Board of Governors of the Federal Reserve System

International Finance Discussion Papers

ISSN 1073-2500 (Print)
ISSN 2767-4509 (Online)

Number 1440

July 2026

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

Domenico Ferraro, Giuseppe Fiori, and Damian Pierri

Please cite this paper as:
Domenico, Ferraro, Giuseppe Fiori, and Damian Pierri (2026). "Multi-Plant Firms, Variable Capacity Utilization, and the Aggregate Hours Elasticity," International Finance Discussion Papers 1440. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/IFDP.2026.1440.

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

Domenico Ferraro Giuseppe Fiori Damián Pierri

June 1, 2026

Abstract

We develop a business cycle model with perfectly competitive product and labor markets in which production requires a minimum labor input, generating endogenous capacity utilization. The aggregate production function is kinked, featuring constant returns to scale below capacity—typically in recessions—and decreasing returns at capacity in expansions. Motivated by new empirical evidence that narratively identified labor tax shocks have significantly larger effects on hours and output when capacity utilization is below trend, we calibrate the model to U.S. data and show that the aggregate hours elasticity is higher in recessions, differing markedly from the micro elasticity implied by preferences.

JEL Classification: E22; E23; E24; E32; E62; H24; H25.
Keywords: Minimum labor requirements; Hours constraints; Capacity utilization; State dependence; Labor taxation; Elasticity of aggregate hours.

*Ferraro: Department of Economics, W. P. Carey School of Business, Arizona State University, United States (email: domenico.ferraro@asu.edu); Fiori: Division of International Finance, Board of Governors of the Federal Reserve System, United States (e-mail: giuseppe.fiori@frb.gov); Pierri: ESCP Business School and Departamento de Teoría e Historia Económica, Universidad Autónoma de Madrid, Spain, and Facultad de Ciencias Económicas, Universidad de Buenos Aires, Argentina (e-mail: damian.pierri@gmail.com). We thank the Editor, Morten Ravn, three anonymous referees, Florin Bilbiie, Antonia Díaz, Evi Pappa, Kevin Reffett, Juan Pablo Rincón-Zapatero, and Hernán Seoane for valuable comments. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System (or of any other person associated with the Federal Reserve System).

1. Introduction

In the United States, capacity utilization exhibits pronounced cyclical variation, remaining close to an upper bound associated with full utilization of productive resources during expansions and declining sharply during recessions. These patterns suggest that economic activity is frequently constrained by capacity in expansions but temporarily operates below capacity after adverse shocks. Motivated by this evidence, this paper asks whether the effects of distortionary taxes on aggregate hours worked and output are state-dependent, varying systematically with the degree of capacity utilization.

This paper makes three contributions to the literature discussed in Section 2. First, Section 3 documents new empirical evidence of state-dependent labor tax effects using narratively identified tax shocks (Romer and Romer, 2009; Mertens and Ravn, 2013, 2014; Mertens and Montiel Olea, 2018). Tax policy has significantly stronger effects on market hours worked when capacity utilization is below trend and substantially weaker effects when it is above trend.

Second, we show that pronounced state dependence in the effects of labor taxes can arise in a fully competitive economy without frictions that impede market clearing, non-competitive pricing, nominal rigidities, or asymmetric adjustment costs. The mechanism is a technological minimum labor requirement at the plant level, which gives rise to two margins of adjustment: the number of active plants and hours per plant. Hours below the minimum produce no output, while hours above the threshold yield output under diminishing returns, as in Lucas (1978). Although production is nonconvex at the plant level, the capacity choices of the firms convexify the aggregate production set, generating an endogenously kinked aggregate production function that is constant returns to scale below capacity, as in the standard RBC model, and exhibits aggregate diminishing returns at full capacity.

Section 4 develops this mechanism in a static model. Below capacity—typically in recessions—the minimum hours constraint binds for some plants, firms leave capacity idle, and adjustment occurs primarily along the extensive margin, making aggregate hours highly responsive to shocks. At capacity—typically in expansions—all plants operate and adjustment occurs through hours per plant, a margin subject to decreasing returns, which dampens the response of hours. As a result, proportional tax changes generate nonlinear, state-dependent effects, and occasionally binding capacity constraints break the standard separation between preferences and technology in determining aggregate labor supply elasticities (Hornstein and Prescott, 1993).

The third contribution is methodological. Occasionally binding capacity constraints in a dynamic general equilibrium model with stochastic distortionary taxes make the competitive equilibrium non–Pareto optimal and rule out a social planner formulation. In our setting, a kinked aggregate production function and a discontinuous marginal product of labor further render local approximation methods unsuitable. We address these challenges by adapting the monotone-map (Coleman–Reffett) method (Coleman, 1990, 1991; Reffett, 1996) to kinked production environments. The approach exploits that the kink's location is price-independent and applies an equilibrium selection rule under which, at the kink, hours are determined by labor demand and wages by the household's marginal rate of substitution. At capacity, this implies a wedge between wages and the marginal product of labor, generating quasi-rents beyond those from decreasing returns.

Our main theorem, based on Greenwood, Hercowitz and Huffman (1988) preferences, establishes that the competitive equilibrium is unique, ergodic, and constructive, with an existence proof yielding an iteratively convergent algorithm for accurate computation. This ensures well-defined generalized impulse response functions (GIRFs) for numerical analysis.

Section 5 embeds this mechanism in an infinite-horizon economy with capital accumulation, encompassing a standard RBC model when capacity is slack and a Lucas-style span-of-control model in which all plants operate under diminishing returns to labor. As a result, the propagation of tax shocks differs qualitatively across business cycle phases.

Sections 6 and 7 calibrate the model to U.S. data and quantify how the effects of labor tax shocks vary with capacity utilization. We focus on labor taxes, which directly distort labor supply and generate most U.S. tax revenue. To capture nonlinear propagation, we compute GIRFs to tax cuts, conditioning on whether the economy is below or at full capacity. Labor tax shocks follow a symmetric AR(1) process estimated from post-1950 U.S. effective average labor tax rates (Mendoza, Razin and Tesar, 1994), so state dependence arises endogenously rather than from the shock process.

We find that aggregate hours respond more strongly to tax cuts in recessions than in expansions, implying a state-dependent aggregate labor supply elasticity distinct from the micro-elasticity implied by preferences. The nonlinearity stems from transitions between a constant-returns regime below capacity and a diminishing-returns regime at full capacity; within each regime, responses are nearly linear.

2. Related Literature

Our work relates to three strands of literature. The first studies the general equilibrium effects of distortionary taxes in RBC models (Aiyagari, Christiano and Eichenbaum, 1992; Baxter and King, 1993; Braun, 1994; Burnside, Eichenbaum and Fisher, 2004; McGrattan, 1994; McGrattan, Rogerson and Wright, 1997; McGrattan and Prescott, 2005; McGrattan and Ohanian, 2010; McGrattan, 2012). In these models, proportional taxes distort relative prices and induce intertemporal substitution: hours respond immediately, while capital adjusts gradually, propagating shocks over time. Although this literature attributes a nontrivial share of U.S. output and hours volatility to tax variation, it typically relies on local approximations that rule out nonlinear propagation and state dependence (Aruoba, Fernandez-Villaverde and Rubio-Ramirez, 2006).

In our framework, tax changes can cause capacity constraints to bind, making their effects depend on whether the economy operates below or at full capacity. The closest related paper is Hansen and Prescott (2005), which studies cyclical asymmetries in hours using a similar production structure but considers only productivity shocks and relies on the social planner's solution. By contrast, distortionary taxation in our setting requires us to solve directly for the competitive equilibrium and prices in an economy with a kinked production function.

The second strand examines whether the effects of fiscal policy vary over the business cycle. Empirical work has focused mainly on government spending and has reached mixed conclusions (Auerbach and Gorodnichenko, 2012; Ramey and Zubairy, 2018; Barnichon, Debortoli and Matthes, 2022). Theoretical models typically generate such nonlinearities through search frictions, asymmetric adjustment costs, or occasionally binding nominal constraints (Christiano, Eichenbaum and Rebelo, 2011; Ferraro and Fiori, 2023; Ghassibe and Zanetti, 2022; Michaillat, 2014; Pizzinelli, Theodoridis and Zanetti, 2020). We show instead that pronounced state dependence arises even in frictionless, competitive economies when production is subject to a minimum labor requirement at the plant level: adjustment occurs along the extensive margin in recessions and along the intensive margin in expansions, generating intrinsically state-dependent responses of hours and output to tax shocks.

Finally, our paper relates to the literature on computational methods for nonlinear dynamic economies. While Rendahl (2015) studies Pareto-optimal allocations, we analyze a decentralized, non–Pareto-optimal equilibrium with distortionary taxation. We reduce equilibrium characterization to a single Euler equation, allowing us to apply the Coleman–Reffett framework. This reduction relies on an equilibrium selection rule that preserves continuity of income and capital returns despite a discontinuous marginal product of labor. A closely related approach is Mertens and Ravn (2014), who propose a block-recursive method for computing decentralized, nonoptimal equilibria but do not establish convergence. We prove convergence, albeit under a more restrictive equilibrium structure, again characterized by a single Euler equation.

3. State-Dependent Effects of Labor Taxes

This section presents new evidence that the effects of labor taxes are state-dependent. Using narratively identified tax shocks and local projections, we show that the aggregate hours elasticity is larger when the economy is operating below capacity.

3.1. Empirical Strategy

Our approach to estimating the effects of labor tax shocks proceeds in three steps. First, we calculate the average marginal labor tax rates. Second, drawing on narrative sources, we classify legislated tax changes as exogenous to contemporaneous economic conditions or endogenous responses to ongoing developments. Third, we estimate the effects of the narratively-identified labor tax shocks on output, hours worked, and employment, and examine how these effects vary systematically with capacity utilization. We now describe each step in turn.

Average Marginal Tax Rates. We begin by detailing our construction of the average marginal tax rate (AMTR). Following Barro and Redlick (2011) and subsequent work, we define labor income broadly to include wages, self-employment income, partnership distributions, and S-corporation earnings. Individual-level microdata are drawn from the March Supplement of the CPS (ASEC), which links income to demographic variables.

The AMTR incorporates both federal individual income and payroll (FICA) taxes. Using NBER TAXSIM, we simulate individual-level marginal tax rates for income and payroll taxes separately. We then aggregated these rates by computing a weighted average of the average marginal individual income tax rate (AMIITR) and the average marginal payroll tax rate (AMPTR), with weights based on adjusted gross income shares.

Tax Shocks and Identification. Identifying the causal effects of tax changes on macroeconomic outcomes is challenging because most tax rate movements are countercyclical policy responses, creating endogeneity. We address this issue by following Mertens and Montiel Olea (2018) and Ferraro and Fiori (2020) and adopting the narrative identification strategy of Romer and Romer (2009). In this framework, changes in total tax liabilities are classified as "exogenous" when the stated legislative motivation reflects long-run objectives unrelated to contemporaneous business cycle conditions or the need to address inherited budget deficits. This classification isolates genuine policy innovations rather than systematic responses to current conditions.

Implementation lags create an additional complication by generating anticipation effects, that is, aggregate activity can move before statutory changes take effect (Mertens and Ravn, 2012). To limit such contamination, we restrict attention to individual income-tax liability changes that were both legislated and implemented within the same calendar year, following Mertens and Ravn (2013).

Applying these criteria yields four exogenous tax reforms in our sample period. As detailed in subsection 3.2, we use these shocks both as direct measures of policy change and as instruments for AMTR in our empirical specifications. We measure the impact of a year-t reform as the difference between two counterfactual AMTRs computed on the same income distribution (that of year $t-1$). The first applies the year t statutory rates and brackets; the second applies the year $t-1$ schedule. Their difference isolates the effect of the reform implemented in year t on the AMTR. A complication is the CPI indexation of the federal income tax beginning in 1985.[1] To address this, we rescale incomes by the automatic adjustments in bracket widths embedded in the federal tax code.

Capacity Utilization as a Proxy for Economic Slack. We measure economic slack using the aggregate capacity utilization index from the Federal Reserve Board of Governors' G.17 Industrial Production and Capacity Utilization release. The index measures the extent to which the economy operates relative to "sustainable maximum output," defined as the highest level of production that can be maintained under normal operating conditions, allowing for routine downtime and full use of the installed capital stock.[2] By comparing actual production to sustainably attainable output given existing resources, capacity utilization provides a particularly informative indicator of cyclical conditions.

[1] The Economic Recovery Tax Act of 1981 mandated automatic CPI-based increases in personal exemptions, the standard deduction, and bracket widths starting in 1985.

[2] For methodological details, see g17/capnotes.htm.

This gap between actual and potential output plays a central role in our theoretical framework, developed in Sections 4 and 5.

3.2. Empirical Implementation

To implement our empirical strategy, we employ local projections, which have become a standard tool in fiscal policy analysis (Jordà, 2005). Our baseline specification follows Romer and Romer (2009) and relates the outcome variable of interest y_{t+h}—measured h years ahead and given by real GDP, hours per capita, or employment—to narrative labor tax shocks, τ_t^{shock}, their interaction with detrended capacity utilization, $\tau_t^{\text{shock}}\widehat{CU}_{t-1}$, detrended capacity utilization itself, and other controls, over a sample spanning 1988 to 2019:

$$y_{t+h} = \alpha_h + \beta_h\, y_{t-1} + \left(\gamma_{h,1} + \gamma_{h,2}\,\widehat{CU}_{t-1}\right)\tau_t^{\text{shock}} + \delta_h\,\widehat{CU}_{t-1} + \eta_h\,\text{controls}_{t-1} + \varepsilon_t. \quad (1)$$

This specification allows us to estimate the dynamic effects of labor tax shocks on real GDP, hours per capita, and employment, and how these effects vary systematically with the degree of capacity utilization across the business cycle. The tax shocks described in subsection 3.1 are unanticipated policy innovations—unexpected given current and past information and exogenous to contemporaneous macroeconomic conditions—ensuring that OLS estimation of equation (1) delivers consistent and unbiased estimates of the coefficients of interest. Variables are at an annual frequency, except for capacity utilization, which corresponds to the last quarter of the year.

As discussed in Ramey and Zubairy (2018) and Jordà (2005), the specification requires careful attention to variable timing to preserve the causal interpretation of the results. The narrative tax shock τ_t^{shock} is dated at period t, capturing the moment at which the policy innovation occurs rather than its realization, ensuring orthogonality with respect to information available at $t-1$. The state variable $\widehat{CU}_{t-1}$, the linearly detrended capacity utilization, pre-determined relative to the shock: dating it at $t-1$ prevents the simultaneity problem that would arise if economic slack responded endogenously to τ_t^{shock}, which is essential for the interaction term $\left(\gamma_{h,1} + \gamma_{h,2}\,\widehat{CU}_{t-1}\right)\tau_t^{\text{shock}}$ to recover a clean estimate of state-dependence in the tax multiplier. The same logic applies to the controls, which enter with at least one lag to avoid reintroducing the simultaneity that the local projection framework is designed to sidestep. Following Ramey and Zubairy (2018), we include one lag of the yearly outcome variable and relevant fiscal aggregates to absorb serial correlation and pre-existing trends that might otherwise be misattributed to the shock.

Following Mertens and Montiel Olea (2018), the control vector includes the log of real government spending per capita, the change in the log of real federal debt per capita and the real S&P 500 index; full details on variable construction are provided in Appendix A. The inclusion of these variables serves two distinct purposes. First, government spending and federal debt control for concurrent fiscal policy actions that may coincide with tax changes, ensuring that the estimated effects of τ_t^{shock} are not confounded by other policy instruments operating simultaneously. Second, and more subtly, the S&P 500 index addresses anticipation effects: equity markets aggregate forward-looking information and tend to respond to tax policy news before its macroeconomic consequences materialize, so that omitting stock prices could allow anticipated future conditions to contaminate the estimated contemporaneous impact of the shock. Together, these controls guard against omitted variable bias along both the policy and expectations dimensions, reinforcing the identifying assumption that τ_t^{shock} is orthogonal to information available at $t-1$ conditional on the included regressors.

Alternative Estimation Methods and Specifications. We also employ the instrumental variable approach of Mertens and Ravn (2013) and Mertens and Montiel Olea (2018), using the narrative tax shocks—and their interaction with detrended capacity utilization—as instruments for the average marginal tax rate and its interaction with $\widehat{CU}_{t-1}$, respectively. The IV estimates, reported in Appendix B.1, confirm the main findings and reinforce the robustness of our empirical conclusions.

Appendix B.2 presents two additional robustness checks that relax the parametric assumptions on the form of nonlinearity. Together, they provide reassurance that the main findings are not an artifact of the linear interaction assumed in equation (1). The motivation for these checks is straightforward: while the baseline specification imposes that the fiscal multiplier varies linearly and continuously with $\widehat{CU}_{t-1}$, the true relationship between economic slack and the size of the multiplier may be better approximated by a discrete regime change, or alternatively by a non-linear continuous function that amplifies large deviations from trend more than proportionally. The two checks are designed to probe these possibilities without materially increasing the number of free parameters, which is an important consideration given the limited number of narrative shock observations.

3.3. Empirical Findings

The specification (1) allows us to estimate how the outcome variables respond to changes in tax policy conditional on the level of utilization of the productive capacity relative to its trend. The interpretation is straightforward. For a given horizon h, when capacity utilization is at trend ($\widehat{CU}_{t-1} = 0$), the effect of tax policy is given by the coefficient γ_1. When capacity utilization deviates from the trend ($\widehat{CU}_{t-1} \neq 0$), the total effect of the policy is equal to the sum of the baseline effect γ_1 and the interaction term $\gamma_2 \widehat{CU}_{t-1}$. Conditional on higher taxes reducing GDP (that is $\gamma_1 < 0$), a positive estimate of γ_2 implies that tax policy effects are stronger when capacity utilization is below trend and weaker when the economy operates above it, regardless of whether the underlying tax change is a cut or an increase.

Table 1 supports the view that capacity utilization shapes the effects of tax policy on real GDP and hours per capita. Columns A and B in Panel A shows that when capacity utilization is at trend, reductions in marginal tax rates are associated with higher real GDP, consistent with Barro and Redlick (2011) and the broader empirical literature on tax shocks (Mertens and Ravn, 2013, 2014). Despite our more restricted sample, the estimated magnitudes closely align with those reported by Mertens and Montiel Olea (2018), both on impact and one year ahead.

The interaction coefficient γ_2 is consistently positive across specifications, confirming that the effects of tax policy are amplified when capacity utilization falls below its trend and dampened when capacity is above trend. Panel B reports the implied semi-elasticities based on our estimates. For example, the impact semi-elasticity when capacity utilization is below trend is equal to -0.75. One year ahead, the estimated semi-elasticities reach about -1.2.

These estimates imply that a one–percentage point reduction in the marginal tax rate raises real GDP by more than 0.7 percent on impact and by more than 1 percent after one year when the economy is operating below capacity. By contrast, both the impact and one-year-ahead elasticities are smaller, though still statistically significant, when the economy is operating at trend, and are statistically insignificant when capacity utilization is above trend. Taken together, these results indicate that the effects of tax policy are state dependent, varying systematically with the degree of capacity utilization in the economy.

The state dependence of tax policy extends beyond output to labor market outcomes, including hours worked per capita and employment. Columns C and D of Table 1 shows that the semi-elasticities of per capita hours worked are substantially larger when the

Table 1: Effects of Tax Shocks on Real GDP and Hours

	log (Real GDP)		log (Hours per Capita)	
	$h = 0$	$h = 1$	$h = 0$	$h = 1$
	(A)	(B)	(C)	(D)
Panel A:				
Tax shock (γ_1)	-0.356^{***}	-0.755^{***}	-0.198^{*}	-0.474^{*}
	(0.12)	(0.25)	(0.10)	(0.25)
Tax shock $\times$ $\widehat{CU}$ (γ_2)	0.136^{***}	0.157^{*}	0.093^{**}	0.171^{***}
	(0.04)	(0.08)	(0.04)	(0.04)
Panel B:				
Semi-elasticity when capacity utilization is:				
1 std. dev. below trend	-0.746^{***}	-1.206^{***}	-0.465^{**}	-0.967^{***}
at trend	-0.356^{***}	-0.755^{***}	-0.198^{*}	-0.474^{*}
1 std. dev. above trend	0.034	-0.304	0.070	0.018
Testing for differences between "below trend" vs. "at trend":				
p-value	0.003	0.061	0.017	0.001

Notes: The sample period runs from 1988 to 2019. Each specification includes the lag of the dependent variable, the lag of linearly detrended capacity utilization, the log of real government spending per capita, the change in the log of real federal debt per capita, and the change in the log of the real S&P 500 index. Variables are at an annual frequency, except for capacity utilization, which corresponds to the last quarter of the year. Newey-West standard errors, computed with two lags, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

economy is operating below capacity than when capacity utilization is at trend. For example, the impact semi-elasticity is -0.47 and the one-year-ahead semi-elasticity is -0.97 when capacity utilization is below trend, compared to much smaller values of -0.20 and -0.47, respectively, when utilization is at trend. These differences are statistically significant.

Table 2: Effects of Tax Shocks on Real Investment and Employment

	log (Real Investment)		log (Employment Full-time)	
	$h = 0$ (A)	$h = 1$ (B)	$h = 0$ (C)	$h = 1$ (D)
Panel A:				
Tax shock (γ_1)	-1.461^{**}	-2.705^{**}	-0.314^{**}	-0.784^{**}
	(0.59)	(1.29)	(0.13)	(0.33)
Tax shock $\times$ $\widehat{CU}$ (γ_2)	0.762^{***}	0.714	0.181^{***}	0.174^{***}
	(0.19)	(0.45)	(0.03)	(0.05)
Panel B:				
Semi-elasticity when capacity utilization is:				
1 std. dev. below trend	-3.653^{***}	-4.760^{***}	-0.835^{***}	-1.285^{***}
at trend	-1.461^{**}	-2.705^{**}	-0.314^{**}	-0.784^{**}
1 std. dev. above trend	0.731	-0.649	0.207	-0.283
Testing for differences between "below trend" vs. "at trend":				
p-value	0.001	0.128	0.000	0.003

Notes: The sample period runs from 1988 to 2019. Each specification includes the lag of the dependent variable, the lag of linearly detrended capacity utilization, the log of real government spending per capita, the change in the log of real federal debt per capita, and the change in the log of the real S&P 500 index. Variables are at an annual frequency, except for capacity utilization, which corresponds to the last quarter of the year. Newey-West standard errors, computed with two lags, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

As reported in Table 2, a similar pattern holds for investment and employment: differences in the estimated semi-elasticities are statistically significant on impact, although for investment they are not statistically significant at the one-year horizon.

Additional results reported in the appendix further corroborate the central finding of this section: the response of macroeconomic aggregates to labor tax shocks is systematically larger when capacity utilization is below trend. This pattern remains robust to an instrumental variable specification in which narrative tax shocks serve as instruments for the average marginal tax rate rather than entering directly as regressors, as well as to alternative specifications that replace the continuous interaction with a binary characterization of the state or allow for a non-linear response of the multiplier to the degree of capacity utilization.

Finally, we also address the methodological concerns raised by Gonçalves et al. (2024), in particular the risk of *threshold crossing*—the possibility that a shock is large enough to reverse the state of the economy, which would bias the local projection estimator. This concern is most limited for the impact horizon ($h = 0$), where the state variable is fixed by construction and no dynamic feedback has yet operated; it becomes progressively more relevant at longer horizons. To assess its empirical relevance, we estimate the impulse response of capacity utilization to the tax shock. We then conduct an observation-by-observation assessment for each year with a non-zero tax shock, computing the predicted impact change in capacity utilization and checking whether any shock is large enough to move the economy across the trend threshold. The calculations, reported in Appendix B.3, suggest that threshold crossing is negligible in practice. This is further corroborated by the insignificance of the interaction term between the tax shock and the CU determinant, which provides no evidence that the response of capacity utilization to the tax shock is itself state-dependent. One qualification deserves acknowledgment, however: the impact coefficient used to construct $\Delta\widehat{CU}$ is itself estimated from the state-dependent local projection, so the diagnostic is not fully independent of the estimator whose validity it is intended to support. The evidence is therefore best interpreted as consistent with threshold crossing being immaterial, rather than as conclusive proof.

4. A Static Model Illustrating the Mechanism

This section introduces a static general equilibrium model that illustrates the central mechanism of the paper. Unlike standard production theory, production is subject to a technological constraint requiring a minimum amount of labor services to generate output. Although the environment is closely related to Hansen and Prescott (2005), our focus on the decentralized competitive equilibrium in the presence of distortionary taxation raises additional issues that must be explicitly addressed.

4.1. Production with Minimum Labor Requirements

Production takes place at the level of an individual plant or production line. Within the context of our model, the terms *plant* and *production line* are used interchangeably. The key assumption is that labor in the present setting—and both capital and labor in the quantitative model developed in the next section—can be seamlessly reallocated across plants or production lines at will and without cost.

Each production unit, if active, operates a diminishing-returns technology, as in the span-of-control model of Lucas (1978). A plant employing h units of labor produces $f(h) = zh^{\phi}$, with $0 < \phi < 1$. To capture a minimum labor requirement, we assume that $f(h) = 0$ for $0 \leq h < \overline{h}$ and $f(h) = zh^{\phi}$ for $h \geq \overline{h}$. This formulation differentiates between hours worked and labor services: labor services are zero for $h < \overline{h}$ and equal to hours worked for $h \geq \overline{h}$.[3]

The minimum labor requirement parsimoniously captures technological and organizational features of production, such as coordination costs, setup time, and minimum efficient scale, and is consistent with the empirical observation that firms rarely employ very low-hours workers. This assumption renders the production set nonconvex and implies a discontinuous marginal product of labor: $MPL = 0$ for $h < \overline{h}$ and $MPL = \phi z h^{\phi-1} > 0$ for $h \geq \overline{h}$. With decreasing returns, the average product of labor is $APL \equiv f(h)/h = zh^{\phi-1}$, which exceeds the marginal product for all $h \geq \overline{h}$. Absent the minimum labor requirement ($\overline{h} = 0$), profit maximization in a competitive labor market implies $MPL = W$, that is, a downward-sloping labor demand curve $h = (\phi z/W)^{1/(1-\phi)}$ and positive profits $\Pi = (1-\phi)zh^{\phi} > 0$.

4.2. Single-Plant and Multi-Plant Firms

We consider two alternative production environments. In the first, firms are *single-plant*: there is a unit measure of identical firms, each operating a single technology $f(h) = zh^{\phi}$ subject to the minimum labor requirement. This case serves as a benchmark, as the standard competitive equilibrium implies that the economy operates at full capacity at all times. With identical firms, idle capacity—defined as unused technologies—cannot arise in equilibrium.

In the second environment, firms are *multi-plant*. A large number of identical firms

[3] Prescott, Rogerson and Wallenius (2009) use a related mechanism to generate extensive and intensive labor supply margins.

operate multiple identical plants and jointly choose the measure of active plants $m \leq M$ and labor input per plant h. In this setting, idle capacity ($m < M$) arises naturally as a competitive equilibrium outcome.

4.2.1. Competitive Equilibrium with Single-Plant Firms

There is a continuum of identical firms indexed by $i \in [0, M]$, each maximizing profits $\Pi = zh^{\phi} - Wh$. The first-order condition for labor is $z\phi h^{\phi-1} \leq W$, with equality for $h \geq \overline{h}$. It is convenient to define two wage cutoffs. The lower cutoff $\underline{w} \equiv MPL(\overline{h}) = z\phi\overline{h}^{\phi-1}$ is such that if $W < \underline{w}$, the minimum labor requirement does not bind and labor demand satisfies the standard condition $z\phi h^{\phi-1} = W$, yielding $h = (\phi z/W)^{1/(1-\phi)}$. In this region, profits are strictly positive. The upper cutoff $\overline{w} \equiv APL(\overline{h}) = z\overline{h}^{\phi-1} > \underline{w}$ satisfies that if $W > \overline{w}$, profits are negative for all $h \geq \overline{h}$ and firms optimally choose $h = 0$.

The resulting firm-level labor demand is

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

which features a kink at $(\overline{h}, \underline{w})$, hence the term *kinked labor demand*.

Aggregating across firms yields total labor demand

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

which inherits the kink from the firm-level demand.

A competitive equilibrium with positive real wages requires labor-market clearing, $L(W) = H(W)$, where $L(W)$ denotes labor supply. When $\overline{h} = 0$, a unique equilibrium exists under standard assumptions. However, when $\overline{h} > 0$, an equilibrium wage does not need to exist, for example with a sufficiently steep labor supply schedule, as may occur under the GHH preferences (Greenwood, Hercowitz and Huffman, 1988), or with a vertical labor supply schedule, as under the KPR preferences (King, Plosser and Rebelo, 1988), if households supply less than $M\overline{h}$ hours. Even when an equilibrium exists, it implies full capacity utilization at all times, regardless of productivity or taxes.

4.2.2. Competitive Equilibrium with Multi-Plant Firms

As shown in the previous subsection, a minimum labor requirement combined with uncoordinated production decisions generates a nonconvex aggregate production set. Such nonconvexities are problematic, as they undermine general existence and uniqueness results for competitive equilibrium and complicate computation, concerns that are amplified in dynamic stochastic settings (Brown, 1991). In addition, the single-plant equilibrium implies a *floor* on aggregate hours worked, which is inconsistent with the observed asymmetry of hours over the U.S. business cycle (Hansen and Prescott, 2005; McKay and Reis, 2008). Because hours cannot fall below this floor, the kinked labor demand curve implies implausibly large wage declines in recessions. Finally, the model cannot account for observed time variation in idle capacity, as identical firms imply full capacity utilization at all times.

To address these shortcomings, we consider an alternative environment with coordinated production decisions. There is a unit mass of identical firms, each allocating total labor input H across $m \leq M$ identical plants, subject to the feasibility constraint $mh \leq H$ and a minimum labor requirement $h \geq \overline{h}$ per plant. Firms jointly choose the number of active plants and labor per plant, allowing for endogenous variation in capacity utilization.

The firm's problem can be decomposed into two margins of adjustment. First, given total labor input H, the firm chooses the number of active plants m, the *extensive margin* of labor demand. Second, given m, the firm chooses H, which determines labor per plant $h = H/m$, the *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

$H \leq \overline{H}$.

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

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

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.[4]

As illustrated in Figure 1, the firm's production set remains convex, in spite of the non-convexity at the plant level.

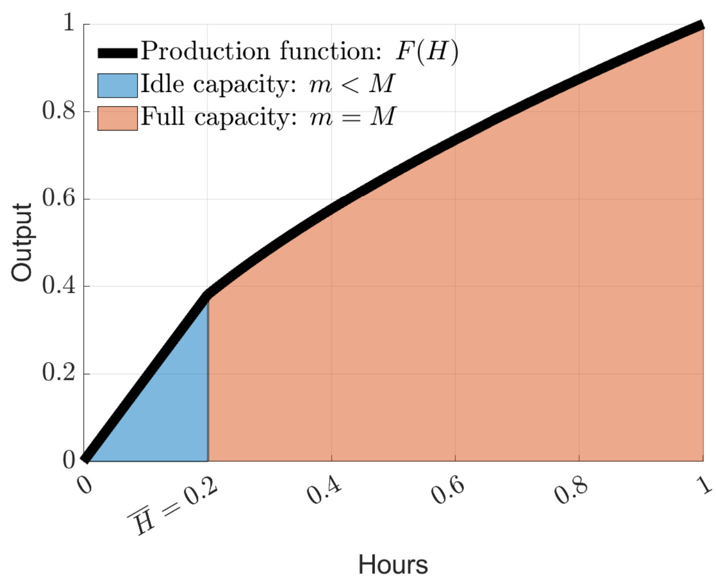


Figure 1: Linear-Concave Production Function

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 $z = 1$.

[4] In Section 5, 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 2(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. Notice that the production function and thus the average product of labor are continuous in $H > 0$, which implies that the left and right limits are equal. This is not the case for the marginal product of labor, as this function is not smooth in H. The shape of the aggregate labor demand is endogenous and depends on the level of the labor input: (i) it is horizontal and perfectly elastic for all $H < \overline{H}$; (ii) it consists of two dots and a hole for $H = \overline{H}$; and (iii) it is decreasing and convex for all $H > \overline{H}$.

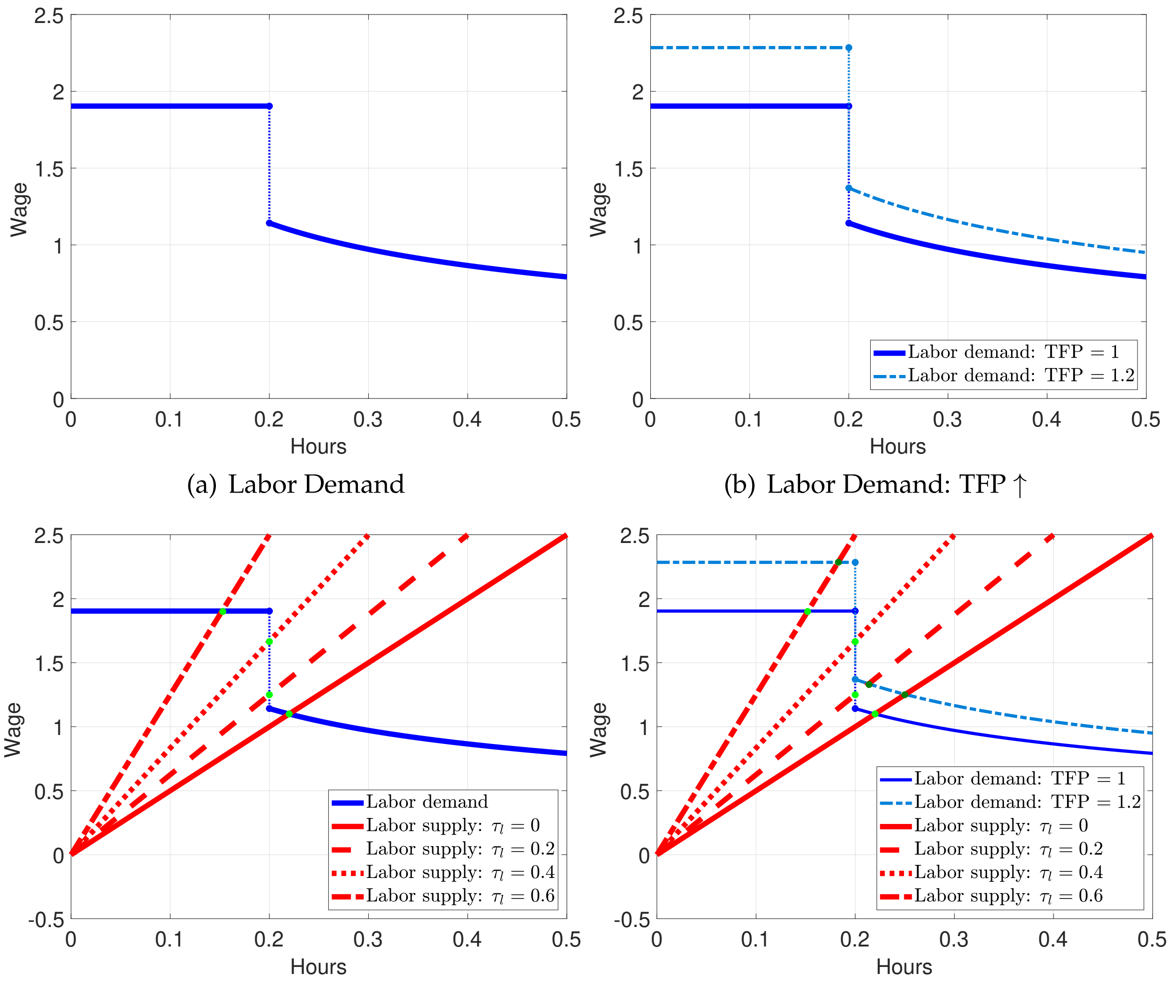
Equilibrium with a Discontinuous Labor Demand. Equipped with such generalized labor demand, we are ready to study the equilibrium of the labor market. Labor demand features a hole so that the wage would be discontinuous under standard competitive pricing, i.e., wage $= MPL$. Thus, we must take a stand 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. Furthermore, under this selection rule, there are no idle labor services or unemployed hours, making it directly comparable to labor allocation in RBC theory.

Figure 2(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. 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 case arises when the labor supply curve intersects the vertical-dotted segment of the labor demand curve. In this region, standard equilibrium logic breaks down: aggregate hours are pinned down at $H = \overline{H}$ by labor demand, while labor supply determines the wage. A leftward shift in labor supply—for example, due to an increase in the flat-rate labor tax on labor income—raises the wage without affecting hours worked. At such equilibria, workers are indifferent between working and not working, as the wage equals the marginal rate of substitution, while firms earn positive profits. Notably, the equilibrium wage lies below the marginal product of labor at the kink.

4.3. Inspecting the Mechanism

We use a numerical example to illustrate how productivity and labor tax changes affect the static labor market equilibrium under competitive equilibrium with multi-plant firms. The responses are highly nonlinear and depend critically on the equilibrium's position along the labor demand curve. In particular, hours worked are more sensitive to tax shocks in recessions, when capacity is idle, than in expansions.



(a) Labor Demand

(b) Labor Demand: TFP ↑

(c) Equilibrium

(d) Equilibrium: TFP ↑

Figure 2: 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 2(a) shows the aggregate labor demand implied by the model where the discontinuity associated with the kink in the production function occurs at $H = \overline{H} \equiv M\bar{h}$. Figure 2(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 2(d) shows how the labor market equilibrium with different configurations of the labor tax rate in Figure 2(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. As shown in Figure 2(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. Infinite-Horizon Model with Physical Capital

In this section, we embed multi-plant firms subject to a minimum labor requirement into a stochastic neoclassical growth model. Aggregate uncertainty arises from shocks to total factor productivity and proportional tax rates. Because distortionary taxation renders the economy nonoptimal, we characterize and solve directly for the competitive equilibrium rather than relying on a planner's problem.

5.1. 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 assumed to be numéraire and its price is set to one. Households have three sources of income: (i) labor income, $W_t l_t$, where W_t is the real wage rate; (ii) capital income, $R_t k_t$, by renting out k_t units of capital at the real rental rate, R_t; and (iii) dividends from the ownership of firms, Π_t. The government levies proportional tax rates $\tau_{l,t}$ on labor income and rebates the proceeds in a lump-sum fashion, T_t, so that the taxes net of transfers are $\mathcal{T}_t \equiv \tau_{l,t} W_t l_t - T_t$. 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, $\{\tau_{l,t}\}_{t=0}^{\infty}$, 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:

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

Finally, the transversality condition applies.

5.2. Business Sector

The business sector is populated by a unit mass of identical firms, each with M production units. In each period, a firm chooses the mass $m_t \in [0, M]$ of plants to operate and the allocation of labor, h_t, and capital, k_t, per plant. 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 $\phi + \theta < 1$.

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 4, the marginal product of labor, $F_H(H_t, K_t)$, is discontinuous at $F_H(\overline{H}, K_t)$. At the kink, the marginal product of labor has two directional derivatives, which are finite due to the concavity of the production function. As we show in the appendix, the marginal product of capital, $F_K(H_t, K_t)$, is instead continuous, so that the firm's first-order condition for renting capital and the determination of the capital rental rate follow standard intuition.

5.3. Government Sector

The government budget is balanced period by period, so that government spending equals tax revenues, $G_t + T_t = \tau_{l,t} W_t l_t$, for all t, where we specify government purchases as a share of output, $G_t \equiv gY_t$, 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.

5.4. Competitive Equilibrium

In this subsection, we define competitive equilibrium in sequential and recursive form and discuss the associated computational challenges. The equilibrium definition is nonstandard, as it incorporates a selection rule that applies when the economy operates at the kink of the aggregate production function. As we have two possible marginal products of labor at the kink, both well-defined, we must refine the competitive equilibrium. We then establish its existence, uniqueness, and ergodicity.

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, $L_t \equiv \int_0^1 l_{i,t} di = H_t$. Under the selection rule that 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, $C_t + X_t + G_t = Y_t$, where $X_t = K_{t+1} - (1 - \delta)K_t$ denotes investment.

5.4.1. Wage Determination at the Kink

Before defining the Recursive Competitive Equilibrium (RCE), we introduce the notation and key definitions, beginning with the equilibrium selection rule. In our price-taking environment, individual decisions do not affect prices, so the equilibrium wage and interest rate depend only on aggregate states. We denote individual choice and state variables by lowercase letters and aggregate state variables by uppercase letters, and collect exogenous states—technology and tax shocks—in $\mathcal{Z}$. Because the marginal product of labor differs from the equilibrium wage when the economy operates at capacity, we distinguish quasi-rents, denoted by B, from profits due to decreasing returns to scale, denoted by P. Making the dependence of prices on aggregate states explicit, the household's budget constraint can be written as

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

The model economy exhibits three regimes, denoted by $s = \{1, 2, 3\}$, where one is "expansion," two is "recession," and three is "at capacity." In the appendix, we show that it is possible to partition the state space for capital as a function of solely exogenous shocks. We denote the partition for the regime s as $\mathbb{K}_s(\mathcal{Z})$. Formally, 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:

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

This selection rule is the new object that ensures the continuity of 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 (that is, $L_3(\mathcal{Z}, K) = \bar{h}M$ 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. Then (19) implies that 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 capital remuneration is $R(\mathcal{Z}, K)k + P(\mathcal{Z}, K) + B(\mathcal{Z}, K)$. Note also that the continuum of admissible selection rules suggests that the functional distribution of income, that is, the division between capital and labor, is generally indeterminate at the kink.

5.4.2. Recursive Representation of the Equilibrium

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 the 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, $\mathcal{I}$, such that:

- Households solve the recursive problem (21)-(22).
- Firms solve the capacity choice and profit maximization problems (15) and (17).
- Under the selection rule defined in subsection 5.4.1, the labor market clears.
- 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.

In addition, we call an RCE constructive if its existence proof delivers an iteratively convergent algorithm, and ergodic if the induced dynamical system admits an invariant probability measure.

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

Proof. See Appendix C. ∎

Role of GHH Preferences. GHH preferences are central to tractability because they deliver a closed-form partition of the state space that depends only on exogenous shocks, parameters, and the minimum labor requirement. In particular, we derive an explicit partition of the capital stock into three disjoint regions that exhaust the endogenous state space. This partition allows us to characterize the equilibrium Euler equation and to establish constructive existence and uniqueness.

Ergodicity follows from isolating the discontinuities in the wage equation through a specific equilibrium selection rule. Combined with GHH preferences, this rule yields a block-recursive equilibrium structure that is essential for isolating discontinuities in the Coleman–Reffett operator and proving ergodicity, conditional on continuity and compactness of the block that excludes wages. Wages are then recovered from the selection rule, using the fact that the marginal product of labor is a correspondence with well-defined directional derivatives because the aggregate production function is concave. The continuum-of-plants assumption is crucial, as it convexifies the production set and ensures continuity and concavity in the aggregate.

Under the selection rule, when the economy is at the capacity threshold, aggregate labor is determined by labor demand, while the wage adjusts to place households on their labor supply curve. Because under GHH preferences labor supply depends only on the after-tax wage, wage determination at the kink remains well behaved despite the discontinuity in the marginal product of labor. This matters for two reasons. First, the kink can be handled by fixing labor at its demand-determined full-capacity level and choosing the wage to satisfy the static labor-supply condition, which preserves continuity of allocations, the return to capital, and hence the Euler equation across regimes. Second, because labor choice is static, the recursive problem can be solved with an Euler-equation operator that maps candidate consumption (or composite) policy functions into themselves, allowing global and accurate solutions near the kink without local approximations or artificial smoothing.

When wealth effects are present, as under King–Plosser–Rebelo preferences, consumption enters directly into the labor supply condition. At the kink, wage adjustments then feed back into consumption and saving, altering the object that governs labor supply. This feedback may enlarge the effective state space, weaken the monotonicity and contraction properties of the Euler operator, and generate multiplicity or nonconstructive equilibria unless additional restrictions are imposed.

5.5. Computing the Recursive Competitive Equilibrium

This subsection describes the time-iteration procedure used to compute the RCE. We adapt the monotone-map method of Coleman (1991), Reffett (1996), and Mirman, Morand and Reffett (2008) to an economy with a kinked aggregate production function. Because distortionary taxes render the equilibrium non–Pareto optimal, a planner's problem as in Hansen and Prescott (2005) is unavailable; instead, we solve directly for the decentralized equilibrium, including prices. The labor-side kink poses an additional challenge by making local approximation methods unreliable near the capacity threshold.

The algorithm exploits two features of the model. First, the kink's location is independent of prices. Second, at the kink, the equilibrium selection rule determines the aggregate hours from labor demand and the wage from the household's marginal rate of substitution. This preserves the continuity of allocations and the return to capital despite the discontinuity in the marginal product of labor.

Operationally, we develop a modified Coleman–Reffett operator based on the Euler equation (14). Combined with the budget constraint (20), the Euler equation depends only on the interest rate, which Theorem 1 shows is *continuous* under the selection rule (19). This implies continuous policy functions $\{C(\mathcal{Z},K), L(\mathcal{Z},K), K'(\mathcal{Z},K)\}$ and yields constructive existence and uniqueness of the RCE. We also establish the compactness and ergodicity of the state space using the concept of sustainable capital stock in Stokey, Lucas and Prescott (1989). Wages and quasi-rents then follow from the selection rule.

5.5.1. State-Space Partition into Regimes

Before presenting the computational algorithm, it is useful to briefly discuss how the state space is partitioned into regimes. Recall that $\mathcal{Z} = (z, \tau_\ell)$ collects the TFP and labor tax rate shocks, while K denotes the aggregate capital stock and $\overline{H}$ the capacity threshold for aggregate hours. At the threshold, define the wage that places households on their labor supply schedule as

$$\mathcal{W}(\mathcal{Z}) \equiv \frac{\alpha \overline{H}^{1/\eta}}{1-\tau_\ell}.$$

Next, define the two one-sided marginal products of labor at the kink as

$$W^{UB}(\mathcal{Z},K) \equiv zK^\theta(1-\theta)\overline{H}^{-\theta}\left(\frac{M}{\overline{H}}\right)^{1-\phi-\theta}, \quad \text{and} \quad W^{LB}(\mathcal{Z},K) \equiv zK^\theta \phi \overline{H}^{\phi-1} M^{1-\phi-\theta},$$

where W^{UB} is the left derivative of the production function at the kink, while W^{LB} is the right derivative. Because the marginal product of labor falls discretely at the capacity threshold, the interval $\left[W^{LB}(\mathcal{Z},K),\, W^{UB}(\mathcal{Z},K)\right]$ is the gap in the labor demand schedule.

The state space is then partitioned as follows:

$$\begin{aligned} &\mathcal{W}(\mathcal{Z}) > W^{UB}(\mathcal{Z},K) && \Rightarrow \quad H < \overline{H} \quad \text{(idle-capacity regime)} \\ &W^{LB}(\mathcal{Z},K) \leq \mathcal{W}(\mathcal{Z}) \leq W^{UB}(\mathcal{Z},K) && \Rightarrow \quad H = \overline{H} \quad \text{(at-capacity regime)} \\ &\mathcal{W}(\mathcal{Z}) < W^{LB}(\mathcal{Z},K) && \Rightarrow \quad H > \overline{H} \quad \text{(full-capacity regime)} \end{aligned} .$$

This partition is the key computational simplification. Conditional on the regime, static equilibrium conditions determine labor, wages, output, and the return to capital, leaving only a single Euler equation.

5.5.2. Time Iteration in an Economy with a Kinked Production Function

Our algorithm hinges on Theorem 1, which guarantees the convergence of the associated iterative procedure. In its numerical implementation, each grid point is classified by regime, the corresponding static block is solved, and a Coleman–Reffett time-iteration step is applied until the relative sup-norm distance between successive policy functions falls below a prescribed tolerance.

Figure 3 provides a schematic representation of the iterative algorithm. The algorithm proceeds through the following steps.

Step 1. Discretize the state space. Construct a finite Markov chain for the exogenous state $\mathcal{Z} = (z, \tau_\ell)$ and a grid for capital, $K \in [0, K_{\max}]$.

Step 2. Compute the kink cutoffs. For every grid point $(\mathcal{Z}, K)$, compute the two directional marginal products of labor at the capacity threshold:

$$W^{UB}(\mathcal{Z},K) \quad \text{and} \quad W^{LB}(\mathcal{Z},K).$$

These two objects encode the discontinuity in labor demand.

Step 3. Classify the regime. Compare the wage implied by labor supply at capacity, $\mathcal{W}(\mathcal{Z})$, with the two cutoffs. If $\mathcal{W}(\mathcal{Z}) > W^{UB}(\mathcal{Z},K)$, the economy is below capacity. If $\mathcal{W}(\mathcal{Z}) < W^{LB}(\mathcal{Z},K)$, the economy is above capacity. If $\mathcal{W}(\mathcal{Z})$ lies between the two cutoffs, the economy is exactly at capacity.

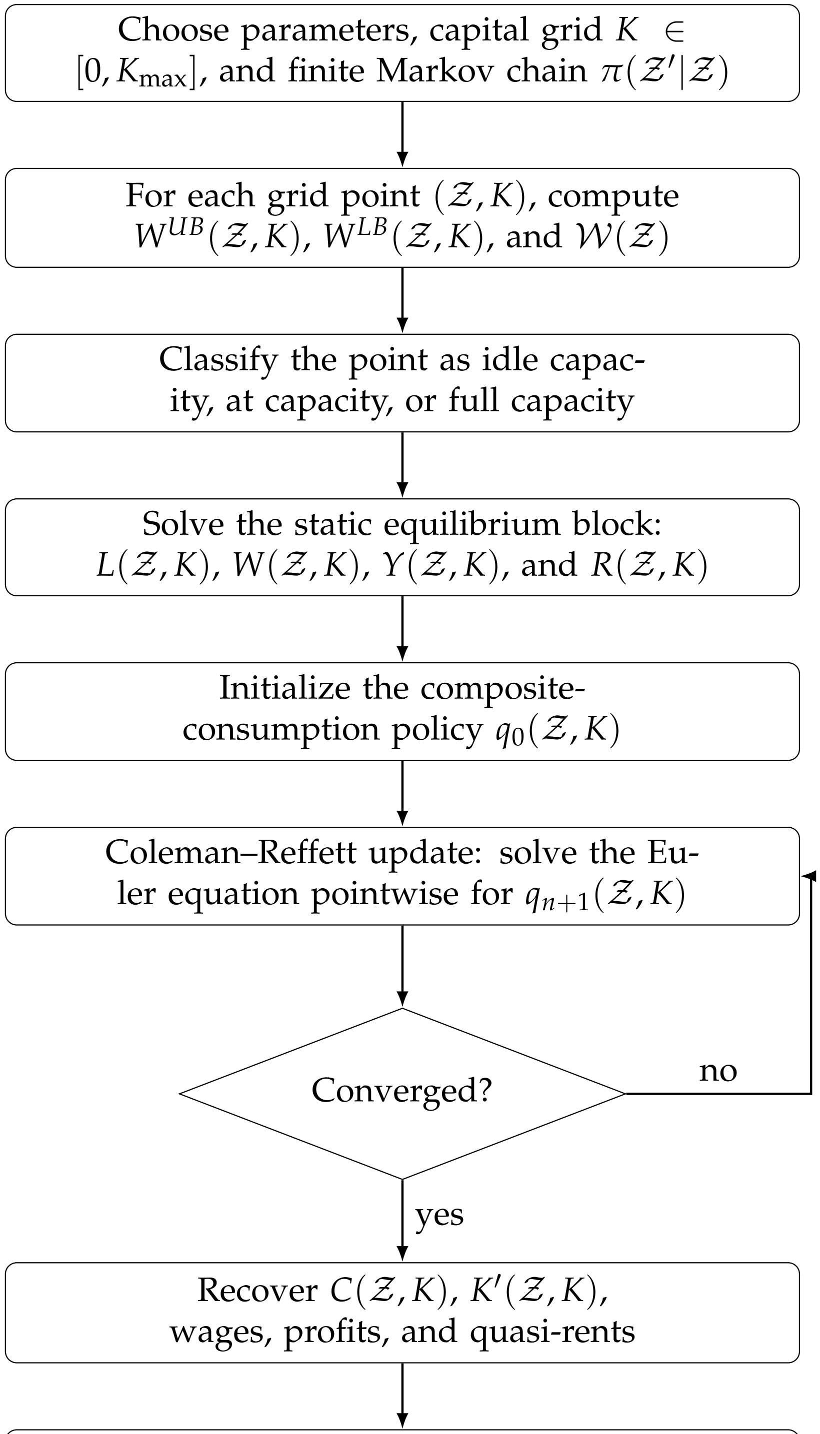


Figure 3: Time-Iteration Algorithm

Step 4. Solve the static equilibrium block. Conditional on the regime, compute aggregate labor $L(\mathcal{Z}, K)$ and the wage $W(\mathcal{Z}, K)$. In the idle- and full-capacity regimes, the wage equals the relevant marginal product of labor. At capacity, the selection rule sets

$$L(\mathcal{Z}, K) = \overline{H},$$

and the wage equals the household's marginal rate of substitution:

$$W(\mathcal{Z}, K) = \frac{\alpha \overline{H}^{1/\eta}}{1 - \tau_\ell}.$$

Output and the return to capital are then computed from the corresponding branch of the aggregate production function.

Step 5. Initialize the time-iteration problem. Define the GHH composite

$$q(\mathcal{Z}, K) \equiv C(\mathcal{Z}, K) - \alpha \frac{L(\mathcal{Z}, K)^{1+1/\eta}}{1 + 1/\eta}.$$

The algorithm starts from an initial guess $q_0(\mathcal{Z}, K)$. In the constructive proof, the iteration is initialized at the upper envelope of the candidate function space.

Step 6. Apply the Coleman–Reffett update. Given q_n, solve pointwise for $q_{n+1}(\mathcal{Z}, K)$ using the Euler equation:

$$q_{n+1}(\mathcal{Z}, K)^{-\sigma} = \beta \sum_{\mathcal{Z}' \in \mathcal{Z}} \pi(\mathcal{Z}' | \mathcal{Z})\, q_n(\mathcal{Z}', K')^{-\sigma} \left[R(\mathcal{Z}', K') + 1 - \delta \right].$$

The next-period capital stock K' is implied by the aggregate resource constraint:

$$K' = Y(\mathcal{Z}, K) + (1 - \delta)K - G(\mathcal{Z}, K) - C(\mathcal{Z}, K).$$

At each grid point, the numerical problem reduces to solving a scalar root-finding problem, which we do using Broyden's method.

Step 7. Iterate until convergence. Update q_n and repeat the Coleman–Reffett step until

$$\max_{\mathcal{Z}, K} \left| \frac{q_{n+1}(\mathcal{Z}, K) - q_n(\mathcal{Z}, K)}{q_n(\mathcal{Z}, K)} \right| < \varepsilon.$$

This is the finite-grid analogue of convergence of the Coleman–Reffett operator.

Step 8. Recover equilibrium objects. Once the fixed point is obtained, recover consumption, investment, output, labor, wages, returns, profits, and quasi-rents. In the at-capacity regime, quasi-rents are computed as the gap between the relevant directional marginal product of labor and the wage implied by the selection rule.

Step 9. Simulate the economy. The policy functions define a Markov process for $(\mathcal{Z}, K)$. Since the recursive equilibrium is ergodic, generalized impulse response functions are well-defined and can be computed by simulating benchmark and counterfactual paths under common future shocks.

The key operational point is that the kink need not be smoothed. The algorithm separates static regime classification from the dynamic Euler-equation problem, and the selection rule keeps labor and the return to capital continuous across regimes. This allows the modified Coleman–Reffett iteration to be applied globally despite the discontinuity in the marginal product of labor at capacity.

6. Calibration

Having established the theoretical case for state dependence, we now bring the model to the data. We calibrate the model at an annual frequency, so that one period corresponds to one year. The parameter vector is divided into two groups. The first consists of *externally calibrated* parameters, whose values are either standard in the literature or based on empirical targets estimated outside the model. The second consists of *internally calibrated* parameters, which are jointly chosen to match key moments in the data through model simulations.

As is well understood in dynamic general equilibrium models, no single parameter maps one-to-one to a specific empirical moment. Nevertheless, it is useful to organize the calibration strategy into two conceptually distinct steps.

6.1. Externally Calibrated Parameters

We begin by discussing the two sources of aggregate uncertainty—productivity shocks and labor tax shocks—before turning to preferences and technology. Table 3 reports the externally calibrated parameters.

Table 3: Externally Calibrated Parameters

Parameter	Description	Value
β	Time discount factor	0.960
σ^{-1}	Intertemporal elasticity of substitution	0.500
η	Frisch elasticity of labor supply	2.000
δ	Capital depreciation rate	0.100
$\bar{z}$	Mean: TFP shock	1.000
ρ_z	Autocorrelation: TFP shock	0.919
σ_z	Std. dev.: TFP shock	0.014
g	Government spending to output	0.200
$\bar{\tau}_l$	Mean: Labor tax shock	0.210
ρ_l	Autocorrelation: Labor tax shock	0.883
σ_l	Std. dev.: Labor tax shock	0.009

Notes: The table reports the externally calibrated parameter values, which are either standard in the literature or based on empirical targets estimated outside the model.

Productivity Shocks. 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 are 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 Tax Shocks. Following the literature, we remain agnostic about the determinants of tax policy and instead use historical data to estimate a statistical process for labor tax rates. This requires constructing a time series of effective tax rates and specifying how agents form expectations about future policy.

We calculate the average effective labor tax rates following Mendoza, Razin and Tesar (1994), aggregating federal, state, and local governments into a single sector. The average labor tax rate over 1950–2020 is 21% with a standard deviation of 0.04. Although many tax changes were initially legislated as permanent, subsequent reforms often reversed them, generating substantial variation in rates (Barro and Redlick, 2011; Mertens and Ravn, 2013; Mertens and Montiel Olea, 2018; Romer and Romer, 2009, 2010).

To model private sector expectations about future tax policy, we assume that labor tax rates follow a log AR(1) process:

$$\log(\tau_{l,t}) = (1-\rho_l)\log(\overline{\tau}_l) + \rho_l \log(\tau_{l,t-1}) + \sigma_l \epsilon_{l,t}. \tag{24}$$

At the quarterly frequency, $\rho_l = 0.9694$ and $\sigma_l = 0.0048$. Using standard conversion formulas, the corresponding annual parameters are $\rho_l = 0.9694^4 = 0.8831$ and $\sigma_l = \sqrt{0.0048^2(1+0.9694^2+0.9694^4+0.9694^6)} = 0.0092$.

Preferences and Technology. We set the time discount factor to $\beta = 0.96$. Preferences are given by the GHH utility specification:

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

This specification eliminates the wealth effect on labor supply by shutting down the standard channel through which changes in consumption affect hours worked. As a result, variations in labor tax rates influence labor supply exclusively through intertemporal substitution, by altering the relative price of working today versus tomorrow.

We set $\sigma = 2$, implying an intertemporal elasticity of substitution of 0.5, a standard value in the literature. We set $\eta = 2$, corresponding to a Frisch elasticity of labor supply equal to two, also standard in the RBC literature. We emphasize that the individual-level Frisch elasticity does not coincide with the aggregate labor supply elasticity in our model. Because capacity constraints occasionally bind, the usual separation between preference and technology parameters in determining aggregate labor supply responses breaks down.

Finally, we set the capital depreciation rate to $\delta = 0.10$ per year, following King and Rebelo (1999).

6.2. Internally Calibrated Parameters

Next, we calibrate the remaining parameters to match moments in the data. Because the model is nonlinear, average outcomes need not coincide with deterministic steady-state values. We therefore use a simulated method of moments (SMM) procedure based on artificial data. In this exercise, technology and labor tax shocks jointly affect the economy, and model-implied moments are computed from long simulated time series. Concretely, we choose the parameters to minimize the quadratic loss function

$$L(\Theta) = (\widehat{m}(\Theta) - m)' \, W \, (\widehat{m}(\Theta) - m) \, ,$$

where $\widehat{m}(\Theta)$ denotes simulated model moments under parameter vector Θ, m denotes the corresponding data moments, and W is a diagonal weighting matrix. All weights are set to one.

The targeted data moments are the fraction of time spent working (0.30), the capital-output ratio (3.30), the consumption-to-GDP ratio (0.64), and the cyclical volatilities of per capita hours worked (1.43) and real GDP (1.42). These empirical targets are drawn from standard U.S. macroeconomic data sources. Volatility is measured by the standard deviation of HP-filtered series with smoothing parameter 6.25, following Ravn and Uhlig (2002), where variables are expressed as log deviations from trend multiplied by 100.

Our objective is not to explain the aggregate volatility in hours or output, since the ability of this class of models to reproduce U.S. business-cycle volatility is already well understood. Rather, the calibration is meant to discipline the model along standard dimensions and to assess whether, conditional on matching aggregate volatility, it can reproduce the state dependence of the responses to labor tax shocks observed in the data. Importantly, we do not target moments related to the response to tax shocks, let alone their state dependence, so the model's implications along these dimensions are not imposed by construction and can be interpreted as an informative test of the theory.

Table 4 reports the internally calibrated parameters, while Table 5 compares simulated and empirical moments. The model matches the targeted moments closely.[5]

Table 4: Internally Calibrated Parameters

Parameter	Description	Value
α	Disutility of work	2.75
ϕ	Output elasticity to labor	0.27
θ	Output elasticity to capital	0.25
$\overline{h}$	Minimum labor requirement	0.26
M	Mass of plants	4.70

Notes: The table reports the internally calibrated parameter values, which are jointly chosen to match key moments in the data based on model simulations.

[5] An alternative calibration fixes the mass of available plants at $M = 1$ and instead uses the technology level z as the scale parameter.

Table 5: Model and Data Moments

Moment	Model	Data
Time spent working	0.27	0.30
Capital/GDP	3.21	3.30
Consumption/GDP	0.58	0.64
Volatility of hours	1.21	1.43
Volatility of GDP	1.41	1.42

Notes: The table reports sample averages in the data alongside the corresponding model averages computed from artificial data simulated from the model. Volatility is measured by the standard deviation of HP-filtered series with smoothing parameter 6.25, following Ravn and Uhlig (2002), where variables are expressed as log deviations from trend multiplied by 100.

7. Quantitative Analysis

This section examines the quantitative predictions of the calibrated model for the U.S. economy. We begin by comparing the notion of capacity utilization in the model and in the data. We then turn to the model's nonlinear propagation mechanism, analyzing the dynamic responses to tax shocks through impulse response functions.

7.1. Capacity Utilization in the Data and the Model

Before turning to the quantitative results, it is helpful to address how one can think of the notion of capacity and capacity utilization in the model compared with the data on capacity utilization we used to estimate the state-dependent effects of labor tax shocks.

Capacity Utilization in the U.S. From the survey administered by the U.S. Census Bureau, whose responses form the basis for the Federal Reserve Board's G.17 Industrial Production and Capacity Utilization release, two provisions are particularly relevant here. First, the unit of analysis is the plant rather than the firm, and full production capacity should consider "only machinery and equipment in place and ready to operate." Second, respondents are instructed to account for "the hours of operation and overtime pay that can be sustained under normal conditions and a realistic work schedule in the long run." Thus, hours per plant enter directly into the definition of production capacity.

We interpret this definition as referring to the maximum level of output a plant can produce when all machinery and equipment in place are operated under a normal and sustainable work schedule.

Capacity Utilization in the Model. In the model, capacity utilization operates along two distinct margins. The first is the extensive margin—the number of active plants (or production units). The second is the intensive margin—the number of hours operated per plant. Together, these margins determine total effective utilization in the economy.

Importantly, the variable $m \leq M$ need not be interpreted literally as the number of physical plants. More generally, it can be viewed as the number of production lines within a plant. Under this interpretation, a firm may operate a single physical plant that contains multiple production lines, each of which can be activated or deactivated independently. What matters for the model is not the physical definition of a "plant," but the existence of production units that can be switched on or off.

A key assumption is that existing capital in place can be freely and instantaneously reallocated across plants or production lines without cost. This abstraction allows the model to capture variations in capacity utilization through changes in the number of active production units and hours per unit, while abstracting from frictions in reallocating capital across them. Under this interpretation, the economy is operating at full capacity whenever $m = M$, that is, when all production lines are active.

Turning to the data, the U.S. Census Bureau's survey-based measures of capacity utilization refer to production under "normal conditions" and a "realistic work schedule." The model does not contain a direct analogue of a "normal work schedule" unless one interprets normal conditions as corresponding to $m = M$ and $h = \overline{h}$, where $\overline{h}$ denotes some benchmark level of hours per unit. Under this interpretation, the model delivers a notion of below-capacity operation when $H < \overline{H}$ and above-capacity operation when $H > \overline{H}$, where H denotes aggregate hours.

7.2. State Dependence in the Propagation of Shocks

We now turn to the quantification of the model and study the effects of tax shocks—understood as unexpected and unanticipated changes in the labor tax rate—on hours worked and output and the implied elasticities of aggregate hours and output. To characterize how these effects depend on the state of the economy, we rely on generalized impulse response functions (GIRFs), which allow responses to vary with initial conditions.[6]

We consider tax shocks with magnitudes comparable to those observed historically in the United States. Importantly, tax shocks are assumed to be symmetric by construction; therefore, any state dependence in the resulting responses reflects the model's internal nonlinear propagation mechanism rather than differences in the shocks themselves.

Generalized Impulse Response Functions. In linear or linearized models, impulse response functions are fully characterized by a single response function that is invariant to the sign, size, and timing of shocks. By contrast, GIRFs in nonlinear models do not yield a unique, state-independent response. Instead, the effects of a shock depend not only on its magnitude and sign, but also on the state of the economy at the time the shock occurs, as summarized by the history of past shocks and initial conditions. Consequently, the response to a negative shock cannot be obtained by simply reversing the sign of the response to a positive shock, nor can the response to a small shock be interpreted as a scaled-down version of the response to a large shock. Rather, both the magnitude and the dynamic evolution of the marginal effect of a given shock vary systematically with the economic environment in which the shock hits. This inherent state dependence creates a well-known reporting challenge for GIRFs, which we address by presenting impulse responses under a set of carefully chosen alternative scenarios.

Computing GIRFs requires simulating sequences of labor tax rates and using the model's solution to generate the corresponding equilibrium paths of endogenous variables. An impulse response is defined as the difference between the average equilibrium paths of two economies, which we label the *benchmark* and the *counterfactual*.

Operationally, the procedure consists of two steps. First, in the benchmark economy, we simulate 10,000 stochastic paths for each variable of interest—such as hours worked—over T model periods (interpreted as years). At time $t = 0$, the labor tax rate is initialized at its median value, $\tau_{l,0} = \overline{\tau}_l = 0.21$, across all replications. For $t > 0$, tax rates evolve according to an identical AR(1) process but differ across replications due to distinct shock realizations. This generates a collection of realized tax paths $\{\tau^i_{l,t}\}^T_{t=0}$, for $i = 1, \ldots, 10{,}000$, with $\tau^i_{l,0} = \overline{\tau}_l$ for all i. Conditional on each tax path, the model produces a corresponding time series for every endogenous variable.

Second, we repeat the same simulation exercise for the counterfactual economy. The only difference is the initial condition: the labor tax rate at $t = 0$ is set below the median

[6]See Gallant, Rossi and Tauchen (1993), Koop, Pesaran and Potter (1996), Potter (2000), and Gonçalves et al. (2024), for discussions of impulse responses in nonlinear time-series models.

when analyzing a tax cut. The GIRF to a labor tax shock is then computed as the average difference between the simulated equilibrium paths in the counterfactual and benchmark economies.

By construction, the benchmark and counterfactual economies are subject to identical sequences of technology shocks and differ only in their initial labor tax rate and thereby the sequences of labor tax shocks thereafter. This isolates the effect of the initial tax shock. In a linear model, the resulting responses are invariant to initial conditions. In contrast, in the nonlinear environment studied here, the same tax shock can generate markedly different adjustment dynamics depending on whether the economy is initially operating at capacity or below capacity, as reflected in capacity utilization. Averaging across many stochastic realizations reveals how nonlinearities in the model's propagation mechanism map differences in initial capacity utilization into state-dependent impulse responses.

Capacity Utilization and the Aggregate Hours Elasticity. To quantify the degree of state dependence in the model, we compute the GIRFs to labor tax cuts, conditioning on the level of capacity utilization at the time the shock hits. Crucially, because the stochastic process generating the model's artificial data is ergodic (Theorem 1), these impulse response functions are well-defined equilibrium objects.

More specifically, we distinguish between a *below-capacity* state, in which part of the productive capacity is idle and some plants are not operated ($m < M$) and $h = \overline{h}$, and an *above-capacity* state, in which all available plants are active ($m = M$) and $h > \overline{h}$. Operationally, these states are generated by choosing the initial capital stock such that, in the below-capacity economy, the number of operating plants is 2.5% below its maximum ($m = 0.975\,M$), with an analogous deviation above capacity. These magnitudes are chosen to match the empirical specification in Section 3, in which the state-dependent effects of tax shocks are estimated conditioning on capacity utilization one standard deviation below and above its trend.

This procedure highlights the key role of initial capacity utilization in determining the regime in which the model economy operates when the labor tax shock occurs. It also formalizes the intuition underlying the empirical regression specification, in which the tax shock is interacted with capacity utilization. In the model—just as in the empirical analysis—capacity utilization serves as the conditioning variable that selects the relevant state of the economy and, consequently, implies state-dependent elasticities with respect to tax shocks.

Table 6: The State-Dependent Aggregate Hours Elasticity

Panel A: Impact Response

	Output		Hours	
	Data	Model	Data	Model
2.5% Below-capacity	0.746***	0.266	0.465***	0.355
2.5% Above-capacity	0.034	0.125	0.070	0.260

Panel B: 1-Year Ahead Response

	Output		Hours	
	Data	Model	Data	Model
2.5% Below-capacity	1.206***	0.125	0.967***	0.260
2.5% Above-capacity	0.304	0.125	0.018	0.260

Notes: The table reports semi-elasticities with respect to a labor tax cut. The columns labeled "Data" present estimates based on narratively identified tax shocks. *** denotes statistical significance at the one percent level. Columns labeled "Model" report the semi-elasticities implied by the model's generalized impulse response functions.

Panel A of Table 6 compares the model-implied impact semi-elasticities of output and hours per capita with their empirical counterparts. In the model, these statistics are computed from the corresponding impulse responses to a 1.7-percentage-point labor tax cut.

Importantly, these moments are not targeted in the calibration. Hence, any agreement between the model-generated and empirically estimated semi-elasticities is not mechanically imposed or hard-wired through parameter choices. Instead, the comparison serves as an informative diagnostic—a test of the model's ability to account for the state dependence observed in the responses with respect to the identified labor tax shocks in the data.

Turning to per capita hours, the model's implications are broadly consistent with the empirical evidence, both in the magnitude of the estimated semi-elasticities and in the degree of state dependence. When the economy operates below capacity, the model implies an impact semi-elasticity of 0.355, compared with 0.465 in the data. By contrast, when the economy operates above capacity, the response of hours is more muted, again pointing to state dependence.

A similar pattern emerges for the output responses. The model captures the extent of state dependence in the output response to labor tax shocks, although the magnitude of the response when the economy is below capacity is somewhat smaller in the model than in the data. In this case, the impact semi-elasticity is 0.266 in the model, compared with the estimate of 0.746 in the data.

Panel B of Table 6 shows that the model falls short of generating the one-year ahead semi-elasticities in the data. The model-implied elasticities are both smaller in magnitude and show no asymmetry between whether the model economy was below or above capacity when hit by the tax shock.

These results largely reflect the use of GHH preferences, which eliminate wealth effects on labor supply, and the global stability properties of the dynamic system. In this sense, the model's inability to match the longer-horizon asymmetry in the data is informative rather than surprising: the framework deliberately abstracts from the features—such as adjustment costs—that would generate the additional inertia needed to sustain state-dependent dynamics beyond impact. That a single frictionless mechanism reproduces the on-impact evidence underscores its quantitative relevance as a first-order driver of state dependence in the transmission of tax shocks.

The model is calibrated to match average hours of 0.30, with an aggregate minimum labor requirement of 0.26. This threshold separates recession from expansion and, together with the capacity constraint, generates three regimes: recession (below-capacity), at capacity, and expansion (above-capacity). Thus, the state space is partitioned in three regions. Targeting average hours of 0.30 implies stability in the expansion regime. For any initial condition, the simulated average of hours must converge to its long run value.

Given the continuity of the policy functions (except for wages) and equilibrium uniqueness, convergence to a level above the minimum labor requirement can only occur within the expansion region of the state space. Under GHH preferences and the annual calibration, the model economy transitions from the below-capacity regime to the stable above-capacity regime within one model period—that is, within one year.

GHH preferences impose sufficient structure to deliver a closed-form characterization of this partition. While this analytical partition of the state space allows us to seamlessly apply the Euler-equation operator and to establish existence, uniqueness, and ergodicity, it comes at the cost of limiting the scope for endogenous nonlinearities.

The results in panel B indicate that elasticities vary over time only when the economy switches regimes—for example, when transitioning from below capacity to above capacity, the stable regime. Conditional on being in the above-capacity state, the combination of the global stability properties of the dynamical system and GHH preferences implies an isoelastic response of hours and output. As a result, short- and long-run impulse responses coincide when the economy starts in expansion. Finally, long-run elasticities are invariant to initial conditions, reflecting ergodicity: once the economy converges to its stable regime—above-capacity or expansion regime—its long-run behavior no longer depends on its starting point.

Finally, the magnitude of the model-implied semi-elasticities is virtually unchanged if one considers an equally sized tax increase or varies the magnitude of the tax shock. Conditional on the regime in which the model economy operates, elasticities are largely symmetric and nearly invariant to the size of the shock. State dependence in the propagation of tax shocks arises primarily from changes in regime. Within a given regime, the model exhibits little to no nonlinearity, much like a textbook neoclassical growth framework. This is consistent with our empirical specification, in which differences in responses and elasticities to tax shocks arise solely from variation in capacity utilization.

8. Conclusion

This paper formalizes, quantifies, and empirically documents the view that the effects of tax policy depend on the extent of idle capacity. Using narratively identified tax shocks and local projections, we provide new evidence that the aggregate hours elasticity is larger when the economy is operating below capacity. The theoretical framework we develop to account for this finding does not rely on frictions that impede market clearing, noncompetitive pricing, nominal rigidities, or asymmetric adjustment costs. Instead, it builds on a technological nonconvexity: production requires a minimum labor input.

Embedding minimum labor requirements in a multi-plant firm setting generates endogenous idle capacity and time-varying capacity utilization. In response to shocks that affect labor demand and supply, firms optimally choose whether to operate all or only a subset of available plants or to adjust labor input per plant along a margin subject to decreasing returns.

When the economy operates below capacity, firms accommodate increases in labor demand primarily by activating idle plants. When the economy operates at capacity, adjustment occurs through hours per plant, which entails sharply diminishing returns. These distinct adjustment mechanisms generate strong state dependence: aggregate hours elasticities differ markedly between recessions and expansions. In our calibrated model, for example, hours respond more to a labor tax cut when the economy is below capacity than when it is at capacity, consistent with the empirical evidence based on narratively identified labor tax shocks.

Several directions for future research are promising. Introducing labor market search frictions would allow a richer distinction between extensive and intensive margins of adjustment at the worker, plant, and firm levels. More broadly, the framework developed here may prove useful for studying the aggregate implications of other nonstandard production constraints affecting factor utilization beyond labor.

References

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.

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.

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.

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.

Ferraro, Domenico, and Giuseppe Fiori. 2020. "The Aging of the Baby Boomers: Demographics and Propagation of Tax Shocks." *American Economic Journal: Macroeconomics*, 12(2): 167–93.

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

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.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95(1): 161–182.

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.

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

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. 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks." *American Economic Journal: Economic Policy*, 4(2): 145–81.

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.

Mertens, Karel, and Morten Ravn. 2014. "Fiscal Policy in an Expectations-Driven Liquidity Trap." *Review of Economic Studies*, 81(4): 1637–1667.

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.

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.

Ravn, Morten O., and Harald Uhlig. 2002. "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations." *Review of Economics and Statistics*, 84(2): 371–376.

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

Rendahl, Pontus. 2015. "Inequality Constraints and Euler Equation-based Solution Methods." *Economic Journal*, 125(585): 1110–1135.

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.* Harvard University Press.

Appendix

A. Variables Used in Regression Analysis

Data for real GDP, investment, full-time employment, and total capacity utilization are obtained from the FRED database using the series GDPC1, GPDIC1, LNS12500000, and TCU, respectively. Quarterly series for GDP, investment, and employment are annualized by averaging series for each year. To measure the cyclical stance of capacity utilization, we construct $\widehat{CU}_t$ as the linearly detrended capacity utilization rate, using the value observed in the final quarter of year t.

Hours per capita are constructed from CPS data. As in Mertens and Montiel Olea (2018), Government debt is federal debt held by the public, measured by Table L.106 line 19 (federal government, liabilities, credit market instruments) in the US Financial Accounts (release Z.1 of the Federal Reserve Board), deflated using the BLS CPI Research Series Using Current Methods (CPI-U-RS) and potential tax units. Government spending per tax unit is the sum of federal government purchases, net interest rate expenditures and transfers (NIPA 3.2 line 46 less lines 3,4,7,10 and 11 plus NIPA 3.12U line 25), divided by the CPI-U-RS and potential tax units.

The real stock price is constructed using the U.S. share price index (SPASTT01USM661N), published by the OECD and available through FRED (Federal Reserve Economic Data), deflated by the CPI. The annual stock price change is defined as the growth rate from the last month of year $t-1$ to the last month of year t.

B. Additional Empirical Results

Appendix B.1 reports additional IV results. Appendix B.2 examines alternative nonlinear specifications for the interaction between tax shocks and capacity utilization, showing that the baseline results are not driven by the assumed linear specification. Appendix B.3 studies the potential threshold-crossing problem, whereby capacity utilization itself—the conditioning variable used to capture the state of the economy—may respond endogenously to the tax shock.

B.1. GDP, Hours, Employment, and Investment—IV Approach

Tables B.1 and B.2 report the corresponding IV estimates.

B.2. Alternative Nonlinear Specifications

The baseline specification imposes a linear interaction between capacity utilization and the tax shock. While this parameterization is parsimonious and facilitates direct interpretation of the state-dependent multiplier, it raises the question of whether the estimated heterogeneity reflects a genuine feature of the data or is instead an artifact of the assumed functional form. To address this concern, we consider two alternative specifications that relax the linearity assumption.

The first check, inspired by Ramey and Zubairy (2018), replaces the continuous cyclical component $\widehat{CU}_{t-1}$ with a binary indicator equal to minus one when capacity utilization is below trend and plus one otherwise, so that the specification becomes:

$$y_{t+h} = \alpha_h + \beta_h y_{t-1} + (\gamma_{1,h} + \gamma_{2,h}\,\mathbb{1}_{t-1})\,\tau_t^{\text{shock}} + \delta_h\,\mathbb{1}_{t-1} + \eta_h\,\text{Controls}_{t-1} + \varepsilon_{t+h}. \tag{B.1}$$

This sign convention preserves the direction of the original cyclical component and facilitates direct interpretation of the multiplier coefficients. By construction, equation (B.1) identifies the state-dependent fiscal multiplier without imposing any functional form on how the multiplier varies *within* each regime, at the cost of discarding the continuous information in $\widehat{CU}_{t-1}$. It therefore provides a natural first check: evidence of state dependence should remain detectable even under this coarser characterization of the state.

The second check addresses curvature within the continuous specification. Rather than augmenting the linear interaction with a separate squared term—which would intro-

Table B.1: Effects of Tax Shocks on GDP and Hours: IV

	log (GDP)		log (Hours per capita)	
	$h = 0$	$h = 1$	$h = 0$	$h = 1$
	(A)	(B)	(C)	(D)
Panel A:				
Tax Rate (γ_1)	-0.639^{**}	-1.252^{***}	-0.374	-0.948^{**}
	(0.31)	(0.41)	(0.24)	(0.38)
Tax Rate $\times$ $\widehat{CU}_{t-1}$ (γ_2)	0.345^{*}	0.480^{**}	0.244	0.443^{***}
	(0.17)	(0.18)	(0.15)	(0.15)
F-statistic	5.51	5.51	4.58	15.68
Panel B:				
Semi-elasticity when capacity utilization is:				
1 std.dev. below trend	-1.632^{**}	-2.632^{***}	-1.075^{*}	-2.223^{***}
at trend	-0.639^{**}	-1.252^{***}	-0.374	-0.948^{**}
1 std.dev. above trend	0.354	0.128	0.328	0.327
Testing for differences between "below trend" vs. "at trend":				
p-value	0.059	0.016	0.116	0.006

Notes: The sample period runs from 1988 to 2019. The tax rate and its interaction term with detrended capacity utilization are instrumented using two instruments: exogenous tax shocks and the interaction between tax shocks and lagged detrended capacity utilization. Each specification also includes the lagged dependent variable, the lag of linearly detrended capacity utilization, the log of real government spending per capita, the change in the log of real federal debt per capita, and the change in the log of the real S&P 500 index. Heteroskedasticity-consistent standard errors with small-sample correction are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.2: Effects of Tax Shocks on Investment and Employment: IV

	Log (Real Investment)		log (Full-time Employment)	
	$h = 0$	$h = 1$	$h = 0$	$h = 1$
	(A)	(B)	(C)	(D)
Panel A:				
Tax Rate (γ_1)	-2.819^{*}	-4.637^{**}	-0.596	-1.310^{**}
	(1.57)	(1.93)	(0.36)	(0.51)
Tax Rate $\times$ $\widehat{CU}_{t-1}$ (γ_2)	1.874^{*}	2.092^{*}	0.423^{*}	0.510^{***}
	(1.00)	(1.06)	(0.21)	(0.16)
F-statistic	5.05	5.05	4.44	4.44
Panel B:				
Semi-elasticity when capacity utilization is:				
1 std.dev. below trend	-8.211^{*}	-10.657^{**}	-1.814^{*}	-2.778^{***}
at trend	-2.819^{*}	-4.637^{**}	-0.596	-1.310^{**}
1 std.dev. above trend	2.573	1.383	0.622	0.158
Testing for differences between "below trend" vs. "at trend":				
p-value	0.072	0.061	0.057	0.004

Notes: The sample period runs from 1988 to 2019. The tax rate and its interaction term with detrended capacity utilization are instrumented using two instruments: exogenous tax shocks and the interaction between tax shocks and lagged detrended capacity utilization. Each specification also includes the lagged dependent variable, the lag of linearly detrended capacity utilization, the log of real government spending per capita, the change in the log of real federal debt per capita, and the change in the log of the real S&P 500 index. Heteroskedasticity-consistent standard errors with small-sample correction are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

duce an additional free parameter and raise degrees-of-freedom concerns—we replace the linear interaction $\widehat{CU}_{t-1} \cdot \tau_t^{\text{shock}}$ with the sign-preserving quadratic $\widehat{CU}_{t-1}^2 \cdot \text{sign}(\widehat{CU}_{t-1}) \cdot \tau_t^{\text{shock}}$, yielding:

$$y_{t+h} = \alpha_h + \beta_h y_{t-1} + \left(\gamma_{1,h} + \gamma_{2,h}\, \widehat{CU}_{t-1}^2\, \text{sign}(\widehat{CU}_{t-1})\right) \tau_t^{\text{shock}} \\ + \delta_h\, \widehat{CU}_{t-1}^2\, \text{sign}(\widehat{CU}_{t-1}) + \eta_h\, \text{Controls}_{t-1} + \varepsilon_{t+h}. \quad \text{(B.2)}$$

The term $\widehat{CU}_{t-1}^2 \cdot \text{sign}(\widehat{CU}_{t-1})$ magnifies deviations from trend non-linearly while pre-

Table B.3: Effects of Tax Shocks on Real GDP and Hours: Discrete Indicator

	log (Real GDP)		log (Hours per Capita)	
	$h=0$ (A)	$h=1$ (B)	$h=0$ (C)	$h=1$ (D)
Panel A:				
Tax shock (γ_1)	-0.456^{***}	-0.904^{***}	-0.219	-0.605^{**}
	(0.13)	(0.19)	(0.13)	(0.27)
Tax shock $\times \widehat{CU}_{t-1}^{binary}$ (γ_2)	0.408^{***}	0.402	0.252^{**}	0.441^{***}
	(0.13)	(0.27)	(0.12)	(0.11)
Panel B:				
Semi-elasticity when capacity utilization is:				
1 std. dev. below trend	-0.862^{***}	-1.305^{***}	-0.471^{**}	-1.046^{***}
at trend	-0.456^{***}	-0.904^{***}	-0.219	-0.605^{**}
1 std. dev. above trend	-0.049	-0.503	0.033	-0.165
Testing for differences between "below trend" vs. "at trend":				
p-value	0.004	0.150	0.049	0.001

Notes: The sample period runs from 1988 to 2019. Each specification also includes the lag of the dependent variable, the lag of linearly detrended capacity utilization discretized into a binary indicator $\{-1, +1\}$, the log of real government spending per capita, the change in the log of real federal debt per capita, and the change in the log of the real S&P 500 index. Newey–West standard errors, computed with two lags, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

serving their sign, so that the direction of state dependence—a larger multiplier when the economy operates below trend—is maintained by construction. This specification therefore nests the same economic hypothesis as the baseline while allowing for a more flexible relationship between the degree of utilization and the size of the multiplier, without increasing the number of interaction parameters.

Both alternatives are reported in Tables B.3 and B.4. In each case, the estimated state-dependent component $\gamma_{2,h}$ is statistically significant and of a magnitude consistent with the baseline results in the main text: the fiscal multiplier is significantly larger when the economy operates below trend, and the difference between states is of a similar order to the baseline. Taken together, the two checks suggest that the main finding is not an artifact of the linear interaction assumption, but reflects a pattern of state dependence that is detectable under both coarser and more flexible characterizations of the state variable.

Table B.4: Effects of Tax Shocks on Real GDP and Hours: Nonlinear Capacity Utilization

	log (Real GDP)		log (Hours per Capita)	
	$h = 0$ (A)	$h = 1$ (B)	$h = 0$ (C)	$h = 1$ (D)
Panel A:				
Tax shock (γ_1)	-0.260^{**}	-0.639^{**}	-0.150	-0.369
	(0.12)	(0.31)	(0.09)	(0.26)
Tax shock $\times \widehat{CU}^2_{t-1} \cdot sign(\widehat{CU}_{t-1})$ (γ_2)	0.032^{***}	0.045^{**}	0.019^{*}	0.043^{***}
	(0.01)	(0.02)	(0.01)	(0.01)
Panel B:				
Semi-elasticity when capacity utilization is:				
1 std. dev. below trend	-0.779^{***}	-1.370^{***}	-0.459^{**}	-1.056^{***}
at trend	-0.260^{**}	-0.639^{**}	-0.150	-0.369
1 std. dev. above trend	0.258	0.092	0.159	0.318
Testing for differences between "below trend" vs. "at trend":				
p-value	0.002	0.033	0.083	0.002

Notes: The sample period runs from 1988 to 2019. Each specification also includes the lag of the dependent variable, the lag of the square of linearly detrended capacity utilization multiplied by -1 when negative, the log of real government spending per capita, the change in the log of real federal debt per capita, and the change in the log of the real S&P 500 index. Newey–West standard errors, computed with two lags, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.3. Threshold Crossing

We now examine the "threshold crossing" concern raised in Gonçalves et al. (2024) in the context of measuring state-dependent effects of tax policy: the possibility that a shock is large enough to reverse the state of the economy, which would bias the local projection estimator because the state variable responds endogenously to the shock and the economy moves across the threshold. This concern is also most limited at $h = 0$. Threshold crossing is inherently a dynamic problem: it requires the shock to have propagated sufficiently to alter the state of the economy, and this propagation takes time. At the impact horizon, the economy has not yet had the opportunity to cross the threshold in response to the shock—the state is fixed at $\widehat{CU}_{t-1}$ by the predeterminedness assumption, and no dynamic feedback has yet operated. The concern therefore becomes progressively more relevant as the horizon h increases, since at longer horizons the shock has had more time to move the economy across the threshold, potentially contaminating the local projection estimator. To assess the empirical relevance of this concern at $h = 1$, we estimate the impulse response of capacity utilization to the tax shock. The estimated impact coefficient is -0.871 (95% confidence interval: $[-1.240, -0.504]$), implying that a tax increase (cut) reduces (raises) capacity utilization, as expected.

To further assess the empirical relevance of the threshold-crossing concern, we examine whether any shock in our sample is large enough to move the economy across the capacity utilization threshold on impact. As a benchmark, a one-percentage-point tax shock is estimated to move capacity utilization by 0.871 percentage points on impact, corresponding to approximately three-tenths of the standard deviation of $\widehat{CU}_{t-1}$. While this normalization provides useful context for the typical magnitude of the shock's effect, the relevant question for threshold crossing is not whether the shock is large relative to the distribution of $\widehat{CU}_{t-1}$, but whether it is large enough to change the sign of $\widehat{CU}_{t-1}$ for any particular observation—a condition that depends on the level of the pre-shock state variable rather than its dispersion. We therefore turn to a direct, observation-by-observation assessment.

To this end, and to account for uncertainty around the estimated impact coefficient, Table B.5 reports, for each year with a non-zero tax shock, the pre-shock cyclical component $\widehat{CU}_{t-1}$, the point-estimated impact change $\Delta\widehat{CU} = -0.871 \times \tau^{shock}$, and the implied post-shock value $\widehat{CU}_{t-1} + \Delta\widehat{CU}$, together with bounds obtained by replacing the point estimate with the endpoints of its 95% confidence interval $[-1.240, -0.504]$. Threshold crossing would require the economy to start below (above) trend and the shock to push it above (below) trend.

Table B.5: Threshold Crossing Assessment: Predicted Impact Effect of Tax Shocks on Capacity Utilization

Year	τ^{shock}	$\Delta\widehat{CU}$	$\widehat{CU}_{t-1}$	$\widehat{CU}_{t-1} + \Delta\widehat{CU}$ Point	CI upper
1991	+0.526	−0.458	−1.480	−1.938	−1.745
1993	+0.113	−0.098	−0.946	−1.044	−1.003
2003	−2.576	+2.244	−4.203	−1.959	−1.010
2018	−3.637	+3.168	+2.053	+5.221	+3.887

Notes: τ^{shock} denotes the narrative tax shock. $\Delta\widehat{CU} = -0.871 \times \tau^{shock}$ is the predicted impact change in the cyclical component of capacity utilization, where −0.871 is the estimated impact response coefficient (95% confidence interval: [−1.240, −0.504]). The "CI upper" column reports the implied post-shock value using the confidence interval endpoint that maximises the absolute predicted effect: $-1.240 \times \tau^{shock}$ for tax cuts ($\tau^{shock} < 0$) and $-0.504 \times \tau^{shock}$ for tax increases ($\tau^{shock} > 0$). $\widehat{CU}_{t-1}$ is the cyclical component of capacity utilization in the last quarter of year $t-1$ obtained via linear detrending of the capacity utilization series. Threshold crossing requires $\widehat{CU}_{t-1} < 0$ (economy below trend) and $\widehat{CU}_{t-1} + \Delta\widehat{CU} > 0$ (economy above trend after the shock), or vice versa. Years with zero tax shocks are omitted. The standard deviation of $\widehat{CU}_{t-1}$ is approximately 2.9 percentage points, implying that a one percentage point shock moves capacity utilization by roughly 0.30 standard deviations on impact.

The calculations suggest that threshold crossing is unlikely to arise for any shock in the sample. The most demanding case is 2003, where the largest tax cut ($\tau^{shock} = -2.576$) generates a point-estimated impact increase in capacity utilization of +2.244 percentage points, rising to +3.193 percentage points at the upper end of the confidence interval. Nevertheless, with $\widehat{CU}_{t-1} = -4.203$—the economy was substantially below trend, at a distance of more than 1.4 standard deviations—the implied post-shock value remains below trend under both the point estimate and the confidence interval bounds, ranging from −1.959 to −1.010. The pre-shock distance from the threshold is thus the key factor: even under the most adverse parametric scenario, the shock effect falls well short of the gap that would need to be bridged for crossing to occur. While the remaining margin in the most adverse scenario is moderate, it is consistent across both detrending methods reported in the tables. The 2018 tax cut, the largest in absolute terms ($\tau^{shock} = -3.637$), does not raise a crossing concern since the economy was already above trend in 2017 ($\widehat{CU}_{t-1} = +2.053$, approximately 0.7 standard deviations above trend), and the shock moves it further above trend in all scenarios. For the two tax increases in 1991 and 1993, the economy was below trend and is pushed further below trend in all cases.

One important clarification also deserves acknowledgment. The impact coefficient used to construct $\Delta\widehat{CU}$ is itself estimated from the state-dependent local projection, so the assessment is not fully independent of the estimator whose validity it is intended to support. Nevertheless, two pieces of evidence suggest threshold crossing is unlikely to be a material source of bias. First, using the estimated coefficients to directly analyse the extent of threshold crossing, we find it to be negligible in practice. Second, the interaction term between the tax shock and the CU determinant is statistically insignificant, providing no evidence that the response of capacity utilization to the tax shock is itself state-dependent. With these qualifications in mind, the calculations provide supporting, if not conclusive, evidence that threshold crossing does not materially contaminate our estimates.

C. Proofs

The aggregate production function is

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

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. As a result, 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

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

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 $F_H^-(\overline{h}, K, z) \equiv W^{UB}(K,z) > F_H^+(\overline{h}, K, z) \equiv W^{LB}(K,z)$.

Next, recall that $\mathcal{Z} = (z, \tau_l)$ and that the economy is in regime s whenever $K \in \mathbb{K}_s(\mathcal{Z})$, as we will show below. The selection rule,

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

implies positive quasi-rents for the firm and 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. It is then possible to partition the state space using a stationary and static mechanism:

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

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 $\alpha\overline{h}^{1/\eta}/(1-\tau_l) = W^{UB}(K^{WUB},z)$ and $\alpha\overline{h}^{1/\eta}/(1-\tau_l) = W^{UB}(K^{WLB},z)$ we obtain

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

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 $\mathcal{Z} \in \mathbb{Z}$, $K \in (K^{WUB}, K^{WLB})$ 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, $K \in [0, K^{WUB}]$, $F_H^0(H,K) = \alpha L^{1/\eta}/(1-\tau_l)$, $L = H$, such that

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

Similarly, for Regime 2,

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

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 (K^{WUB}, K^{WLB}) from (C.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 $K \in [0, K^{WUB})$ in Regime 1 and $F_K^0(L_2(\mathcal{Z}, K), K)$ for $K \in (K^{WLB}, +\infty)$ 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 $K \in [K^{WUB}, K^{WLB}]$ in Regime 3. Then, we obtain

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

which implies

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

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

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

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

It is easy to see that

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

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

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

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 Euler operator under GHH preferences,

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

let us define the *composite* as

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

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:

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

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 $\mathbf{D} \equiv \eta/(1+\eta)$:

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

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. Let z_+ and $\mathbb{Z} \times \mathbf{K} \equiv S$ denote the TFP level "tomorrow" and the state space, respectively. We will represent an element in the state space by $s \in S$:

$$\begin{aligned} &[A(C_1)(s)]^{-\sigma} = \\ &\beta E\left\{\left[C_1\left(h(s) - A(C_1)(s)\mathbf{D}^{-1}, z_+\right)\right]^{-\sigma}\left[R\left(h(s) - A(C_1)(s)\mathbf{D}^{-1}, z_+\right) + 1 - \delta\right]\right\}. \end{aligned} \tag{C.17}$$

Given the definition of $\mathbb{C}_1$ and *if* h, which we will define below, is *continuous, bounded and increasing* in K, (C.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 (C.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 (C.17), using (C.15), (C.5), and (C.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 (C.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 (C.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

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

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:

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

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, (C.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 (C.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^i_H and the selection rule.