Multi-Plant Firms, Variable Capacity Utilization, and the Aggregate Hours Elasticity*

Domenico Ferraro Damián Pierri

October 10, 2024

Abstract

We develop an equilibrium business cycle model of multi-plant firms with perfectly competitive product and labor markets. Plant-level production features a minimum labor requirement, leading to occasionally binding capacity constraints at the firm level. The aggregate production function is kinked, displaying constant returns to scale when the economy is below capacity and decreasing returns when at capacity. We calibrate the model to U.S. data and show that the effects of distorting taxes are highly nonlinear and state-dependent, varying systematically with the state of the business cycle. The aggregate hours elasticity is higher in recessions and decreases with the size of the labor tax cut. Moreover, it differs from the structural preference parameter determining the individual-level labor supply elasticity.

JEL Classification: E22; E23; E24; E32; E62; H24; H25.

Keywords: Minimum labor requirement; Hours constraints; Capacity utilization; State dependence; Labor taxes; Aggregate hours elasticity.

---

*Ferraro: Department of Economics, W. P. Carey School of Business, Arizona State University, 501 E Orange St., Tempe, AZ 85281, United States (e-mail: domenico.ferraro@asu.edu); Pierri: ESCP Business School, C. de Arroyofresno, 1, Moncloa - Aravaca, 28035 Madrid, Spain and Departamento de Teoría Económica e Historia Económica, Universidad Autónoma de Madrid, C. Francisco Tomás y Valiente, 5, 28049 Madrid, Spain (e-mail: damian.pierri@gmail.com). We thank Florin Bilbiie, Evi Pappa, Morten Ravn, Kevin Reffett, Juan Pablo Rincón-Zapatero, and Hernan Seoane for valuable comments.

1. Introduction

In the United States, capacity utilization varies greatly over economic contractions and expansions, moving along a ceiling, corresponding to the full utilization of productive resources and temporarily plucked downwards by recessionary shocks (see Figure 1). Motivated by these patterns, this paper addresses the question of whether the effects of distorting taxes on aggregate hours worked and output are state-dependent, varying systematically based on the extent of capacity utilization.

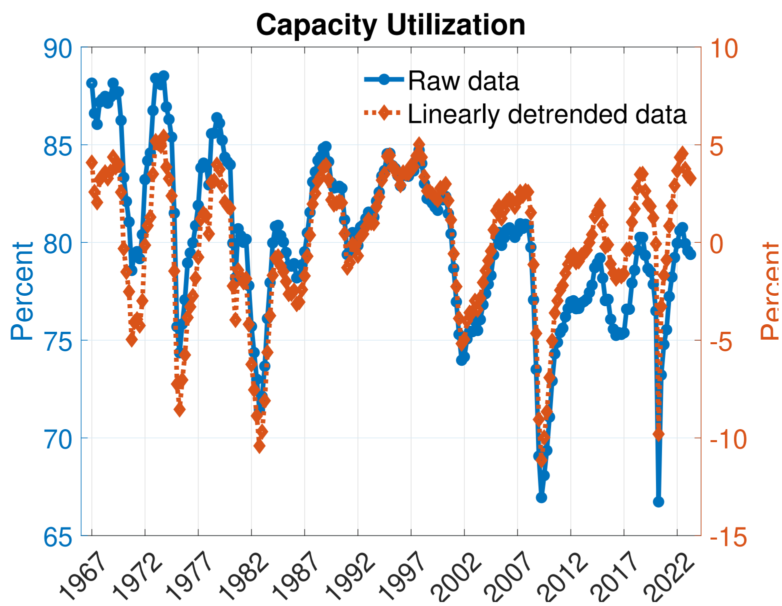


Figure 1: U.S. Capacity Utilization, 1967:Q1-2023:Q2

Notes: Board of Governors of the Federal Reserve System (US), Capacity Utilization: Total Index [TCU], percent of capacity, quarterly, seasonally adjusted, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TCU, October 1, 2023.

This paper makes two contributions related to the literature in Section 2. The first is to show that contrary to the conventional view and theorizing, one does not need to appeal to "frictions" that prevent market clearing, noncompetitive pricing, price stickiness, nor asymmetric adjustment costs as a source of nonlinear propagation and state dependence.[1]

[1] Parker (2011) summarizes the conventional wisdom as: "In this (old) Keynesian view, in recessions, markets are somehow failing, and these failures can be (probably imperfectly) rectified through fiscal policy that will return idle resources to work. In an expansion however, markets are somehow working and there are no benefits to expansionary fiscal policy as it would primarily raise interest rates and crowd out private consumption and investment, as in the neoclassical model of economic fluctuations."

Instead, we argue technological constraints imposing a minimum labor requirement on plant-level production can naturally lead to occasionally binding capacity constraints and, ultimately, to a *kinked* aggregate production function. The aggregate production set is thus a convex cone when the economy operates below capacity as in the standard real business cycle (RBC) model. In contrast, it bends, featuring aggregate diminishing returns when the economy operates at full capacity. In response to productivity or tax shocks that disturb labor demand and supply, firms decide whether to use all or a fraction of the available plants, leave some capacity idle temporarily, and change the labor input per plant under decreasing returns. We further develop and quantify this idea in an equilibrium business cycle model of multi-plant firms with perfectly competitive product and labor markets and show it embeds nonlinear propagation of tax shocks: the impact of tax changes is highly nonlinear and state-dependent, varying systematically with the state of the business cycle.

The second contribution is methodological. Occasionally binding capacity constraints embedded into a dynamic general equilibrium model in which distorting taxes evolve stochastically over time raise well-known challenges for characterizing and computing a non-Pareto optimal competitive equilibrium, adding new ones related to the additional complication that the aggregate production function is kinked and the marginal product of labor is discontinuous in our setup. We overcome such a challenge by adapting the monotone-map method or Coleman-Reffett operator (Coleman, 1990, 1991; Reffett, 1996) for the case of kinked production functions. Intuitively, for now, our approach capitalizes on the property of our setup that the location of the kink does not depend on prices, and the choice of an appropriate *equilibrium selection rule* according to which, at the kink, the quantity of labor is determined by the labor demand, whereas the marginal rate of substitution between consumption and leisure pins down the wage. Such a selection rule implies when the economy is at capacity, the wage is below the marginal product of labor, thus giving rise to quasi-rents, which differ from the rents due to decreasing returns.

Section 3 illustrates the main idea using a static model with multi-plant firms where labor is the only input. The model formalizes a theory of capacity constraints, which yields convex supply curves consistent with the evidence in Boehm and Pandalai-Nayar (2022). Firms choose how many plants to operate and how much labor to allocate to each plant. We assume that a firm is *not* free to select any number of work hours as typically the case; instead, it faces a technological constraint at the plant level that puts a floor on the number of hours required to produce output. There is a difference between work hours and labor services. For hours of work below a minimum level, labor services are insufficient to produce output. Conversely, output is produced under diminishing returns for work hours above the minimum level as in the Lucas (1978)'s span-of-control model.

Such minimum labor requirement or hours constraint can arise from technological factors or forces related to the organization of production, such as coordination of work schedules, as in an assembly line, or more generally in team production, time spent in setting up machines, and calibrating equipment (Prescott, Rogerson and Wallenius, 2009; Rogerson, 2011). This formulation of technology gives an extensive and intensive margin of labor demand. The former relates to the firm's capacity choice of the mass of active plants and the latter to the labor allocation per plant.

While the plant's production set is nonconvex, firms' capacity choices convexify the aggregate production set. The aggregate production function is thus endogenous, linear-concave with a kink: it exhibits constant returns to scale for low values of the labor input when the economy operates below capacity and decreasing returns for high values of the labor input when the economy operates at full capacity. Since the hours constraint is more likely to bind during recessions, state-dependent effects of proportional tax changes naturally arise as an equilibrium phenomenon.

During recessions, the minimum labor requirement binds so that firms leave some plants idle. In this scenario, the extensive margin of labor demand that operates through the mass of active plants soaks up the adjustments to shocks. The aggregate production function is linear, and hours worked are more responsive to shocks. During expansions, all plants are utilized; the adjustment to shocks comes from the intensive margin of labor demand, which operates through changes in hours per plant. The aggregate production function displays diminishing returns, and total hours are less responsive to shocks. As a result, the aggregate hours elasticity is state-dependent. Moreover, an important point is that occasionally-binding capacity constraints break the separation between preferences and technology in determining the aggregate elasticity of labor supply as in Hornstein and Prescott (1993).

Section 4 extends the static model to infinite-horizon with capital accumulation. The extended model looks like a standard RBC model when the economy operates below capacity, displaying overall constant returns to scale in production, and Lucas's span-of-control model when the economy operates at capacity, displaying overall diminishing returns. Recessions and expansions are not alike: the propagation mechanisms of shocks change qualitatively depending on whether the economy is at or below capacity.

A few challenges arise when moving to a dynamic environment subject to stochastic shocks to technology and tax rates. As taxes are distorting, the competitive equilibrium is not Pareto optimal, so we cannot use the social planner's formulation to calculate the economy's equilibrium allocations. Moreover, we cannot rely on local approximation methods such as perturbations, as the capacity constraints bind only occasionally, and we want to preserve nonlinearities.

Sections 5 and 6 calibrate the model to U.S. data and use it to study the propagation mechanism and quantify the extent to which the effects of proportional labor tax shocks vary depending on capacity utilization. We focus on labor taxes, which directly impinge on labor allocations by disturbing labor supply. They also represent the vast majority of tax revenues for the U.S. government sector at the federal, state, and local levels. To gauge the model's nonlinear propagation mechanism, we compute *generalized impulse response functions* (GIRFs) by simulating several artificial economies hit by labor tax cuts and hikes of varying magnitude, further conditioning on whether the economy is below or at full capacity. Unlike the more widely used linear IRFs, GIRFs preserve nonlinearities and state dependence in the propagation mechanism. To be sure, meaningful computation of such GIRFs critically hinges on the existence, uniqueness, and ergodicity of the model's equilibrium, which we prove in Section 4.

Labor tax rate shocks are symmetric as generated by a first-order retrogressive AR(1) process. Hence, the resulting nonlinear effects of the tax shocks are not hard-wired into the model; instead, they endogenously arise as an equilibrium phenomenon due to the model's highly nonlinear propagation mechanism. We calculate effective average tax rates on labor income using Mendoza, Razin and Tesar (1994)'s method and use their time series post-1950 to estimate the parameters of the AR(1) process governing tax shocks.

The effects of such labor tax shocks are highly nonlinear, displaying pronounced state dependence. For example, the aggregate hours response to a tax cut is considerably larger in recessions than in expansions when the economy is at full capacity. The aggregate hours elasticity is state-dependent, higher in recessions, consistent with the evidence in Attanasio et al. (2018). Moreover, its magnitude varies considerably based on the sign and size of the labor tax rate shock. Across all experiments, the aggregate or macro hours elasticity does not coincide with the structural preference parameter determining labor supply's individual-level or micro elasticity. They turn out to be quite different, highlighting the role of hours constraints and technological non-convexities in understanding how work incentives and relative price distortions affect aggregate labor supply.

2. Related Literature

This paper contributes to further our understanding of how distorting taxes affects hours worked when capacity utilization varies over contractions and expansions. In doing so, it also makes a methodological contribution, developing a novel approach to characterize and compute nonoptimal economies when the aggregate production possibility sets are kinked, a nontrivial departure from standard competitive equilibrium theory.

Our work closely relates to two strands of literature in addition to the work cited in the previous section. The first strand focuses on the general equilibrium effects of distorting taxes in RBC models in which disturbances to labor and capital tax rates induce variation in aggregate quantities and prices (Aiyagari, Christiano and Eichenbaum, 1992; Baxter and King, 1993; Braun, 1994; Burnside, Eichenbaum and Fisher, 2004; McGrattan, 1994, 2012; McGrattan and Prescott, 2005; McGrattan, Rogerson and Wright, 1997; McGrattan and Ohanian, 2010). In these models, proportional tax rates impinge on allocations by distorting relative prices. As a result, temporary yet persistent disturbances to labor and capital tax rates activate intertemporal substitution. In a spot labor market, time spent working and total hours worked respond instantaneously to changes in work incentives, while the physical capital stock responds sluggishly. Capital accumulation propagates tax rate disturbances, adding to the persistence of fluctuations. While magnitudes differ, RBC models predict that the observed variation in tax rates explains a significant share of U.S. output and hours variability.

This literature typically studies equilibrium dynamics using approximation methods that deliver a first-order approximation to the model's solution around the deterministic steady state, which by construction precludes nonlinearities in the propagation of tax shocks, let alone state dependence. However, it is well-known that RBC models yield approximately log-linear stochastic processes for aggregate variables even when solved with methods that preserve potential nonlinearities (Aruoba, Fernandez-Villaverde and Rubio-Ramirez, 2006).

What distinguishes our work from previous studies in the RBC tradition is the focus on time-varying capacity utilization as a propagation mechanism of tax rate disturbances. We provide a quantitative theory in which tax rate shocks occasionally induce capacity constraints to bind. The effects of tax policy changes are then naturally state-dependent. The marginal impact of a given tax rate change critically depends on the state of the business cycle, whether the economy is operating below or at capacity.

Perhaps the closest work to ours is Hansen and Prescott (2005). This paper studies cyclical asymmetry in U.S. hours using a model that shares our production structure but with solely productivity shocks. They compute the competitive equilibrium indirectly by capitalizing on the equivalence between the decentralized equilibrium and the social planning solution. By contrast, our model economy is nonoptimal due to distorting taxes, forcing us to solve for the competitive equilibrium and prices directly. To this aim, we adapt the Coleman-Reffett operator to an environment with a kinked production function without relying on local approximation methods.

The second strand of literature directly addresses whether the effects of tax policy vary over the business cycle. This long-standing question has received renewed interest in recent years, leading to important empirical work (Auerbach and Gorodnichenko, 2012; Gonçalves et al., 2024; Ramey and Zubairy, 2018) and theorizing. The theoretical literature on this topic that uses general equilibrium models typically resorts to search frictions that prevent markets from clearing in the Walrasian sense. Moreover, frictions worsen during recessions, giving rise to state dependence (Ferraro and Fiori, 2023*a*; Ghassibe and Zanetti, 2022; Michaillat, 2014; Pizzinelli, Theodoridis and Zanetti, 2020). In these models, hiring costs are convex in aggregate employment akin to models with asymmetric adjustment costs, inducing sluggish and asymmetric employment dynamics (Dupraz, Nakamura and Steinsson, 2020; Ferraro, 2018, 2023; Ferraro and Fiori, 2023*b*).

Reliance on search frictions or, more generally, asymmetric adjustment costs rests on the premise that equilibrium business cycle models in which product and labor markets function well cannot generate significant nonlinearities. Models with nominal rigidities share the same property unless the zero lower bound on the nominal interest rate binds (Christiano, Eichenbaum and Rebelo, 2011), through state-dependent pricing (Burstein, 2006; Devereux and Siu, 2007), or downward nominal wage rigidities (Abbritti and Fahr, 2013; Barnichon, Debortoli and Matthes, 2022)

Our substantive contribution relative to this literature is to show that nonlinearities arise without search frictions, price rigidities, or asymmetric adjustment costs if plants are subject to a minimum labor requirement. During recessions, the extensive margin of labor demand through the number of active plants soaks up all the adjustments. By contrast, firms use all plants during expansions, so the intensive margin of labor demand of hours per plant takes the lion's share of the economy's dynamic adjustment. The latter margin is subject to decreasing returns; the former is not.

3. Illustrating the Idea in a Static Model

This section lays out a static general equilibrium model to illustrate our central idea, which we later embed into a stochastic neoclassical growth model. The model's non-standard feature is that a firm is not free to choose any number of hours as typically in standard production theory; instead, it faces a technological constraint that puts a *floor* on the number of hours required to produce output. With this simple model, we establish some qualitative results and discuss features of the model that will later be of quantitative importance in our numerical analysis.

3.1. The Minimum Labor Requirement

The technology has diminishing returns as in the Lucas (1978)'s span-of-control model. A production unit, say, a "plant," with h units of labor input produces $f(h) = zh^{\phi}$ units of output, with $0 < \phi < 1$. To capture the idea of a minimum labor requirement, we assume that $f(h) = 0$ for $0 \leq h < \overline{h}$ and $f(h) > 0$ for $h \geq \overline{h}$, where $\overline{h} \in (0, 1)$ is the minimum amount of labor input required to produce output. This formulation hinges on the notion that working hours differ from labor services. In particular, we assume that labor services are zero for $h < \overline{h}$ and equal to h for all $h \geq \overline{h}$.[2]

This assumption is meant to capture in a tractable way technological factors and forces related to the organization of production, such as coordination of workers within the firm, time spent to set up machines, and minimum efficient scale. As an empirical matter, note that the property of a minimum labor requirement is consistent with the observation that for many activities, firms do not consider hiring part-time workers.

The technological requirement of a minimum labor input makes the production set of the plant *non-convex*, as linear combinations of zero and $\overline{h}$ are not allowed. At the plant level, the marginal product of labor (MPL) is discontinuous at $h = \overline{h}$ as $f_h(h) = 0$ for $0 \leq h < \overline{h}$ and $f_h(h) > 0$ for $h \geq \overline{h}$. In addition, because of diminishing returns, the average product of labor, $APL \equiv f(h)/h = zh^{\phi-1}$, exceeds $MPL = \phi z h^{\phi-1}$, for all $h \geq \overline{h}$. In a competitive labor market without the minimum labor requirement ($\overline{h} = 0$), profit maximization requires $MPL = W$, which gives a downward-sloping labor demand

[2] Prescott, Rogerson and Wallenius (2009) use a similar argument to generate an intensive and extensive margin of labor supply. Note also that a formulation in which labor services are concave in hours of work above the minimum level is subsumed in the current setup. To see this, think of the technology $f(g(h))$ in which g is strictly increasing and concave.

curve, $h = (\phi z/W)^{\frac{1}{1-\phi}}$. Since $APL > MPL$, the average cost of production is smaller than the marginal cost so that profits are positive and equal to $\Pi = (1-\phi)zh^{\phi} > 0$.

3.2. Single-Plant vs. Multi-Plant Firms

We consider two alternative environments characterized by a different organization of production. In the first environment with *single-plant* firms, there is a measure $M \equiv 1$ of identical firms, each operating the technology $f(h) = zh^{\phi}$, subject to the minimum labor requirement described above. This case is a helpful benchmark as the standard definition of a competitive equilibrium renders a model of an economy operating at "full capacity" at all times. Unless one deviates from the assumption of identical firms, idle capacity, in the sense that some technologies remain unused, cannot arise as an equilibrium.

In the second environment with *multi-plant* firms, instead, there is a large number of identical firms, each operating multiple, again, identical, technologies or plants. In such a model, production decisions are coordinated in the sense that the measure $m \leq M$ of active plants and the labor input h allocated to each plant are jointly chosen. In this setting, the possibility of "idle capacity," defined as $m < M$, naturally arises as a potential competitive equilibrium outcome.

3.2.1. Equilibrium with Single-Plant Firms

There is a continuum of identical firms indexed by $i \in [0, M]$, each maximizing profits $\Pi \equiv zh^{\phi} - Wh$. The first-order condition (FOC) for h is

$$z\phi h^{\phi-1} \leq W, \tag{1}$$

with equality for $h \geq \overline{h}$. To understand the solution to the profit maximization problem, it is helpful to define two cutoffs on the wage. The first cutoff $\underline{w} \equiv MPL(\bar{h}) = z\phi\overline{h}^{\phi-1}$ is such that if $W < \underline{w}$, the FOC (1) holds with an equal sign and $\pi > 0$ for all $h > \overline{h}$. Since in this case the minimum labor requirement constraint does not bind, the equation $z\phi h^{\phi-1} = W$ gives a downward-sloping labor demand curve $h = (\phi z/W)^{\frac{1}{1-\phi}}$. The second cutoff $\overline{w} \equiv APL(\overline{h}) = z\overline{h}^{\phi-1} > \underline{w}$ is such that if $W > \overline{w}$, then $\pi < 0$ for all $h \geq \overline{h}$. In this latter case, profit maximization requires $h = 0$.

Thus, the labor demand curve for firm i is

$$h_i(W) = \begin{cases} 0 & \text{if} \quad W > \overline{w} \\ \overline{h} & \text{if} \quad \overline{w} \geq W \geq \underline{w} \\ \left(\frac{\phi z}{W}\right)^{\frac{1}{1-\phi}} & \text{if} \quad W < \underline{w} \end{cases}. \tag{2}$$

For future reference, we note that the labor demand curve (2) features a kink at $(\overline{h}, \underline{w})$, thus we refer to it as "kinked labor demand."

Summing over firms, and using the fact that the labor demand is the same across firms, we obtain the aggregate demand for labor as

$$H(W) \equiv \int_0^M h_i(W)di = Mh(W) = \begin{cases} 0 & \text{if} \quad W > \overline{w} \\ M\overline{h} & \text{if} \quad \overline{w} \geq W \geq \underline{w} \\ M\left(\frac{\phi z}{W}\right)^{\frac{1}{1-\phi}} & \text{if} \quad W < \underline{w} \end{cases}, \tag{3}$$

which inherits the kink at $(\overline{h}, \underline{w})$ from the firm's labor demand curve.

Market clearing requires that labor supply, $L(W)$, equals labor demand, $H(W)$, so that $L(W) = H(W) = Mh(W)$ determines the wage. Note that if $\overline{h} = 0$, i.e., in the standard case of no minimum labor requirement, using standard arguments, one can show that a competitive equilibrium always exists and is unique. If instead $\overline{h} > 0$, an equilibrium wage equating labor demand and supply need not exists. For example, this could happen if the labor supply curve is upward-sloping and sufficiently steep, a possibility with GHH preferences (Greenwood, Hercowitz and Huffman, 1988), or in the case of a vertical labor supply with KPR preferences (King, Plosser and Rebelo, 1988) if households are willing to work less than $M\overline{h}$ hours. Note also that even in the case in which an intersection existed, the equilibrium of the model would describe an economy operating at full capacity at all times, irrespective of the level of technology or taxes.

3.2.2. Equilibrium with Multi-Plant Firms

As shown in the previous subsection, a minimum labor requirement with uncoordinated production decisions yields a non-convex aggregate production set. As is well-known, a non-convexity at the aggregate level is troubling for several reasons, including the fact that there is no general proof of existence and uniqueness of a competitive equilibrium, let alone a reliable algorithm to compute it (Brown, 1991). Of course, these issues get even more challenging when one turns to a dynamic stochastic general equilibrium model.

Also, the equilibrium with single-plant firms features a *floor* on total hours worked, which is unpalatable if one aims to account for the observed asymmetry in hours worked over the U.S. business cycle (Hansen and Prescott, 2005; McKay and Reis, 2008). The kinked labor demand curve (2) implies that wages must fall a great deal during recessions to compensate for the fact that hours cannot fall below the floor, a pattern we do not see in the data.

Finally, in the U.S., the extent of idle capacity changes over time, arguably in response to a number of shocks. The model with identical firms cannot account for that variation in the data, as it gives a model economy that operates at full capacity at all points in time.

Against this background, we now consider an alternative environment that is immune to such criticisms. Specifically, there is a large number of firms, say, a unit mass, each solving the problem of allocating H hours of labor services among m identical plants, subject to the capacity constraint on the number of plants, $m \leq M$, and the feasibility constraint $mh \leq H$, where $h \geq \overline{h}$ is the minimum labor requirement per plant. Production decisions are thus coordinated in the sense that the mass of producing plants and the labor allocated to each plant are jointly chosen.

The firm's allocation problem can be split in two parts. First, the firm chooses how many plants to operate, m, given the total labor input, H, a choice to which we refer to as "extensive margin" of labor demand. Second, the firm chooses H given m, taking the wage as given, which gives the labor input per plant as $h = H/m \geq \overline{h}$, a choice to which we refer to as "intensive margin" of labor demand.

Extensive Margin of Labor Demand. The optimal capacity choice of m is determined by maximizing output taking H and the capacity constraint $m \leq M$ as given:

$$F(H) = \max_{m \leq \min\left\{M, \frac{H}{\overline{h}}\right\}} mz\left(\frac{H}{m}\right)^{\phi} = \max_{m \leq \min\left\{M, \frac{H}{\overline{h}}\right\}} zH^{\phi}m^{1-\phi}, \tag{4}$$

where we used the assumption that the technology is the same across plants. As the right-hand side of (4) is increasing in m, the firm's production set exhibits two regions:

(i) "full capacity," i.e., $Y \leq F_{\text{full}}(H) = zH^{\phi}M^{1-\phi}$, for $m = M \leq H/\overline{h}$, or equivalently $H \geq \overline{H} \equiv M\overline{h}$;

(ii) "idle capacity," i.e., $Y \leq F_{\text{idle}}(H) = z\overline{h}^{\phi-1}H$, for $m = H/\bar{h} \leq M$, or equivalently

$H \leq \overline{H}$.

As illustrated in Figure 2, the firm's production set remains convex, in spite of the non-convexity at the plant level.

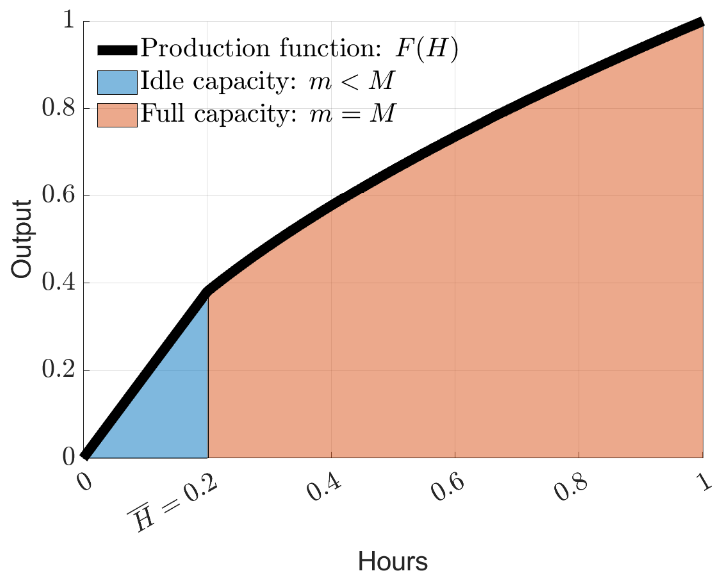


Figure 2: Linear-Concave Production

Notes: The figure depicts the firm's production set. To construct it, we use (5) with $\phi = 0.6$, $\overline{h} = 0.2$, $M = 1$, so that $\overline{H} = M\overline{h} = 0.2$, and $z = 1$.

The firm's production function $F(H)$ is *linear-concave* with a *kink* at $H = \overline{H}$:

$$F(H) = \begin{cases} z\overline{h}^{\phi-1}H & \text{for} \quad H \leq \overline{H} \\ zH^{\phi}M^{1-\phi} & \text{for} \quad H \geq \underline{H} \end{cases} . \tag{5}$$

In a nutshell, a binding capacity constraint on the number of plants induces decreasing returns to labor. Notably, the shape of the aggregate production function is endogenous: constant returns to scale (CRS) for low levels of labor input when the capacity constraint is slack and decreasing returns to scale (DRS) for high levels of labor input when the capacity constraint is binding.[3]

[3]In Section 4, we show that the same logic applies when we add capital to the model. When the capacity choice is maximized out, the production function with capital and labor as inputs is CRS for $m < M$ and DRS for $m = M$.

Intensive Margin of Labor Demand. Given (5), the firm's MPL exhibits a discontinuity at $H = \overline{H}$:

$$F_H(H) = \begin{cases} z\overline{h}^{\phi-1} & \text{for} \quad H \leq \overline{H} \\ z\phi M^{1-\phi} H^{\phi-1} & \text{for} \quad H \geq \overline{H} \end{cases} . \tag{6}$$

Such discontinuity makes the firm's profit maximization problem nonstandard. Indeed, to characterize the profit-maximizing choice of hours, a generalization of the notion of optimality is needed, to which we turn next.

To begin, we write explicitly the firm's profit maximization problem:

$$\max_{\{H\}} \Pi(H; W) \equiv \begin{cases} z\overline{h}^{\phi-1} H - WH & \text{for} \quad H \leq \overline{H} \\ zH^{\phi} M^{1-\phi} - WH & \text{for} \quad H \geq \overline{H} \end{cases} . \tag{7}$$

Absent the minimum labor requirement ($\overline{h} = 0$), the firm's optimal choice of hours is characterized as $H(W) = \arg\max \{\Pi(H; W) | \Pi_H(H; W) = 0\}$, where $\Pi_H(H; W)$ is the derivative of the profit function with respect to H. For $\overline{h} > 0$, this characterization does no longer work as the kink in MPL gives a "hole" in the labor demand curve. As implied by the linear-concave production function (5), MPL is constant for all $H \leq \overline{H}$, decreasing and convex for all $H \geq \overline{H}$, with a discontinuity at $H = \overline{H}$. (Figure 3(a) shows a numerical example of labor demand that illustrates these points.)

To characterize labor demand in the presence of such discontinuity, we define two cutoffs on the wage. The first cutoff, $\overline{w} \equiv z\overline{h}^{\phi-1}$, is the marginal and the average product of labor on the left of the kink: $APL^-(\overline{h}) = MPL^-(\overline{h}) = \overline{w}$, given the linearity of the production function for $H \leq \overline{H}$. The second cutoff $\underline{w} \equiv z\phi M^{1-\phi}\overline{H}^{\phi-1} < \overline{w}$ is the MPL at the kink for the concave part of the production function.

At the kink, the following relationships hold:

$$APL^-(\overline{h}) = z\overline{h}^{\phi-1} = APL^+(\overline{h}) = z\overline{H}^{\phi} M^{1-\phi} = \overline{w} > \underline{w}. \tag{8}$$

Thus, for all wages that fall into the hole of the labor demand curve, i.e., for all $W \in [\underline{w}, \overline{w}]$, the firm makes non-negative profits. So, the shape of the aggregate labor demand is endogenous and depends on the level of the labor input: (i) horizontal, and so perfectly elastic, for all $H < \overline{H}$; (ii) vertical, and so perfectly inelastic, at $H = \overline{H}$; and (iii) decreasing and convex for all $H > \overline{H}$.

Equilibrium with a Discontinuous Labor Demand. Equipped with such generalized labor demand, we are ready to study the equilibrium of the labor market. Labor demand features a hole so that the wage would be discontinuous under standard competitive pricing, i.e., wage $= MPL$. Thus, we must take a stand on wage determination to bypass this issue. Here, we assume that workers are always on their labor supply curve so that the marginal rate of substitution between consumption and leisure pins down the wage, whereas the labor demand pins down the quantity of labor. This choice amounts to an "equilibrium selection rule," which, as shown in the following section, has nice theoretical properties, preserving the existence, uniqueness, and ergodicity of the equilibrium in the infinite-horizon model with physical capital accumulation and aggregate uncertainty.

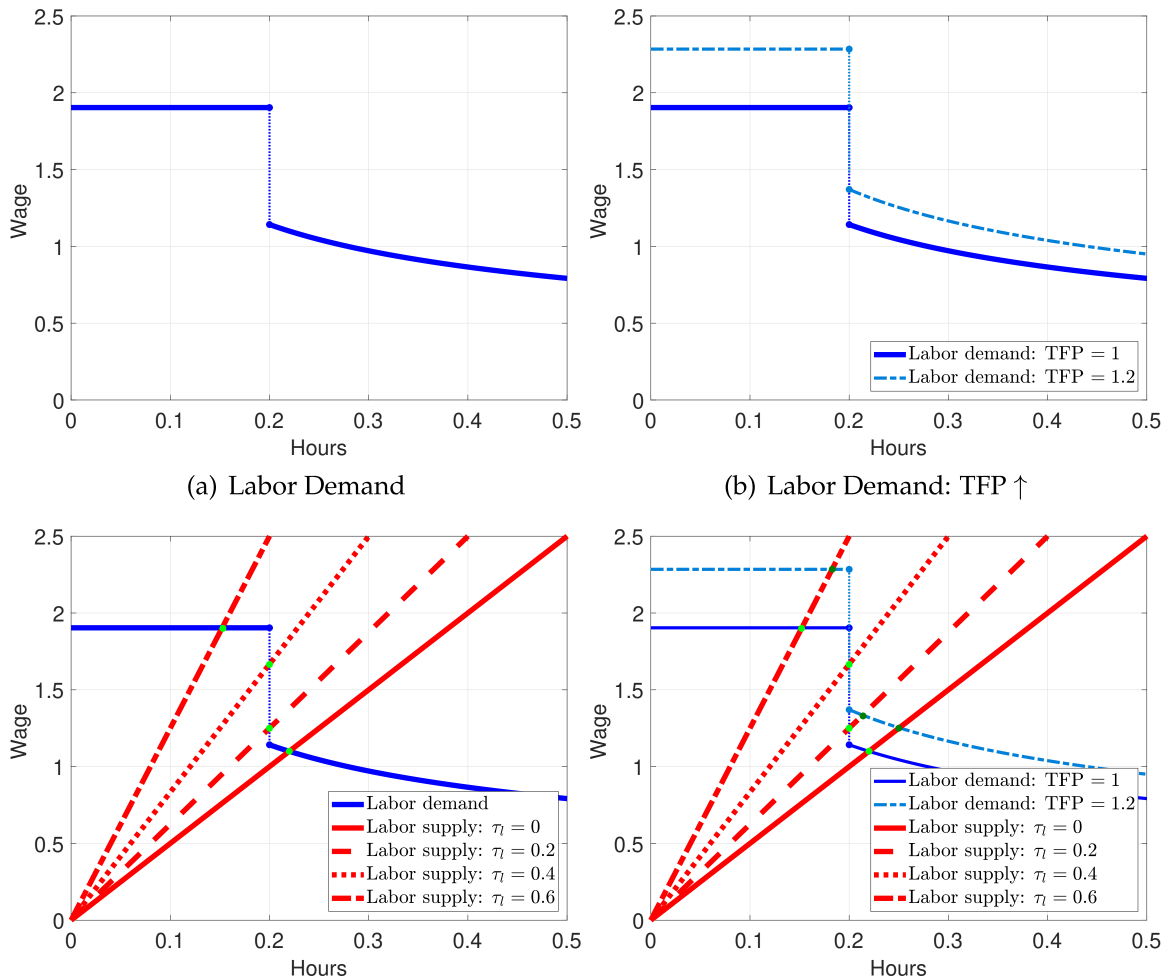
Figure 3(c) illustrates the labor market equilibrium with different labor supply curves, indexed by labor tax rate, τ_l. Linear supply curves derive from GHH preferences in which labor supply depends only on the after-tax wage, and assuming a unitary labor supply elasticity ($\eta = 1$). For $\tau_l = 0$, the equilibrium is the textbook case of the intersection between an upward-sloping labor supply curve and a downward-sloping labor demand curve, jointly determining wage and hours. The equilibrium is again standard looking for $\tau_l = 0.6$, with the difference that the labor demand curve is now horizontal, such that the wage is determined by the labor demand and hours by the labor supply.

The more interesting and novel cases are those ones where the labor supply curves go through the hole of the labor demand curve. In these scenarios, standard logic breaks down. Labor demand at $H = \overline{H}$ pins down hours worked, whereas labor supply pins down the wage. Specifically, a leftward shift in labor supply, caused by an increase in the flat-rate tax on labor income, leads to a higher wage, leaving hours unchanged. At these equilibria, workers are indifferent between working and not working as the wage equals the marginal rate of substitution. Firms make positive profits.[4]

3.3. Inspecting the Mechanism

We now illustrate the effects of changes in productivity and labor tax rates on the static labor market equilibrium with a numerical example. Perhaps not surprisingly, given the discussion above, we show that such effects are highly nonlinear. Notably, the impact of a change in TFP and the labor tax rate critically depends on the equilibrium's location on the labor demand curve. Overall, hours worked are more sensitive to tax shocks during recessions when the economy features idle capacity than during expansions.

[4]Note that the wage is below the marginal product of labor at the kink.



(a) Labor Demand

(b) Labor Demand: TFP ↑

(c) Equilibrium

(d) Equilibrium: TFP ↑

Figure 3: Static Labor Market Equilibrium

Notes: In all panels, labor demand is constructed using (6), with $\phi = 0.6$, $\overline{h} = 0.2$, and $M = 1$, so that $\overline{H} = M\overline{h} = 0.2$. In panels (a) and (b), $z = 1$. In panels (c) and (d), $z = 1.2$. Labor supply is $l = [(1 - \tau_l)\, W/\alpha]^{\eta}$, with $\eta = 1$, $\alpha = 5$, and $\tau_l = \{0, 0.2, 0.4, 0.6\}$.

Productivity Shocks as Labor Demand Shifters. Figure 3(a) shows the aggregate labor demand implied by the model where the discontinuity associated with the kink in the production function occurs at $H = \overline{H} \equiv M\overline{h}$. Figure 3(b) further illustrates how the labor demand shifts if TFP increases. As the kink's location does *not* depend on TFP, there is an upward shift for $H < \overline{H}$, and a rightward shift for $H > \overline{H}$, with the kink remaining at $H = \overline{H}$. Of course, one can also use the figure to consider the effects of a TFP decrease in which the initial labor demand position is the outer dot-dashed line and the new labor demand is the solid line associated with a lower TFP level.

Figure 3(d) shows how the labor market equilibrium with different configurations of the labor tax rate in Figure 3(c) changes when TFP increases. Alternatively, one can view the figure as showing how the effects of tax rate changes depend on the location of the labor demand curve, a proxy for the state of the business cycle. The impact of TFP changes critically depend on the equilibrium location on the labor demand curve, which naturally depends on the value of the tax rate.

First, if the initial equilibrium's location is in the hole of the labor demand curve, $H = \overline{H}$, neither the wage nor hours change as long as TFP shocks represent relatively small perturbations around the initial equilibrium. If TFP shocks are large enough to push the economy in the $H < \overline{H}$ or $H > \overline{H}$ regions, then wages and hours will adjust. Suppose the TFP shock is sufficiently large and positive. In that case, the labor supply curve might intersect the new labor demand curve in the decreasing returns to scale region, mandating increased hours worked and wages. By contrast, if the TFP shock is sufficiently large and negative, the labor supply curve might intersect the labor demand curve in the constant returns to scale part, mandating a drop in the wage and hours. Second, if the initial equilibrium is in the full capacity region, $H > \overline{H}$, changes in TFP induce small changes in wages and hours. If, instead, the initial equilibrium is in the idle capacity region with $H < \overline{H}$, TFP changes wages and hours significantly. This latter scenario resembles a standard RBC model's adjustment with constant returns to scale technology.

Tax Rate Shocks as Labor Supply Shifters. We now turn to the effects of changes in the labor tax rate and how they depend on the location of the labor demand curve.[5] As shown in Figure 3(c), a cut in the labor tax rate shifts the labor supply curve clockwise, implying that households are willing to supply more work hours for a given wage.

There are three different scenarios to analyze. First, suppose the initial equilibrium is in the idle capacity region, $H < \overline{H}$. In that case, wages are constant so that hours worked take up all the adjustment needed to restore the equilibrium in the labor market. In this scenario, hours worked are extremely sensitive to tax rate changes as the labor demand curve is infinitely elastic. Second, suppose the initial equilibrium is in the full capacity region, $H \geq \overline{H}$. In that case, there are two scenarios: (i) for $H = \overline{H}$, changes in tax rates have no effect on hours worked so that wages must adjust to equate labor demand and supply; (ii) for $H > \overline{H}$, both wages and hours adjust in response to tax rate changes.

[5] In a one-period model with GHH preferences, proportional capital income taxes have no effect on equilibrium labor and wages.

4. Infinite-Horizon Model with Physical Capital

In this section, we embed multi-plant firms subject to a minimum labor requirement in production into a stochastic neoclassical growth model. Uncertainty takes the form of stochastic shocks to the level of TFP and proportional tax rates. Since the economy is nonoptimal, we cannot rely on the centralized solution of the planning problem; instead, we need to solve for the competitive equilibrium directly. The main goal is to understand how the technological requirement of a minimum labor input at the plant level affects the economy's response to labor tax shocks. Of particular interest is the extent to which the effects of tax policy are state-dependent.

4.1. Household Sector

Time is discrete and continues forever, indexed by $t \geq 0$. The economy is inhabited by a unit mass of identical infinitely lived households, so aggregate values are interpreted as per capita. Each household has one unit of time per period, so the total available time for labor supply equals one.

Preferences and Budget Set. The households' preferences over (random) sequences of consumption, $\{c_t\}_{t=0}^{\infty}$, and time spent working or labor supply, $\{l_t\}_{t=0}^{\infty}$, are described by

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t), \tag{9}$$

where the utility function u is twice continuously differentiable, strictly increasing and concave in c_t and strictly decreasing in l_t, $0 < \beta < 1$ is the time discount factor, and E_0 denotes the mathematical expectation over the probability distributions of aggregate shocks conditional on information at time zero, which is the present time.

The final good is the numéraire so that its price is set to one. Households have three sources of income: (i) labor income, $W_t l_t$, where W_t is the wage rate; (ii) capital income, $R_t k_t$, from renting the capital stock, k_t, at the capital rental rate, R_t; and (iii) dividends from the ownership of firms, Π_t. The government levies proportional tax rates on labor and capital income $(\tau_{l,t}, \tau_{k,t})$, and rebates the proceeds in a lump-sum fashion, T_t, so that taxes net of transfers are $\mathcal{T}_t \equiv \tau_{l,t} W_t l_t + \tau_{k,t} R_t k_t - T_t$. Income net of taxes and transfers can be used for consumption, c_t, and next period capital, k_{t+1}.

The household's flow budget constraint is

$$c_t + k_{t+1} \leq W_t l_t + R_t k_t + (1-\delta)k_t + \Pi_t - \mathcal{T}_t, \tag{10}$$

where $0 < \delta < 1$ is the capital depreciation rate.

Household's Problem. Households maximize expected lifetime utility (9), subject to the budget constraint (10), taking as given prices, $\{W_t, R_t\}_{t=0}^{\infty}$, dividends $\{\Pi_t\}_{t=0}^{\infty}$, tax rates, $\{\tau_{l,t}, \tau_{k,t}\}_{t=0}^{\infty}$, and transfers $\{T_t\}_{t=0}^{\infty}$, with initial capital stock, k_0.

Let u_1 and u_2 denote the partial derivatives of the utility function with respect to c_t and l_t, respectively. The FOCs for consumption and labor supply are, respectively:

$$\beta^t u_1(c_t, l_t) = \lambda_t, \tag{11}$$

$$-\beta^t u_2(c_t, l_t) = \lambda_t (1-\tau_{l,t}) W_t, \tag{12}$$

where λ_t is the Lagrange multiplier associated with (10). Combining (11) and (12) gives the familiar intratemporal labor supply condition that the marginal rate of substitution equals the after-tax wage rate:

$$-\frac{u_2(c_t, l_t)}{u_1(c_t, l_t)} = (1-\tau_{l,t}) W_t. \tag{13}$$

Combining (11) with the FOC for next period capital gives the Euler equation:

$$u_1(c_t, l_t) = \beta \cdot E_t \left\{ u_1(c_{t+1}, l_{t+1}) \left[(1-\tau_{k,t+1}) R_{t+1} + 1 - \delta \right] \right\}. \tag{14}$$

Finally, the standard transversality condition (TVC) applies.

4.2. Business Sector

The business sector is populated by a unit mass of identical firms, each with a continuum of measure M of plants. In each period, a firm chooses the mass $m_t \leq M$ of plants to operate and the allocation of labor, h_t, and capital, k_t, per plant.[6] A plant with h_t units of labor and k_t units of capital produces $y_t = z_t h_t^{\phi} k_t^{\theta}$ units of output, with $0 < (\phi, \theta) < 1$, and $\phi + \theta < 1$. We describe the stochastic properties of the technology shock, z_t, when we parameterize the model in the following section.

[6] We normalize the mass of plants to one so that m_t is the fraction of plants a firm operates at time t.

Variable Capacity Utilization. Firms solve the capacity choice problem and maximize profits given the implied production function. Using the assumption that the production technology is the same across plants, we write the firm's capacity choice problem as

$$F(H_t, K_t) = \max_{m_t \leq \min\left\{M, \frac{H_t}{\overline{h}}\right\}} m_t z_t \left(\frac{H_t}{m_t}\right)^{\phi} \left(\frac{K_t}{m_t}\right)^{\theta} = \max_{m_t \leq \min\left\{M, \frac{H_t}{\overline{h}}\right\}} z_t H_t^{\phi} K_t^{\theta} m_t^{1-\phi-\theta}, \tag{15}$$

where $H_t = m_t h_t$ and $K_t = m_t k_t$ are total labor and capital inputs at the firm level. As $z_t H_t^{\phi} K_t^{\theta} m_t^{1-\phi-\theta}$ on the right-hand side of (15) is increasing in m_t, the firm's production set exhibits two regions: (i) one with "full capacity," i.e., $Y_t \leq F_{\text{full}}(H_t, K_t) = z_t H_t^{\phi} K_t^{\theta} M^{1-\phi-\theta}$, for $m_t = M \leq H_t/\overline{h}$, or, equivalently, $H_t \geq \overline{H} \equiv M\overline{h}_t$; (ii) another with "idle capacity," i.e., $Y_t \leq F_{\text{idle}}(H_t, K_t) = z_t \overline{h}^{\phi+\theta-1} K_t^{\theta} H_t^{1-\theta}$, for $m_t = H_t/\overline{h} \leq M$, or, equivalently, $H_t \leq \overline{H}$.

Profit Maximization. The firm's production function $F(H_t, K_t)$ is constant returns to scale for $H_t \leq \overline{H}$, and decreasing returns to scale for $H_t \geq \overline{H}$:

$$F(H_t, K_t) = \begin{cases} z_t \overline{h}^{\phi+\theta-1} K_t^{\theta} H_t^{1-\theta} & \text{for} \quad H_t \leq \overline{H} \\ z_t H_t^{\phi} K_t^{\theta} M^{1-\phi-\theta} & \text{for} \quad H_t \geq \underline{H} \end{cases}. \tag{16}$$

Given (16), firm's profit maximization takes the familiar form:

$$\max_{\{K_t, H_t\}} \Pi \equiv F(H_t, K_t) - W_t H_t - R_t K_t. \tag{17}$$

As in Section 3, the marginal product of labor, $F_H(H_t, K_t)$, is discontinuous at $F_H(\overline{H}, K_t)$. For future reference, note that the marginal product of capital, $F_K(H_t, K_t)$, is continuous so that the firm's first-order condition for renting capital and the determination of the capital rental rate follow standard intuition.

4.3. Government Sector

The government balances the budget on a period-by-period basis so that government spending equals tax revenues, $G_t + T_t = \tau_{l,t} W_t l_t + \tau_{k,t} R_t k_t$, for all t, where we specify government purchases as a share of output, $G_t \equiv g Y_t$, with $0 < g < 1$. Time-varying tax rates distort relative prices, whereas, as commonly assumed, government purchases are "wasteful" as they do not facilitate production or provide utility to households. Here, we zoom in on the effects of relative price distortions from distorting taxes, abstracting from wealth effects induced by changes in government consumption. For future reference, we set $g = 0.2$, equal to the average ratio of government consumption expenditures and gross investment, including federal, state, and local government levels, to gross domestic product (GDP) in the postwar United States.

4.4. Competitive Equilibrium

In this subsection, we first define the model economy's Competitive Equilibrium (CE), both the sequential and recursive formulation, and then discuss challenges to computing it and how we overcome them. The equilibrium definition is nonstandard as it includes the specification of a "selection rule" that applies when the economy is at the kink of the aggregate production function. Here, we also provide a theorem proving existence, uniqueness, and ergodicity. These results ensure we can take the model to the data and compute moments from the joint probability distribution over outcomes implied by the model, including means, standard deviations, and impulse response functions.

Definition: Sequential Competitive Equilibrium (SCE). An SCE is a list of allocations, $\{C_t, K_{t+1}, X_t, H_t, Y_t\}_{t=0}^{\infty}$, rental rates, $\{R_t, W_t\}_{t=0}^{\infty}$, profits, $\{\Pi_t\}_{t=0}^{\infty}$, government purchases $\{G_t\}_{t=0}^{\infty}$, and taxes $\{\mathcal{T}_t\}_{t=0}^{\infty}$ such that:

- Households maximize lifetime utility (9), subject to the budget constraint (10), with k_0 as initial capital stock.
- Firms solve the capacity choice and profit maximization's problems in (15) and (17).
- The labor market clears, $L_t \equiv \int_0^1 l_{i,t} di = H_t$. Assuming an appropriate selection rule when the economy is at capacity, $L_t = M\bar{h}$, the equilibrium wage equals the marginal rate of substitution between consumption and leisure, and the equilibrium labor satisfies labor demand.
- The capital rental market clears, $k_t = K_t$.
- The government budget is balanced.
- Market clearing in the product market implies the aggregate resource constraint, $C_t + X_t + G_t = Y_t$.

Next, we turn to Recursive Competitive Equilibrium (RCE), which is ultimately what we compute and whose solution we use to simulate equilibrium quantities and prices. Before doing so, we must set notation and a few definitions, beginning with the selection rule. Here, we discuss the key essential points and refer to Appendix B for more details, including two lemmas and a proposition.

Selection Rule. In a price-taking environment like ours, individual agents' decisions do not affect prices, and the equilibrium interest rate and wage solely depend on aggregate states. Henceforth, we denote individual choice and state variables with lowercase letters and aggregate state variables with uppercase letters. Exogenous states (technology and tax shocks) are denoted by $\mathcal{Z}$. Further, as the marginal product of labor differs from the equilibrium wage when the economy is at capacity, we introduce the notion of "quasi-rents," denoted by B, which differ from the profit due to decreasing returns to scale, P. So, making explicit the dependence of prices on aggregate state variables, we rewrite the household's budget constraint as

$$c + k' = W(\mathcal{Z},K)l + R(\mathcal{Z},K)k + (1-\delta)k + \overbrace{P(\mathcal{Z},K) + B(\mathcal{Z},K)}^{\Pi(\mathcal{Z},K)} - \mathcal{T}(\mathcal{Z},K). \quad (18)$$

The model economy exhibits three regimes, denoted by $s = \{1,2,3\}$, where one is "expansion," two is "recession," and three is "at capacity." Appendix B shows that it is possible to partition the state space for capital as a function of solely exogenous shocks. We denote this partition $\mathbb{K}_s(\mathcal{Z})$. Given $\mathcal{Z}$, if $K \in \mathbb{K}_s(\mathcal{Z})$, the economy is in regime s. Equipped with this partition, it is possible to define an indicator function: $\mathcal{I}(\mathbb{K}_s)(\mathcal{Z},K) = 1$ if $K \in \mathbb{K}_s(\mathcal{Z})$, and $\mathcal{I}(\mathbb{K}_s)(\mathcal{Z},K) = 0$, otherwise. We can now define the *selection rule* as follows:

$$\mathcal{I}(\mathbb{K}_3)(\mathcal{Z},K)B(\mathcal{Z},K) > 0 \text{ and } \mathcal{I}(\mathbb{K}_1)(\mathcal{Z},K)B(\mathcal{Z},K) = \mathcal{I}(\mathbb{K}_2)(\mathcal{Z},K)B(\mathcal{Z},K) = 0. \quad (19)$$

Such selection rule (19) is the new object that ensures the continuity of the competitive equilibrium despite a discontinuous labor demand. We assume that when the economy is in regime three, wages are determined according to the labor supply, maintaining full employment (i.e., $L_3(\mathcal{Z},K) = \overline{h}M$ for all $K \in \mathbb{K}_3(\mathcal{Z})$ and all $\mathcal{Z}$). And that the difference between the marginal product of labor and the equilibrium wage accrues to the firm. Note that there are two possible marginal products of labor for each $K \in \mathbb{K}_3(\mathcal{Z})$ in regime three:

one is above the equilibrium wage $W_3(\mathcal{Z}, K)$; the other is below (see Appendix B). Then, (19) implies we choose the supremum of these two for all $K \in \mathbb{K}_3(\mathcal{Z})$ and all $\mathcal{Z}$. As a result, the total remuneration accruing to capital equals $R(\mathcal{Z}, K)k + P(\mathcal{Z}, K) + B(\mathcal{Z}, K)$. Note also there is a continuum of possible selection rules, suggesting that the functional distribution of income—the split between capital and labor—is indeterminate at the kink.

Recursive Equilibrium Representation. In an appropriately selected equilibrium, little k equals big K, and the household's budget constraint implies the aggregate resource constraint:

$$C + G + K' = Y + (1 - \delta)K. \tag{20}$$

Households then solve the following recursive problem:

$$V(\mathcal{Z}, k, K) = \max_{\{c,l,k'\}} u(c, l) + \beta \cdot EV(\mathcal{Z}', k', K') \tag{21}$$

$$\text{s.t.} \quad (18), \quad c \geq 0, \quad 0 \leq l \leq 1. \tag{22}$$

Now, we are ready to define an RCE.

Definition: Recursive Competitive Equilibrium. An RCE is a list of policy functions, $\{C, K', L\}$, a value function, V, rental rates, $\{R, W\}$, profits, P, quasi-rents, B, government purchases, G, taxes, $\mathcal{T}$, a selection rule, $\mathcal{I}$, and a state space $\mathbb{Z} \times \mathbb{K}$ such that:

- Households solve the recursive problem (21)-(22).
- Firms solve the capacity choice and profit maximization problems (15) and (17).
- The labor market clears using the selection rule (19).
- The capital rental market clears.
- The government budget is balanced.
- The aggregate resource constraint (20) holds.
- The individual state k equals the aggregate state K and the state space is $\mathbb{Z} \times \mathbb{K}$.

In addition, we say that an RCE is *constructive* if its existence proof induces an iteratively convergent algorithm. We say that an RCE is *ergodic* if the dynamical system generated by it has an invariant probability measure.

The theorem below proves the existence of an RCE, its constructive uniqueness, and ergodicity. We compute and simulate it with the operator used to prove the theorem.

Theorem 1 *Assume Greenwood, Hercowitz and Huffman (1988) GHH preferences. The RCE is constructive, unique, and ergodic.*

Proof. See Appendix B. ■

Computation. Computing the RCE defined above requires adapting the monotone-map method of Coleman (1991), Reffett (1996), and Mirman, Morand and Reffett (2008) for kinked aggregate production functions. Recall that ours is a non-Pareto optimal economy due to distorting taxes. Thus, we cannot rely on the social planner's solution as in Hansen and Prescott (2005); instead, we must solve for the decentralized equilibrium and so prices directly. In addition, our setting poses another challenge due to the kink in the aggregate production function along the labor dimension.

Operationally, we use a modified version of the Coleman-Reffett operator based on a recursive version of (14). Using (20), the Euler equation (14) only depends on the interest rate, which we prove in Theorem 1 to be *continuous* under the selection rule (19). This is the key result ensuring that the equilibrium policy functions $\{C(\mathcal{Z}, K), L(\mathcal{Z}, K), K'(\mathcal{Z}, K)\}$ are also continuous. The appropriately modified Coleman-Reffett operator then provides constructive existence of a unique RCE. We then use the definition of sustainable capital stock from Stokey, Lucas and Prescott (1989) to prove that the state space is compact and ergodic. Finally, we use the selection rule to back out simulated wages and associated quasi-rents.

5. Calibrating the Model Economy to U.S. Data

We now discuss the calibration of the parameters describing preferences, technology, and government policy. We adopt a conservative approach to calibration, choosing parameter values that preserve comparability with previous work whenever possible.

Calibrating the model involves two steps. First, we assume technology and tax rate shocks follow AR(1) processes whose parameters we set to reproduce the variance and autocorrelation properties of measured TFP or Solow residuals and the proportional tax rates on labor and capital income in the United States.

Second, we pin down the parameter values describing preferences and technology requiring the model to reproduce U.S. data targets. Operationalizing this method-of-moments exercise requires a different approach than calibrating a log-linearized RBC model. In our model, the average values of endogenous variables do not coincide with their respective values associated with the model's deterministic steady state. Hence, we are to implement a simulated method of moments (SMM) algorithm based on simulating artificial data, calculate model-implied moments, and compare them with those in the data. When implementing this procedure, we assume that technology, capital, and labor tax shocks simultaneously hit the model economy.

5.1. Aggregate Uncertainty

The model has three sources of aggregate uncertainty: productivity, labor, and capital tax rate shocks.

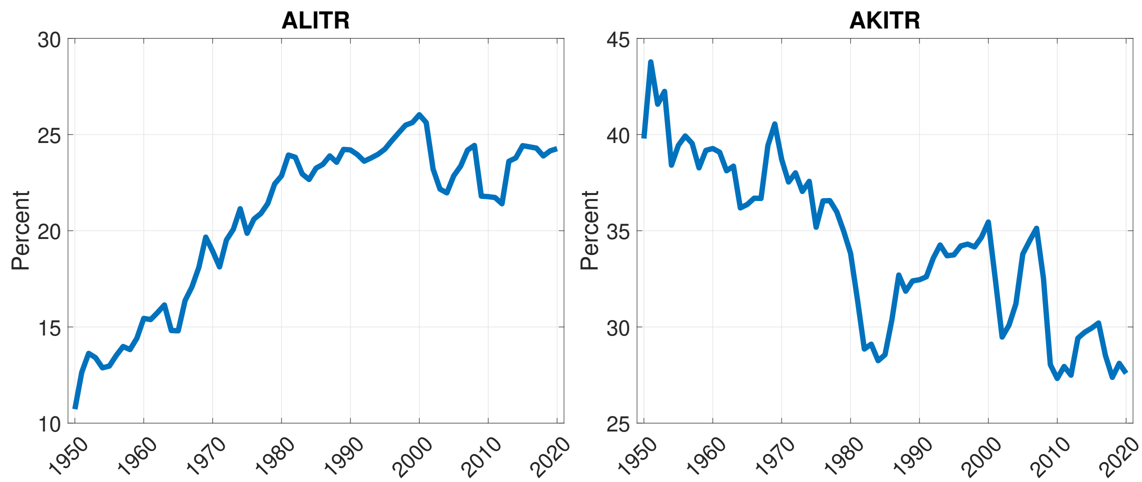
Productivity Shocks. We assume productivity shocks are AR(1) in logs:

$$\log(z_t) = (1-\rho_z)\log(\bar{z}) + \rho_z \log(z_{t-1}) + \sigma_z \epsilon_{z,t}, \tag{23}$$

where $0 < \rho_z < 1$ governs the persistence of productivity shocks and σ_z the standard deviation of the innovations, $\epsilon_{z,t}$, that we assume i.i.d. Normal with zero mean and unit variance. Consistent with King and Rebelo (1999), $\rho_z = 0.979$, and $\sigma_z = 0.0072$, quarterly. Using textbook conversion formulas, the corresponding annual parameters are $\rho_z = 0.979^4 = 0.9186$ and $\sigma_z = \sqrt{0.0072^2(1+0.979^2+0.979^4+0.979^6)} = 0.014$.

Labor and Capital Tax Rate Shocks. A realistic representation of the U.S. tax policy that accounts for the historical variation in tax rate data is essential to quantitative analysis. Following the literature, we do not take a stand on the determinants of tax policy changes. Instead, we use historical data to develop a statistical model that captures the underlying probability distributions of the realized tax rates. To do so, we must first construct time series for tax rates and then specify how the private sector in the model sets expectations about future tax policy.

Figure 4 shows effective tax rates for the United States from 1950 to 2020.[7] The average labor and capital tax rates are 21% and 34% during this period, respectively. Many of the tax changes were initially legislated as permanent; however, subsequent legislation partly or wholly overturned previously enacted changes, leading to substantial variation in tax rates (Barro and Redlick, 2011; Mertens and Ravn, 2013; Mertens and Montiel Olea, 2018; Romer and Romer, 2009, 2010). Standard deviations are approximately 0.04 for both series. Such tax rate variability activates intertemporal substitution in time spent working and capital investment, leading to hours worked and output fluctuations.



(a) Average Labor Income Tax Rate (b) Average Capital Income Tax Rate

Figure 4: U.S. Average Effective Tax Rates, 1950-2020

Notes: Effective average tax rates are calculated based on Mendoza, Razin and Tesar (1994). We aggregate all levels of the government (federal, state, and local) into one general government sector. We categorize individual income as labor income and capital income, which includes dividends, rental income, interest income, and profits. See Appendix A for more details on data sources, variables' definitions, and construction.

To gauge private sector's expectations about future tax policy, we assume tax rates are AR(1) processes in logs:

$$\log(\tau_{j,t}) = (1-\rho_j)\log(\bar{\tau}_j) + \rho_j \log(\tau_{j,t-1}) + \sigma_j \epsilon_{j,t}, \quad \text{with} \quad j \in \{l,k\}. \tag{24}$$

For the labor tax rate, $\rho_l = 0.9694$, and $\sigma_l = 0.0048$, at the quarterly frequency. Using standard conversion formulas, the corresponding annual parameters are $\rho_l = 0.9694^4 =$

[7] See Appendix A for details on data sources, definitions of variables, and construction.

0.8831 and $\sigma_l = \sqrt{0.0048^2(1+0.9694^2+0.9694^4+0.9694^6)} = 0.0092$. For the capital tax rate, $\rho_k = 0.9437$, and $\sigma_k = 0.0077$, whereas the corresponding annual parameters are $\rho_k = 0.9437^4 = 0.7931$ and $\sigma_k = \sqrt{0.0077^2(1+0.9437^2+0.9437^4+0.9437^6)} = 0.0142$.

5.2. Utility Function, Preferences, and Technology Parameters

As is well understood, in dynamic general equilibrium models, none of the parameters has a one-to-one relationship to a specific moment. Nonetheless, describing the model calibration as a few distinct steps is helpful. Before choosing parameter values, we discuss the functional form of the utility function.

Utility Function. We adopt the utility function specification in Greenwood, Hercowitz and Huffman (1988):

$$u(c,l) = \frac{1}{1-\sigma}\left(c - \alpha\frac{l^{1+\frac{1}{\eta}}}{1+\frac{1}{\eta}}\right)^{1-\sigma}, \quad \sigma > 0, \quad \alpha > 0, \quad \eta > 0. \tag{25}$$

These preferences feature no wealth effect on labor supply, eliminating the dampening effect of wealth on hours worked via consumption. To see this, combine the FOCs for c and l to obtain the labor supply curve, $l = (\widetilde{W}/\alpha)^\eta$, where $\widetilde{W} \equiv (1-\tau_l)W$ is the after-tax wage rate. As is evident, c does not enter the equation, which illustrates the well-known property that GHH preferences sterilize the wealth effect on labor supply. Hence, labor tax rate changes directly impinge on the labor allocation solely via intertemporal substitution effects by disturbing the relative price of working today versus tomorrow.

In addition to their well-known appeal for reproducing business cycle comovements (Jaimovich and Rebelo, 2009), GHH preferences afford another advantage that is specific to our setting. They allow us to compute the sequential equilibrium directly from the set of Euler equations. As mentioned before, computing these type of models using recursive methods remains an open question as neither the minimal state space literature nor the "enlarged state space" methods can be applied directly.

Preferences Parameters. A model period corresponds to a year. We set the discount factor β equal to 0.99. There are three other parameters related to preferences, (σ, α, η). We set $\sigma = 2$, which implies an intertemporal elasticity of substitution (IES) of 0.5, a standard value in the literature. We set $\eta = 1$, which pins down the individual-level or micro elasticity of labor supply to one. For future reference, we stress that the individual-level unitary labor supply elasticity does not coincide with the aggregate elasticity of labor supply. In our setup, occasionally binding capacity constraints break the neat separation between preferences and technology parameters in determining the aggregate elasticity of labor supply. Indeed, as we will see below, the aggregate hours elasticity depends on the sign and size of the labor tax shock, and it is highly state-dependent. In all experiments, it takes values below one, with the highest value being 0.5, which falls in the range of plausible estimates surveyed by Chetty et al. (2013).[8] Finally, given the calibrated values for σ and η, we choose $\alpha = 0.14$ so that average hours worked are 0.33.

Technology Parameters. We set δ to 0.1, implying a 10% capital depreciation rate a year (King and Rebelo, 1999). Next, three other parameters remain, $(\theta, \phi, \bar{h})$. We set the output elasticity to capital θ to 0.4 and the output elasticity to labor ϕ to 0.25 so that the model reproduces the average labor share of 0.65 (Koh, Santaeulàlia-Llopis and Zheng, 2020) and the average consumption expenditures to output ratio in the postwar United States.

The new parameter of the theory is the minimum labor or hours requirement. Recall that if $\bar{h} = 0$, the model collapses to an RBC model with distorting taxes. For $\bar{h} > 0$, the model's basic properties change qualitatively, allowing for the possibility of occasionally binding capacity constraints. While this parameter is central to the theory, direct data or measurement informing its value is nonexistent. So, we set $\bar{h} = 0.15$ in the benchmark calibration and show that the main insights hold for two alternative parameterizations $\bar{h} = 0.13$ and $\bar{h} = 0.17$ in Appendix C. Note also that the 0.15 figure implies a minimum hours requirement that is approximately 45.5% of the average hours worked target of 0.33. In the U.S., an employed person works on average 40 hours a week, thus spending $40/(5 \times 24) = 0.33$ of available time working. Hence, $\bar{h} = 0.15$ amounts to approximately 18 hours worked per week, a value lower than the less than 35 hours per week threshold used by the Bureau of Labor Statistics (BLS) to classify part-time work. Bick, Blandin and Rogerson (2022) document that less than 35 hours per week are relatively uncommon, accounting for only about four percent of the observations in the Current Population Survey (CPS) outgoing rotation group (ORG) for 1995-2007. Our calibration strategy is conservative in this sense as we do not impose an hours constraint binding at the individual worker level.

[8]Chetty et al. (2011) show that micro and macro elasticities of labor supply differ when firms are subject to hours constraints.

6. Nonlinear Propagation of Tax Shocks

We now turn to the model's quantification and study the effects of labor tax shocks—unexpected and unanticipated labor tax rate changes—on hours worked and capacity utilization, as well as the implied aggregate hours elasticity.[9] We rely on generalized impulse response functions (GIRFs) to describe the nonlinear propagation mechanism embodied in the model. We focus on tax shocks whose magnitudes are comparable with those observed historically in the United States. Further, we stress that tax shocks are symmetric, so any asymmetry in outcomes is solely the result of the internal propagation mechanism at play.[10]

Properties of GIRFs. Contrary to linear IRFs, GIRFs are generally *not* invariant to the sign and the size of shocks, nor the realized sequences of past and future shocks or initial conditions. As a result, one cannot infer the shape of the IRF to a negative shock from that to a positive shock by simply flipping the sign of the response, nor can one think of an IRF to a small shock as a scaled-down version of the response to a big shock. The magnitude of the marginal effect of a given shock and its dynamic implications critically depends on whether the shock is positive or negative, whether it is large or small, and whether a history of positive or negative realizations preceded it. This property creates a well-known reporting problem, which we address by producing several IRFs under alternative scenarios.

Computation of GIRFs. Calculating GIRFs requires simulating sequences of tax rates and using the model's solution to compute equilibrium time paths of the endogenous variables. An impulse response is the difference between the average equilibrium paths of two economies, which we refer to as "benchmark" and "counterfactual."

Operationally, the exercise involves two steps. First, starting with the benchmark, we simulate $10,000$ paths or replications for the variable of interest, say, hours worked, over T model periods ("years"). At $t = 0$, the initial tax rate across all replications rests at the median state, $\tau_{l,0} = \bar{\tau}_l = 0.21$. From $t > 0$, the tax rate sequences vary across replications as they are different realizations of the same AR(1) process. This procedure yields time series of realized tax rates $\{\tau^i_{l,t}\}^T_{t=0}$, where $i = 1, 2, \ldots, 10,000$, and $\tau^i_{l,0} = \bar{\tau}_l$

[9] Appendix Figures C.1-C.3 show results for output.

[10] See Gallant, Rossi and Tauchen (1993), Gonçalves et al. (2024), Koop, Pesaran and Potter (1996), Potter (2000) for a discussion of the subtleties in calculating IRFs for nonlinear time series models.

for all replications. Associated with these sequences of tax rates, the model generates a $10,000$ time series for each endogenous variable of interest.

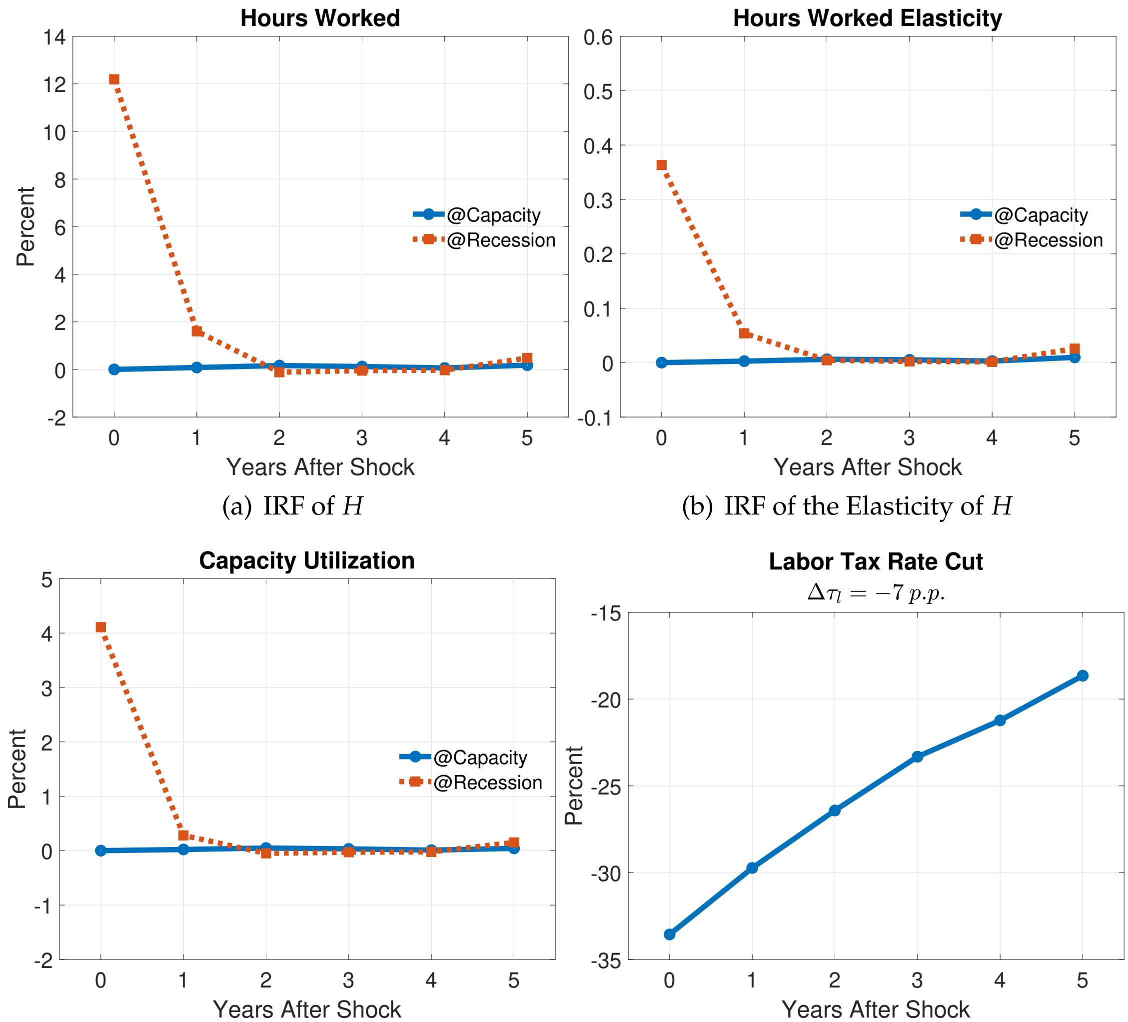
Second, we repeat the same steps for the counterfactual economy, with the only difference being that the initial tax rate is not the median value but a value below the median if we are interested in a labor tax cut or a value above the median if we are interested in a tax increase. The average difference between the simulated paths of benchmark and counterfactual is the IRF to a $\pm\Delta\tau_l$ tax rate shock.

State-Dependent Effects of Tax Changes. To quantify the extent of state dependence, we compute IRFs for labor tax shocks by further conditioning on whether the economy is *at capacity* (@Capacity) or *below capacity* (@Recession) at the time of the shock, $\pm\Delta\tau_l$. Such conditioning amounts to appropriately selecting the initial capital stock.

Figure 5 shows the IRFs to an equally-sized labor tax cut when the economy is below capacity or in recession versus at capacity or expansion. The tax shock is "large," inducing a temporary, rather persistent, more than 30% reduction in the labor tax rate (from 21% to 14%). The differences between the IRFs are striking. When the economy is at capacity, hours worked do not change despite a sizeable increase in work incentives. By contrast, hours worked considerably rise on impact (approximately 12%) when the economy is below capacity, with an implied aggregate hours elasticity of slightly higher than 0.35. Capacity utilization increases, too. These results show that the hours elasticity is state-dependent, virtually zero at capacity, and positive in recession.

Large vs. Small Tax Changes. Figure 6 shows the IRFs for a much smaller tax cut that reduces the labor tax rate by approximately 13% (from 21% to 18%), instead of the 33% in Figure 5. Not surprisingly, hours worked rise by less, approximately 6%. A notable result is that the aggregate hours elasticity is now 0.5, significantly higher than the 0.35 figure for the large tax cut. In this sense, the aggregate hours elasticity is not only state-dependent but also depends on the size of the shock.

Tax Cuts vs. Tax Hikes. Figure 7 shows the IRFs to a large tax rate increase (from 21% to 31.6%). As expected hours worked and capacity utilization fall. However, the hours worked elasticity is significantly smaller than that implied by a tax cut.



(a) IRF of H

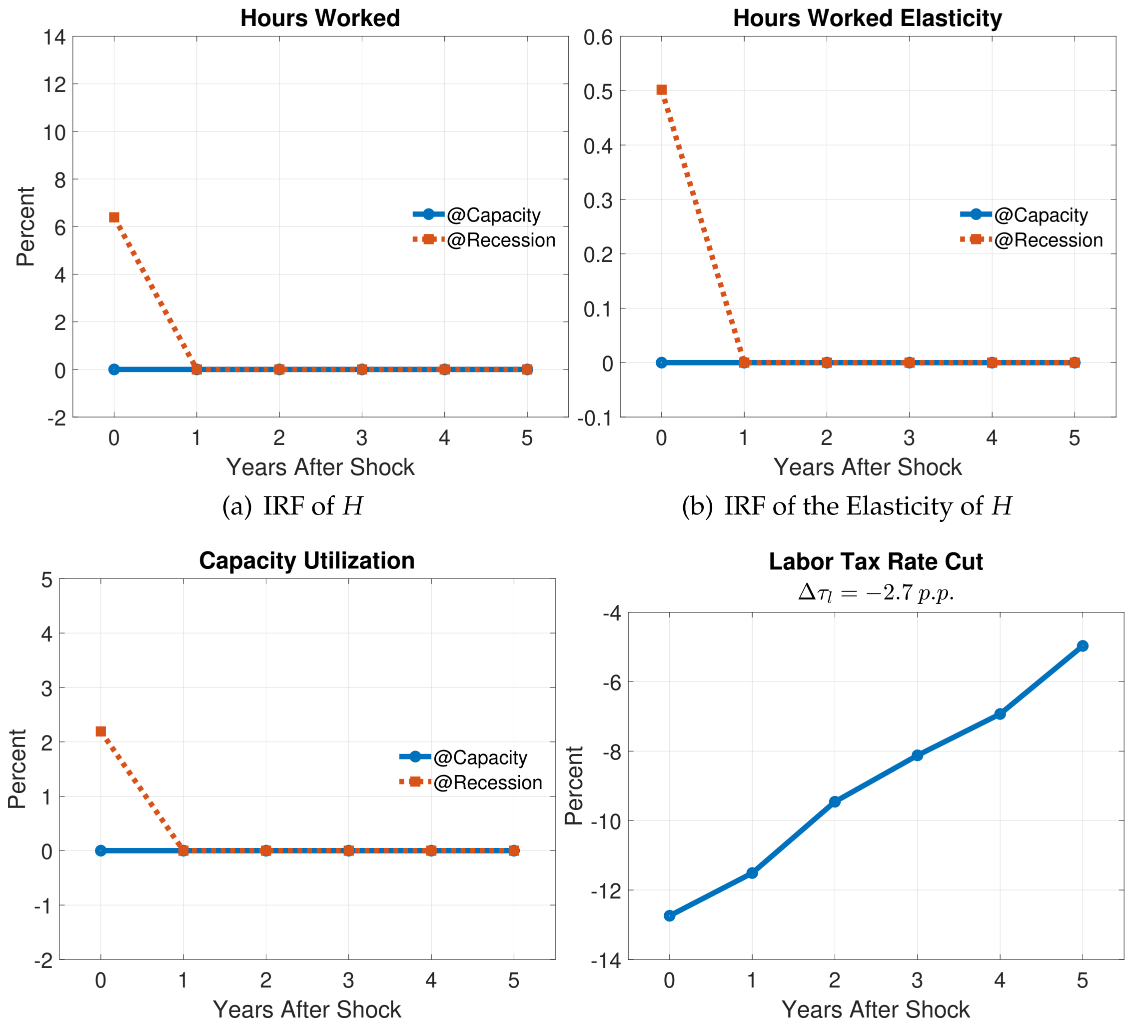
(b) IRF of the Elasticity of H

(c) IRF of $H/\overline{H}$

(d) IRF of τ_l

Figure 5: Impulse Responses to a Large Tax Cut

Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "large" labor tax cut (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.

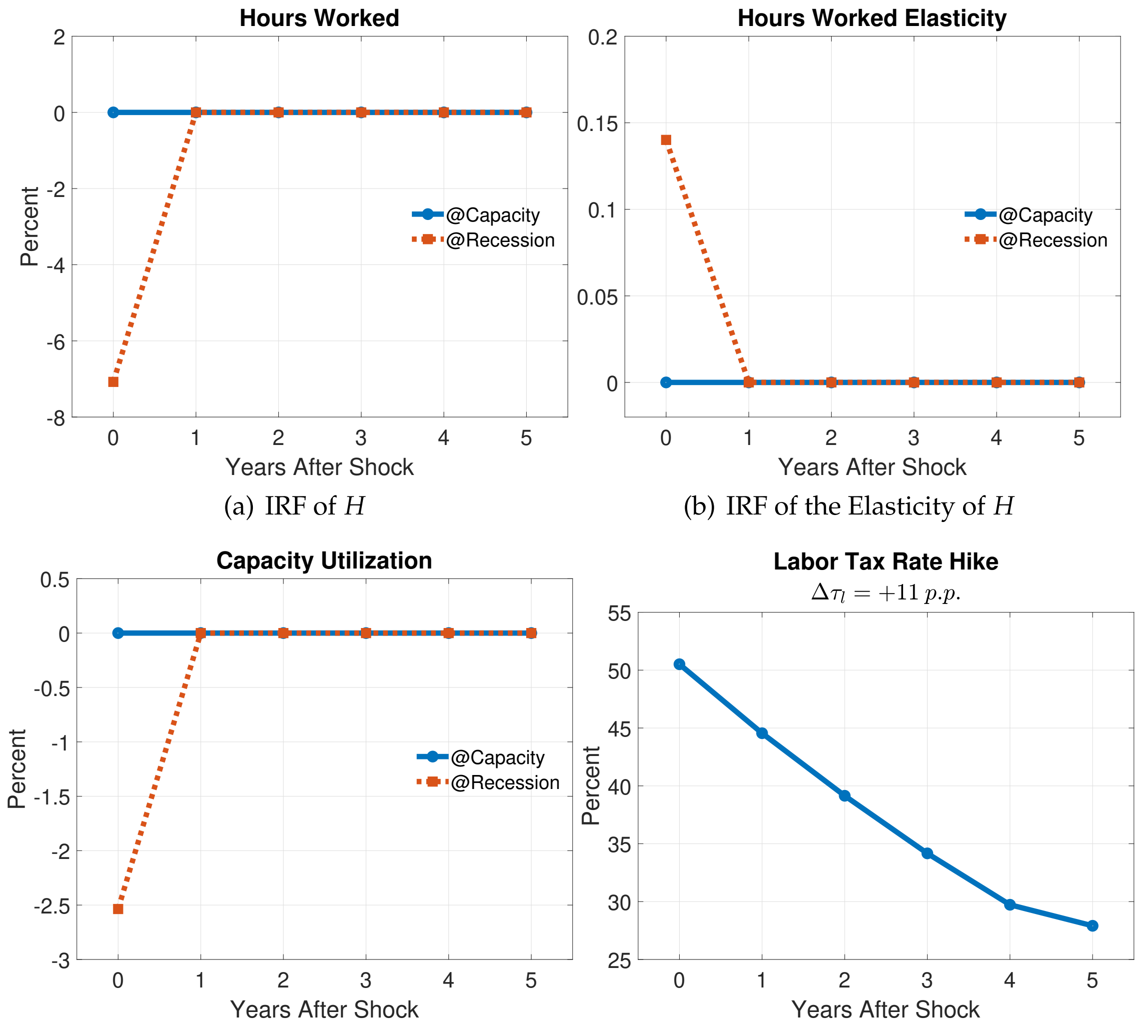


(a) IRF of H (b) IRF of the Elasticity of H

(c) IRF of $H/\overline{H}$ (d) IRF of τ_l

Figure 6: Impulse Responses to a Small Tax Cut

Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "small" labor tax cut (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.



(a) IRF of H

(b) IRF of the Elasticity of H

(c) IRF of $H/\overline{H}$

(d) IRF of τ_l

Figure 7: Impulse Responses to a Large Tax Hike

Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "large" labor tax hike (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.

7. Conclusion

This paper formalizes and quantifies the view that the effects of tax policy depend on the extent of idle capacity. Contrary to conventional wisdom, the quantitative theory we develop does not rely on "frictions" that prevent market clearing, noncompetitive pricing or price stickiness, nor asymmetric adjustment costs; instead, it leverages a technology feature according to which plant-level production requires a minimum labor input.

Embedding such minimum labor requirements into a multi-plant firm setting gives a model of idle capacity and time-varying capacity utilization. In response to shocks that disturb labor demand and supply, firms decide whether to operate all or a fraction of the available plants and leave some capacity idle temporarily or to change the labor input per plant under decreasing returns.

When the economy operates below capacity, firms accommodate shocks that increase labor demand by activating idle plants; increasing labor per plant is more costly as firms face decreasing returns. When the economy operates at capacity instead, shocks that increase labor demand push firms to increase hours worked across all active plants, thus facing sharp decreasing returns. Such qualitatively different adjustment to the same type of shock generates strong state dependence, leading to aggregate hours elasticities that vary significantly in recessions and expansions. For example, in our calibrated model, the response of hours worked to a labor tax cut is considerably higher when the economy is below capacity than at capacity.

There are several exciting avenues for future research. For instance, extending the current model with search frictions in the labor market would provide an arguably more realistic description of aggregate labor-market fluctuations, further distinguishing between the extensive margin of the number of workers at the plant and firm level and the intensive margin of hours per worker. Further, using the insight and approach developed here might prove helpful in studying the aggregate implications of other nonstandard features of plant- and firm-level technologies imposing constraints on the use of factors of production other than labor.

References

Abbritti, Mirko, and Stephan Fahr. 2013. "Downward Wage Rigidity and Business Cycle Asymmetries." *Journal of Monetary Economics*, 60(7): 871–886.

Aiyagari, S. Rao, Lawrence J. Christiano, and Martin Eichenbaum. 1992. "The Output, Employment, and Interest Rate Effects of Government Consumption." *Journal of Monetary Economics*, 30(1): 73–86.

Aruoba, S Borağan, Jesus Fernandez-Villaverde, and Juan F. Rubio-Ramirez. 2006. "Comparing Solution Methods for Dynamic Equilibrium Economies." *Journal of Economic dynamics and Control*, 30(12): 2477–2508.

Attanasio, Orazio, Peter Levell, Hamish Low, and Virginia Sánchez-Marcos. 2018. "Aggregating Elasticities: Intensive and Extensive Margins of Women's Labor Supply." *Econometrica*, 86(6): 2049–2082.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy*, 4(2): 1–27.

Barnichon, Regis, Davide Debortoli, and Christian Matthes. 2022. "Understanding the Size of the Government Spending Multiplier: It's in the Sign." *Review of Economic Studies*, 89(1): 87–117.

Barro, Robert J., and Charles J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics*, 126(1): 51–102.

Baxter, Marianne, and Robert G. King. 1993. "Fiscal Policy in General Equilibrium." *American Economic Review*, 83(3): 315–334.

Bick, Alexander, Adam Blandin, and Richard Rogerson. 2022. "Hours and Wages." *Quarterly Journal of Economics*, 137(3): 1901–1962.

Boehm, Christoph E., and Nitya Pandalai-Nayar. 2022. "Convex Supply Curves." *American Economic Review*, 112(12): 3941–3969.

Braun, R. Anton. 1994. "Tax Disturbances and Real Economic Activity in the Postwar United States." *Journal of Monetary Economics*, 33(3): 441–462.

Brown, Donald J. 1991. "Equilibrium Analysis with Non-Convex Technologies." *Handbook of Mathematical Economics*, 4: 1963–1995.

Burnside, Craig, Martin Eichenbaum, and Jonas D.M. Fisher. 2004. "Fiscal Shocks and Their Consequences." *Journal of Economic Theory*, 115(1): 89–117.

Burstein, Ariel T. 2006. "Inflation and Output Dynamics with State-Dependent Pricing Decisions." *Journal of Monetary Economics*, 53(7): 1235–1257.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2013. "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." *NBER Macroeconomics Annual*, 27(1): 1–56.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics*, 126(2): 749–804.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. 2011. "When is the Government Spending Multiplier Large?" *Journal of Political Economy*, 119(1): 78–121.

Coleman, Wilbur John. 1990. "Solving the Stochastic Growth Model by Policy-Function Iteration." *Journal of Business & Economic Statistics*, 8(1): 27–29.

Coleman, Wilbur John. 1991. "Equilibrium in a Production Economy with an Income Tax." *Econometrica*, 59(4): 1091.

Devereux, Michael B., and Henry E. Siu. 2007. "State Dependent Pricing and Business Cycle Asymmetries." *International Economic Review*, 48(1): 281–310.

Dupraz, Stephane, Emi Nakamura, and Jón Steinsson. 2020. "A Plucking Model of Business Cycles." *Available at SSRN 3523653*.

Ferraro, Domenico. 2018. "The Asymmetric Cyclical Behavior of the U.S. Labor Market." *Review of Economic Dynamics*, 30: 145–162.

Ferraro, Domenico. 2023. "Fast Rises, Slow Declines: Asymmetric Unemployment Dynamics with Matching Frictions." *Journal of Money, Credit and Banking*, 55(2-3): 349–378.

Ferraro, Domenico, and Giuseppe Fiori. 2023*a*. "Nonlinear Employment Effects of Tax Policy." *Journal of Money, Credit and Banking*, 55(5): 1001–1042.

Ferraro, Domenico, and Giuseppe Fiori. 2023*b*. "Search Frictions, Labor Supply, and the Asymmetric Business Cycle." *Journal of Money, Credit and Banking*, 55(1): 5–42.

Futia, Carl A. 1982. "Invariant Distributions and the Limiting Behavior of Markovian Economic Models." *Econometrica*, 50(2): 377–408.

Gallant, A. Ronald, Peter E. Rossi, and George Tauchen. 1993. "Nonlinear Dynamic Structures." *Econometrica*, 61(4): 871–907.

Ghassibe, Mishel, and Francesco Zanetti. 2022. "State Dependence of Fiscal Multipliers: The Source of Fluctuations Matters." *Journal of Monetary Economics*, 132: 1–23.

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento. 2024. "State-Dependent Local Projections." *Journal of Econometrics*, 105702.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review*, 78(3): 402.

Hansen, Gary D., and Edward C. Prescott. 2005. "Capacity Constraints, Asymmetries, and the Business Cycle." *Review of Economic Dynamics*, 8(4): 850–865.

Hornstein, Andreas, and Edward C. Prescott. 1993. "The Firm and the Plant in General Equilibrium Theory." In *General Equilibrium, Growth, and Trade*. 393–410. Elsevier.

Jaimovich, Nir, and Sergio Rebelo. 2009. "Can News About the Future Drive the Business Cycle?" *American Economic Review*, 99(4): 1097–1118.

Joines, Douglas H. 1981. "Estimates of Effective Marginal Tax Rates on Factor Incomes." *Journal of Business*, 191–226.

Jones, John Bailey. 2002. "Has Fiscal Policy Helped Stabilize the Postwar U.S. Economy?" *Journal of Monetary Economics*, 49(4): 709–746.

King, Robert G., and Sergio T. Rebelo. 1999. "Resuscitating Real Business Cycles." *Handbook of Macroeconomics*, 1: 927–1007.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo. 1988. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics*, 21(2-3): 195–232.

Koh, Dongya, Raül Santaeulàlia-Llopis, and Yu Zheng. 2020. "Labor Share Decline and Intellectual Property Products Capital." *Econometrica*, 88(6): 2609–2628.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. "Impulse Response Analysis in Nonlinear Multivariate Models." *Journal of Econometrics*, 74(1): 119–147.

Leeper, Eric M., Michael Plante, and Nora Traum. 2010. "Dynamics of Fiscal Financing in the United States." *Journal of Econometrics*, 156(2): 304–321.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9(2): 508–523.

McGrattan, Ellen R. 1994. "The Macroeconomic Effects of Distortionary Taxation." *Journal of Monetary Economics*, 33(3): 573–601.

McGrattan, Ellen R. 2012. "Capital Taxation During the U.S. Great Depression." *Quarterly Journal of Economics*, 127(3): 1515–1550.

McGrattan, Ellen R., and Edward C. Prescott. 2005. "Taxes, Regulations, and the Value of U.S. and U.K. Corporations." *Review of Economic Studies*, 72(3): 767–796.

McGrattan, Ellen R., and Lee E. Ohanian. 2010. "Does Neoclassical Theory Account for the Effects of Big Fiscal Shocks? Evidence from World War II." *International Economic Review*, 51(2): 509–532.

McGrattan, Ellen R., Richard Rogerson, and Randall Wright. 1997. "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy." *International Economic Review*, 267–290.

McKay, Alisdair, and Ricardo Reis. 2008. "The Brevity and Violence of Contractions and Expansions." *Journal of Monetary Economics*, 55(4): 738–751.

Mendoza, Enrique G., Assaf Razin, and Linda L. Tesar. 1994. "Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption." *Journal of Monetary Economics*, 34(3): 297–323.

Mertens, Karel, and José Luis Montiel Olea. 2018. "Marginal Tax Rates and Income: New Time Series Evidence." *Quarterly Journal of Economics*, 133(4): 1803–1884.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review*, 103(4): 1212–1247.

Michaillat, Pascal. 2014. "A Theory of Countercyclical Government Multiplier." *American Economic Journal: Macroeconomics*, 6(1): 190–217.

Mirman, Leonard J., Olivier F. Morand, and Kevin L. Reffett. 2008. "A Qualitative Approach to Markovian Equilibrium in Infinite Horizon Economies with Capital." *Journal of Economic Theory*, 139(1): 75–98.

Parker, Jonathan A. 2011. "On Measuring the Effects of Fiscal Policy in Recessions." *Journal of Economic Literature*, 49(3): 703–18.

Pizzinelli, Carlo, Konstantinos Theodoridis, and Francesco Zanetti. 2020. "State Dependence in Labor Market Fluctuations." *International Economic Review*, 61(3): 1027–1072.

Potter, Simon M. 2000. "Nonlinear Impulse Response Functions." *Journal of Economic Dynamics and Control*, 24(10): 1425–1446.

Prescott, Edward C., Richard Rogerson, and Johanna Wallenius. 2009. "Lifetime Aggregate Labor Supply with Endogenous Workweek Length." *Review of Economic dynamics*, 12(1): 23–36.

Ramey, Valerie A, and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy*, 126(2): 850–901.

Reffett, Kevin. 1996. "Production-Based Asset Pricing in Monetary Economies with Transactions Costs." *Economica*, 63(251): 427–443.

Rogerson, Richard. 2011. "Individual and Aggregate Labor Supply with Coordinated Working Times." *Journal of Money, Credit and Banking*, 43(s1): 7–37.

Romer, Christina D., and David H. Romer. 2009. "A Narrative Analysis of Postwar Tax Changes." *Unpublished paper, University of California, Berkeley.*

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review*, 100(3): 763–801.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott. 1989. "Recursive Methods in Economic Dynamics."

Appendix

A. Data Sources, Variables' Definitions, and Construction

This appendix provides details on data sources, definitions, and how to construct average effective tax rates. The main source of data is the National Income and Product Account (NIPA) tables by the Bureau of Economic Analysis (BEA). All data items are indexed by table and line numbers. Our approach closely follows that of Mendoza, Razin and Tesar (1994). We aggregate all levels of the government (federal, state and local) into one general government sector.

APITR. The **Average Personal Income Tax Rate** is defined as

$$\text{APITR} \equiv \frac{\text{PIT}}{\text{WSA} + \text{PRI}/2 + \text{CI}}, \tag{A.1}$$

where PIT is **personal income taxes**, that consists of federal personal income taxes (NIPA Table 3.2 line 3) and state and local personal income taxes (NIPA Table 3.3 line 4), WSA is **wage and salaries** (NIPA Table 1.12 line 3), PRI is **proprietors' income** (NIPA Table 1.12 line 9), CI $\equiv$ PRI/2 + RI + DI + NI is **capital income**, RI is **rental income** (NIPA Table 1.12 line 12), DI is **net dividends** (NIPA Table 1.12 line 16), and NI is **net interest** (NIPA Table 1.12 line 18). As discussed in Joines (1981), the imputation of proprietor's income to capital and labor income is somewhat arbitrary. Here, we follow Jones (2002) and split proprietor's income evenly between capital and labor income.

ALITR. The **Average Labor Income Tax Rate** is defined as

$$\text{ALITR} \equiv \frac{\text{APITR} \times (\text{WSA} + \text{PRI}/2) + \text{CSI}}{\text{CEM} + \text{PRI}/2}, \tag{A.2}$$

where CSI is **contributions for government social insurance** (NIPA Table 3.1 line 7), and CEM is **compensation of employees** (NIPA Table 1.12 line 2). The calculations of APITR and ALITR are based on Jones (2002). Leeper, Plante and Traum (2010) follow a similar methodology but they restrict their calculations to the federal government.

AKITR. The **Average Capital Income Tax Rate** is defined as

$$\text{AKITR} \equiv \frac{\text{APITR} \times \text{CI} + \text{FCT\_NET} + \text{PRT}}{\text{CI} + \text{PRT}}, \tag{A.3}$$

where FCT_NET = (FCT − FCT_FED) is **federal corporate taxes** (NIPA Table 3.2 Line 8) minus **federal corporate taxes of federal reserve banks** (NIPA Table 3.2 Line 16), and PRT is **property taxes** to state and local government (NIPA Table 3.3 Line 9).

GRATIO. The **Government Spending to GDP Ratio** is defined as GRATIO ≡ GOV/GDP, where GOV is **government consumption expenditures and gross investment**, including federal (national defense plus nondefense), state and local government level (NIPA Table 1.1.5 line 22) and GDP is **gross domestic product** (NIPA Table 1.1.5 line 1).

B. Proofs

We begin with some basic definitions and notation. To keep the appendix self-contained, we reproduce some concepts that have already been introduced in the body of the paper.

To start, the aggregate production function is

$$F(H,K) = \begin{cases} z\overline{h}^{\phi+\theta-1} K^{\theta} H^{1-\theta} & \text{if} \quad H \leq \overline{H} \\ zK^{\theta} H^{\phi} & \text{if} \quad H \geq \underline{H} \end{cases}, \tag{B.1}$$

where $\overline{H} = \overline{h}$, as we normalize the mass of plants to one, $M = 1$. When $H = \overline{H}$, both branches of the function are equal to $z\overline{h}^{\phi} K^{\theta}$. We refer to this regime as "at capacity." We will show that the continuity of the aggregate production function is essential to prove the continuity of the policy function for labor, which is stationary and intratemporal, thanks to GHH preferences. This latter observation implies that we can apply an almost standard Coleman-Reffett operator to prove the existence of the recursive equilibrium. However, a minimum labor requirement generates a discontinuity in the wage. Thus, we need to derive our results using directional derivatives.

To proceed, let $i \in \{-, 0, +\}$ indicate the derivative from the left, both sides, and the right, respectively. Below, we present the marginal product of labor as

$$F_H^i(H,K) = \begin{cases} F_H^0(H,K) = z\overline{h}^{\phi+\theta-1}K^\theta(1-\theta)H^{-\theta} & \text{if} \quad H < \overline{H} \\ F_H^0(H,K) = zK^\theta \phi H^{\phi-1} & \text{if} \quad H > \overline{H} \\ \{F_H^-(H,K), F_H^+(H,K)\} = \left\{z\overline{h}^{\phi+\theta-1}K^\theta(1-\theta)\overline{h}^{-\theta}, zK^\theta\phi\overline{h}^{\phi-1}\right\} & \text{if} \quad H = \overline{H} \end{cases}. \tag{B.2}$$

As F is concave, it has well-defined directional derivatives. Note that the cardinality of the image of $F_H^i(H,K)$ is not equal to one when the economy is at capacity, which implies that the marginal product of labor is a correspondence. Moreover, as $(1-\theta) > \phi$, we know that $F_H^-(\overline{h},K,z) \equiv WUB(K,z) > F_H^+(\overline{h},K,z) \equiv WLB(K,z)$. Equation (19) generates positive quasi-rents for the firm and, at the same time, keeps workers on the supply curve, making labor contracts compatible with optimality on both sides of the market (profit and utility maximization). Finally, the discontinuity in the marginal product of labor will not affect the continuity of the Coleman-Reffett operator as the sum of wages, profits, and the return of capital, after taxes and transfers from the government, is always equal to F.

Next, from Figure 3(c), it is clear that it is possible to partition the state space using a stationary and static mechanism:

$$\mathbb{K}(z,\tau_l) = \begin{cases} \left[0, K^{WUB}\right) & \text{if} \quad \alpha\overline{h}^{1/\eta}/(1-\tau_l) > WUB(K,z) & \text{Regime 1} \\ \left(K^{WLB}, +\infty\right) & \text{if} \quad \alpha\overline{h}^{1/\eta}/(1-\tau_l) < WLB(K,z) & \text{Regime 2} \\ \left[K^{WUB}, K^{WLB}\right] & \text{otherwise} & \text{Regime 3} \end{cases}. \tag{B.3}$$

There are three things to notice about $\mathbb{K}(z,\tau_l)$: (i) it is time-independent and a function of the space of shocks, $\mathbb{Z}$; (ii) it is not bounded above; (iii) $K^{WUB} < K^{WLB}$. To see this last remark, notice that by setting $\alpha\overline{h}^{1/\eta}/(1-\tau_l) = WUB(K^{WUB},z)$ and $\alpha\overline{h}^{1/\eta}/(1-\tau_l) = WUB(K^{WLB},z)$ we obtain

$$K^{WUB} = \left[\frac{\alpha\overline{h}^{1/\eta+1-\phi}}{z(1-\theta)(1-\tau_l)}\right]^{1/\theta} < K^{WLB} = \left[\frac{\alpha\overline{h}^{1/\eta+1-\phi}}{z\phi(1-\tau_l)}\right]^{1/\theta}. \tag{B.4}$$

Later, we will define the set of sustainable capital stocks and impose an upper bound on $\mathbb{K}(z,\tau_l)$. We are ready to prove the first result.

Lemma 2 ***(Continuity of labor and of the interest rate)*** *Let $\mathcal{Z} \equiv (z,\tau_l) \in \mathbb{Z}$ be the vector of shocks, $L(\mathcal{Z},K)$ be the equilibrium policy function for aggregate labor, and $r(\mathcal{Z},K)$ be the* *equilibrium interest rate. If for any $\mathcal{Z} \in \mathbb{Z}$, $K \in \left(K^{WUB}, K^{WLB}\right)$ implies $L(\mathcal{Z}, K) = \overline{h}$ (i.e., the selection rule guarantees full employment when the economy is in the interior of Regime 3), then: (i) $L(\mathcal{Z}, K)$ is continuous in $K \in \mathbb{K}(\mathcal{Z})$ for any $\mathcal{Z} \in \mathbb{Z}$, and (ii) $r(\mathcal{Z}, K)$ is continuous in $K \in \mathbb{K}(\mathcal{Z})$ for any $\mathcal{Z} \in \mathbb{Z}$.*

Proof. The proof consists of two parts.

Part (i). For Regime 1, $K \in \left[0, K^{WUB}\right]$, $F_H^0(H, K) = \alpha L^{1/\eta}/(1-\tau_l)$, $L = H$, such that

$$L_1(\mathcal{Z}, K) = \left[\frac{(1-\tau_l)}{\alpha} z K^{\theta} (1-\theta) \overline{h}^{\theta+\phi-1}\right]^{\frac{1}{1/\eta+\theta}}. \tag{B.5}$$

Similarly, for Regime 2,

$$L_2(\mathcal{Z}, K) = \left[\frac{(1-\tau_l)}{\alpha} z K^{\theta} (\phi) \overline{h}^{\theta+\phi-1}\right]^{\frac{1}{1/\eta+1-\phi}}. \tag{B.6}$$

To show continuity, it suffices to prove that $L_1(\mathcal{Z}, K^{WUB}) = L_2(\mathcal{Z}, K^{WLB}) = \overline{h}$. By replacing the values of $\left(K^{WUB}, K^{WLB}\right)$ from (B.4) in L_1 and L_2, respectively, we obtain the desired result. This completes the first part of the proof.

Part (ii). We now turn to the equilibrium interest rate. We can define the equilibrium marginal product of capital as we did with the marginal product of labor. $F_K^0(L_1(\mathcal{Z}, K), K)$ for $K \in \left[0, K^{WUB}\right)$ in Regime 1 and $F_K^0(L_2(\mathcal{Z}, K), K)$ for $K \in \left(K^{WLB}, +\infty\right)$ in Regime 2. Then, it suffices to show that $F_K^-(L_1(\mathcal{Z}, K^{WUB}), K) = F_K^+(L_2(\mathcal{Z}, K^{WLB}), K) = F_K^0(\overline{h}, K)$ for any $K \in \left[K^{WUB}, K^{WLB}\right]$ in Regime 3. Then, we obtain

$$F_K^-(L_1(\mathcal{Z}, K^{WUB}), K) = F_K^+(L_2(\mathcal{Z}, K^{WLB}), K) = F_K^0(\overline{h}, K) = z K^{\theta} \overline{h}^{\phi}, \tag{B.7}$$

which implies

$$F_K^-(L_1(\mathcal{Z}, K^{WUB}), K^{WUB}) = F_K^0(\overline{h}, K^{WUB}), \tag{B.8}$$

$$F_K^+(L_2(\mathcal{Z}, K^{WLB}), K^{WLB}) = F_K^0(\overline{h}, K^{WLB}). \tag{B.9}$$

This completes the second part of the proof. ■

Notice that $\mathbb{K}(\mathcal{Z})$ maybe unbounded above. To apply the standard Coleman-Reffett operator, we need to compactify the state space. The following proposition defines the notion of *sustainable capital stock* borrowed from Stokey, Lucas and Prescott (1989). Given the production function and depreciation rate, it is possible to define a function $g(\mathcal{Z}, K) = F(L(\mathcal{Z}, K), K) - \delta K$ that contains the levels of capital consistent with $K_{+} = K$, where K_{+} represent the capital stock "tomorrow," and $C \geq 0$. Using the function g we can define a set $G(\mathcal{Z}) = \{K : g(\mathcal{Z}, K) \geq 0\}$, which contains the sustainable capital levels. Given the concavity of the production, it is possible to find a maximum sustainable level of capital for each $\mathcal{Z} \in \mathbb{Z}$. This level of capital, $\overline{K}(\mathcal{Z})$, satisfies $g(\mathcal{Z}, \overline{K}) = 0$.

The proposition below proves that the set of upper bounds for sustainable capital levels has a uniform bound in $\mathbb{Z}$. That is, $\sup_{\mathcal{Z}\in\mathbb{Z}} \overline{K}(\mathcal{Z}) < +\infty$.

Proposition 3 ***(Uniform bound on the capital stock)*** *Assume $(1-\theta) > \phi$. Let $G(\mathcal{Z}) = \{K : g(\mathcal{Z}, K) \geq 0\}$, with $g(\mathcal{Z}, K) = F(L(\mathcal{Z}, K), K) - \delta K$. Then, (i) for each $\mathcal{Z} \in \mathbb{Z}$, $G(\mathcal{Z})$ is pointwise bounded by $\overline{K}(\mathcal{Z})$, (ii) $K_{max} \equiv \sup_{\mathcal{Z}\in\mathbb{Z}} \overline{K}(\mathcal{Z}) < +\infty$.*

Proof. Notice that we can write $F(L(\mathcal{Z}, K), K)$ as

$$F(L(\mathcal{Z},K),K) = \begin{cases} z\overline{h}^{\phi+\theta-1}\left[\frac{(1-\tau_l)}{\alpha} z(1-\theta)\overline{h}^{\phi+\theta-1}\right]^{\frac{1-\theta}{1/\eta+\theta}} K^{\theta\left(1+\frac{1-\theta}{1/\eta+\theta}\right)} & \text{Recession} \\ z\left[\frac{(1-\tau_l)}{\alpha} z\phi\right]^{\frac{\phi}{1/\eta+\phi+1}} K^{\theta\left(\frac{1/\eta+1}{1/\eta+1-\phi}\right)} & \text{Expansion} \\ z\overline{h}^{\phi} K^{\theta} & \text{Otherwise} \end{cases} . \tag{B.10}$$

It is easy to see that

$$0 < \theta\left(1 + \frac{1-\theta}{1/\eta+\theta}\right) < 1. \tag{B.11}$$

As $\theta < 1$, the above inequality proves the equilibrium production function in recession and at capacity are strictly concave. Then, as we assume $(1-\theta) > \phi$, we obtain

$$0 < \theta\left(\frac{1/\eta+1}{1/\eta+1-\phi}\right) < 1. \tag{B.12}$$

This proves that $F(L(\mathcal{Z}, K), K)$ is *strictly* concave in $\mathbb{K}(\mathcal{Z})$. The results in Stokey, Lucas and Prescott (1989) show that $\overline{K}(\mathcal{Z}) < +\infty$ exists for any $\mathcal{Z} \in \mathbb{Z}$. As the set of shocks is finite, we can order the pointwise upper bounds $\overline{K}(\mathcal{Z})$. This completes the proof. ■

Corollary 4 ***(Compactness of the state space)*** *The results in Proposition 3 imply that $\mathbb{K}(\mathcal{Z})$ is compact for any $\mathcal{Z} \in \mathbb{Z}$, with $\mathbb{K}(\mathcal{Z}) \subseteq [0, K_{max}]$. Given that $\mathbb{Z}$ is compact, the state space*

$[0, K_{max}] \times \{\mathbb{Z}\}$ *is compact.*

We now turn to the existence of the recursive equilibrium. We will prove the existence of a *recursive composite* based on the instantaneous utility function $u(c, l)$ and the Euler operator implied by (14). Given GHH preferences,

$$u(c, l) = \frac{1}{1-\sigma} \left(\frac{c - \alpha l^{1+\frac{1}{\eta}}}{1 + \frac{1}{\eta}} \right)^{1-\sigma}, \tag{B.13}$$

let us define the *composite* as

$$C_1(c, l) \equiv \left(\frac{c - \alpha l^{1+\frac{1}{\eta}}}{1 + \frac{1}{\eta}} \right). \tag{B.14}$$

Next, we introduce the candidate space of functions for the composite. We will assume that any possible C_1 is a function of the minimal state space $[0, K_{max}] \times \{\mathbb{Z}\}$. By the definition of the composite and equilibrium labor being a function of the same state space, $L(\mathcal{Z}, K)$, it follows we can describe consumption C using the minimal state space:

$$C(\mathcal{Z}, K) = C_1(\mathcal{Z}, K)((1+\eta)/\eta) + \alpha L(\mathcal{Z}, K)^{1+\frac{1}{\eta}}. \tag{B.15}$$

Following Coleman (1991), we can define a space of functions for the composite. We will use an Euler operator to prove existence, which will map from and to this space. Let $\mathbf{K} \equiv [0, K_{max}]$ and $\mathbf{D} \equiv \eta/(1+\eta)$:

$$\mathbb{C}_1(\mathbb{Z} \times \mathbf{K}) = \begin{cases} C_1 : \mathbb{Z} \times \mathbf{K} \longrightarrow \left[-\alpha \mathbf{D}^{-1}, \mathbf{D}(F(1, K_{max}) + (1-\delta)K_{max})\right] & \text{is continuous} \\ -\alpha \mathbf{D}^{-1} L(\mathcal{Z}, K)^{1+\frac{1}{\eta}} \leq C_1(\mathcal{Z}, K) \leq \mathbf{D} h(\mathcal{Z}, K), & h \text{ bounded-continuous} \\ 0 \leq [C_1(\mathcal{Z}, y) - C_1(\mathcal{Z}, x)] \leq \mathbf{D}\left[h(\mathcal{Z}, y) - h(\mathcal{Z}, x)\right] & y \geq x; (x, y) \in \mathbf{K} \end{cases}. \tag{B.16}$$

Two minor differences exist between $\mathbb{C}_1$ and the canonical space for the Euler operator defined in Coleman (1991). First, the uniform lower and upper bounds are different. As we solve the model using the composite and then recover its associated consumption level, $\mathbb{C}_1$ is uniformly bounded below by the maximum dis-utility of labor and uniformly bounded above by $\mathbf{D}$ times maximal available resources. Proposition 3 guarantees that consumption is non-negative. Second, the Lipschitz constant that defines the equicontinuous family of functions differs from the canonical space but still generates bounded variations, which preserves the compactness of the space.

We define the Euler operator directly from (14) using (25). Let $\tau_{k,+}, z_+$ and $\mathbb{Z} \times \mathbf{K} \equiv S$ denote capital tax rate and TFP level "tomorrow" and the state space, respectively. We will represent an element in the state space by $s \in S$:

$$[A(C_1)(s)]^{-\sigma} = \beta E \left\{ \left[C_1 \left(h(s) - A(C_1)(s)\mathbf{D}^{-1}, z_+ \right) \right]^{-\sigma} \left[(1 - \tau_{k,+}) r \left(h(s) - A(C_1)(s)\mathbf{D}^{-1}, z_+ \right) + 1 - \delta \right] \right\}. \tag{B.17}$$

Given the definition of C_1 and *if* h is *continuous, bounded and increasing* in K, (B.17) defines the standard Euler operator in Coleman (1991). The following lemma proves that h has the desired properties.

Lemma 5 ***(Properties of*** h***)*** *Let* h *be the function used to define the Euler operator in* (B.17) *and* $\mathbb{Z} \times \mathbf{K} \equiv S$. *Then,* $h : S \longrightarrow [-\alpha, F(1, K_{max}) + (1 - \delta)K_{max}]$, *is continuous and increasing in* K *for any* $\mathcal{Z} \in \mathbb{Z}$.

Proof. Let $h(\mathcal{Z}, K) \equiv F(L(\mathcal{Z}, K), K) - \alpha L(\mathcal{Z}, K)^{1+\frac{1}{\eta}}$, $\Phi_1 \equiv \frac{(1-\tau_l)}{\alpha} z(1-\theta)\bar{h}^{\phi+\theta-1}$, $\Phi_2 \equiv \frac{(1-\tau_l)}{\alpha} z\phi$, and R1, R2, R3 be expansion, recession and "at capacity," respectively. Then, the continuity and boundness of h follows from Lemma 2 and Proposition 3. It remains to be proven that it is increasing. We can define h for each regime as follows:

$$h(\mathcal{Z}, K) = \begin{cases} z\bar{h}^{\phi+\theta-1} \Phi_1^{\frac{1-\theta}{1/\eta+\theta}} K^{\theta\left(1+\frac{1-\theta}{1/\eta+\theta}\right)} + (1-\delta)K - \alpha \Phi_1^{\frac{1+1/\eta}{1/\eta+\theta}} K^{\theta\left(\frac{1+1/\eta}{1/\eta+\theta}\right)} & \text{in R1} \\ z\Phi_2^{\frac{\phi}{1/\eta-\phi-1}} K^{\theta\left(\frac{1+1/\eta}{1/\eta-\phi+1}\right)} + (1-\delta)K - \alpha \Phi_2^{\frac{1+1/\eta}{1/\eta-\phi+1}} K^{\theta\left(\frac{1+1/\eta}{1/\eta-\phi+1}\right)} & \text{in R2} \\ z\bar{h}^{\phi} K^{\theta} - \alpha \bar{h}^{1+(1/\eta)} & \text{in R3} \end{cases}.$$

In R1, taking the left derivative with respect to K, it suffices to impose that $1 > (1 - \tau_l)(1-\theta)$, which holds by assumption as $(\tau_l, \theta) \in (0, 1)$. Similarly, in R2, it suffices to assume $(1 - \tau_l)\phi < 1$, which is satisfied as $\phi \in (0, 1)$. Finally, R3 is trivial. This completes the proof. ■

As $\mathbb{Z}$ has finite cardinality, we can equip it with a transition matrix, π. These two elements then define an infinite horizon process (Ω, π), where Ω is the infinite Cartesian product of $\mathbb{Z}$, $\Omega \equiv \mathbb{Z} \times \mathbb{Z} \times ...$ that we will use to simulate the model. For any fixed-point of the Coleman-Reffett operator in (B.17), using (B.15), (B.5), and (B.6), we can derive the policy function for consumption $C([0, K_{max}] \times \{\mathbb{Z}\})$. Then, given the state space, $[0, K_{max}] \times \{\mathbb{Z}\}$, we can define a (vector-valued) function J mapping $[0, K_{max}] \times \{\mathbb{Z}\} \longrightarrow [C, L, F]$. Using the definition of GDP, $F = C + I$, and J, we can define a transition function T mapping $(\mathcal{Z}, K) \longrightarrow K'$ as $K' = I + (1 - \delta)K$. Finally, using T, we can define a Markov kernel $P_T([\mathcal{Z}, K], A \times B) = \{\pi(\mathcal{Z}, B) : T(\mathcal{Z}, K) \in A\}$.

From Theorem 9.13 in Stokey, Lucas and Prescott (1989), P_T is a Markov kernel and $([0, K_{max}] \times \{\mathbb{Z}\}, P_T)$ is a Markov Process. It follows that if T is continuous and $[0, K_{max}] \times \{\mathbb{Z}\}$ is compact, $([0, K_{max}] \times \{\mathbb{Z}\}, P_T)$ has an ergodic invariant measure (Futia, 1982). The theorem below will prove this result after showing that the modified Coleman-Reffett operator has a fixed point. Note that given an initial condition $(\mathcal{Z}_0, K_0)$ and using P_T, we can simulate the evolution of the minimal state space, then use J to recover C, L, and F, and F_K to compute the interest rate, r. Finally, we use the partition of the state space (B.4) and the following selection rule to simulate wages. Let $[0, K^{WUB})(\mathcal{Z}) \equiv \mathbb{K}_1(\mathcal{Z})$, $(K^{WLB}, K_{max}](\mathcal{Z}) \equiv \mathbb{K}_2(\mathcal{Z})$ and $[K^{WUB}, K^{WLB}](\mathcal{Z}) \equiv \mathbb{K}_3(\mathcal{Z})$, with K^{WUB}, K^{WLB} given by (B.4). Then, $I_m(\mathcal{Z}, K) = 1$ if $K \in \mathbb{K}_m(\mathcal{Z})$, with $m = 1, 2, 3$, and $L(\mathcal{Z}, K) = I_m(\mathcal{Z}, K) L_m(\mathcal{Z}, K)$, where $L_3 = \bar{h}$. Finally, remember the wage is given by

$$w_m(\mathcal{Z}, K) = \frac{\alpha L_m(\mathcal{Z}, K)^{1/\eta}}{(1 - \tau_l)}. \tag{B.18}$$

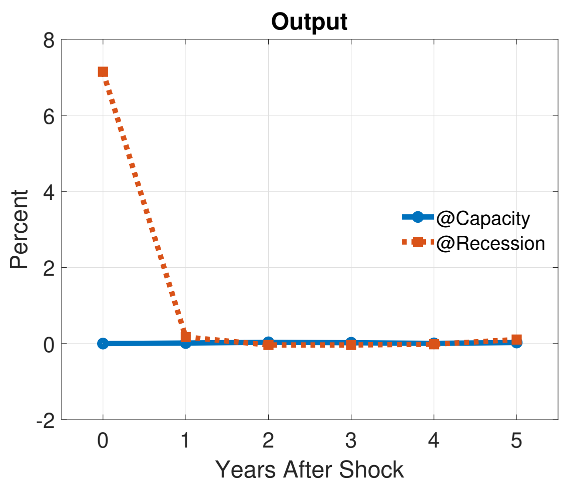
Further, if $K \in \mathbb{K}_3(\mathcal{Z})$, then $F^i_{H,3}(\mathcal{Z}, L(\mathcal{Z}, K)) \neq w_3(\mathcal{Z}, K)$. In Regime 3, as wages are not equal to the marginal product of labor, we must define quasi-rents. We denote the quasi-rent in regime m by B_m, which is given by:

$$B_m(\mathcal{Z}, K) = \max_i \left\{ F^i_{H,M}(L_m(\mathcal{Z}, K), K) \right\} - w_m(\mathcal{Z}, K), \tag{B.19}$$

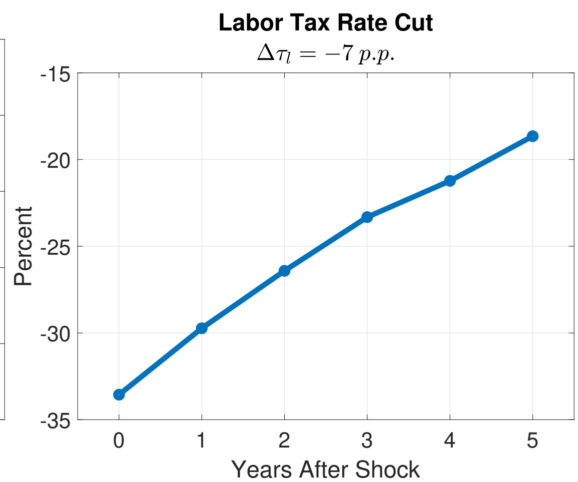
where $F^i_{H,m}$ is the directional derivative for labor in regime m. Note that $B_3 > 0$ and $B_1 = B_2 = 0$. Then, total income accruing to capital is $r_m(\mathcal{Z}, K)K + P_m(\mathcal{Z}, K) + B_m(\mathcal{Z}, K)$, where $r_m(\mathcal{Z}, K) = F^0_{K,m}(\mathcal{Z}, K)$ and $P_m(\mathcal{Z}, K) = F_m(L_m(\mathcal{Z}, K)) - r_m(\mathcal{Z}, K)K - w_m(\mathcal{Z}, K)L_m(\mathcal{Z}, M)$. Once we define the selection rule for wages, using Lemmas 2 and 5, Proposition 3 together with Corollary 4, we prove the main theorem of the paper.

Proof of Theorem 1. Under Lemmas 2 and 5, and Corollary 4, (B.17) defines a slightly modified Coleman-Reffett operator. Proposition 6 in Coleman (1991) implies existence of a continuous fixed point for the composite in minimal state space. The equilibrium is unique if we initialize iterations in the supremum of $\mathbb{C}_1$. Then, Proposition 3 guarantees consumption is non-negative and bounded. Moreover, this equilibrium is constructive as we can iteratively converge to it by starting operator (B.17) in the supremum of $\mathbb{C}_1$. The compactness and continuity of equilibrium labor, consumption, and investment imply the ergodicity of the recursive equilibrium in minimal state space (Futia, 1982; Stokey, Lucas and Prescott, 1989) as we can compute equilibrium wages using F_H^i and the selection rule (19).

C. Additional Results



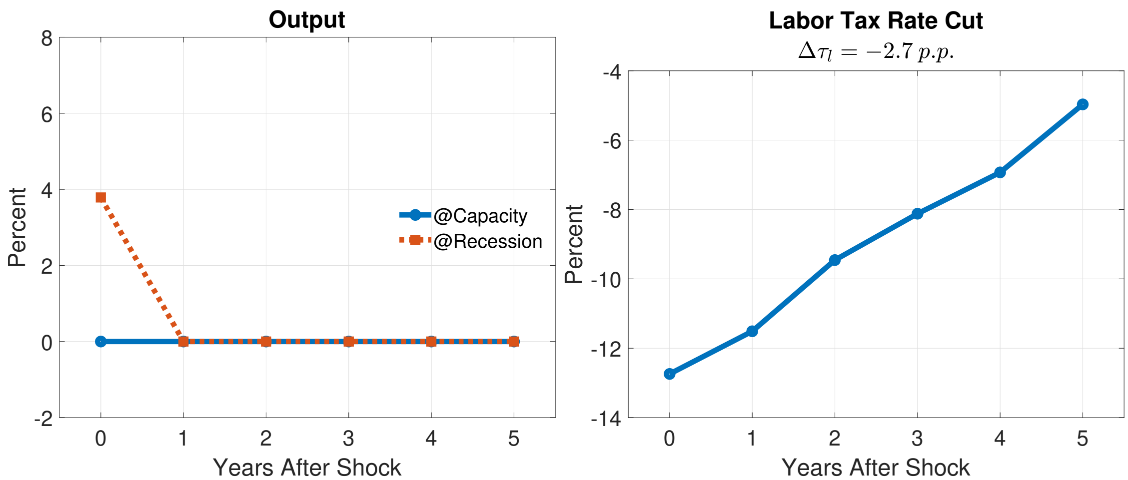
(a) IRF of Y



(b) IRF of τ_l

Figure C.1: Output Response to a Large Tax Cut

Notes: The figure shows the IRF for output (left panel) to a "large" labor tax cut (right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.

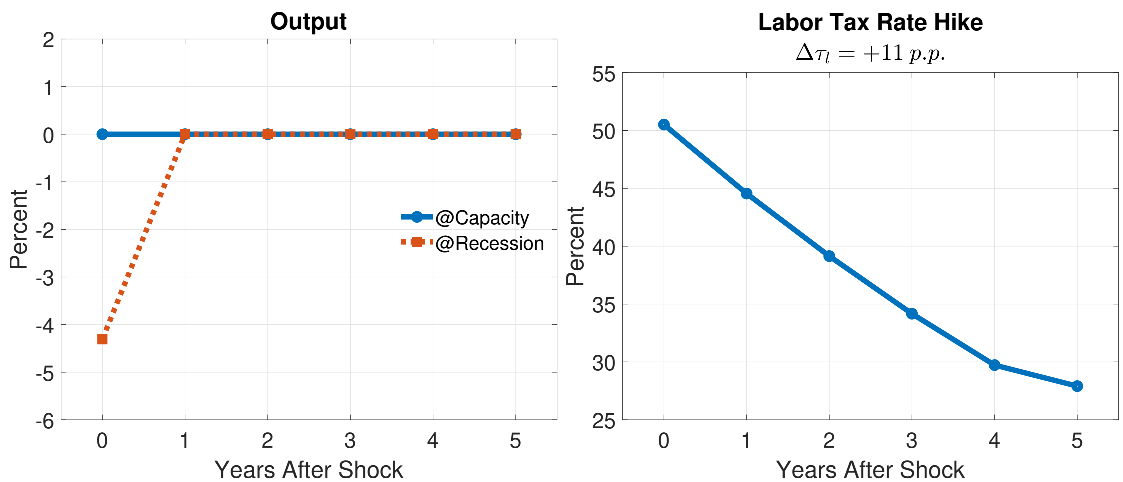


(a) IRF of Y

(b) IRF of τ_l

Figure C.2: Output Response to a Small Tax Cut

Notes: The figure shows the IRF for output (left panel) to a "small" labor tax cut (right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.



(a) IRF of Y

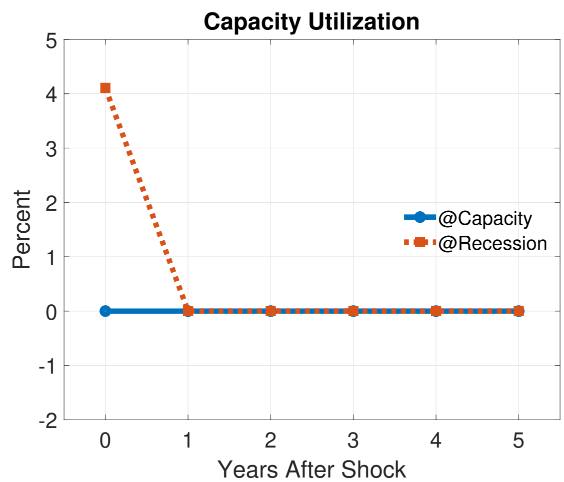
(b) IRF of τ_l

Figure C.3: Output Response to a Large Tax Hike

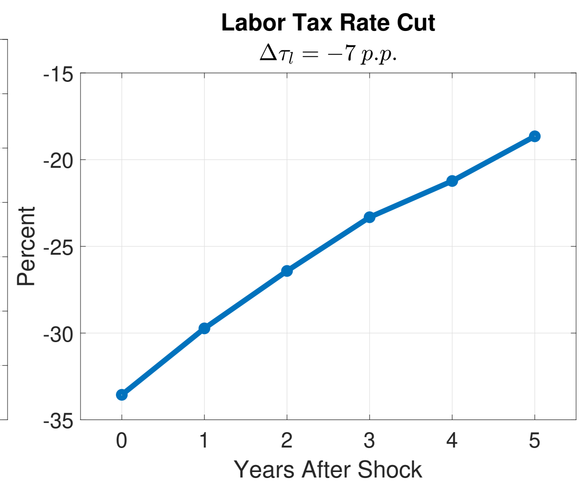
Notes: The figure shows the IRF for output (left panel) to a "large" labor tax hike (right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.

(a) IRF of H

(b) IRF of the Elasticity of H



(c) IRF of $H/\overline{H}$



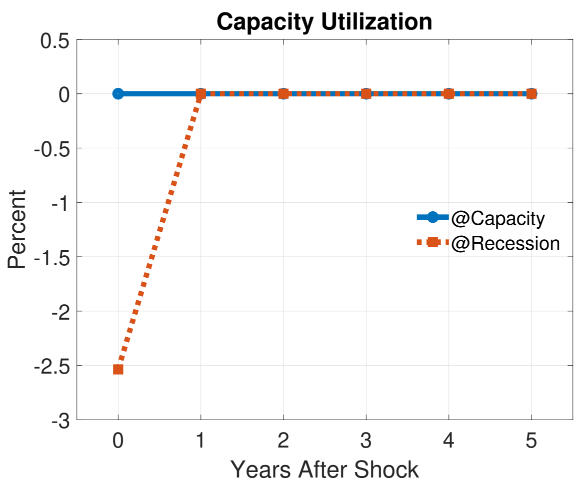
(d) IRF of τ_l

Figure C.4: IRFs to a Large Tax Cut with Labor Requirement $\overline{h} = 0.13$

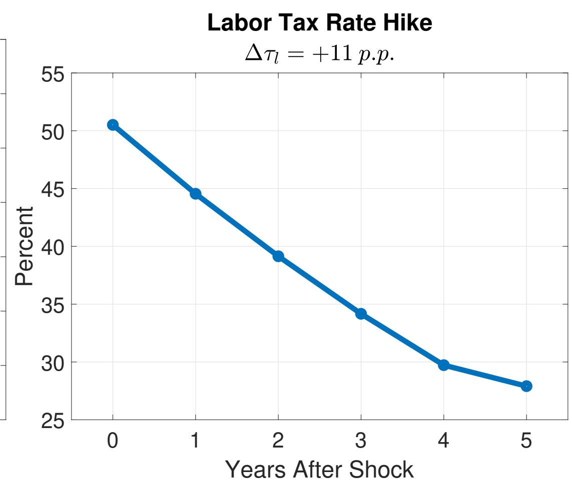
Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "large" labor tax cut (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.

(a) IRF of H

(b) IRF of the Elasticity of H



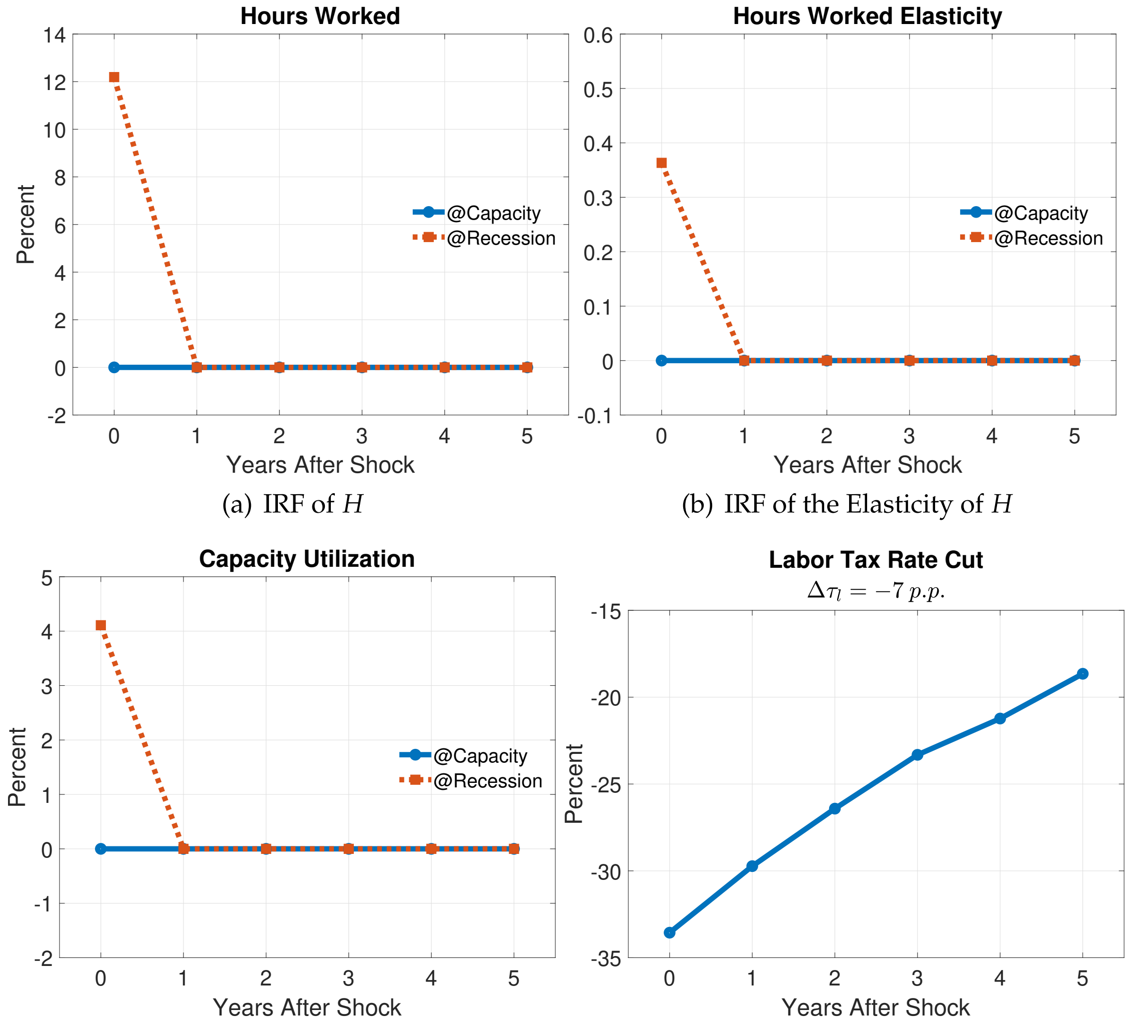
(c) IRF of $H/\overline{H}$



(d) IRF of τ_l

Figure C.5: IRFs to a Large Tax Hike with Labor Requirement $\overline{h} = 0.13$

Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "large" labor tax hike (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.



(a) IRF of H

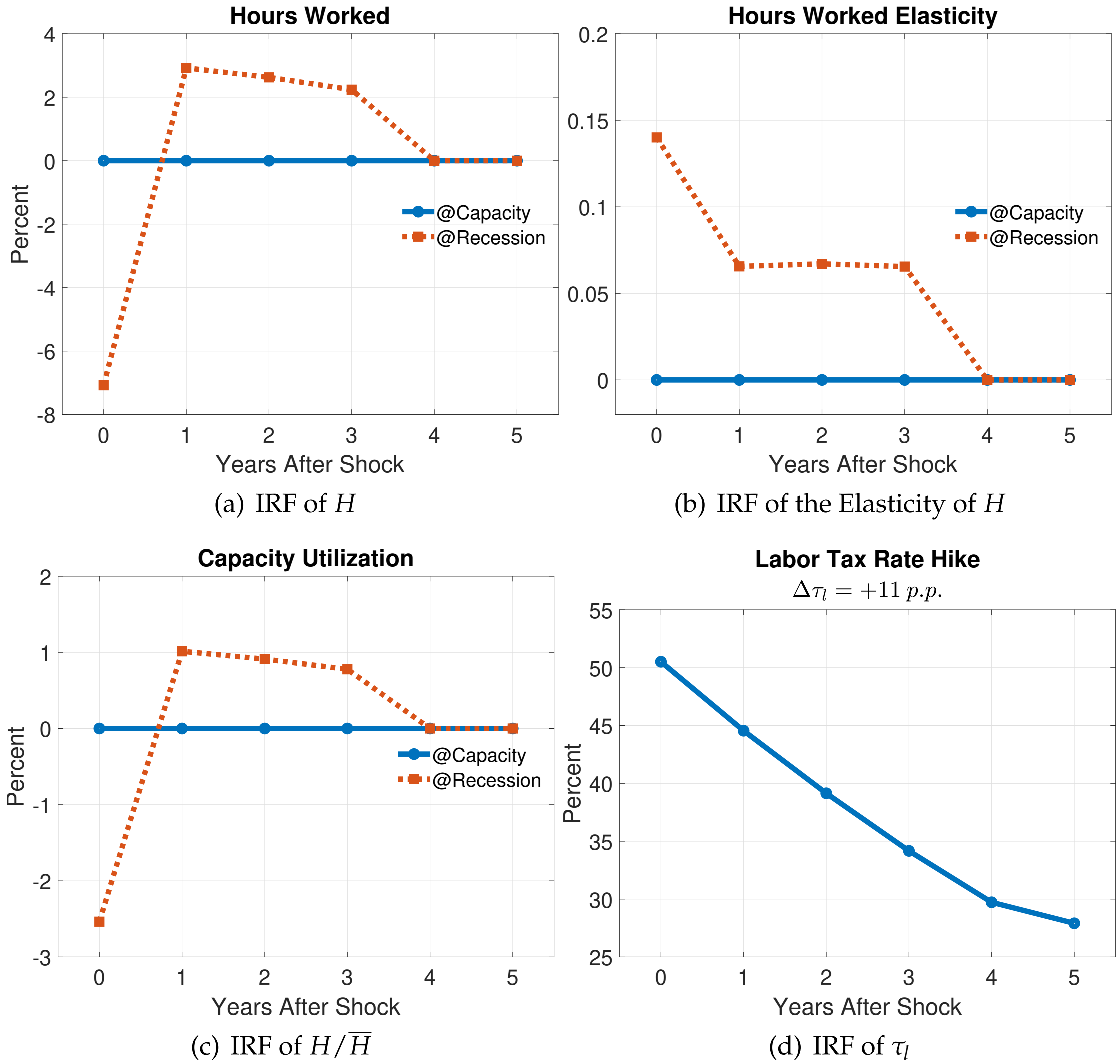
(b) IRF of the Elasticity of H

(c) IRF of $H/\overline{H}$

(d) IRF of τ_l

Figure C.6: IRFs to a Large Tax Cut with Labor Requirement $\overline{h} = 0.17$

Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "large" labor tax cut (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.



(a) IRF of H

(b) IRF of the Elasticity of H

(c) IRF of $H/\overline{H}$

(d) IRF of τ_l

Figure C.7: IRFs to a Large Tax Hike with Labor Requirement $\overline{h} = 0.17$

Notes: The figure shows the IRFs for hours worked (top-left panel), hours worked elasticity (top-right panel), and capacity utilization (bottom-left panel) to a "large" labor tax hike (bottom-right panel) when the economy is at capacity (solid line with circles) and below capacity (dotted line with squares). We compute the IRFs by simulating the equilibrium paths of two economies—"benchmark," and "counterfactual"—and taking the difference between the two. The benchmark economy starts with the tax rate at its median value on the grid, and, going forward, tax shocks continuously hit it. The counterfactual economy begins instead with a tax rate below the median and, as the benchmark, is hit by shocks drawn from the same AR(1) process. The tax shock realizations differ indexed by the initial conditions. In addition to the tax shocks, technology shocks hit both economies, forcing the technology shock realizations to be identical. We report the average IRFs across 10,000 replications.