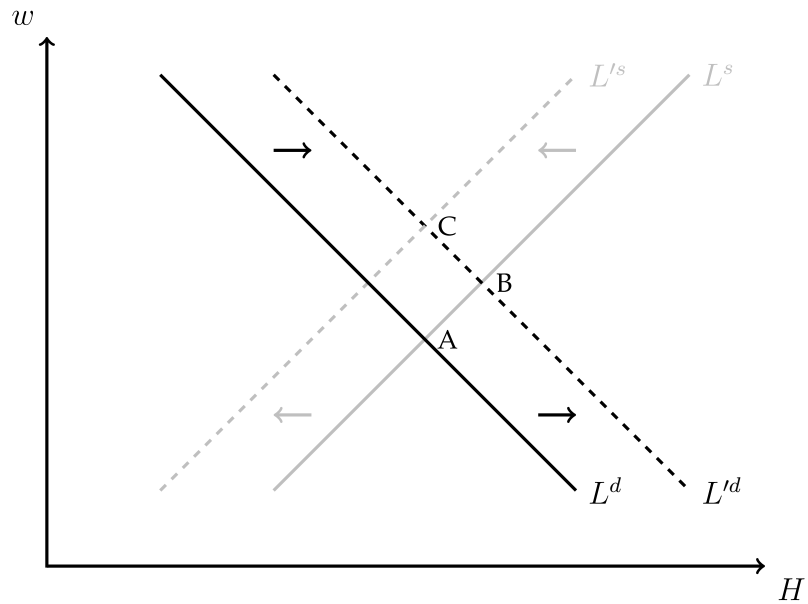

Figure 18: The Effect of an Increase in Productivity on the Labor Market

which Marx seemed to assume in *Wage-Labor and Capital* – the real wage will equal the marginal product of labor. An increase in productivity will, therefore, increase real wages.

Figure 18 illustrates this point. Suppose the economy starts off at point A, where the initial labor demand and labor supply curves cross. (Recall from section 1 that $L = H$, because $L = NH$ and we normalized N to 1.) Then suppose that productivity increases. Let's first consider how this affects the labor demand curve. An increase in productivity increases the marginal product of labor for any level of labor supplied (i.e., no matter what L is, $(1-a)AL^{-a}K^a$ is higher the higher is A in this expression). This means that an increase in productivity shifts the labor demand curve up and to the right (i.e., the points that satisfy the equation $w_t = (1-a)AL^{-a}K^a$ are further to the north-east in Figure 18 the higher is the value of A). If this were the only effect of the increase in productivity, the economy would move from point A to point B in the figure. Wages would rise and so would the quantity of labor.

But the increase in productivity also affects the labor supply curve through an income effect. Recall that the labor supply curve plots points for which

$$w = \frac{V'(H)}{U'(C)}. \tag{19}$$

The increase in wages and the quantity of labor that results from the shift in the labor demand curve discussed above (the economy moving from point A to point B in Figure 18) implies that households have higher income. Higher income results in higher consumption, which affects the right-hand-side of equation (19). Since C goes up, $U'(C)$ goes down (recall that $U(\cdot)$ in concave). Since $U'(C)$ goes down and is in the denominator of the right-hand-side of equation (19), a higher wage is needed to satisfy equation (19) for any given value of H. This implies that the labor supply curve shifts up and back. As a result, the economy doesn't move from point A to point B, but rather moves from point A to point C.

The combined effect of the shifts in the labor demand curve and the labor supply curve are unambiguously to increase the wage. Whether the quantity of labor increases or decreases depends on the strength of the income effect on labor supply. I have drawn Figure 18 such that the income effect and substitution effect are equal in strength. In this case, the increase in productivity has no effect on the quantity of labor. Our analysis of data over the past 200 years in section 2 suggests that if anything income effects on labor supply have been a little bit stronger than substitution effects over the long term implying that perhaps the labor supply curve should shift back even more in Figure 18 (resulting in an even larger increase in the wage).

This analysis shows that Marx's ideas about the effect of productivity on wages are diametrically opposed to the mainstream model we have developed. But more importantly, Marx turned out to be dead wrong empirically. Figure 19 plots real wages of laborers in England from 1700 to 2000. After stagnating over the course of the 18th century, real wages in England rose sharply over the course of the 19th and 20th centuries. The cumulative increase in real wages over this period was a staggering 1400%. Clearly, increases in productivity over the past 200 years have increased, not decreased, the wages of workers.

5.1 Was Marx Simply Confused?

I find it hard to read *Wage-Labour and Capital* without coming to the conclusion that Marx was simply confused. As we discuss above, Marx argues that increases in productivity lower wages. But in other parts of the essay Marx argues forcefully that increases in productivity lower the prices of goods. Marx is quite clear on this point:

> If, now, by a greater division of labour ... one capitalist has found a means of producing ... a whole yard of linen in the same labour time in which his competitors weave half a yard, how will this capitalist operate?

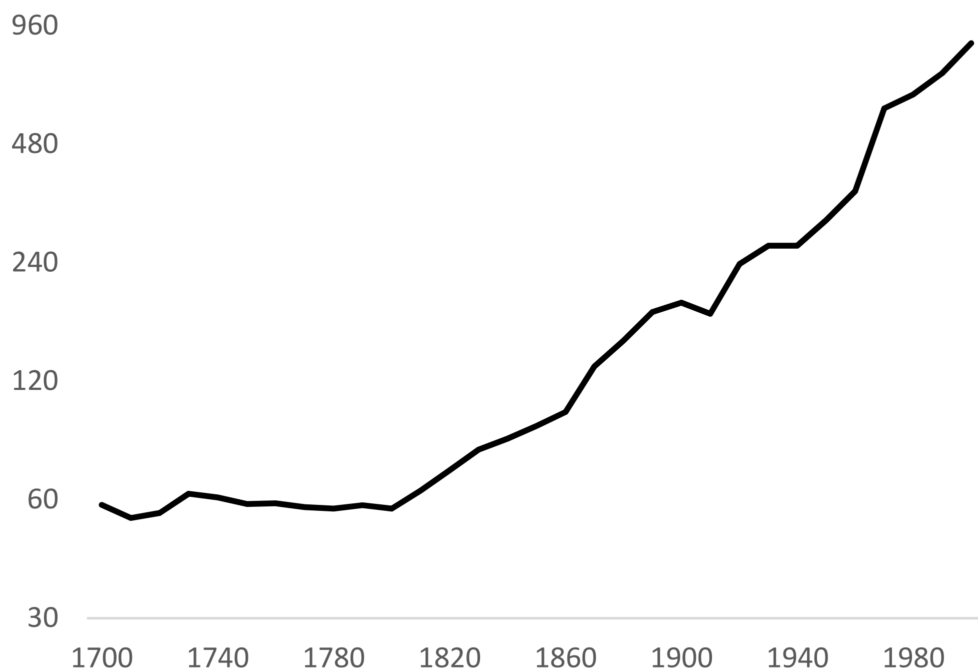

Figure 19: Real Wages of Laborers in England from 1700 to 2000

Note: This series is constructed by splicing together data from Clark (2010) for the period 1700 to 1860 and Clark (2005) for the period 1860 to 2000. The series is plotted on a logarithmic scale (base 2) and is scaled to be equal to 100 in the 1860s.

> He could continue to sell half a yard of linen at the old market price; this would, however, be no means of driving his opponents from the field and of enlarging his own sales. ... The more powerful and costly means of production that he has called into life enable him, indeed, to sell his commodities more cheaply, they *compel* him, however, at the same time *to sell more commodities*, to conquer a much *larger* market for his commodities; consequently, our capitalist will sell his half yard of linen more cheaply than his competitors.
>
> However, the *privileged position* of our capitalist is not of long duration; other competing capitalists introduce the same machines, the same division of labour, introduce them on the same or on a larger scale, and this introduction will become so general that the price of linen *is reduced* not only *below its old*, but *below its new cost of production*.

In other words, increases in productivity lower prices. But doesn't this act to improve the purchasing power of workers? Can't they then buy more goods for a given amount of wages? In that case, does productivity growth make workers worse off, or better off?

It is, of course, neither the nominal wage nor the nominal price that determines the purchasing power of the worker. It is the real wage – the nominal wage divided by the price level, W/P – that determines the purchasing power of workers. If an increase in productivity lowers both W and P, it is crucial to determine which falls by more. If P falls by more, W/P will rise. If W falls by more, W/P will fall. This elementary point is not acknowledged in Marx's essay. It seems Marx may not have understood this point.

5.2 Can We Make Sense of Marx?

Marx may well have been confused. But it is also possible that Marx was implicitly making different assumptions than we make in the model we have developed earlier in this chapter. Exploring this possibility is a useful way to understand more deeply some of the assumptions our "mainstream" conclusions rest upon.

5.2.1 Produce More Goods or Hire Less Workers?

In the model we have developed, an increase in productivity results in higher output. The better technology makes it possible to produce more output per hour of labor employed. This leads the firms to demand more labor at a given wage since it is more profitable to hire workers at a given wage when productivity is higher. As a result, the labor demand curve in Figure 18 shifts outward.

But is this so obvious? In some passages of *Wage-Labour and Capital*, Marx seems to assume that what will happen instead is that higher productivity will result in less workers being hired. Recall that he says

> "the greater *division of labour* enables *one* worker to do the work of five, ten or twenty; it therefore multiplies competition among the workers fivefold, tenfold and twentyfold.

Here, it sounds as though workers are competing for the right to produce a fixed number of goods. If this is the case, higher productivity results in fewer workers being hired, since less workers are needed to produce the fixed set of goods. This means there will be more competition among workers, and lower wages. This idea seems to resonate strongly with many people today just as it did in the 19th century. These people are worried that new technologies destroy jobs and makes workers worse off. Some go so far as to predict a distopian future of mass unemployment and immiseration of workers.

(I should note that in other passages of *Wage-Labour and Capital* – in particular, the one quoted in section 5.1 – Marx seems to argue that greater division of labor will actually result in greater output. But let's ignore this apparent inconsistency across different parts of Marx's essay.)

Which view is correct? Does higher productivity result in more output or less work? This depends crucially on demand in the product market. A crucial question is: can the firm's sell the extra output? Consider a firm that has suddenly become twice as productive as before. Should it produce twice as much with the same amount of labor, or the same amount with half the labor? Suppose the firm decides to produce more. Who is going to buy these extra goods?

One concern is that people don't have enough income to buy the larger quantity of goods. But if someone does buy the goods, this provides extra income to the seller and the employees of the seller. This is an example of the circular flow of payments in a market economy. One person's expenditures are another person's income. This means that if expenditures go up, so does income. But there seems to be a chicken and egg problem: if the goods sell, then people have income to buy them, but which comes first the income or the expenditure?

An important assumption we have implicitly made in the model we developed is the assumption that *markets clear*. This assumption implies that everything that is produced will sell (at some price). Since everything sells, the sum of everyone's income is enough to buy all the goods produced. In this way, the chicken and egg problem is effectively assumed away in much of neoclassical economics. You may ask: what if the good sells at a really low price? Doesn't this mean the sellers have really low income? Yes, but this also means that buying the goods requires very low expenditure. As long as markets clear, there is no problem.

Or what? Actually, there is a large branch of macroeconomics devoted to the problem of insufficient demand: Keynesian macroeconomics. The model we developed earlier in this chapter is neoclassical. In this model, markets work well enough that insufficient demand does not arise as a problem. But this might not be the world we live in.

The crucial mechanism through which markets clear in neoclassical models is that prices and wages change. Consider what happens after an increase in productivity, starting from the prices and wages that previously prevailed. As Marx noted in the passage quoted in section 5.1, firms have an incentive to expand production since unit costs have fallen. Expanding production lowers P. Firms also have an incentive to hire more workers since the marginal product of labor is above the wage. Hiring more workers raises W. Both of these forces imply that workers purchasing power increases.

But what if wages and prices are stuck at prior levels for some reason. For example, consider the case where some friction prevents firms from lowering their prices. This will mean they can't expand demand for their goods. But if demand for their goods is fixed, they don't need as many workers since each worker produces more goods than before. Models with sticky prices or wages are therefore examples of models where demand can be insufficient to fully employ all workers. These models are called Keynesian models since they capture important elements of the market pathologies discussed in Keynes' *General Theory* (Keynes, 1936). We will discuss Keynesian models in quite a bit of detail later in this book.

This suggests that one way to make sense of Marx's arguments is that he was making a Keynesian argument long before Keynes did. But insufficient demand due to sticky prices is most plausible as an explanation for short run pathologies such as recessions. Marx was thinking of longer-run secular changes. Over longer periods, it is harder to justify the notion that prices and wages are sticky. However, the short-run stickiness of prices and wages may make it seem as though increases in productivity destroy jobs because this may be what happens in the short run.

5.2.2 Can Wages Rise Above the Cost of Production of Labor?

Another (charitable) way to interpret some of Marx's thinking in *Wage-Labour and Capital* is that he was making a "Malthusian" argument. Thomas Malthus famously argued in 1798 that wages would be forever stuck at extremely low levels – close to subsistence – because population growth outstrips increases in productivity. Without developing Malthus' ideas in detail – this will be done in chapter XX [Malthus chapter] – we can briefly explore how Malthusian logic (which was common among economists in the 19th century) plays a part in the argument Marx is making in *Wage-Labour and Capital*.

This line of Marx' thinking starts with the notion that wage-labor is just like any other commodity. Marx furthermore believed that the price of all commodities is determined by their cost of production. Since wage-labor is a commodity like any other, the price of labor (the wage) should then be determined by the "cost of production" of labor. Marx is quite explicit on this point:

> *[T]he price of labor will be determines by the cost of production, by the labour time necessary to produce this commodity—labour power.*
>
> *What, then is the cost of production of labour power?*
>
> *It is the cost required for maintaining the worker as a worker and of developing him into a worker.*
>
> The less the period of training, therefore, that any work requires the smaller is the cost of production of the worker and the lower is the price of his labour, his wages. In those branches of industry in which hardly any period of apprenticeship is required and where the mere bodily existence of the worker suffices, the cost necessary for his production is almost confined to the commodities necessary for keeping him alive and capable of working. The *price of his labour* will, therefore, be determined by the *price of the necessary means of subsistence.*
>
> Another consideration, however, also comes in. ... [I]n calculating the cost of production of simple labour power, there must be included the cost of reproduction, whereby the race of workers is enabled to multiply and to replace worn-out workers by new ones. Thus the depreciation of the worker is taken into account in the same way as the depreciation of the machine.
>
> The cost of production of simple labour power, therefore, amounts to the *cost of existence and reproduction of the worker.* The price of this cost of existence and reproduction constitutes wages.

If the wages of workers are determined by the "*cost required for maintaining the worker as a worker and of developing him into a worker*" then division of labor will lead wages to fall because it replaces skilled workers – which are more costly to train – with unskilled workers – which are cheaper to train.

How do these ideas differ from those in the model we developed earlier in this chapter? The main difference is that we took the size of the population as given and (implicitly) assumed that increases in productivity do not cause the population to change. In stark contrast, Malthus argued that increases in real wages led to faster population growth. Higher income allowed people to feed and clothe more children. More children, therefore, survived to adulthood and the population grew faster when wages were higher. (Malthus argued that the "passions between the

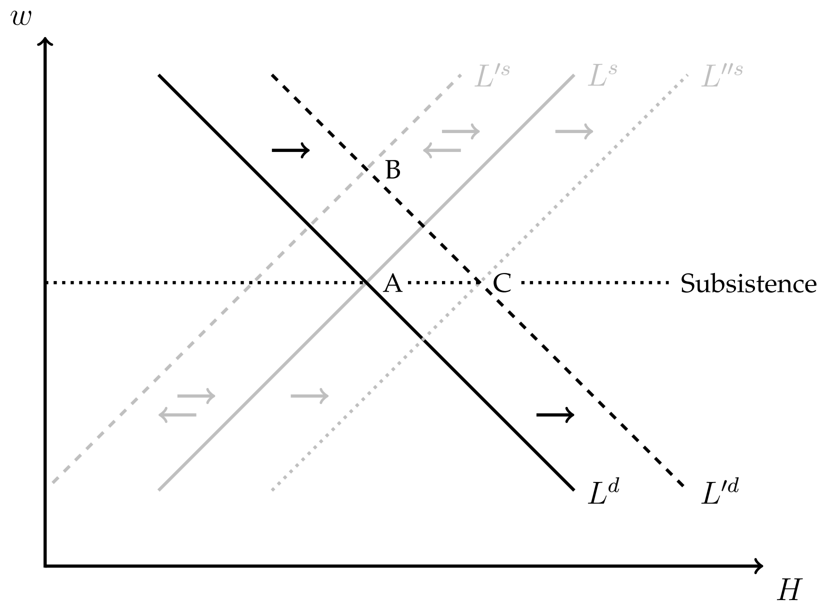

Figure 20: The Effect of an Increase in Productivity on the Labor Market

sexes" made it inevitable that many children would be born. The real question was how many survived to adulthood.)

Figure 20 illustrates the difference that the Malthusian assumption about population growth makes. An increase in productivity shifts the labor demand curve out, as in Figure 18. In the short run – before the population has time to adjust – the labor supply curve shifts back, as in Figure 18, and the economy moves from point A to point B. In the long run, however, higher wages lead the population to increase. This shifts out the labor supply curve (not because each person works more but because there are more people) and pushes wages back down. The population keeps increasing until real wages have shifted all the way back to their original point. In the long run, the economy therefore moves to point C and wages remain stagnant.

Malthusian ideas can seem strange to the point of being ridiculous to modern readers. Wages have risen for two hundred years. Furthermore, the relationship between income and fertility has switched sign: richer people tend to have fewer children today both within and across countries. But, arguably, Malthus' ideas provide a good description of the world prior to the year 1800 as we discuss in chapter XX [Malthus chapter]. It is therefore (arguably) understandable that Marx would make a Malthusian argument around 1850.

5.2.3 Immiserizing Technical Progress

In an economy where everyone produces the same good (a one-good economy), technical progress cannot make workers worse off for the simple reason that the workers always retain the option to use the old technology. Consider an economy where the only good produced is coconuts. Suppose there is a traditional technology for harvesting coconuts that yields 5 coconuts per hour of labor. Then at some point a newer technology is invented, which yields 10 coconuts per hour of labor. Suppose only young workers have the ability to learn the new technology. Will this technology make the older workers worse off. No! They can continue to use the old technology. If the labor market is competitive, their wage will continue to be 5 coconuts per hour. The young workers will earn more, since they will be using the new technology. But that doesn't affect the wage of the older worker, which remains 5 coconuts per hour.

This simple idea breaks down in an economy that produces more than one good. Suppose there are two goods that are produced in the economy: coconuts and textiles. Each person works either in the coconut industry or the textile industry. But all people consume both coconuts and textiles. In both industries there is a traditional technology. Early on, all workers use these technologies. Suppose for simplicity that the marginal product of workers in the coconut industry is 5 coconuts per hour, while the marginal product of workers in the textile industry is one yard of woven textile per hour. Suppose for simplicity that all prices and wages in the economy are denoted in coconuts and that the price of textiles is 5 coconuts per yard. In this case, workers in both industries are paid the same amount – 5 coconuts per hour – early on.

Now suppose a new technology is invented in the textile industry (think: spinning jenny or power-loom). Again, only younger workers are able to learn this new technology. But their productivity is much higher – 10 yards per hour worked. Does the introduction of this technology reduce the wages of the older workers that can only use the traditional textile technology? In this case, the answer is 'Yes, it does!'. The reason is that the increased supply of textiles reduces the price of textiles. The older textile workers continue to be able to produce one yard of textiles per hour. But since the price of textiles in terms of coconuts has fallen, the wages of the older textile workers in terms of coconuts has also fallen.

This is an example of what economists call a *terms of trade effect*. The textile workers sell textiles, but buy all goods. The fall in the price of textiles relative to coconuts implies that the price of the goods the textile workers sell has fallen relative to the price of the goods they buy. We say that their terms of trade have deteriorated. This makes the textile workers worse off. Terms of trade effects are important in international economics where the prices of a country's exports may vary relative to the prices of its imports. A famous paper by Jagdish Bhagwati showed that terms of trade effects can (in principle) be severe enough that technological progress at the country level can lead the country that improves its technology to become poorer. Bhagwati termed this immiserizing growth (Bhagwati, 1958).

A prominent example discussed by Marx in *Capital* of technical progress leading to a fall in wages was the introduction of the power-loom in weaving. The power loom yielded a huge increase in productivity in weaving. As a consequence, the price of woven cloth fell sharply and the wages of hand-loom weavers fell sharply as well. But the fall in the price of woven cloth benefited all other workers, who could now use their wages to purchase more clothes at lower prices. In other words, the fall in the price of woven cloth raised the real wages (W/P) of all workers other than hand-loom weavers.

The fate of hand-loom weavers is an example of the "perennial gale of creative destruction" that often characterizes capitalism (Schumpeter, 1942). Allen (2018) discusses how the invention of the power loom was itself a reaction to high wages of hand-loom weavers after the invention of the spinning jenny and water frame which both dramatically increased productivity in spinning. The traditional bottleneck in the production of textiles had been spinning. The new inventions in spinning cleared that bottleneck and led to a large increase in the supply of cheap yarn in the 1780s in England.

This dramatically increased demand for weaving. The number of hand-loom weavers and their wages shot up over the following 30 years. The period from 1780 to about 1820 was a golden age for hand-loom weavers. But the high wages of hand-loom weavers created strong incentives to invent a machine that replaced these workers. The power loom was that machine. Thus, as Allen puts it

> [T]he cottage mode of production was destroyed at its zenith not by the immiseration of the workforce but by its prosperity. Immiseration quickly followed, however, for wages collapsed in the 1820s and employment levels followed in the 1830s and 1840s.

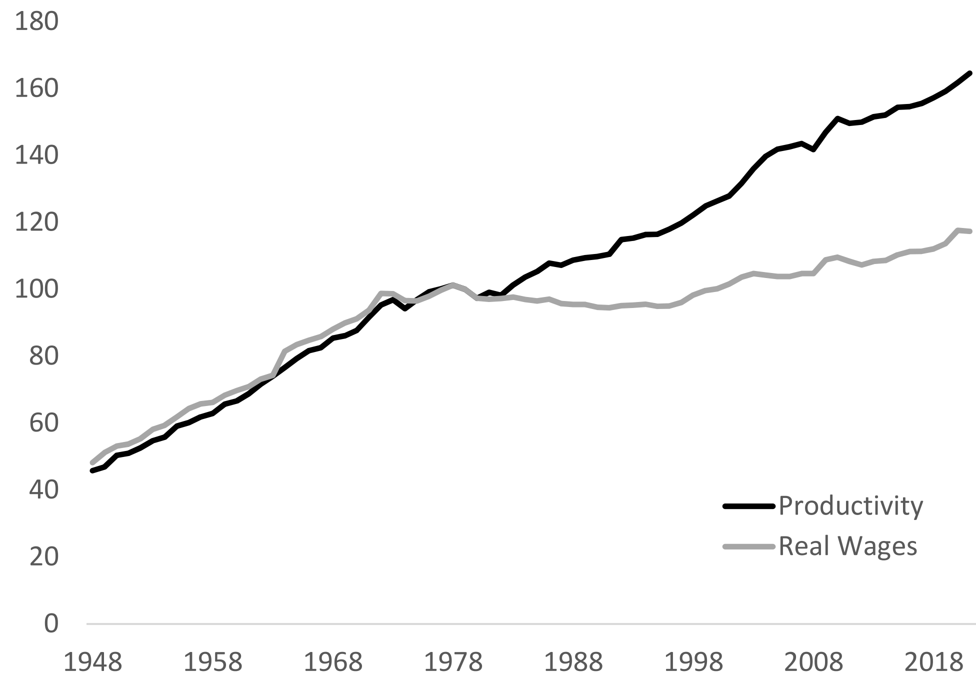

Figure 21: Productivity and Real Wages in the United States

Note: These data series were constructed by the Economic Policy Institute using methodology described in Bivens and Mishel (2015). The black line is productivity measured as output of goods and services less depreciation per hour worked. The gray line is hourly compensation (wages and benefits) of production and non-supervisory workers in the private sector. Data is from 1948 to 2021.

5.3 Was Marx Right After All?

For about 150 years Marx was dead wrong empirically about the relationship between productivity and real wages. Productivity rose and real wages rose along with it. But over the past 40 years, things have looked quite different. Figure 21 plots productivity and the real wages of production and non-supervisory workers in the U.S. from 1948 to 2021. Between 1948 and 1980, productivity and real wages increased in lock step. This is what the model we developed earlier in the chapter predicts should happen. But after 1980 the two lines diverge. Productivity continued to increase – albeit at a somewhat slower pace – but real wages stopped increasing and an ever widening gap opened up between changes in productivity and changes in real wages.

The period between 1980 and 1996 was particularly dismal. Over this period, productivity rose by 21% but real wages fell by 2%. Since 1996, real wages have started to rise again. But the rate at which they have risen has not kept up with productivity growth. Over the period between 1996 and 2021, productivity growth was 1.3% per year, while growth in real wages was 0.8%. Over this period, growth in real wages was about 60% as fast as growth in productivity.

Many details are important when making a comparison between productivity and real wages. First, it is important to consider whose real wages we are looking at. In Figure 21, it is private-sector production and non-supervisory workers. This group comprises over 80% of private sector employment, but excludes groups that tend to have very high income (supervisors). The idea is to consider typical workers. Second, it is important to consider a broad measure of compensation that includes not only wages but also benefits (such as health and pension benefits). The measure in Figure 21 includes both wages and benefits. A third important choice is how to convert nominal wages into real wages. For the real wage series in Figure 21, this is done using the consumer price index (CPI).

Figure 21 thus shows that increases in the real compensation of typical workers in the United States has lagged far behind increase in productivity since 1980. While Marx's prediction that productivity would lead to lower real wages has thankfully not come true, the simple model we presented earlier in this chapter needs some modification or augmentation to be able to account for the evolution of the labor market over the last 40 years. This is an important area of research as this is written.

References

Acosta, M., A. Mueller, E. Nakamura, and J. Steinsson (2023): "Macroeconomic Effects of UI Extensions at Short and Long Durations," NBER Working Paper No. 31784.

Alesina, A., E. Glaeser, and B. Sacerdote (2005): "Work and Leisure in the United States and Europe: Why So Different," *NBER Macroeconomics Annual*, 20, 1–64.

Allen, R. C. (2018): "The hand-loom weaver and the power loom: a Schumpeterian perspective," *European Review of Economic History*, 22, 1–22.

Angrist, J. D. and J.-S. Pischke (2015): *Mastering Metrics*, Princeton, NJ: Princeton University Press.

Atkeson, A. (2020): "What Will Be the Economic Impact of Covid-19 in the US? Rough Estimates of Disease Scenarios," NBER Working Paper No. 26867.

——— (2023): "The Impact of Vaccines and Behavior on U.S. Cumulative Deaths from COVID-19," NBER Working Paper No. 31525.

Bhagwati, J. (1958): "Immiserizing Growth: A Geometrical Note," *Review of Economic Studies*, 25, 201–205.

Bivens, J. and L. Mishel (2015): "The Historical Divergence between Producitivy and a Typical Worker's Pay," EPI Briefing Paper No. 406.

Boppart, T. and P. Krusell (2020): "Labor Supply in the Past, Present, and Future: A Balanced-Growth Perspective," *Journal of Political Economy*, 128, 118–157.

Chitwood, M. H., M. Russi, K. Gunasekera, J. Havumaki, F. Klaassen, V. E. Pitzer, J. A. Salomon, N. A. Swartwood, J. L. Warren, D. M. Weinberger, T. Cohen, and N. A. Menzies (2022): "Reconstructing the Course of the COVID-19 Epidemic over 2020 for US States and Counties: Results of a Bayesian Evidence Synthesis Model," *PLOS Computational Biology*, 18, e1010465.

Clark, G. (2005): "The Condition of the Working Class in England, 1209-2004," *Journal of Political Economy*, 113, 1307–1340.

——— (2010): "The Macroeconomic Aggregates for England 1209-1869," *Research in Economic History*, 27, 51–140.

Fenton, G. and F. Koenig (2023): "Labor Supply and Entertainment Innovations: Evidence from U.S. TV Rollout," Working Paper, Carnegie Mellon University.

Ferguson, N. M., D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. Okell, S. V. Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. Walker, C. Walters, P. Winskill,

C. Whittaker, C. Donnelly, S. Riley, and A. Ghani (2020): "Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcar Demand," Report by Imperial College COVID-19 Response Team.

Ganong, P., F. Greig, P. Noel, D. M. Sullivan, and J. Vavra (2023): "Spending and Job-Finding Impacts of Expanded Unemployment Benefits: Evidence from Administrative Micro Data," NBER Working Paper No. 30315.

Ganong, P., P. Noel, and J. Vavra (2020): "US Unemployment Insurance Replacement Rates During the Pandemic," *Journal of Public Economics*, 191, 104273.

Greenwood, J. and G. Vandenbroucke (2005): "Hours Worked: Long-Run Trends," NBER Working Paper No. 11629.

Keane, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49, 961–1075.

Keynes, J. M. (1930/1963): "The Economic Possibilities of our Grandchildren," in *Essays in Persuasion*, New York, NY: W.W. Norton & Co., 358–373.

——— (1936): *The General Theory of Employment Interest and Money*, London, UK: MacMillan and Co.

Kreps, D. M. (1990): *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press.

Martínez, I. Z., E. Saez, and M. Siegenthaler (2021): "Intertemporal Labor Supply Substitution? Evidence from the Swiss Income Tax Holidays," *American Economic Review*, 111, 506–546.

Marx, K. (1849/1978): "Wage Labour and Capital," in *The Marx-Engels Reader*, ed. by R. C. Tucker, New York, NY: W.W. Norton & Co., 203–217.

Milstein, E. and D. Wessel (2021): "What did the Fed Do in Response to the COVID-19 Crisis?" Hutchins Center on Fiscal and Monetary Policy, Brookings Institution.

Prescott, E. C. (2002): "Prosperity and Depression," *American Economic Review*, 92, 1–15.

Schumpeter, J. A. (1942): *Capitalism, Socialism, and Democracy*, New York, NY: Harper & Brother.

Sigurdsson, J. (2023): "Labor Supply Responses and Adjustment Frictions: A Tax-Free Year in Iceland," Working Paper, Stockholm University.

U.S. Bureau of the Census (1975): *Historical Statistics of the United States: Colonial Times to 1970*, Washington DC: U.S. Department of Commerce.

——— (2003): *Statistical Abstract of the United States*, Washington DC: U.S. Department of Commerce.

Varian, H. R. (2014): *Intermediate Microeconomics: A Modern Approach*, New York, NY: W.W. Norton & Co.

WHAPLES, R. (1990): "The Shortening of the American Work Week: An Economic and Historical Analysis of Its Context, Causes, and Consequenses," Ph.D. dissertation, University of Pennsylvania.

WILLIAMSON, J. G. (1995): "The Evolution of Global Labor Markets since 1830: Background Evidence and Hypotheses," *Explorations in Economic History*, 32, 141–196.