THE UNIVERSITY OF CHICAGO

CAPITAL GAINS TAXATION IN PRIVATE BUSINESS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ARSHIA HASHEMI

CHICAGO, ILLINOIS

JUNE 2024

Copyright © 2024 by Arshia Hashemi

All Rights Reserved

For my parents, Sonia and Mehrdad.

Thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am indebted to the faculty members in my committee, Chad Syverson (Chair), Eric Zwick, Thomas Wollmann, and Pascal Noel, for their guidance, advice, and encouragement in writing this PhD dissertation. I am deeply grateful to Eric Budish and Matthew Notowidigdo for believing in my potential as an economic researcher.

I thank Greg Kaplan and Esteban Rossi-Hansberg for standing by my side during the most challenging times. I thank Mike Golosov for countless thought-provoking conversations about research. I thank Ufuk Akcigit, Jonathan Dingel, Peter Ganong, Veronica Guerrieri, Erik Hurst, and Joe Vavra for many enlightening conversations. I am grateful to the economists at the Chicago Fed, especially Raffaella Giacomini, for graciously hosting me. Beyond the University of Chicago, I thank Hassan Afrouzi, Dan Benjamin, Job Boerma, Steve Bond, Mariacristina De Nardi, Nate Miller, Ben Moll, Kate Smith, Gianluca Violante, and Nicolas Werquin for their inspiration.

I thank my friends and peers, Aditya Bhandari, Tyler Jacobson, Nidhaan Jain, Jan Morgan, Lucy Msall, Laura Murphy, Ashton Pallottini, Gabriele Romano, Jordan Rosenthal-Kay, James Traina, and Zizhe Xia for their support and generosity. I thank members of the PhD program office at Chicago Booth for their assistance with administrative matters during my graduate studies in the Stevens Doctoral Program. I am grateful to the Institute for Humane Studies for their support.

Most of all, I am grateful to my loving family. My parents, Sonia and Mehrdad, provide a constant source of unconditional love, positivity, and affection. My younger brother Aria instills in me the value of hard work and belief in oneself. And my beautiful son, baby Alexander ("Lexi"), blesses my every moment with his delightful presence, his blissful innocence, and his heart filled with love and kindness. I love you all.

ABSTRACT

This paper studies the taxation of realized capital gains in the decentralized market for private businesses, an asset class with empirically large unrealized capital gains. I develop a dynamic heterogeneous agent model of private business, with incomplete markets, financial frictions, and an over-the-counter market for bilaterally exchanging business productivity levels. By influencing the reservation prices of sellers and buyers, the capital gains tax affects both exchange allocations and exchange asset prices. Quantitative analysis suggests that if financial frictions are severe, then a positive capital gains tax rate maximizes aggregate productivity. Yet the productivity gains stem from a decrease in aggregate input demand, rather than from an increase in aggregate output, which is necessarily lower under any positive tax rate relative to that under a zero tax rate.

1. Introduction

Preamble. Capital taxation encompasses different forms. Traditional forms of capital taxation distort the return to saving or investment and typically manifest in two types. The first type is a capital income tax levied on the flow income from saving or investment. The second type is a wealth tax levied on the stock of wealth accumulated over time from saving or investment.[1] Meanwhile, a different form of capital taxation is the capital gains tax ("CGT") levied on the realized capital gain upon selling an asset.[2] While the body of knowledge about traditional capital taxation is large, the body of knowledge about the CGT is notably smaller.[3]

Research Question and Motivation. This paper contributes to the literature on capital gains taxation by studying its effects in the decentralized market for private businesses. The motivation for this research question is twofold. The first motivation concerns studying the CGT itself. Studying the CGT as a distinct form of capital taxation is important because our insights about the effects of the capital income tax and the wealth tax do not directly carry over to the CGT. The reason is that studying the CGT requires enriching models of traditional capital taxation in at least two respects. First, sellers only incur a CGT upon realizing a capital gain, and so the model must incorporate a market in which sellers and buyers can transact with one another to sell and buy assets. By contrast, both neoclassical models of traditional capital taxation (Chamley, 1986; Chari et al., 1994; Farhi, 2010; Judd, 1985) and modern models of traditional capital taxation (Boar & Midrigan, 2023; Guvenen et al., 2023; Saez & Stantcheva, 2018; Straub & Werning, 2020)

1 Recent research determines the circumstances under which either the capital income tax or the wealth tax is optimal (Boar & Midrigan, 2023; Guvenen et al., 2023).
2 A capital gain is realized upon the sale of an asset if the sale price exceeds the owner's cost-basis, which is the price at which the owner had originally purchased the asset.
3 The original findings on optimal capital taxation include Chamley (1986), Judd (1985), and Chari et al. (1994) while more recent contributions include Farhi (2010), Saez & Stantcheva (2018), and Straub & Werning (2020).

abstract from markets in which agents directly transact with one another. Second, one requires a theory of asset pricing in order to determine realized capital gains or losses endogenously. Asset pricing presents a useful price theory tool with which one can endogenously determine the sale price of an asset. The sale asset price serves as the cost-basis for capital gains taxation once the asset is subsequently sold in the future. In contrast to studying capital gains taxation, studying traditional capital taxation typically does not require using asset pricing. The second motivation concerns the asset class of interest, namely private business. Using data from the 2022 U.S. Survey of Consumer Finances, I document that the unrealized capital gains in private business far exceed those in other asset classes, such as equities and housing, on which the existing empirical literature about capital gains taxation has focused (Agersnap & Zidar, 2021; Dai et al., 2008). In constituting a large tax base for unrealized capital gains, then, private business is an important asset class with respect to which to study the effects of the capital gains taxation.

Model Overview. This paper extends a canonical heterogeneous agent model of private business, à la Cagetti & De Nardi (2006), to account for the two aforementioned margins required to study the CGT. First, to account for a market in which sellers and buyers can interact, I extend the frictional over-the-counter ("OTC") market structure of Duffie et al. (2005) to account for bilateral exchanges of private business assets. The OTC market structure is well-suited because private business is an illiquid asset class that is exchanged on a decentralized market and for which a single price schedule does not exist. Moreover, the OTC market is frictional in the sense that one must wait a strictly positive amount of time before encountering a bilateral exchange opportunity with a counterparty. A key assumption in the quantitative model is that private business owners engage in bilateral exchanges in which the seller and buyer can exchange their private business

productivity levels with one another for a lump-sum payment from the buyer to the seller.[4] Given a bilateral exchange opportunity between two private business owners, with different business productivity levels, the seller is indexed by the higher productivity level and the buyer is indexed by the lower productivity level.[5] Specifically, the seller can downgrade in productivity in exchange for receiving a lump-sum payment from the buyer, minus a CGT payment to the government if the seller realizes a capital gain. The buyer can upgrade in productivity in exchange for paying a lump-sum payment to the seller.

The gains from exchanging productivity levels on the OTC market arise endogenously from a financial friction, manifesting as a collateral constraint on capital demand in production (Buera et al., 2011; Buera & Shin, 2013; Moll, 2014). In the frictional economy with a collateral constraint, wealth affects the efficiency of production in private business. The reason is that, due to the imperfect enforceability of contracts, renting capital in production requires owners to pledge their wealth as collateral to a financial intermediary. The implication is that less wealthy owners are collateral constrained and hence are unable to operate their private businesses at the efficient scale.[6] In equilibrium, less wealthy sellers (indexed by high productivity) and more wealthy buyers (indexed by low productivity) choose to exchange their productivity levels. Relative to the less wealthy seller, the more wealthy buyer can operate the high productivity business at, or closer to, its efficient scale. The more wealthy buyer benefits from the additional flow profit income

4 Business productivity represents any rival, non-excludable, and indivisible asset that positively contributes to the quality of the homogenous consumption good and that is not a primary factor of production, such as capital or labor. Examples include goodwill, reputation, customer base, or management best practices. Characteristics of private business that are inseparable from the owner, such as innate human capital, are not captured in this model.

5 I simplify the model even further by assuming that business productivity is binary. This implies that a private business owner with high productivity pre-exchange is a prospective seller, while one with low productivity pre-exchange is a prospective buyer.

6 For each productivity level, there is a unique and finite efficient scale of production due to decreasing returns to scale.

associated with a high productivity business, while the less wealthy seller benefits from the lump-sum payment received from the buyer.

Second, to determine realized capital gains or losses endogenously, I refine the canonical model of private business in two additional ways. First, I introduce the seller's cost-basis as an additional state variable in the dynamic optimization problem of consumption-saving. The cost-basis evolves endogenously over time as buyers transition into sellers. Adding an additional state variable, especially one with a large support, is computationally challenging. To make progress, I formulate the dynamic optimization problem in continuous time, leveraging recent methodological advancements (Achdou et al., 2021). Second, I characterize the reservation asset prices of sellers and buyers endogenously. The seller's reservation price is the minimum asset price that the seller is willing to accept in order to downgrade in productivity. The buyer's reservation price is the maximum asset price that the buyer is willing to pay in order to upgrade in productivity. Indifference conditions equating pre-exchange values to post-exchange values determine the two reservation asset prices endogenously.

If the buyer's maximum asset price exceeds the seller's minimum asset price, then a bilateral exchange occurs. What remains to be determined is the asset price at which the bilateral exchange occurs. To ensure that both parties are strictly better off from exchanging, the exchange asset price must lie within the compact support bounded from above by the buyer's maximum asset price and bounded from below by the seller's minimum asset price. To determine the exchange asset price uniquely, I assume the exchange asset price is a weighted average of the two reservation prices, regulated by a bargaining parameter. This parameter captures the bargaining power of the seller relative to that of a buyer. A microfoundation for this assumption is a Nash bargaining problem in which the seller and buyer bargain over the asset price that maximizes their

joint surplus from the bilateral exchange. Bargaining is a fitting tool for determining asset prices because the price of a private business depends partly on the value that it confers to the next best owner. The buyer's maximum asset price captures this value, while the exchange asset price reflects both the value of the asset in the hands of the current owner (i.e., the seller) and the value of the asset in the hands of the next best owner (i.e., the buyer).

Theoretical Predictions. A central contribution of this paper lies in characterizing two elasticities that summarize how asset prices respond to a change in the CGT. All else equal, an increase in the CGT induces an increase in the seller's CGT liability upon realizing a capital gain. Bearing a greater tax burden upon realizing a capital gain, the seller stipulates a higher minimum asset price in order to accept a bilateral exchange. This gives rise to the well-known "lock-in" effect of capital gains taxation (Chari et al., 2005). The lock-in effect, in turn, passes through to an increase in the exchange asset price. The pass-through rate depends on the weight attached to the seller's minimum asset price in determining the exchange asset price. This paper characterizes, in closed-form, both the elasticity of the seller's minimum asset price with respect to the CGT and the elasticity of the exchange asset price with respect to the CGT.

In addition, I characterize the theoretical predictions of the lock-in effect on exchange allocations. First, the increase in the seller's minimum asset price may distort exchange allocations. Indeed, there are some bilateral exchanges for which the seller's minimum asset price is strictly less than the buyer's maximum asset price prior to the increase in the CGT, while the seller's minimum asset price is strictly greater than the buyer's maximum asset price after the increase in CGT. Such bilateral exchanges occur before the increase in the CGT, but not afterwards, resulting in a deadweight loss. Second, even if a bilateral exchange occurs after the increase in the CGT, the

asset price at which the exchange occurs is different, due to the increase in the seller's minimum asset price stemming from the lock-in effect.

Quantitative Results. Another central contribution of this paper lies in quantifying the effects of capital gains taxation on efficiency and equity. One measure of efficiency is aggregate productivity, which is endogenous in the frictional economy with a collateral constraint on capital demand in production (Moll, 2014).[7] Formally, aggregate productivity in the quantitative model is endogenous because it depends on the stationary joint distribution function over the state space of wealth and productivity. This joint distribution function summarizes how different productivity levels are allocated across the wealth distribution of private business owners. Maximizing aggregate productivity requires allocating high productivity businesses to wealthy owners (i.e., positive assortative matching between productivity and wealth), thereby minimizing the capital misallocation stemming from the collateral constraint. The OTC market partially achieves this optimal allocation, albeit not entirely due to the frictional aspect of this market.

Capital gains taxation shapes the joint distribution function in the stationary equilibrium by influencing wealth accumulation through bilateral exchanges. This force manifests through not only whether a bilateral exchange occurs, but also through the asset price at which an exchange occurs. I decompose the effects of capital gains taxation on aggregate productivity into three components: (i) a change in the aggregate output of sellers; (ii) a change in the aggregate output of buyers; and (iii) a change in the aggregate input measured by aggregate capital demand. The severity of financial frictions and the severity of bargaining power interact in an important manner. If financial frictions are severe, then a strictly positive CGT rate maximizes aggregate productivity.

7 Collateral constraints on capital demand are a pervasive source of financial frictions and give rise to sizeable aggregate productivity losses (Catherine et al., 2022).

However, the productivity gains stem from a decrease in input use, rather than from an increase in output. In fact, relative to a zero CGT rate, aggregate output is necessarily lower under any strictly positive CGT rate. Having established this distortive effect of capital gains taxation, this paper traces out the efficiency-equity frontier detailing the trade-off between increasing aggregate output (i.e., efficiency) and increasing aggregate tax revenues (i.e., equity). The balance of bargaining power between sellers and buyers affects this trade-off because bargaining power determines the asset prices at which bilateral exchanges occur. In turn, the exchange asset prices determine not only the efficiency with which owners operate their private businesses post-exchange (by affecting the post-exchange wealth stock), but also the tax revenues that the government collects from capital gains taxation.

Related Literature. This paper contributes to three related strands of literature. The first strand is the literature studying financial frictions as a source of capital misallocation and quantifying the ensuing aggregate productivity losses (Buera et al., 2011; Buera & Shin, 2013; Moll, 2014). A subset of this literature adopts a normative focus, studying optimal policy remedies for capital misallocation stemming from financial frictions. For example, Itskhoki & Moll (2019) show optimal industrial policy should initially favor businesses at the expense of workers in order to facilitate faster wealth accumulation in economies with financial frictions. Recently, Guvenen et al. (2023) and Boar & Midrigan (2023) study the trade-off between taxing the flow capital income versus the stock of wealth in economies in which heterogeneous rates of return on wealth arise endogenously from a collateral constraint on capital demand in production. This paper contributes to this literature by studying the nexus between financial frictions and capital gains taxation in the decentralized OTC market for private businesses, and in quantifying the aggregate productivity losses from capital misallocation in such an economy.

The second strand of literature studies the trade of private businesses. The origins of this literature lie in the model of Holmes & Schmitz (1990) that posits a theory of specialization that drives changes in business ownership over time. Chari et al. (2005) offer an important contribution in quantifying the distortive lock-in effects of capital gains taxation when there is heterogeneity in the comparative advantage for starting new businesses. Separately, Bhandari et al. (2021) study the trade of physical capital across private business owners via decentralized markets. Similarly, Guntin & Kochen (2023) study the aggregate implications of trading private firms in decentralized market. The quantitative model of this paper builds on that of Guntin & Kochen (2023) by allowing for a capital gains tax with a non-degenerate cost-basis. Allowing for a non-degenerate cost-basis is computationally challenging because it requires introducing the cost-basis as an additional state variable in the dynamic optimization problem.

The third strand of literature studies the nexus between capital gains (taxes) and asset pricing. Dai et al. (2008) distinguish between two effects of capital gains taxation on asset prices, namely a supply-side lock-in effect and a demand-side capitalization effect. Moreover, Fagereng et al. (2022) provide a quantitative framework for evaluating the welfare effects of changes in asset prices, distinguishing between unrealized and realized capital gains. This paper builds on this literature by characterizing the effects of capital gains taxes on the asset prices of illiquid private business assets. Unlike equities or housing, private businesses do not have a single price schedule at which sellers and buyers can transact. Rather, the asset price of a private business depends on its value in the hands of the next best alternative owner. This value depends not only on the idiosyncratic characteristics (e.g., wealth) of the next best owner, but also on aggregate conditions such as the severity of financial frictions and the balance of bargaining power between sellers and

buyers. This paper studies how the characteristics of sellers and buyers interact with economic primitives to shape reservation prices and exchange asset prices.

Roadmap. The remainder of this paper is organized as follows. Section 2 presents motivating evidence documenting that unrealized capital gains in private business are large relative to those in other asset classes such as equities or housing. Section 3 outlines a stylized model highlighting financial frictions as the economic primitive generating gains from bilateral exchanges. Section 4 outlines the quantitative model with which I evaluate the effects of capital gains taxation in the decentralized market for private businesses. Section 5 characterizes the theoretical predictions of a change in the CGT rate on endogenous outcomes. Section 6 calibrates the quantitative model to perform quantitative analysis that characterizes the efficiency-equity frontier. Section 7 concludes.

2. Motivating Evidence

This section uses data from the 2022 U.S. Survey of Consumer Finances ("SCF") to document stylized facts about private business owners and their unrealized capital gains. The SCF asks survey respondents to report the total value of their unrealized capital gains across three asset classes: (i) equities (stocks and mutual funds); (ii) housing (primary residence and other real estate for non-commercial use); and (iii) private business. Among the three aforementioned asset classes, the largest unrealized capital gains lie in private business. From the perspective of offering a large tax base, then, private business is an important asset class with respect to which to study the effects of capital gains taxation.

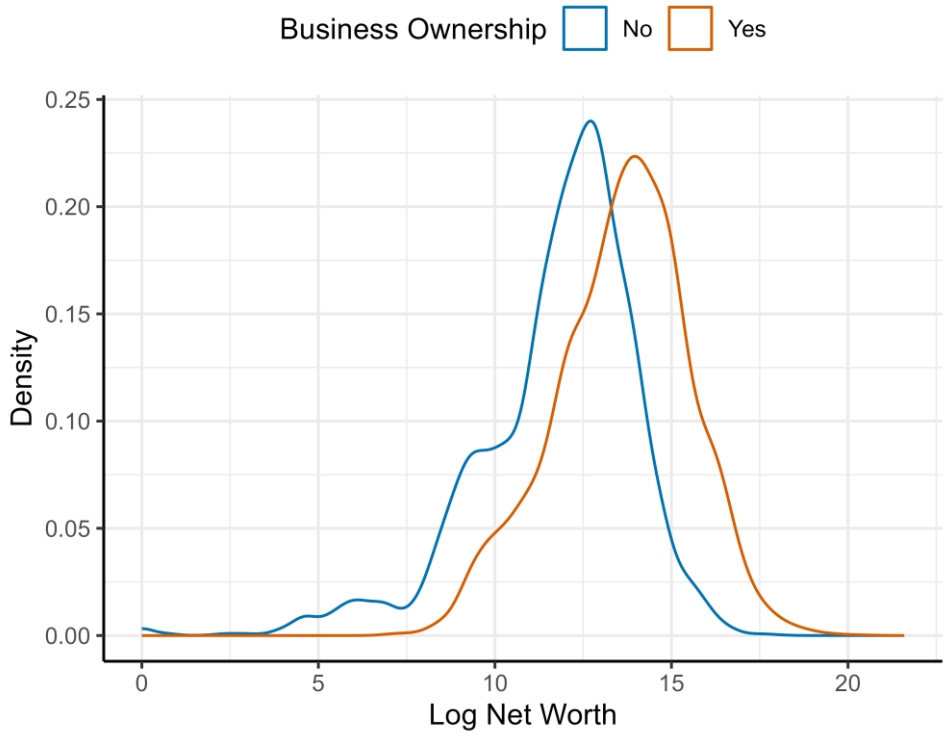
2.1 Wealth Distribution by Business Ownership

Approximately 14.6% of U.S. households own or operate a private business. U.S. households that own private businesses are wealthier than U.S. households that do not own any private businesses. Indeed, the median private business owner has approximately six times more wealth than the median non-owner. The measure of wealth in the SCF is net worth, which is defined as the difference between total assets and total liabilities. Total assets include financial assets, such as checking accounts, certificates of deposit, stocks, and retirement accounts, and non-financial assets, such as vehicles, real estate, and private businesses.

Figure 1 plots the marginal density of logarithmic net worth by private business ownership status. The distribution of wealth for private business owners lies to the right of the distribution of wealth for non-owners. Moreover, as the units of wealth are logarithmic, the rightward shift implies that households at the upper end of the wealth distribution are disproportionately private

business owners. In focusing on private business owners, this paper studies a small segment of the total household population in the United States at the top end of the overall wealth distribution.

Figure 1: Net Worth Distribution by Business Ownership



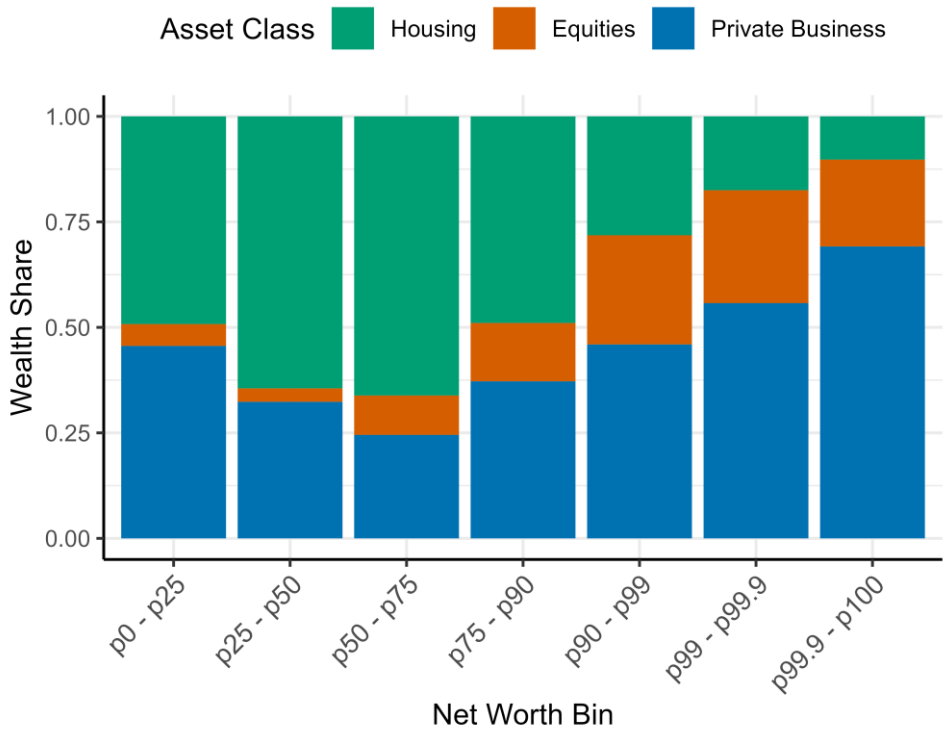
2.2 Unrealized Capital Gains in Private Businesses

Having established that private business owners are disproportionately wealthier than non-owners, I proceed to consider the sources of wealth among private business owners. Figure 2 documents the share of total wealth held by one of three asset classes: housing; equities; and private business. This figure groups private business owners into bins on the basis of their percentile in the net worth distribution. Each vertical bar represents the average value for a particular net worth bin.

Housing comprises more than half of total wealth among net worth bins below the 75th percentile. However, among net worth bins above the 75th percentile, housing comprises a smaller

share of total wealth, especially among net worth bins at the highest percentiles. Rather, private business constitutes a larger share of total wealth at the top end of the wealth distribution. The large share of wealth held in private businesses at the top end of the wealth distribution indicates that private business presents a sizeable source of idiosyncratic risk. This risk is especially pertinent as the market for private businesses is not highly liquid and disposing of private business assets is a major undertaking.

Figure 2: Wealth Share Decomposition by Asset Class



Another variable of interest in the SCF is the value of unrealized capital gains or losses, which is reported by asset class. Figure 3 reports the raw dollar value of unrealized capital gains or losses, among the subset of U.S. households that report being private business owners, across three asset classes: (i) private business; (ii) equities (stocks and mutual funds); and (iii) housing (for non-commercial use). Each dot in each subfigure represents a particular U.S. household in the SCF.

The size of the dot represents the sampling weight attached to the household in the SCF, with larger dots indicating a larger sampling weight. The main takeaway is that the unrealized capital gains in private business are large relative to those in the equities or housing asset classes.

Figure 3: Total Unrealized Capital Gains

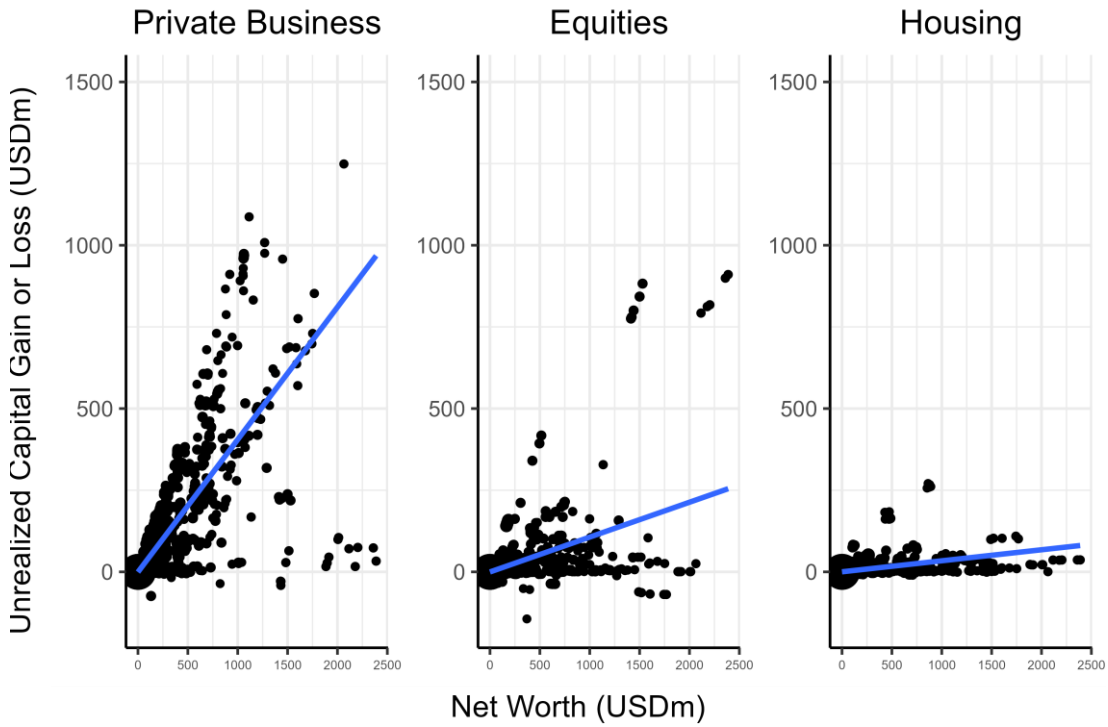
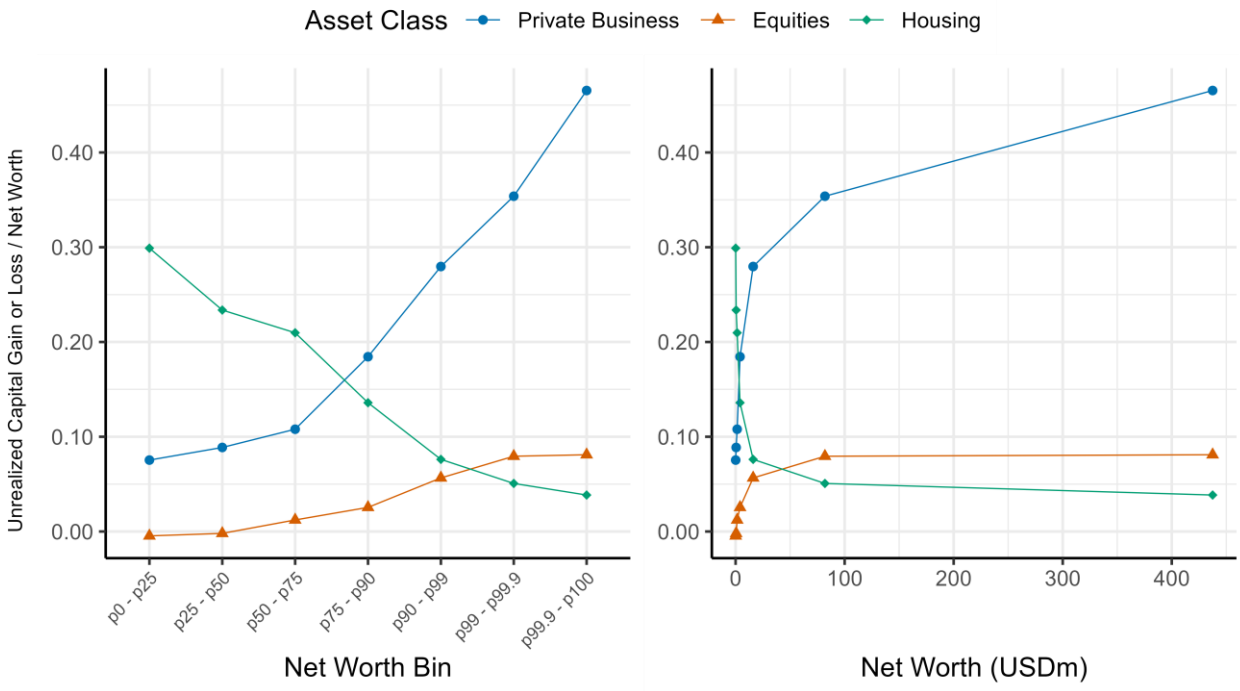


Figure 4: Ratio of Total Unrealized Capital Gains to Net Worth



A potential explanation for the positive relationship between unrealized capital gains and net worth is that unrealized capital gains are proportional to net worth. Figure 4 evaluates this explanation by plotting the ratio of unrealized capital gains to net worth by net worth, after grouping households in bins on the basis of their percentile in the net worth distribution and computing the mean value of unrealized capital gains and net worth for each bin. For equities and private business, the share of unrealized capital gains in net worth is increasing in net worth, suggesting that net worth is not the sole driver of unrealized capital gains.

3. Stylized Model

This section outlines a stylized model that describes key features of the private business economy. By making stylized assumptions, the stylized model admits closed-form solutions for various endogenous variables. In doing so, the stylized model makes apparent the gains from exchanging business productivity levels.

3.1 Static Optimization

3.1.1 Primitives

Production Technology. Private business owners produce a single homogeneous good and are heterogeneous in their business productivity $z \in \{z_\ell, z_h\}$, satisfying $0 < z_\ell < z_h$. The production function (1) is Cobb-Douglas with two factors of production, labor n and capital k, and with constant returns to scale. Business productivity z is capital-augmenting.

$$\mathcal{F}(n, k; z) := (z \cdot k)^\alpha \cdot n^{1-\alpha} \tag{1}$$

Owners rent labor n on a perfectly competitive labor market with a spot unit wage rate $w > 0$, which is a model parameter.

Rental Market for Capital. Owners have access to perfectly competitive financial intermediaries that take in owners' wealth as deposits and rent out capital k to owners at the rental rate $r > 0$, which is a model parameter. Crucially, production is subject to a financial friction manifesting as a collateral constraint (2) on capital demand k.[8]

8 A microfoundation for the collateral constraint (2) stems from imperfect enforceability of contracts. Consider a private business owner with wealth a seeking to rent k units of capital. To do so, the private business owner must deposit its wealth a with a financial intermediary. If the owner reneges on the rental contract by retaining a fraction $1/\lambda$ of the rented capital, then the financial intermediary has the right to garnish the owner's wealth a as a punishment. At the optimum, the financial intermediary will rent capital to the owner until the owner has an incentive to steal the rented capital, yielding the desired collateral constraint $k/\lambda \leq a$.

$$k \in [0, \lambda \cdot a] \tag{2}$$

Specifically, a private business owner's leverage ratio, defined as the ratio of capital demand k to wealth a, can be at most $\lambda \in [1, \infty)$. The leverage ratio parameter λ regulates the severity of the financial friction. In the most severe limiting case, $\lambda \to 1$, one can rent capital only up to the wealth level a. In the least severe limiting case, $\lambda \to \infty$, capital demand is completely unconstrained. The collateral constraint features in papers such as Buera et al. (2011), Buera & Shin (2013), and Moll (2014) that study the aggregate losses stemming from financial frictions.

3.1.2 Profit Maximization Problem

The state variables of a private business owner are wealth a and productivity z. Taking its state variables (a, z) and factor prices (w, r), the private business owner chooses its demand for factors of production (n, k) to maximize its flow profit income, subject to the production function (1) and collateral constraint (2).

$$\pi^*(a, z) := \max_{n,k} \{(z \cdot k)^{\alpha} \cdot n^{1-\alpha} - w \cdot n - r \cdot k\} \tag{3}$$

$$\text{s.t. } k \in [0, \lambda \cdot a]$$

The following lemma characterizes the optimal factor demand policy functions.

Lemma 1: Static Profit Maximization Problem (Stylized Model)

The factor demand policy functions are linear in wealth a.

$$n^*(a,z) = \left[\frac{1-\alpha}{w}\right]^{\frac{1}{\alpha}} \cdot z \cdot k^*(a,z) \tag{4}$$

$$k^*(a,z) = \begin{cases} \lambda \cdot a & MPK(z) > r \\ 0 & MPK(z) \leq r \end{cases} \tag{5}$$

The marginal product of capital, denoted by $MPK(z)$ and evaluated at the labor demand policy function $n = n^*(a,z)$, does not depend on the level of capital demand.

$$MPK(z) := \frac{\partial \mathcal{F}(n,k;z)}{\partial k}\bigg|_{n=n^*(a,z)} = \alpha \cdot \left[\frac{1-\alpha}{w}\right]^{\frac{1}{\alpha}} \cdot z \tag{6}$$

The flow profit income is linear in wealth a.

$$\pi^*(a,z) = \begin{cases} [MPK(z) - r] \cdot \lambda \cdot a & MPK(z) > r \\ 0 & MPK(z) \leq r \end{cases} \tag{7}$$

Proof. See Appendix A.1.

Lemma 1 details that optimal capital demand $k^*(a,z)$ exhibits a "bang-bang" property. That is, capital demand is either at the lower bound of the collateral constraint (2), i.e. $k^*(a,z) = 0$, or at the upper bound, i.e. $k^*(a,z) = \lambda \cdot a$.

This bang-bang property is a consequence of constant returns to scale in production. After optimizing over the choice of labor demand n, the resulting profit objective function is linear in the choice of capital demand k. If the marginal product of capital $MPK(z)$, conditional on the optimal labor demand $n^*(a,z)$, is greater than the user cost of capital r, then the coefficient on the

choice of capital demand k in the objective function is strictly positive, and it is optimal to demand as much capital as is permitted by the collateral constraint. Meanwhile, if $MPK(z) < r$, then it is optimal to demand as little capital as it permitted, i.e., private business owners choose to be inactive in production by demanding zero capital and hence producing zero output.

3.2 Dynamic Optimization

This subsection characterizes the dynamic consumption-saving problem of private business owners. I cast the dynamic aspect of the model in continuous time. I assume private businesses have infinite lives, with a discount rate $\rho > 0$, which is a model parameter. Moreover, I assume that the utility function over consumption is logarithmic.

3.2.1 Excess Rate of Return on Wealth

At any instant of time, the total flow income of private business owners comprises two components. First, there is a flow interest income $r \cdot a$ on wealth a. Second, there is a flow profit income $\pi^*(a, z)$ from owning and operating a private business. Importantly, private business owners that are active in production earn an excess rate of return, denoted by $R(z)$, on their wealth a that exceeds the rate r.

$$R(z) := \frac{\partial \pi^*(a,z)}{\partial a} = \begin{cases} [MPK(z) - r] \cdot \lambda & MPK(z) > r \\ 0 & MPK(z) \leq r \end{cases} \tag{8}$$

Intuitively, private business owners that are sufficiently productive such that their marginal product of capital $MPK(z)$ is greater than the user cost of capital r benefit from additional wealth because more wealth relaxes the collateral constraint (2), enabling more capital to be rented for production. The marginal value of additional capital in generating flow profit income depends on

productivity z, which in turn shapes the marginal product of capital. By constant returns to scale, the marginal product of wealth equals the average product of wealth.

$$\frac{\partial \pi^*(a,z)}{\partial a} = \frac{\pi^*(a,z)}{a}$$

Thus, one may rewrite the flow ex-post profit function as $\pi^*(a,z) = R(z) \cdot a$, which is a multiplicatively separable function of wealth a and productivity z. The excess rate of return $R(z)$ generates an additional savings motive that is isomorphic to a higher interest rate on wealth a.

3.2.2 Hamilton-Jacobi-Bellman Equation

I assume productivity $z \in \{z_\ell, z_h\}$ evolves exogenously over time according to a two-state Poisson process, with transition rates $\gamma_{\ell h} > 0$ and $\gamma_{h\ell} > 0$, where $\gamma_{\ell h}$ denotes the transition rate from z_ℓ to z_h and $\gamma_{h\ell}$ denotes the transition rate from z_h to z_ℓ. Let $V(a,z)$ denote the value function of a private business owner with wealth a and productivity z.[9] The Hamilton-Jacobi-Bellman ("HJB") equation (9) gives a recursive formulation of the private business owner's dynamic consumption-saving problem.

$$\underbrace{\rho \cdot V(a,z)}_{\textit{Annuity Value}} = \max_{c>0} \Bigg\{ \underbrace{\ln c}_{\textit{Flow Utility}} + \underbrace{[r \cdot a + R(z) \cdot a - c] \cdot \frac{\partial V(a,z)}{\partial a}}_{\textit{Consumption-Saving}} + \underbrace{\gamma_{z'z} \cdot [V(a,z') - V(a,z)]}_{\textit{Productivity Transition}} \Bigg\} \tag{9}$$

The left-hand side of the HJB equation (9) denotes the flow annuity value $\rho \cdot V(a,z)$ for a private business owner with wealth a and productivity z. The right-hand side of the HJB equation (9)

9 The value function $V(a,z)$ inherits the units of the flow term in the HJB equation (9). Since the units of the flow term, $\ln c$, are utils, it follows that the units of the value function are also utils.

decomposes the flow annuity value into three constituent components. The first term on the right-hand side denotes the flow utility from consumption. The second and third components capture the change in value from transitions in the state variables (a, z). The second term on the right-hand denotes the change in value from the endogenous consumption-saving decision that governs the evolution of wealth a. Flow saving is equal to total flow income $[r \cdot a + R(z) \cdot a]$ minus flow consumption c. The third term on the right-hand side denotes the change in value from exogenous transitions in productivity z.

Given a logarithmic utility function over consumption and a total flow income function that is linear in wealth a (i.e., $[r + R(z)] \cdot a$), the value function $V(a, z)$ that satisfies the HJB equation (9) admits a closed-form solution, which is detailed in Lemma 2.

Lemma 2: Dynamic Optimization (Stylized Model)

The private business owner's value function $V(a,z)$ is logarithmic in wealth a, the flow consumption policy function $c^*(a,z)$ is linear in wealth a, and the flow saving policy function $\dot{a}^*(a,z)$ is linear in wealth a.

$$V(a,z) = \frac{1}{\rho} \cdot \left[\ln\rho + \left(\frac{r-\rho}{\rho}\right) + \frac{1}{\rho} \cdot \bar{R}(z)\right] + \frac{1}{\rho} \cdot \ln a \tag{10}$$

$$c^*(a,z) = \rho \cdot a \tag{11}$$

$$\dot{a}^*(a,z) = [(r-\rho) + R(z)] \cdot a \tag{12}$$

The expected excess rate of return function, denoted by $\bar{R}(z)$, is a weighted average of the individual excess rate of return functions $R(z)$, with weights that depend on three rate parameters $(\rho, \gamma_{\ell h}, \gamma_{h\ell})$.

$$\bar{R}(z) = \frac{\rho + \gamma_{zz\prime}}{\rho + \gamma_{zz\prime} + \gamma_{z\prime z}} \cdot R(z) + \frac{\gamma_{z\prime z}}{\rho + \gamma_{zz\prime} + \gamma_{z\prime z}} \cdot R(z') \tag{13}$$

Proof. See Appendix A.2.

3.3 Gains from Bilateral Exchanges

While the stylized model does not feature a frictional over-the-counter market for bilateral exchanges, the closed-form solutions that the stylized model admits highlight the primitives from which gains from exchange arise. This paper considers the bilateral exchange of productivity levels. Given binary productivity $z \in \{z_\ell, z_h\}$, a private business owner with high productivity z_h pre-exchange is a prospective seller, while an owner with low productivity z_ℓ pre-exchange is a prospective buyer.

Sellers have the opportunity to downgrade in productivity from z_ℓ to z_h in exchange for receiving a lump-sum payment from buyers. Buyers have the opportunity to upgrade in productivity from z_h to z_ℓ in exchange for paying a lump-sum payment to sellers. The gains from exchange arise from financial frictions manifesting through the collateral constraint on capital demand in equation (2). Private business owners with low wealth and high productivity and owners with high wealth and low productivity find it optimal to exchange. The reason is that the high wealth owners are better positioned to overcome financial frictions and hence operate the high productivity business at a larger scale.

3.3.1 Reservation Asset Price Functions

To determine whether a bilateral exchange occurs, it is necessary to determine two reservation asset prices. The first is the minimum asset price a seller is willing to accept in order to downgrade from a high productivity business z_h to a low productivity business z_ℓ. The second is the maximum asset price a buyer is willing to pay in order to upgrade from a low productivity business z_ℓ to a high productivity business z_h.

Seller's Minimum Asset Price. The seller's minimum asset price function, denoted by $P^*_{min}(a)$, equates the seller's pre-exchange value to the seller's post-exchange value, if the exchange were to occur at the seller's minimum asset price. The seller's indifference condition in equation (14) summarizes this marginal optimality condition and implicitly determines the seller's reservation asset price $P^*_{min}(a)$.

$$V(a + P^*_{min}, z_\ell) = V(a, z_h) \tag{14}$$

The right-hand side of equation (14) denotes the seller's value pre-exchange, namely the value associated with wealth a and operating a private business with a high productivity level z_h. The

left-hand side of equation (14) denotes the seller's value post-exchange, namely the value associated with an increase in wealth to $(a + P^*_{min})$ and a downgrade in productivity to z_ℓ. Given the closed-form solution for the value function (10), one can solve the seller's indifference condition (14) to obtain a closed-form solution for the seller's minimum asset price function $P^*_{min}(a)$, which is detailed in equation (15).

$$P^*_{min}(a) = \left[\exp\left(\frac{1}{\rho} \cdot [\bar{R}(z_h) - \bar{R}(z_\ell)]\right) - 1\right] \cdot a \tag{15}$$

All else equal, the greater is the difference in the expected excess rate of return between a high productivity and a low productivity business, as measured by $[\bar{R}(z_h) - \bar{R}(z_\ell)]$, the greater is the seller's minimum asset price. The reason is that the seller foregoes a greater expected excess rate of return upon exchange, and so requires a larger lump-sum compensation to do so.

Buyer's Maximum Asset Price. The buyer's maximum asset price function, denoted by $P^*_{max}(a')$, equates the buyer's pre-exchange value to the buyer's post-exchange value, if the bilateral exchange occurs at the buyer's maximum asset price. The buyer's indifference condition in equation (16) summarizes this marginal optimality condition.

$$V(a' - P^*_{max}, z_h) = V(a', z_\ell) \tag{16}$$

The right-hand side of equation (16) denotes the buyer's value pre-exchange, namely the value associated with wealth a' and operating a private business with a low productivity level z_ℓ. The right-hand side of equation (16) denotes the buyer's value post-exchange, namely the value associated with a decrease in the wealth stock to $(a - P^*_{max})$, and an upgrade in productivity to z_h. Once more, given the closed-form solution for the value function (10), one can solve the buyer's

indifference condition (16) to obtain a closed-form solution for the buyer's maximum asset price function $P^*_{max}(a)$, which is detailed in equation (17).

$$P^*_{max}(a') = \left[\frac{\exp\left(\frac{1}{\rho}\cdot[\bar{R}(z_h) - \bar{R}(z_\ell)]\right) - 1}{\exp\left(\frac{1}{\rho}\cdot[\bar{R}(z_h) - \bar{R}(z_\ell)]\right)}\right]\cdot a' \tag{17}$$

All else equal, the greater is the buyer's pre-exchange wealth a', the greater is the maximum asset price the buyer is willing to pay. Furthermore, the buyer's maximum asset price is bounded below by zero and bounded above by a', i.e., $P^*_{max}(a') \in (0, a')$. The greater is the difference in the expected excess rate of return across the two productivity levels, i.e., the greater is the value of the difference $[\bar{R}(z_h) - \bar{R}(z_\ell)]$, the greater is the buyer's maximum asset price. Moreover, the largest lump-sum transfer that the buyer can pay is the entire pre-exchange wealth stock a'.

3.3.2 Threshold Rule for Bilateral Exchange

Consider a bilateral exchange opportunity between a seller with wealth a and a buyer with wealth a'. I assume both parties have perfect information about each other's reservation prices, namely $P^*_{min}(a)$ and $P^*_{max}(a')$.[10] I assume an exchange occurs whenever the buyer's maximum asset price is strictly greater than the seller's minimum asset price. This assumption results in a threshold rule for exchange summarized in the following proposition.

10 Under the assumption of perfect information about the seller's and buyer's reservation prices, I abstract from inefficiencies in bilateral trade that arise from imperfect information (Myerson & Satterthwaite, 1983).

Proposition 1: Threshold Rule for Bilateral Exchanges

A seller with wealth a and a buyer with wealth a' find it optimal to exchange if and only if the ratio of the buyer's wealth to the seller's wealth exceeds a threshold.

$$\frac{a'}{a} > \exp\left(\frac{1}{\rho} \cdot [\bar{R}(z_h) - \bar{R}(z_\ell)]\right) \in (1, \infty) \tag{18}$$

Proof. Rearrange the inequality $P^*_{max}(a') > P^*_{min}(a)$ using the solutions for the two reservation asset prices in equations (15) and (17).

The threshold rule for bilateral exchanges in equation (18) underscores two important properties of bilateral exchanges. First, since the threshold on the right-hand side is bounded below by one, it follows that a necessary condition for a bilateral exchange to occur is that the buyer's pre-exchange wealth a' be strictly greater than the seller's pre-exchange wealth a. This is intuitive since the gains from exchange arise from the fact that wealthier buyers are better positioned to overcome financial frictions and operate a high productivity business either at, or closer to, the efficient scale of production. Second, the greater is the difference in the expected excess rates of return, i.e., the greater is the difference $[\bar{R}(z_h) - \bar{R}(z_\ell)]$, the greater must be the ratio of the buyer's wealth to the seller's wealth for a bilateral exchange to occur. This is an implication of the fact that the seller's opportunity cost of downgrading in productivity is greater when the difference in expected excess rates of return is greater. As such, the seller requires a greater minimum asset price that only sufficiently wealth buyers can afford to pay.

4. Quantitative Model

This section presents the quantitative model with which I evaluate the effects of capital gains taxation in the decentralized market for private businesses. The quantitative model generalizes the stylized model in four respects. First, and most importantly, there is a frictional decentralized over-the-counter ("OTC") market in which private business owners encounter bilateral opportunities to exchange their business productivity levels with one another. Second, sellers incur a capital gains tax upon realizing a capital gain under a bilateral exchange on the OTC market. Third, I generalize the returns to scale in production to accommodate decreasing returns to scale, in place of constant returns to scale. Fourth, I generalize the flow utility function from consumption to the Constant Relative Risk Aversion ("CRRA") functional form, in place of a logarithmic function.

4.1 Static Optimization

4.1.1 Decreasing Returns to Scale

The production function (19) generalizes the production function (1) in the stylized model to accommodate decreasing returns to scale in production, which is regulated by an additional scale parameter $\nu \in (0,1)$.

$$\mathcal{F}(n,k;z) := [(z \cdot k)^{\alpha} \cdot n^{1-\alpha}]^{\nu} \tag{19}$$

Under decreasing returns to scale, and in the absence of the collateral constraint (2), there is a unique and finite efficient scale of production associated with each productivity level $z \in \{z_{\ell}, z_h\}$. By contrast, under constant returns to scale, and in the absence of the collateral constraint (2), the efficient scale of production is indeterminate.[11]

11 Under constant returns to scale, and in the absence of the collateral constraint (2), only the optimal capital-labor ratio k/n is determinate.

Under decreasing returns to scale, and in the presence of the collateral constraint (2), an owner can operate the private business at the efficient scale of production if and only if the owner is sufficiently wealthy. The intuition is that, for a sufficiently large wealth stock a, the optimal capital demand associated with the efficient scale of production lies strictly inside the compact set $[0, \lambda \cdot a]$. In this case, the collateral constraint is slack, and the scale of production is efficient.

Meanwhile, for sufficiently small wealth a, the optimal capital demand associated with the efficient scale of production lies strictly outside of the compact set $[0, \lambda \cdot a]$. In this case, the collateral constraint binds (i.e., $k = \lambda \cdot a$), and the scale of production is inefficiently low. In sum, decreasing returns to scale implies that the collateral constraint is slack once a private business owner is sufficiently wealthy. This economic intuition manifests in the solution to the static profit maximization problem outlined and solved below.

4.1.2 Static Profit Maximization Problem

Taking its state variables, wealth a and productivity z, and factor prices (w, r) as given, a private business owner chooses its demand for primary factors of production, labor n and capital k, to maximize flow profit income, subject to the production function (19) and the collateral constraint (2) on capital demand.

$$\pi^*(a, z) := \max_{n,k} \{[(z \cdot k)^{\alpha} \cdot n^{1-\alpha}]^{\nu} - w \cdot n - r \cdot k\} \tag{20}$$

$$\text{s.t. } k \in [0, \lambda \cdot a]$$

The following lemma characterizes the optimal factor demand policy functions.

Lemma 3: Static Profit Maximization (Quantitative Model)

The optimal factor demand policy functions satisfy

$$n^*(a,z) = \left[\frac{\nu(1-\alpha)}{w}\right]^{\frac{1}{1-\nu(1-\alpha)}} \cdot [z \cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \tag{21}$$

$$k^*(a,z) = \begin{cases} \lambda \cdot a & a < \mathcal{A}^*(z) \\ \nu^{\frac{1}{1-\nu}} \cdot \left(\frac{\alpha}{r}\right)^{\frac{1-\nu(1-\alpha)}{1-\nu}} \cdot \left(\frac{1-\alpha}{w}\right)^{\frac{\nu(1-\alpha)}{1-\nu}} \cdot z^{\frac{\nu\alpha}{1-\nu}} & a \geq \mathcal{A}^*(z) \end{cases} \tag{22}$$

where $\mathcal{A}^*(z)$ denotes the threshold wealth level at which the collateral constraint binds marginally.

$$\mathcal{A}^*(z) := \frac{1}{\lambda} \cdot \nu^{\frac{1}{1-\nu}} \cdot \left(\frac{\alpha}{r}\right)^{\frac{1-\nu(1-\alpha)}{1-\nu}} \cdot \left(\frac{1-\alpha}{w}\right)^{\frac{\nu(1-\alpha)}{1-\nu}} \cdot z^{\frac{\nu\alpha}{1-\nu}} \tag{23}$$

Proof. See Appendix A.3.

If wealth a is strictly greater than the threshold level (i.e., $a > \mathcal{A}^*(z)$), then the unconstrained capital demand level associated with the efficient scale of production lies within the compact set $[0, \lambda \cdot a]$, and hence the collateral constraint is slack. Otherwise, if wealth a is weakly less than the threshold level (i.e., $a \leq \mathcal{A}^*(z)$), then the unconstrained capital demand level associated with the efficient scale of production lies outside the compact set $[0, \lambda \cdot a]$, and hence the collateral constraint binds. Consequently, the threshold wealth level $\mathcal{A}^*(z)$ presents a sufficient statistic that determines whether the collateral constraint binds. Moreover, given the capital demand policy function $k^*(a,z)$, the labor demand policy function $n^*(a,z)$ is always unconstrained (i.e., the marginal product of labor equals the marginal cost of labor given by the unit wage rate w).

4.1.3 Complementarity Between Wealth and Productivity

Substituting the labor demand policy function (21) and capital demand policy function (22) into the microeconomic production function (19) yields the ex-post output function (24).

$$y^*(a,z) = \begin{cases} \left[\frac{\nu(1-\alpha)}{w}\right]^{\frac{\nu(1-\alpha)}{1-\nu(1-\alpha)}} \cdot \lambda^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \cdot (a \cdot z)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} & a < \mathcal{A}^*(z) \\ \nu^{\frac{\nu}{1-\nu}} \cdot \left(\frac{\alpha}{r}\right)^{\frac{\nu\alpha}{1-\nu}} \cdot \left(\frac{1-\alpha}{w}\right)^{\frac{\nu(1-\alpha)}{1-\nu}} \cdot z^{\frac{\nu\alpha}{1-\nu}} & a \geq \mathcal{A}^*(z) \end{cases} \tag{24}$$

The ex-post output function details the microeconomic output produced by a private business owner with wealth a and productivity z, conditional on static profit maximization. Crucially, ex-post output exhibits an endogenous complementarity between wealth a and productivity z if the collateral constraint binds (i.e., $a < \mathcal{A}^*(z)$).[12] If the collateral constraint binds, then additional wealth enables more capital to be employed in production, thereby increasing output. Moreover, as the marginal product of capital is increasing in productivity z, the marginal product of wealth is also increasing in productivity z. From an efficiency perspective, then, it is optimal to allocate the high productivity businesses to wealthy owners and the low productivity businesses to less wealthy owners.

4.2 Dynamic Optimization

The ex-post profit function $\pi^*(a,z)$ is the endogenous outcome from the static profit maximization problem that serves as an input into the dynamic optimization problem of consumption-saving. The profit function $\pi^*(a,z)$ measures the flow profit that a private business owner with wealth a

12 The complementarity between wealth and productivity arises endogenously because the ex-post output function is an outcome of the solution to a static profit maximization problem. Formally, the complementarity exists because the cross-partial derivative of ex-post output with respect to wealth and productivity is positive if the collateral constraint binds, i.e., $\frac{\partial y^*(a,z)}{\partial a \partial z} > 0$.

and business productivity z earns in an instant of time. Consequently, the profit function $\pi^*(a, z)$ captures an important source of flow income that affects the consumption-saving decision.

The dynamic optimization problem in this paper shares many common features with the standard heterogeneous agent model cast in continuous time (Achdou et al., 2021), such as incomplete markets, an endogenous consumption-saving decision, and exogenous transitions in business productivity. However, this paper extends the standard incomplete markets model by incorporating a frictional over-the-counter market in which heterogenous private business owners encounter bilateral opportunities to exchange their business productivity levels with one another.

4.2.1 Frictional Over-the-Counter Market

Over-the-Counter Market. I extend the model of Duffie et al. (2005), featuring a frictional over-the-counter market with undirected search, to incorporate a marketplace in which private business owners can exchange their productivity levels bilaterally with one another. This market is frictional because owners must wait some strictly positive amount of time before encountering a bilateral exchange opportunity. Specifically, at a Poisson intensity rate $\eta > 0$, a private business owner encounters a bilateral exchange opportunity with another private business owner in the economy.[13]

Bilateral Exchanges. Business productivity $z \in \{z_\ell, z_h\}$ is binary and uniquely determines whether a private business owner is a prospective seller or a prospective buyer. Sellers are private business owners with high productivity z_h pre-exchange and have the option of downgrading in productivity to z_ℓ in exchange for receiving a lump-sum payment from buyers. Meanwhile, buyers

13 Formally, the time until the next bilateral exchange opportunity is an exponentially distributed random variable with rate parameter $\eta > 0$. The mean waiting time is $1/\eta$, so that large value for the rate parameter η implies less severe search frictions.

are private business owners with low productivity z_ℓ pre-exchange and have the option to upgrade in productivity to z_h in exchange for paying a lump-sum payment to sellers.

Capital Gains Taxation. Crucially, sellers incur a capital gains tax upon realizing a capital gain under a bilateral exchange. Specifically, a seller realizes a capital gain if the lump-sum payment received from the buyer is greater than the seller's cost-basis. Consequently, the seller's cost-basis is an important state variable, of which one must keep track in order to determine whether a capital gain has been realized or not endogenously.

Like-Kind Exchanges. The market structure in which private business owners encounter bilateral exchange opportunities on a frictional over-the-counter market most closely resembles "like-kind" exchanges under Section 1031 of the Internal Revenue Code ("1031 exchange"). A 1031 exchange provides a means of deferring capital gains taxation on the sale of an asset by simultaneously acquiring another asset of a similar nature using the sale proceeds. The leading example of 1031 exchanges is real estate development. Under a 1031 exchange, an investor in commercial real estate can partially (or entirely) defer capital gains taxation by purchasing another investment property simultaneously upon selling an investment property. Recently, the Tax Cuts and Jobs Act of 2018 restricted like-kind exchanges to real property held for use in a trade or investment.

4.2.2 Hamilton-Jacobi-Bellman Equation

The Hamilton-Jacobi-Bellman ("HJB") provides a recursive formula of the dynamic optimization problem. The HJB equation of the quantitative model differs from the HJB equation (9) of the stylized model in two respects. First, the flow utility function over consumption admits a Constant Relative Risk Aversion ("CRRA") functional form, rather than a logarithmic functional form, regulated by the relative risk aversion parameter $\sigma > 0$. Second, the HJB equation of the quantitative model accounts for the additional value of exchanging business productivity levels on

the frictional OTC market. To properly account for this value of exchange on the OTC market, it is useful to present the HJB equation separately for sellers and buyers.

Seller's HJB Equation. A seller is a private business owner whose current productivity level is high (i.e., $z = z_h$). Furthermore, a seller is indexed by two additional state variables: wealth a and cost-basis P_s. Wealth is a state variable for the usual reason of serving as the stock from which one can consume. In this paper, the seller's cost-basis P_s is an additional state variable because it determines the extent to which an owner can claim deductions from the capital gains tax liability upon participating in a bilateral exchange.

Equation (25) details the HJB equation of a seller with wealth a and cost-basis P_s. The left-hand side denotes the annuity value $\rho \cdot V_s(a, P_s; \tau)$ of the seller state with wealth a and cost-basis P_s under a tax regime with CGT rate τ. The right-hand side decomposes the annuity value into its four constituent parts.

$$
\begin{aligned}
&\rho \cdot V_s(a, P_s; \tau) \qquad (25)\\
&= \max_{c>0} \Bigg\{ \frac{c^{1-\sigma} - 1}{1-\sigma} + [r \cdot a + \pi^*(a, z_h) + T(\tau) - c] \cdot \frac{\partial V_s(a, P_s; \tau)}{\partial a}\\
&+ \gamma_{h\ell} \cdot [V_b(a; \tau) - V_s(a, P_s; \tau)]\\
&+ \eta \cdot \int_0^{\infty} D^*(a, P_s, a'; \tau)\\
&\qquad \cdot [V_b(a + P^*(a, P_s, a'; \tau) - \tau \cdot \max\{P^*(a, P_s, a'; \tau) - P_s, 0\}; \tau)\\
&\qquad - V_s(a, P_s; \tau)] \cdot g_b(a'; \tau) \cdot da' \Bigg\}
\end{aligned}
$$

The first term on the right-hand side denotes the flow utility value from consumption. The second term on the right-hand side denotes the change in value from endogenous changes in the wealth

stock a due to the consumption-saving decision. The total flow income comprises three parts: (i) interest income on wealth, $r \cdot a$; (ii) business income from operating a high productivity business, $\pi^*(a, z_h)$; and (iii) a lump-sum transfer $T(\tau)$ from the government that is financed using the tax revenues from capital gains taxation in equilibrium. The third term on the right-hand side denotes the change in value from exogenous transitions in business productivity z according to the two-state Poisson process. At a Poisson rate $\gamma_{h\ell} > 0$, a seller with high productivity z_h exogenously downgrades to low productivity z_ℓ. In this instant of time, the seller's wealth a is unchanged. Since buyers are the private business owners with low productivity z_ℓ, the seller transitions to the buyer state with wealth a and associated value $V_b(a; \tau)$.

The fourth term on the right-hand side denotes the change in value from exchanging business productivity levels on the OTC market and deserved particular attention. At a Poisson rate $\eta > 0$, a seller receives a bilateral exchange opportunity. Upon receiving a bilateral exchange opportunity, the seller randomly draws a counterparty from the distribution of buyers given by the endogenous density function $g_s(a'; \tau)$. Thus, $\eta \cdot g_b(a'; \tau)$ is the total rate at which a seller encounters a bilateral exchange opportunity with a buyer indexed by wealth a'. Let $D^*(a, P_s, a'; \tau) \in \{0,1\}$ denote the exchange policy function indicating whether a seller, indexed by wealth a and cost-basis P_s, and a buyer, indexed by wealth a', choose to exchange productivity levels. If a bilateral exchange occurs (i.e., $D^*(a, P_s, a'; \tau) = 1$), then let $P^*(a, P_s, a'; \tau) > 0$ denote the lump-sum payment that the buyer pays to the seller. In Section 4.3 below, I characterize the two endogenous functions $D^*(a, P_s, a'; \tau)$ and $P^*(a, P_s, a'; \tau)$ in more detail.

If a bilateral exchange occurs, then a seller downgrades in productivity from z_h to z_ℓ, and receives a lump-sum payment $P^*(a, P_s, a'; \tau)$, minus a capital gains tax liability, in return. If the exchange asset price is strictly greater than the seller's cost-basis (i.e., $P^*(a, P_s, a'; \tau) > P_s$), then

a seller is liable to pay a capital gains tax equal to $\tau \cdot [P^*(a, P_s, a'; \tau) - P_s]$, where $\tau \in [0, 1)$ denotes the capital gains tax rate. Otherwise, if the exchange asset price is weakly less than the cost-basis, and so the seller realizes a capital loss upon a bilateral exchange, then the seller is not liable to pay capital gains taxes. The fourth term integrates over all potential bilateral exchange opportunities with buyers of different wealth levels a' and evaluates the change in the seller's value function from a bilateral exchange.

The Role of Cost-Basis. Since the cost-basis P_s is an additional state variable for the seller, a natural question is: how does the seller's value function $V_s(a, P_s; \tau)$ vary with the cost-basis P_s? Figure 23 in Appendix C.1 supplies the answer. This figure plots the difference in the seller's value function evaluated at an arbitrarily large cost-basis and the seller's value function evaluated at an arbitrarily small cost-basis as a function of the seller's wealth a for different CGT rates τ.

There are three takeaways. The first takeaway is that if the CGT rate is zero $\tau = 0$, then the seller's value function does not vary with the cost-basis P_s. The intuition is that the cost-basis only affects the seller's HJB equation (25) if the CGT rate is strictly positive (i.e., $\tau > 0$), as this is the instance in which tax deductions using the cost-basis are meaningful. The second takeaway is that the difference in the value between a large and a small cost-basis is increasing in the CGT rate τ. This is also intuitive because, all else equal, the larger the CGT rate, the greater is the tax deduction that a larger cost-basis affords. The third takeaway is that, for a fixed CGT rate τ, the value of a larger cost-basis is greater at lower wealth levels. This is also intuitive because additional tax deductions from a large cost-basis translate into more wealth, resulting in additional flow consumption. In turn, the marginal value of additional flow consumption is greater when wealth is low.

Buyer's HJB Equation. A buyer is a private business owner whose pre-exchange productivity level is low (i.e., $z = z_\ell$). A buyer's single state variable is wealth a. In contrast to the seller, the buyer does not have a cost-basis as an additional state variable. The reason is due to the assumption of binary business productivity $z \in \{z_\ell, z_h\}$ and bilateral exchanges. This assumption implies that it is not possible to pay a lump-sum payment to acquire a business with the lowest productivity level (i.e., $z = z_\ell$) because, by assumption, there does not exist a lower productivity level to which one can downgrade. In a model with more than two business productivity levels, prospective buyers without the lowest productivity level would have a non-degenerate cost-basis. Hence, the assumption of binary business productivity and bilateral exchanges reduces the dimensionality of the state space across sellers and buyers by eliminating the need to track the cost-basis for buyers.

Equation (26) details the HJB equation of a buyer with wealth a. As usual, the left-hand side denotes the flow annuity value $\rho \cdot V_b(a;\tau)$ of the buyer state with wealth a and under a tax regime with CGT rate τ. The right-hand side decomposes the flow annuity value into its four constituent parts.

$$
\begin{aligned}
&\rho \cdot V_b(a;\tau) \qquad (26)\\
&= \max_{c>0}\Bigg\{ \frac{c^{1-\sigma}-1}{1-\sigma} + [r \cdot a + \pi(a, z_\ell) + T(\tau) - c] \cdot \frac{\partial V_b(a;\tau)}{\partial a}\\
&+\gamma_{\ell h} \cdot [V_s(a, 0;\tau) - V_b(a;\tau)]\\
&+\eta \cdot \int_0^\infty \int_0^\infty D^*(a', P_s, a;\tau) \cdot [V_s(a - P^*(a', P_s, a;\tau), P^*(a', P_s, a;\tau);\tau) - V_b(a;\tau)]\\
&\qquad\qquad \cdot g_s(a', P_s;\tau) \cdot da \cdot dP_s \Bigg\}
\end{aligned}
$$

The first term on the right-hand side denotes the flow utility from consumption. The second term on the right-hand side denotes the flow value from endogenous changes in wealth a due to

consumption-saving. Crucially, note that the buyer's flow profit income from operating a private business, namely $\pi^*(a, z_\ell)$, is indexed by the low productivity level z_ℓ. The third term on the right-hand side denotes the flow value from an exogenous change in business productivity. At a Poisson rate $\gamma_{\ell h}$, a buyer transitions from low productivity z_ℓ to high productivity z_h exogenously, thereby transitioning to the seller state. In this instant of time, the formerly buyer's wealth a is unchanged and the new cost-basis in the seller state is zero because the formerly buyer did not pay any dollars to acquire the high productivity level, but rather acquired the high productivity level exogenously (e.g., due to an innovation breakthrough).

The fourth term on the right-hand side denotes the flow value from endogenous changes in both wealth and productivity from bilateral exchanges on the frictional OTC market. At a Poisson rate $\eta > 0$, a buyer receives a bilateral exchange opportunity. Upon receiving this bilateral exchange opportunity, the buyer randomly draws a counterparty from the endogenous distribution of sellers, given by the density function $g_s(a', P_s; \tau)$. Thus, $\eta \cdot g_s(a', P_s; \tau)$ denotes the total rate at which a buyer encounters a seller with wealth a' and cost-basis P_s for a bilateral exchange opportunity. If a bilateral exchange between a buyer with wealth a and a seller with wealth a' and cost-basis P_s occurs, i.e., $D^*(a', P_s, a; \tau) = 1$, then the buyer transitions to the seller state with a productivity level z_h. In the seller state, the formerly buyer's wealth is $[a - P^*(a', P_s, a; \tau)]$, namely the old wealth value a minus the lump-sum payment $P^*(a', P_s, a; \tau) > 0$ paid to the buyer. Importantly, the new cost-basis in the seller state is equal to the lump-sum payment $P^*(a', P_s, a; \tau)$. This is the sense in which the cost-basis updates endogenously as buyers acquire the high productivity business on the OTC market. The fourth term integrates the expected change in the value of the buyer across all seller types (a', P_s) with which the buyer can perform a bilateral exchange.

Dynamic First Order Condition. The first order consumption of the dynamic problem with respect to consumption $c > 0$ is the standard one for a heterogeneous agent model cast in continuous time. Equations (27) and (28) present the first order conditions for sellers and buyers, respectively.

$$c_s^*(a, P_s; \tau)^{-\sigma} = \frac{\partial V_s(a, P_s; \tau)}{\partial a} \tag{27}$$

$$c_b^*(a; \tau)^{-\sigma} = \frac{\partial V_b(a; \tau)}{\partial a} \tag{28}$$

Intuitively, the first order condition for consumption equates the marginal benefit of consuming an additional unit of wealth, given by the marginal utility value on the left-hand side, to the marginal cost of consuming an additional unit of wealth, given by the marginal loss in value from a lower wealth stock on the right-hand side. Given knowledge of the value functions $V_s(a, P_s; \tau)$ and $V_b(a; \tau)$, one can solve for the flow consumption policy functions $c_s^*(a, P_s; \tau)$ and $c_b^*(a; \tau)$ in closed-form using equations (27) and (28).

Saving Policy Function. Let $\dot{a}_s^*(a, P_s; \tau)$ and $\dot{a}_b^*(a; \tau)$ denote the flow saving policy functions of sellers and buyers, respectively. Flow saving is defined as the difference between total flow income and flow consumption.

$$\dot{a}_s^*(a, P_s; \tau) := r \cdot a + \pi(a, z_h) + T(\tau) - c_s^*(a, P_s; \tau) \tag{29}$$

$$\dot{a}_b^*(a; \tau) := r \cdot a + \pi(a, z_\ell) + T(\tau) - c_b^*(a; \tau) \tag{30}$$

Given knowledge of the flow consumption policy functions $c_s^*(a, P_s; \tau)$ and $c_b^*(a; \tau)$, one can solve for the flow saving policy functions using equations (29) and (30).

4.2.3 Kolmogorov-Forward Equation

The Kolmogorov-Forward ("KF") equation governs the dynamics of the joint distribution function for state variables over time. The seller's state variables comprise wealth a and cost basis P_s. The buyer's single state variable is wealth a. I detail the KF equation separately for sellers and buyers.

Seller's KF Equation. Equation (31) details the KF equation for a seller with wealth a and cost-basis P_s in the stationary equilibrium. The left-hand side of the equation equals zero because the KF equation is evaluated at the stationary equilibrium. The right-hand side of the equation details the various ways in which the seller's joint distribution function, denoted by $g_s(a, P_s; \tau)$, can evolve.

$$
\begin{aligned}
0 = &-\frac{\partial[\dot{a}_s^*(a, P_s; \tau) \cdot g_s(a, P_s; \tau)]}{\partial a} + \gamma_{\ell h} \cdot \mathbf{1}_{P_s=0} \cdot g_b(a; \tau) - \gamma_{h\ell} \cdot g_s(a, P_s; \tau) \\
&+\eta \cdot \int_0^\infty \int_0^\infty \int_0^\infty D^*(a'', P_s'', a'; \tau) \cdot \mathbf{1}_{a'-P^*=a \cap P^*=P_s} \cdot g_s(a'', P_s''; \tau) \cdot g_b(a'; \tau) \cdot da'' \\
&\qquad \cdot dP_s'' \cdot da' \\
&-\eta \cdot \int_0^\infty D^*(a, P_s, a'; \tau) \cdot g_s(a, P_s; \tau) \cdot g_b(a'; \tau) \cdot da'
\end{aligned}
\tag{31}
$$

The first term on the right-hand side denotes the evolution of the seller's distribution function due to the endogenous consumption-saving decision. The second and third terms on the right-hand side denotes the evolution of the seller's joint distribution function due to exogenous transitions in business productivity z according to the two-state Poisson process. Specifically, at a Poisson intensity rate $\gamma_{\ell h}$, a buyer with wealth a upgrades in productivity from z_ℓ to z_h, transitioning into

a seller with wealth a and zero cost-basis $P_s = 0$.[14] This term captures the inflow from the buyer state with wealth a into the seller state with wealth a and cost-basis $P_s = 0$. Similarly, at a Poisson intensity rate $\gamma_{h\ell}$, a seller with wealth a and cost-basis P_s downgrades in productivity from z_h to z_ℓ, transition into a buyer with wealth a. This term captures the outflow from the seller state with wealth a and cost-basis P_s into the buyer state with wealth a.

The fourth and fifth terms on the right-hand side pertain denote the evolution of the seller's joint distribution function due to bilateral exchanges on the frictional OTC market. The fourth term captures the inflow from the buyer state to the seller state. Intuitively, at the intensity rate $\eta \cdot g_s(a'', P_s''; \tau) \cdot g_b(a'; \tau)$, there is a bilateral exchange opportunity between a seller with state (a'', P_s'') and a buyer with wealth a'. If an exchange occurs (i.e., $D^*(a'', P_s'', a'; \tau) = 1$), and if the exchange asset price $P^*(a'', P_s'', a'; \tau)$ is such that the buyer's new wealth level in the seller state is a and the exchange asset price equals the cost-basis P_s, then there is an inflow into the seller state (a, P_s) at rate $\eta \cdot g_s(a'', P_s''; \tau) \cdot g_b(a'; \tau)$. The fourth term captures such inflows by integrating over all possible bilateral exchange opportunities given by the triplet (a'', P_s'', a'). Lastly, the fifth term captures the outflow from the seller state into the buyer state. Intuitively, at the intensity rate $\eta \cdot g_s(a, P_s; \tau) \cdot g_b(a'; \tau)$, there is a bilateral exchange opportunity between a seller with state (a, P_s) and a buyer with wealth a'. If an exchange occurs (i.e., $D^*(a, P_s, a'; \tau) = 1$), then the seller departs the seller state (a, P_s) at the intensity rate $\eta \cdot g_s(a, P_s; \tau) \cdot g_b(a'; \tau)$. The fifth term captures such outflows by integrating over all buyer wealth levels a' with which the seller can engage in a bilateral exchange.

14 The reason the buyer transitions into the seller state with a zero cost-basis (i.e., $P_s = 0$) is that buyers do not pay a lump-sum payment in order to upgrade in business productivity when such a transition occurs exogenously due to the two-state Poisson process.

Buyer's KF Equation. Equation (32) details the KF equation for a buyer with wealth a in the stationary equilibrium. Once more, the left-hand side of the equation equals zero because the KF equation is evaluated at the stationary equilibrium. The right-hand side of the equation details the various ways in which the buyer's joint distribution function $g_b(a;\tau)$ can evolve.

$$0 = -\frac{\partial[\dot{a}_b^*(a;\tau)\cdot g_b(a;\tau)]}{\partial a} + \gamma_{h\ell}\cdot\int_0^\infty g_s(a,P_s;\tau)\cdot dP_s - \gamma_{\ell h}\cdot g_b(a;\tau) \tag{32}$$

$$+\eta\cdot\int_0^\infty\int_0^\infty\int_0^\infty D^*(a'',P_s,a';\tau)\cdot\mathbf{1}_{a''+P^*-\tau\cdot\max\{P^*-P_s,0\}=a}\cdot g_s(a'',P_s;\tau)\cdot g_b(a';\tau)$$

$$\cdot da''\cdot dP_s\cdot da'$$

$$-\eta\cdot\int_0^\infty\int_0^\infty D^*(a',P_s,a;\tau)\cdot g_s(a',P_s;\tau)\cdot g_b(a;\tau)\cdot da'\cdot dP_s$$

The first term on the right-hand side denotes the evolution of the buyer's distribution function due to the endogenous consumption-saving decision. The second and third terms on the right-hand side denotes the evolution of the buyer's joint distribution function due to exogenous transitions in business productivity z according to the two-state Poisson process. Specifically, at a Poisson intensity rate $\gamma_{h\ell}$, a seller with wealth a and cost-basis P_s downgrades in productivity from z_h to z_ℓ, transitioning into a buyer with wealth a.[15] This term integrates over all seller cost-basis levels P_s to captures the inflow from the seller state with wealth a and cost-basis P_s into the buyer state with wealth a. Similarly, at a Poisson intensity rate $\gamma_{\ell h}$, a buyer with wealth a upgrades in productivity from z_ℓ to z_h, thereby transitioning into a seller with wealth a and zero cost-basis

15 Upon transitioning from high productivity z_h to low productivity z_ℓ, the seller's former cost-basis disappears. The reason is that, with binary business productivity $z \in \{z_\ell, z_h\}$, it is never possible to acquire the lowest productivity z_ℓ by paying a lump-sum payment. Moreover, buyers can never deduct a cost-basis from their capital gains taxes as buyers never "sell" their productivity level to another agent because buyers are indexed by the lowest productivity level z_ℓ.

$P_s = 0$. This term captures the outflow from the buyer state with wealth a into the seller state with wealth a and zero cost-basis $P_s = 0$.

The fourth and fifth terms on the right-hand side pertain denote the evolution of g_b due to bilateral exchanges on the OTC market. The fourth term captures the inflow from the seller state to the buyer state. Intuitively, at the intensity rate $\eta \cdot g_s(a'', P_s; \tau) \cdot g_b(a'; \tau)$, there is a bilateral exchange opportunity between a seller with state (a'', P_s) and a buyer with state a'. If an exchange occurs, i.e. $D^*(a'', P_s, a'; \tau) = 1$, and if the exchange asset price $P^*(a'', P_s, a'; \tau)$ is such that the seller's new wealth level in the buyer state is a, then there is an inflow into the buyer state a at rate $\eta \cdot g_s(a'', P_s; \tau) \cdot g_b(a'; \tau)$. The fourth term captures such inflows by integrating over all possible bilateral exchange opportunities. Lastly, the fifth term captures the outflow from the buyer state into the seller state. Intuitively, at the intensity rate $\eta \cdot g_s(a', P_s; \tau) \cdot g_b(a; \tau)$, there is a bilateral exchange opportunity between a seller with state (a', P_s) and a buyer with state a. If an exchange occurs, i.e. $D^*(a', P_s, a; \tau) = 1$, then the buyer departs the buyer state a, capturing an outflow at the rate $\eta \cdot g_s(a', P_s; \tau) \cdot g_b(a; \tau)$.

4.3 Bilateral Exchange Optimization

This subsection outlines the determination of the exchange policy function $D^*(a', P_s, a; \tau)$ and the exchange asset price function $P^*(a, P_s, a'; \tau)$ featuring in the HJB equations (25) and (26), taking the seller's value function $V_s(a, P_s; \tau)$ and the buyer's value function $V_b(a'; \tau)$ as given.

4.3.1 Reservation Asset Price Functions

To determine whether a bilateral exchange occurs, I characterize two reservation asset prices. The first is the minimum asset price the seller is willing to accept and the second is the maximum asset price the buyer is willing to pay.

Seller's Minimum Asset Price. The seller's indifference condition (33) characterizes the seller's minimum asset price $P^*_{min}(a, P_s; \tau)$ as a function of the seller's pre-exchange wealth a and cost basis P_s. The right-hand side of the equation denote the seller's value pre-exchange, while the left-hand side of the equation denotes the seller's value post-exchange, if the exchange were to occur at the seller's reservation asset price P^*_{min}. Post-exchange, the seller downgrades in productivity from z_h to z_ℓ, thereby transitioning to the buyer state.

$$\underbrace{V_b(a + P^*_{min} - \tau \cdot \max\{P^*_{min} - P_s, 0\}; \tau)}_{\textit{Post-Exchange}} = \underbrace{V_s(a, P_s; \tau)}_{\textit{Pre-Exchange}} \tag{33}$$

The seller's wealth post-exchange is equal to pre-exchange wealth a plus the exchange asset price P^*_{min} minus the capital gains tax liability. The capital gains tax is paid only if a capital gain is realized, i.e., the asset price P^*_{min} exceeds the seller's cost-basis P_s. Given the value functions (V_b, V_s), the seller's minimum asset price P^*_{min} adjusts endogenously so as to equate the pre-exchange and post-exchange values and hence satisfy the seller's indifference condition.

Buyer's Maximum Asset Price. The buyer's indifference condition (34) characterizes the buyer's maximum asset price $P^*_{max}(a; \tau)$ as a function of the buyer's pre-exchange wealth a. The left-hand side of the equation denotes the buyer's value pre-exchange, while the right-hand side of the equation denotes the buyer's value post-exchange, if the exchange were to occur at the buyer's reservation asset price P^*_{max}.

$$\underbrace{V_s(a - P^*_{max}, P^*_{max}; \tau)}_{\textit{Post-Exchange}} = \underbrace{V_b(a; \tau)}_{\textit{Pre-Exchange}} \tag{34}$$

Post-exchange, the buyer upgrades in productivity from z_ℓ to z_h, thereby transitioning to the seller state. The buyer's maximum asset price has two opposing effects on post-exchange value. On the one hand, an increase in the buyer's maximum asset price results in lower post-exchange wealth,

thereby reducing value all else equal. On the other hand, an increase in the buyer's maximum asset price results in a higher post-exchange cost basis, thereby increasing value all else equal. Quantitatively, the first effect dominates the second effect, such that the buyer's post-exchange value is strictly decreasing in the buyer's maximum asset price.

4.3.2 Bilateral Exchange Policy Function

Given the two reservation asset price functions, a bilateral exchange between a seller type (a, P_s) and a buyer type a' occurs if and only if the buyer's maximum asset price exceeds the seller's minimum asset price. Otherwise, the bilateral exchange does not occur.

$$D^*(a, P_s, a'; \tau) := \begin{cases} 1 & P^*_{max}(a'; \tau) > P^*_{min}(a, P_s; \tau) \\ 0 & P^*_{max}(a'; \tau) \leq P^*_{min}(a, P_s; \tau) \end{cases} \tag{35}$$

The bilateral exchange policy function in equation (35) summarizes this rule.

4.3.3 Bilateral Exchange Asset Price

Suppose that $P^*_{max} > P^*_{min}$ so that a bilateral exchange occurs. The exchange asset price can lie anywhere between the seller's minimum asset price P^*_{min} and the buyer's maximum asset price P^*_{max}. To determine the exchange asset price uniquely, I assume that the exchange asset price $P^*(a, P_s, a'; \tau)$ in equation (36) is a weighted average of the two reservation prices, regulated by a bargaining parameter $\kappa \in (0,1)$.

$$P^*(a, P_s, a'; \tau) := \kappa \cdot P^*_{max}(a'; \tau) + (1 - \kappa) \cdot P^*_{min}(a, P_s; \tau) \tag{36}$$

Given that the market for private businesses is illiquid, decentralized, and lacking a single price schedule, it is natural presume that bilateral bargaining between sellers and buyers determines the exchange asset prices.

4.4 Government Budget Balance

The government rebates the tax revenues generated from bilateral exchanges on the OTC market back to private business owners as a flow lump-sum transfer $T(\tau)$. This flow lump-sum transfer must satisfy the government budget balance condition (37).

$$T(\tau) = \tau \cdot \eta \cdot \int_0^\infty \int_0^\infty \int_0^\infty D^*(a, P_s, a'; \tau) \cdot \max\{P^*(a, P_s, a'; \tau) - P_s, 0\} \cdot g_s(a, P_s; \tau) \cdot g_b(a'; \tau) \cdot da \cdot dP_s \cdot da' \tag{37}$$

The left-hand side denotes the use of tax revenues, namely a flow lump-sum transfer to private business owners. The right-hand side denotes the source of tax revenues, namely from capital gains taxation. The intuition for the aggregate tax revenues is as follows. The total contact rate for a bilateral exchange opportunity between a seller with wealth a and cost-basis P_s and a buyer with wealth a' is $\eta \cdot g_s(a, P_s; \tau) \cdot g_b(a'; \tau)$. Given this bilateral exchange opportunity, if an exchange occurs, i.e. $D^*(a, P_s, a'; \tau) = 1$, then the tax revenue generated is $\tau \cdot \max\{P^*(a, P_s, a'; \tau) - P_s, 0\}$. Thus, the government only collects tax revenue from a bilateral exchange if the seller realizes a capital gain, i.e., $P^*(a, P_s, a'; \tau) > P_s$. Determining aggregate tax revenues requires integrating expected tax revenues across all possible bilateral exchange opportunities.

4.5 Stationary Equilibrium

Definition 1: Stationary Equilibrium

Given a capital gains tax rate $\tau \in [0, 1)$, a stationary equilibrium comprises: (**1**) threshold wealth function $\mathcal{A}^*(z)$, labor demand policy function $n^*(a, z)$, capital demand policy function $k^*(a, z)$, ex-post output function $y^*(a, z)$, ex-post profit function $\pi^*(a, z)$; (**2**) seller's value function $V_s(a, P_s; \tau)$, buyer's value function $V_b(a; \tau)$; (**3**) seller's consumption policy function $c_s^*(a, P_s; \tau)$,

seller's saving policy function $\dot{a}_s^*(a, P_s; \tau)$, buyer's consumption policy function $c_b^*(a; \tau)$, buyer's saving policy function $\dot{a}_b^*(a; \tau)$; (**4**) seller's joint distribution function $g_s(a, P_s; \tau)$, buyer's joint distribution function $g_b(a; \tau)$; (**5**) seller's minimum asset price function $P_{min}^*(a, P_s; \tau)$, buyer's maximum asset price function $P_{max}^*(a; \tau)$, exchange policy function $D^*(a, P_s, a'; \tau)$, exchange asset price function $P^*(a, P_s, a'; \tau)$; (**6**) lump-sum transfer $T(\tau)$ such that:

1. $\{\mathcal{A}^*, n^*, k^*, y^*, \pi^*\}$ satisfy static optimization: (20), (21), (22), (23), (24).
2. $\{V_s, V_b\}$ satisfy Hamilton-Jacobi-Bellman equations: (25), (26).
3. $\{c_s^*, \dot{a}_s^*, c_b^*, \dot{a}_b^*\}$ satisfy dynamic optimization: (27), (28), (29), (30).
4. $\{g_s, g_b\}$ satisfy Kolmogorov-Forward equations: (31), (32).
5. $\{P_{min}^*, P_{max}^*, D^*, P^*\}$ satisfy bilateral exchange optimization: (33), (34), (35), (36).
6. T satisfies government budget balance constraint: (37).

The definition of the stationary equilibrium concludes the outline of the quantitative model.

5. Theoretical Analysis

This section performs a theoretical analysis of the stationary equilibrium for the quantitative model. The theoretical analysis comprises two parts. First, I analyze theoretical predictions of a change in the CGT rate τ on various endogenous outcomes, such as bilateral exchange asset prices and allocations. Second, I perform an aggregation exercise in which I prove the existence of an aggregate production function in which the aggregate productivity level is endogenous and depends on the allocation of private business productivity levels across the wealth distribution of private business owners.

5.1 Theoretical Predictions of the Capital Gains Tax

This subsection characterizes two tax elasticities of capital gains. The first is the elasticity of the seller's minimum asset price P^*_{min} with respect to the CGT rate τ. The second is the elasticity of the exchange asset price P^* with respect to the CGT rate τ.

Direct Versus Indirect Effects. A change in the CGT rate τ has both direct and indirect effects on endogenous outcomes. I define the direct effects as those stemming from holding the value functions (V_s, V_b) and joint distribution functions (g_S, g_b) fixed at those pertaining to the stationary equilibrium under the prior tax regime. I define the indirect effects as those stemming from allowing the value function and joint distribution function to endogenously adjust to the stationary equilibrium under the new tax regime. I discuss the distinction between the direct and indirect effects of the CGT in more detail below.

5.1.1 Tax Elasticities of Capital Gains

Lock-In Effect Elasticity. A traditional effect of increasing the capital gains tax is the "lock-in" effect whereby, all else equal, the seller chooses to hold the asset rather than sell the asset to a

buyer (Dai et al., 2008). The lock-in effect manifests in the quantitative model via the seller's indifference condition (33). Holding the value functions (V_s, V_b) fixed to capture the direct effects of the CGT rate, an increase in the CGT rate τ implies that the seller's minimum asset price P^*_{min} must increase so as to satisfy the seller's indifference condition, albeit only if a capital gain is realized (i.e., $P^*_{min} > P_s$). The following proposition characterizes the elasticity summarizing the increase in the seller's minimum asset price to an increase in the CGT rate τ.

Proposition 2: Elasticity of Seller's Minimum Asset Price with respect to CGT Rate

The elasticity of the seller's minimum asset price P^*_{min} with respect to the CGT rate τ is

$$\frac{\tau}{P^*_{min}}\frac{\partial P^*_{min}}{\partial \tau} = \begin{cases} \frac{\tau}{1-\tau} \cdot \frac{P^*_{min} - P_s}{P^*_{min}} & P^*_{min} > P_s \\ 0 & P^*_{min} \leq P_s \end{cases} \tag{38}$$

Proof. See Appendix A.4.

I label the tax elasticity of capital gains in equation (38) as the "lock-in effect elasticity." Figure 5 and Figure 6 provide an illustration of the direct lock-in effect captured by the lock-in effect elasticity in equation (38).

Considering a small cost-basis, Figure 5 plots the seller's minimum asset price function $P^*_{min}(a, P_s; \tau)$, under two arbitrary CGT rates τ_ℓ and $\tau_h > \tau_\ell$, as a function of wealth a. For all minimum asset prices above the cost-basis price P_s, there is an upward shift in the seller's minimum asset price schedule. That is, for a given level of the seller's pre-exchange wealth a, the seller demands a higher minimum asset price in order to accept a bilateral exchange. The intuition is that the seller's tax burden has increased upon an increase in the CGT rate, such that the seller

receives fewer dollars upon an exchange. To ensure the seller is indifferent between exchanging and not exchanging, the minimum asset price must increase to offset the additional tax burden.

Figure 5: Lock-In Effect (Small Cost-Basis)

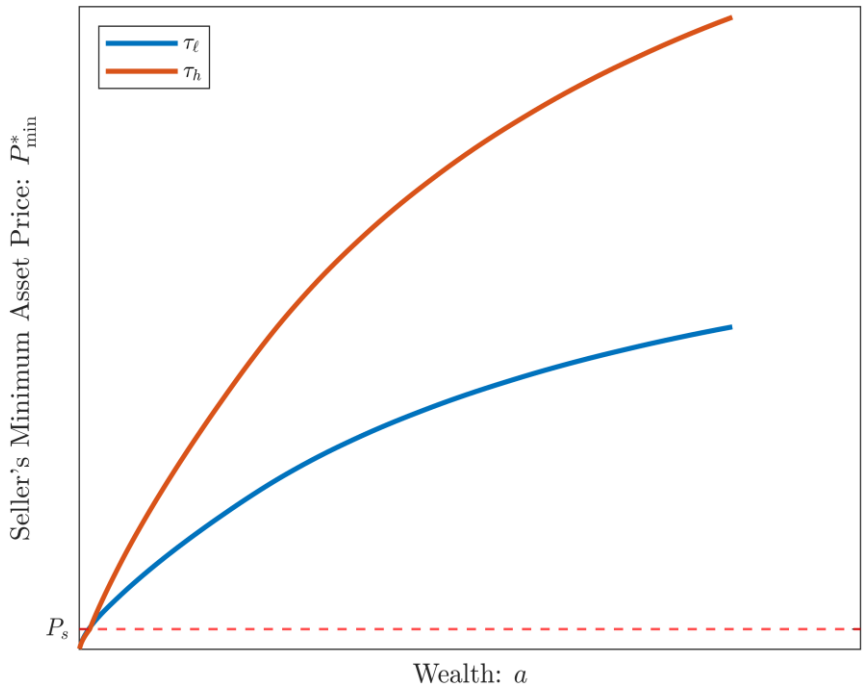
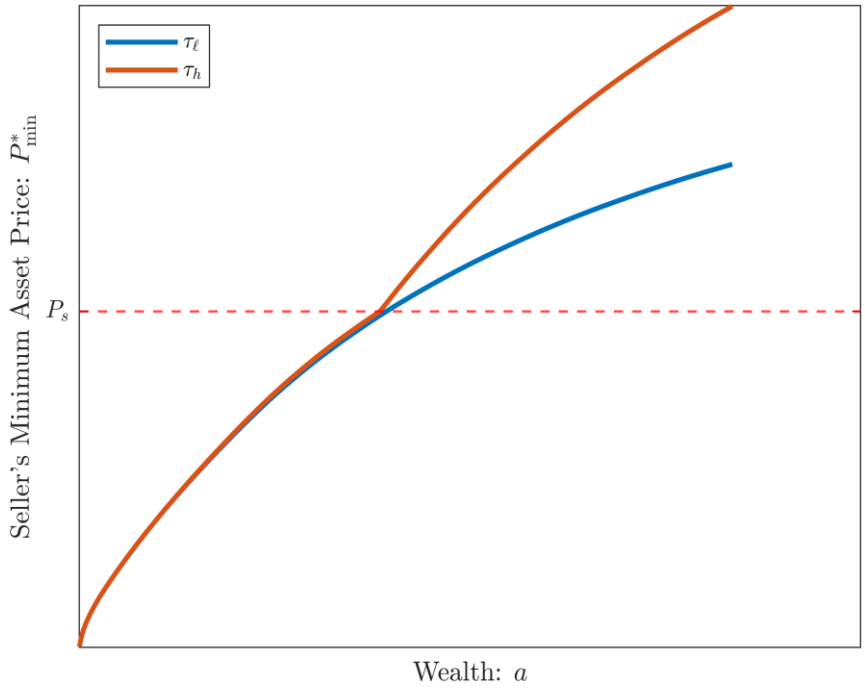


Figure 6: Lock-In Effect (Large Cost-Basis)



Consider a large cost-basis, Figure 6 plots the seller's minimum asset price function $P^*_{min}(a, P_s; \tau)$, under two arbitrary CGT rates τ_ℓ and $\tau_h > \tau_\ell$, as a function of wealth a. Since the cost-basis is large, the seller only realizes a capital gain at minimum asset prices that are sufficiently high. Consequently, the upward shift in the seller's minimum asset price function occurs at minimum asset prices above the seller's cost-basis P_s. The intuition is that the direct effect of the CGT on the seller's minimum asset price is meaningful only if the seller realizes a capital gain. A larger cost-basis enables greater tax deductions, and so reduces the scope to realize capital gains.

Pass-Through Elasticity. The lock-in effect induces an upward shift in the seller's minimum asset price function $P^*_{min}(a, P_s; \tau)$ in response to an increase in the CGT rate τ. In turn, an increase in the seller's minimum asset price passes through to an increase in the exchange asset price $P^*(a, P_s, a'; \tau)$ by equation (36). The following proposition characterizes the pass-through elasticity of a change in the CGT rate τ to a change in the exchange asset price.

Proposition 3: Elasticity of Exchange Asset Price with respect to CGT Rate

The elasticity of the exchange asset price P^* with respect to the CGT rate τ is

$$\frac{\tau}{P^*}\frac{\partial P^*}{\partial \tau} = \begin{cases} (1-\kappa)\cdot\dfrac{\tau}{1-\tau}\cdot\dfrac{P^*_{min}-P_s}{P^*} & P^* > P_s \\ 0 & P^* \leq P_s \end{cases} \tag{39}$$

Proof. See Appendix A.5.

The exchange asset price adjusts to the direct lock-in effect of an increase in the CGT rate if and only if a capital gain is realized (i.e., $P^* > P_s$). Furthermore, the pass-through rate of an increase in the seller's minimum asset price P^*_{min} to an increase in the exchange asset price P^* depends on

the bargaining power parameter κ. The larger is the value of κ, the stronger is the bargaining power of the seller, and hence the greater is the weight attached to the buyer's maximum asset price in determining the exchange price by equation (36). In this scenario, a change in the seller's minimum asset price has a smaller pass-through rate to a change in the exchange asset price. By contrast, the pass-through rate is greater if the buyer has more bargaining power (i.e., κ is small).

5.1.2 Distortion on Bilateral Exchange Allocations

All else equal, this paper establishes that an increase in the CGT rate τ induces an increase in the seller's minimum asset price P^*_{min}. This lock-in effect, in turn, induces a distortion on bilateral exchange allocations. There is a subset of bilateral exchanges for which, prior to the increase in the CGT rate τ, the buyer's maximum asset price exceeds the seller's minimum asset price, while after the increase in the CGT rate τ, the seller's minimum asset price exceeds the buyer's maximum asset price. Such bilateral exchanges occur before the increase in the CGT rate τ, but do not occur afterwards, resulting in a bilateral exchange allocation distortion.

Allocation Distortions. A bilateral exchange on the frictional OTC market comprises a seller type (a, P_s), namely the seller's wealth a and cost-basis P_s, and a buyer type a', namely the buyer's wealth a'. Hence, the triplet (a, P_s, a') uniquely characterizes a bilateral exchange. Figure 7 plots the bilateral exchange policy function (35) in seller-buyer wealth space (a, a'), fixing the seller's cost-basis P_s to an arbitrarily small value. As predicted by the stylized model of Section 3, the bilateral exchanges that occur in equilibrium are those between more wealthy buyers and less wealthy sellers. Relative to less wealthy sellers, more wealthy buyers are better positioned to overcome the collateral constraint (2), and therefore better equipped to operate the high productivity z_h business closer to its efficient scale. Hence, the top-left region denotes the set of

bilateral exchanges that do not occur, while the bottom-right region denotes the set of bilateral exchanges that do occur.

The interior region denotes the set of bilateral exchanges that do occur prior to an increase in the CGT rate τ, but that do not occur after an increase in the CGT rate τ. To build intuition, fix a seller's wealth level a on the vertical axis, and read horizontally from left to right. A point in the interior region includes a buyer's pre-exchange wealth level a' such that, if the seller were to meet this buyer prior to the increase in the CGT rate τ, then the two parties would agree to the bilateral exchange, as the buyer's maximum asset price exceeds the seller's minimum asset price. However, after the increase in the CGT rate τ, the seller's minimum asset price increases due to the lock-in effect and exceeds the buyer's maximum asset price for points in the interior region. Consequently, the seller would reject a bilateral exchange with buyers in the interior region after the increase in the CGT rate τ. Rather, the seller accepts bilateral exchange only with the wealthier buyers to the right of the wealth space, as such buyers exhibit a higher maximum asset price that exceeds the seller's minimum asset price after the increase in the CGT rate τ.

The Role of Cost-Basis. Furthermore, the distortion induced on bilateral exchange allocations from an increase in the CGT rate τ depends crucially on the seller's cost-basis P_s. In particular, if the seller's cost-basis is small, then the distortions are greater, all else equal. The intuition is that a low cost-basis implies smaller tax deductions from capital gains taxes when an exchange occurs. In turn, a smaller tax deduction results in a greater increase in the seller's minimum asset price P^*_{min} in response to an increase in the CGT rate τ, as detailed by the elasticity in equation (38). Meanwhile, if the seller's cost-basis is large, then the seller can claim a larger tax deduction upon realizing a capital gain, inducing a smaller increase in the seller's minimum asset price P^*_{min} in response to an increase in the CGT rate τ. Figure 8 plots the bilateral exchange policy function

(35) in seller-buyer wealth space (a, a'), fixing the seller's cost-basis P_s to an arbitrarily large value. Unlike the interior region characterizing distortions in Figure 7, the interior region characterizing distortions in Figure 8 is notably smaller. This reflects the role of a larger cost-basis in offsetting the effect of an increase in the CGT rate on the seller's minimum asset price. That said, the cost-basis itself is endogenous to the CGT rate τ because the cost-basis is determined by the bilateral exchange asset prices (36) that buyers pay to the sellers in equilibrium.

Figure 7: Bilateral Exchange Allocation Heatmap (Small Cost-Basis)

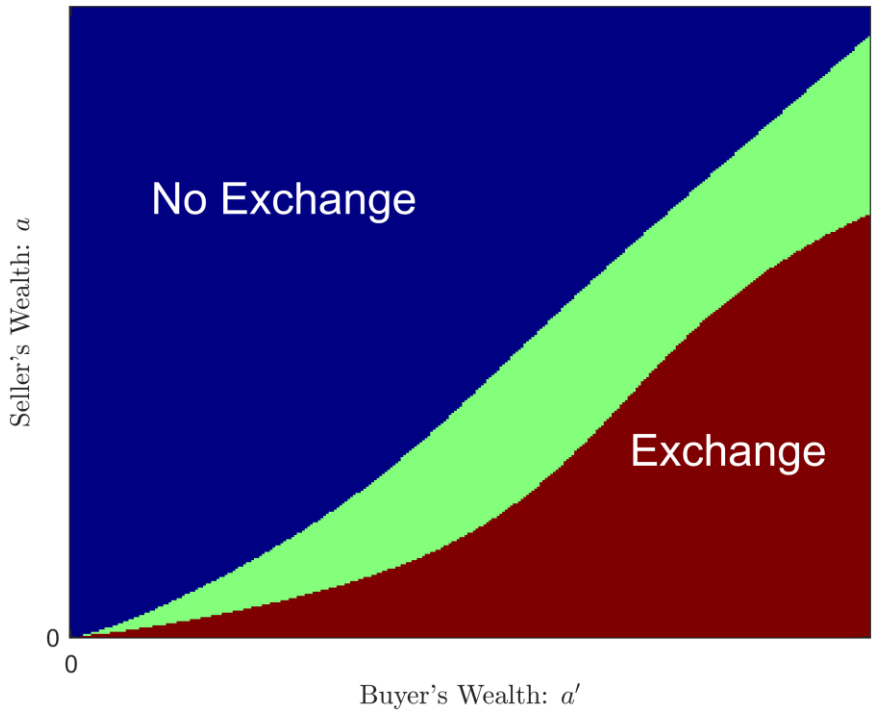
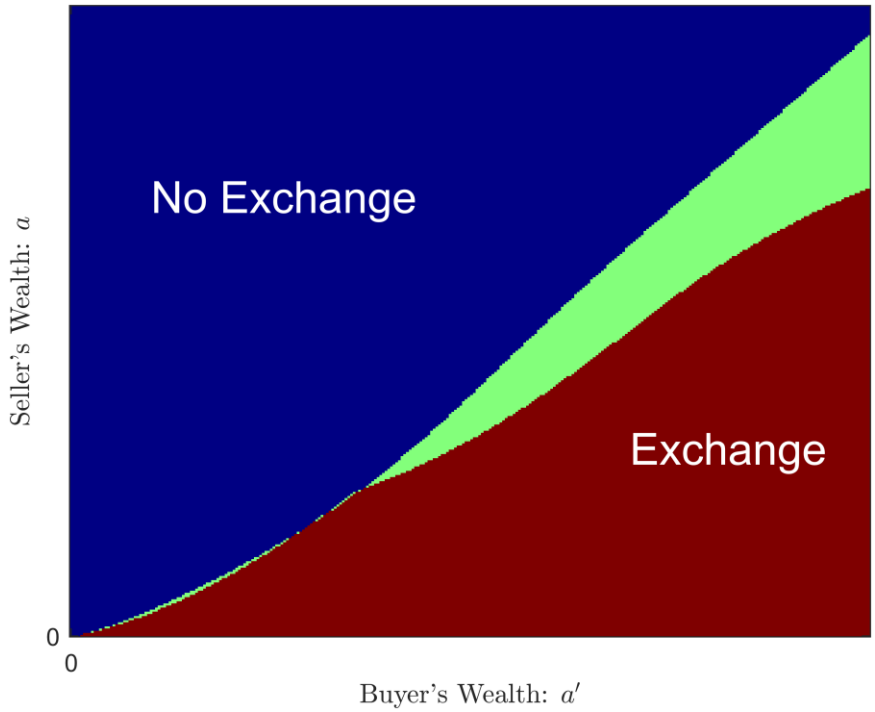


Figure 8: Bilateral Exchange Allocation Heatmap (Large Cost-Basis)



5.1.3 Indirect Effects of the Capital Gains Tax

In sum, the direct effects of an increase in the CGT rate τ include the lock-in effect manifesting as an increase in the seller's minimum asset price P^*_{min}. In turn, an increase in the seller's minimum asset price passes through to an increase in the exchange asset price P^*, with the pass-through rate being smaller when the value of the bargaining parameter κ is larger. Importantly, the direct effects change both exchange allocations and exchange asset prices. Exchange allocations change because the increase in the seller's minimum asset price means that, for some bilateral exchanges, the seller's minimum asset price is below the buyer's maximum asset price prior to the increase in the CGT rate, but above the buyer's maximum asset price after the increase in the CGT rate. Such exchanges occur prior to the increase in the CGT rate, but not afterwards. Exchange asset prices change due to the lock-in effect on the seller's minimum asset price and the resulting pass through to the exchange asset price.

That the direct effects alter both exchange allocations and exchange asset prices imply that the HJB equations and the KF equations are no longer satisfied at the value functions and joint distribution functions pertaining to the old tax regime. The indirect effects of an increase in the CGT rate capture how the value functions and joint distribution functions adjust endogenously so as to satisfy the HJB equations and KF equations, respectively, given that the direct effects induce changes in exchange allocations and exchange asset prices. Since neither the value functions nor the joint distribution functions admit closed-form solutions, one can only characterize the indirect effects numerically. Appendix C.2 presents figures that detail the indirect effects of changes in the CGT rate τ on the wealth distribution, which exhibits a Pareto distribution in the right tail.[16]

16 Formally, the wealth distribution is given by the marginal density function $g_a(a;\tau) := \int_0^\infty g(a,z;\tau) \cdot da$, where the joint density function $g(a,z;\tau)$ is that defined in equation (40).

5.2 Aggregate Productivity

5.2.1 Aggregate Production Function

To determine the aggregate productivity in the economy, I proceed in two steps. First, I define aggregate output and aggregate factors of production, using the microeconomic policy functions of Lemma 3 and the stationary joint distribution function. Second, I prove that the aggregate production function in the economy inherits the functional form of the microeconomic production function (19).

Aggregation. Using the stationary joint distribution functions for the seller, namely $g_s(a, P_s; \tau)$, and for the buyer, namely $g_b(a; \tau)$, I define the joint distribution function $g(a, z; \tau)$ for wealth a and productivity z, for a given CGT rate τ.

$$g(a, z; \tau) := \begin{cases} g_b(a; \tau) & z = z_\ell \\ \int_0^\infty g_s(a, P_s; \tau) \cdot dP_s & z = z_h \end{cases} \tag{40}$$

Using the joint distribution function $g(a, z; \tau)$ defined in equation (40), I define aggregate output $Y(\tau)$, aggregate labor demand $N(\tau)$, and aggregate capital demand $K(\tau)$.

$$Y(\tau) := \sum_{z \in \{z_\ell, z_h\}} \int_0^\infty y^*(a, z) \cdot g(a, z; \tau) \cdot da$$

$$N(\tau) := \sum_{z \in \{z_\ell, z_h\}} \int_0^\infty n^*(a, z) \cdot g(a, z; \tau) \cdot da$$

$$K(\tau) := \sum_{z \in \{z_\ell, z_h\}} \int_0^\infty k^*(a, z) \cdot g(a, z; \tau) \cdot da$$

Intuitively, the aggregate output and aggregate factors are given by integrating the microeconomic output and factor demand policy functions over the state space of wealth a and productivity z,

using the stationary joint distribution function $g(a,z;\tau)$ as the integrating weights. By affecting the stationary joint distribution function $g(a,z;\tau)$, the CGT rate τ shapes aggregate output and aggregate factors. The intuition is that the CGT rate influences how sellers accumulate wealth through participating in bilateral exchanges on the frictional OTC market. In doing so, the CGT rate shapes the wealth distribution of both sellers and buyers.

The following proposition proves the existence of an aggregate production function that inherits the functional form of the microeconomic production function (19).

Proposition 4: Aggregate Production Function

Given aggregate output $Y(\tau)$ and aggregate factors of production $\{N(\tau), K(\tau)\}$, the aggregate production function inherits the functional form of the microeconomic production function (19).

$$Y(\tau) = [[Z(\tau) \cdot K(\tau)]^{\alpha} \cdot N(\tau)^{1-\alpha}]^{\nu} \tag{41}$$

The aggregate productivity level $Z(\tau)$ is endogenous and depends on the CGT rate τ.

$$\underbrace{Z(\tau)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}}_{\text{Aggregate Productivity}} = \frac{\underbrace{\sum_{z\in\{z_\ell,z_h\}} \int_0^{\infty} [z \cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \cdot g(a,z;\tau) \cdot da}_{\propto \text{Aggregate Output}}}{\underbrace{\left[\sum_{z\in\{z_\ell,z_h\}} \int_0^{\infty} k^*(a,z) \cdot g(a,z) \cdot da\right]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}}_{\text{Aggregate Input}}} \tag{42}$$

Proof. See Appendix A.6.

Aggregate productivity (42) is a residual accounting for variation in aggregate output that is not explained by variation in the aggregate input (Syverson, 2011). Adhering to this notion, the aggregate productivity (42) in the quantitative model is a residual manifesting as a ratio of

aggregate output to an aggregate input, namely the contribution of aggregate capital to production. To understand why the numerator on the right-hand side of equation (42) is a measure of aggregate output, it is useful to note the following property of the ex-post output function (24).

Lemma 4: Ex-Post Output as a function of Capital Demand

The ex-post output function $y^*(a,z)$ is proportional to a power function of effective capital demand $z \cdot k^*(a,z)$.

$$y^*(a,z) \propto [z \cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \tag{43}$$

Proof. By definition, the ex-post output function is given by substituting the optimal factor demand policy functions of Lemma 3 into the production function (19).

$$y^*(a,z) = [[z \cdot k^*(a,z)]^{\alpha} \cdot n^*(a,z)^{1-\alpha}]^{\nu}$$

Substituting the labor demand policy function (21) into the expression for ex-post output yields the following expression of ex-post output as a function of the capital demand policy function $k^*(a,z)$.

$$y^*(a,z) = \left[\frac{\nu(1-\alpha)}{w}\right]^{\frac{\nu(1-\alpha)}{1-\nu(1-\alpha)}} \cdot [z \cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}$$

Since (ν, α, w) are constant model parameters, the desired result is implied.

Given that $y^*(a,z) \propto [z \cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}$, it follows that the numerator on the right-hand side of equation (42) is proportional to aggregate output $Y(\tau)$. The denominator on the right-hand side of equation (42) measures the contribution of the aggregate capital input $K(\tau)$ to production.

Specifically, the exponent $\nu\alpha/\big(1-\nu(1-\alpha)\big)$ denotes the elasticity of output with respect to capital, after accounting for the role of labor in production.

Aggregate productivity $Z(\tau)$, in the frictional economy with a leverage ratio parameter satisfying $\lambda \in [1,\infty)$, is endogenous because it depends on the microeconomic capital demand policy function $k^*(a,z)$ and the stationary joint distribution function $g(a,z;\tau)$. Moreover, by altering wealth accumulation through bilateral exchanges on the OTC market, the CGT rate τ influences the stationary joint distribution $g(a,z;\tau)$, and hence in turn affects aggregate productivity $Z(\tau)$.

First-Best Economy. To characterize aggregate productivity in the first best economy without the collateral constraint (2) (i.e., in the limit $\lambda \to \infty$), it is worth noting that the stationary marginal distribution of productivity, denoted by $g_z(z) := \int_0^\infty g(a,z)\cdot da$, is exogenous as a result of the assumption that transactions between sellers and buyers on the frictional OTC market amount to bilateral exchanges.

Lemma 5: Marginal Distribution of Productivity

The stationary marginal distribution of productivity is exogenously given in terms of the Poisson transition rate parameters $(\gamma_{h\ell},\gamma_{\ell h})$.

$$g_z(z) = \begin{cases} \dfrac{\gamma_{h\ell}}{\gamma_{\ell h}+\gamma_{h\ell}} & z = z_\ell \\ \dfrac{\gamma_{\ell h}}{\gamma_{\ell h}+\gamma_{h\ell}} & z = z_h \end{cases} \tag{44}$$

Proof. See Appendix A.7.

The main takeaway is that the presence of a frictional OTC market for bilateral exchanges does not affect the marginal distribution of business productivity z. Intuitively, if a bilateral exchange occurs, then the relative fraction of low productivity and high productivity businesses does not change because productivity levels are simply reallocated from one owner to another. Rather, the relative rate at which business productivity levels exogenous change according to the two state Poisson process is the sole determinant of the stationary marginal distribution of productivity. The following corollary characterizes aggregate productivity in the frictionless economy without a collateral constraint.

Corollary 1: Aggregate Productivity in Frictionless Economy

Aggregate productivity in the frictionless economy with $\lambda \to \infty$ is exogenously given as a CES aggregator of the microeconomic productivity levels $z \in \{z_\ell, z_h\}$, with exogenous weights given by the marginal distribution of productivity $g_z(z)$.

$$Z = \left[\frac{\gamma_{h\ell}}{\gamma_{\ell h} + \gamma_{h\ell}} \cdot z_\ell^{\frac{\nu\alpha}{1-\nu}} + \frac{\gamma_{\ell h}}{\gamma_{\ell h} + \gamma_{h\ell}} \cdot z^{\frac{\nu\alpha}{1-\nu}} \right]^{\frac{1-\nu}{\nu\alpha}} \tag{45}$$

Proof. Substitute the unconstrained capital demand policy function (22) into the aggregate productivity level equation (42), and simplify the resulting expression for aggregate productivity to obtain the desired result.

In the frictionless economy, wealth a has no influence in production. Therefore, the allocation of wealth across heterogeneous private business owners does not affect the efficiency with which capital is allocated across heterogeneous private business owners. Rather, aggregate productivity is exogenously given by the fraction of low productivity and high productivity types and the microeconomic productivity values.

5.2.2 Decomposing Change in Aggregate Productivity

There are three channels through which a change in the CGT rate τ affects aggregate productivity, albeit each channel operates through a change in the stationary joint distribution function $g(a, z; \tau)$. The first channel is through a change in the aggregate output of buyers indexed by low productivity $z = z_\ell$. The second channel is through a change in the aggregate output of sellers indexed by high productivity $z = z_h$. The third channel is through a change in the aggregate input given by aggregate capital demand. To decompose the change in aggregate productivity to a change in the CGT rate τ, it is useful to define the following generalized aggregate productivity function.

$$\bar{Z}(\tau_\ell, \tau_h, \tau_K)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} := \frac{\underbrace{\int_0^\infty [z_\ell \cdot k^*(a, z_\ell)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \cdot g(a, z_\ell; \tau_\ell) \cdot da}_{\propto \textit{Aggregate Output: Buyers}}}{\underbrace{\left[\sum_{z \in \{z_\ell, z_h\}} \int_0^\infty k^*(a, z) \cdot g(a, z; \tau_K) \cdot da\right]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}}_{\textit{Aggregate Input}}} + \frac{\underbrace{\int_0^\infty [z_h \cdot k^*(a, z_h)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \cdot g(a, z_h; \tau_h) \cdot da}_{\propto \textit{Aggregate Output: Sellers}}}{\underbrace{\left[\sum_{z \in \{z_\ell, z_h\}} \int_0^\infty k^*(a, z) \cdot g(a, z; \tau_K) \cdot da\right]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}}_{\textit{Aggregate Input}}} \tag{46}$$

The generalized aggregate productivity function $\bar{Z}(\tau_\ell, \tau_h, \tau_K)$ is indexed by three CGT rates as arguments of the function. The argument τ_ℓ affects the joint distribution function featuring in the aggregate output of buyers, the argument τ_h affects the joint distribution function featuring in the aggregate output of sellers, and the argument τ_K affects the joint distribution function featuring in aggregate capital demand. Importantly, setting these three CGT rates to the same value yields the aggregate productivity level in equation (42), i.e. $Z(\tau) = \bar{Z}(\tau, \tau, \tau), \forall \tau \in [0, 1)$.

Consider a change in the CGT rate from an old rate $\tau = \tau_\ell = \tau_h = \tau_K$ to a new rate $\tau' = \tau'_\ell = \tau'_h = \tau'_K$. One can decompose the change in aggregate productivity $[Z(\tau') - Z(\tau)]$ into three primary components plus a residual component.

$$Z(\tau') - Z(\tau) = \underbrace{[\bar{Z}(\tau'_\ell, \tau_h, \tau_K) - \bar{Z}(\tau_\ell, \tau_h, \tau_K)]}_{\textit{Change in Aggregate Output: Buyers}} + \underbrace{[\bar{Z}(\tau_\ell, \tau'_h, \tau_K) - \bar{Z}(\tau_\ell, \tau_h, \tau_K)]}_{\textit{Change in Aggregate Output: Sellers}} \qquad (47)$$

$$+ \underbrace{[\bar{Z}(\tau'_\ell, \tau'_h, \tau'_K) - \bar{Z}(\tau'_\ell, \tau'_h, \tau_K)]}_{\textit{Change in Aggregate Input}}$$

$$+ \underbrace{[\bar{Z}(\tau'_\ell, \tau'_h, \tau_K) + \bar{Z}(\tau_\ell, \tau_h, \tau_K) - \bar{Z}(\tau'_\ell, \tau_h, \tau_K) - \bar{Z}(\tau_\ell, \tau'_h, \tau_K)]}_{\textit{Residual}}$$

The left-hand side of equation (47) denotes the change in aggregate productivity $Z(\tau)$ to a change in the CGT rate τ. The right-hand side of equation (47) provides the decomposition of the change in aggregate productivity $Z(\tau)$ to a change in the CGT rate τ. The first term on the right-hand side denotes the contribution stemming from the change in the aggregate output of buyers with a low productivity level $z = z_\ell$, holding all else equal. The second term on the right-hand side denotes the contribution stemming from the change in the aggregate output of sellers with a high productivity level $z = z_h$, holding all else equal. The third term on the right-hand side denotes the contribution stemming from a change in the aggregate input, namely aggregate capital demand $K(\tau)$, given the change in both the aggregate output of buyers and the aggregate output of sellers. The fourth term on the right-hand side is a residual component that imposes the constraint that the sum of the components on the right-hand side equals the left-hand side, i.e. $[Z(\tau') - Z(\tau)] = [\bar{Z}(\tau'_\ell, \tau'_h, \tau'_K) - \bar{Z}(\tau_\ell, \tau_h, \tau_K)]$.[17]

17 In the quantitative analysis of Section 6, I show that the value of this residual component is close to zero.

6. Quantitative Analysis

This section calibrates the quantitative model of Section 4 to perform quantitative analysis that supplies answers to two quantitative questions. What is the effect of increasing the CGT rate from a zero rate $\tau = 0$ to a strictly positive rate $\tau > 0$ on aggregate productivity and aggregate tax revenues? What are the properties of the efficiency-equity frontier that characterizes the trade-off between maximizing aggregate productivity versus maximizing aggregate tax revenues? There are three main takeaways.

First, whether a strictly positive CGT rate maximizes aggregate productivity $Z(\tau)$ depends on both the severity of financial frictions, as regulated by the leverage ratio parameter λ, and the imbalance of bargaining power between sellers and buyers, as regulated by the bargaining power κ. If financial frictions are sufficiently severe, then a strictly positive CGT rate maximizes aggregate productivity, independently of bargaining power. However, if financial frictions less severe, then a strictly positive CGT rate maximizes aggregate productivity if bargaining power favors sellers, and a zero CGT rate maximizes aggregate productivity if bargaining power is more equally distributed. In all cases, however, an increase in the CGT rate from zero to a strictly positive value necessarily induces a decrease in aggregate output $Y(\tau)$.

Second, in the instances in which a strictly positive CGT rate maximizes aggregate productivity, the driver of the productivity gain depends on the distribution of bargaining power. If bargaining power is equally distributed, then the productivity gain stems from using a small aggregate input more efficiently, rather than from producing more aggregate output. Meanwhile, if bargaining power favors sellers, then the productivity gains stem from an increase in the aggregate output of buyers.

Third, there is an efficiency-equity trade-off inherent in the model because the CGT rate that maximizes aggregate tax revenues is strictly greater than that which maximizes aggregate productivity. Thus, maximizing aggregate tax revenues for the purpose of attaining more equity among private business owners comes at the expense of aggregate productivity losses.

6.1 Computational Algorithm

The quantitative model does not admit a closed-form solution for the endogenous variables that summarize the stationary equilibrium. Thus, I solve for the stationary equilibrium numerically. To do so, I discretize the state space over wealth a and productivity z and make use of frontier numerical techniques for solving heterogeneous agent models in continuous time (Achdou et al., 2021).

Discretization. Let $\mathbb{A} := \{a_1, a_2, \dots, a_I\}$ denote the non-uniform grid for wealth a, with I grid points. Similarly, let $\mathbb{P} := \{P_{s,1}, P_{s,2}, \dots, P_{s,J}\}$ denote the non-uniform grid for the seller's cost-basis P_s, with J grid points. The dimension of the state space for sellers is $I \cdot J$, namely a value for wealth a and a value for cost-basis P_s, while the dimension of the state space for buyers is I, namely a value for wealth a. Thus, the dimension of the total state space across sellers and buyers is $S := I \cdot J + I$.

Let $\boldsymbol{V}(\tau)$ denote the stacked value function vector, with dimension $(I \cdot J + I) \times 1$, and let $\boldsymbol{g}(\tau)$ denote the stacked joint distribution function vector, also with dimension $(I \cdot J + I) \times 1$.

$$\boldsymbol{V}(\tau) := \begin{pmatrix} V_s(a_1, P_{s,1}; \tau) \\ \vdots \\ V_s(a_I, P_{s,J}; \tau) \\ V_b(a_1; \tau) \\ \vdots \\ V_b(a_I; \tau) \end{pmatrix}$$

$$(\tau) := \begin{pmatrix} g_s(a_1, P_{s,1}; \tau) \\ \vdots \\ g_s(a_I, P_{s,J}; \tau) \\ g_b(a_1; \tau) \\ \vdots \\ g_b(a_I; \tau) \end{pmatrix}$$

Henceforth, I suppress the dependence of the value function vector and the joint distribution function on the CGT rate τ to ease notation. The following lemma provides the characterizes of the discretized dynamic system of Hamilton-Jacobi-Bellman equations (25) and (26) and Kolmogorov-Forward equations (31) and (32).

Lemma 6: Discretized System of HJB and KF Equations

The discretized dynamic system of HJB and KF equations can be expressed a system of nonlinear matrix equations in two endogenous vectors $\boldsymbol{V}$ *and* $\boldsymbol{g}$.

$$\rho \cdot \boldsymbol{V} = U[\boldsymbol{V}] + [A + B(\boldsymbol{V}) + C(\boldsymbol{V}, \boldsymbol{g})] \cdot \boldsymbol{V} \tag{48}$$

$$\mathbf{0} = [A + B(\boldsymbol{V}) + C(\boldsymbol{V}, \boldsymbol{g})]^T \cdot \boldsymbol{g} \tag{49}$$

Proof. See Appendix B.1.

Intuitively, equation (48) represents the system of HJB equations stacked across the state space. The left-hand side denotes the value annuity value $\rho \cdot \boldsymbol{V}$ associated with each point in the state space. The right-hand side decomposes the annuity value into different sources. The first term on the right-hand side, namely $U[\boldsymbol{V}]$, denotes the value stemming from flow utility from consumption. This value depends on the value function vector $\boldsymbol{V}$ because knowledge of the value function derivative is required to determine the consumption policy function satisfying the dynamic first order conditions (27) and (28).

The second term on the right-hand side, namely $[A + B(\boldsymbol{V}) + C(\boldsymbol{V}, \boldsymbol{g})] \cdot \boldsymbol{V}$, denotes the value from exogenous and endogenous transitions in the state variables. This term is the product of a composite transition matrix $[A + B(\boldsymbol{V}) + C(\boldsymbol{V}, \boldsymbol{g})]$ and the value function vector $\boldsymbol{V}$. The three transition matrices in the composite transition matrix summarize the different ways in which the state variables can evolve. The first matrix A summarizes exogenous transitions in productivity z according to the two-state Poisson process and depends only on the Poisson transition rate parameters $(\gamma_{\ell h}, \gamma_{h\ell})$. The second matrix $B(\boldsymbol{V})$ summarizes endogenous transitions in wealth a due to the consumption-saving decision. This matrix depends on the value function vector $\boldsymbol{V}$ because knowledge of the value function derivative is required to determine the consumption policy function and hence the saving policy functions in equations (29) and (30).

The third matrix $C(\mathbf{V}, \mathbf{g})$ summarizes endogenous transitions in wealth a and productivity z arising from bilateral exchanges on the frictional OTC market. This matrix depends on both the value function vector $\boldsymbol{V}$ and the joint distribution vector $\boldsymbol{g}$. The dependence on the value function arises from the fact that the bilateral exchange policy function $D^*(a, P_s, a'; \tau)$ in equation (35) depends on the reservation prices, namely $P^*_{min}(a, P_s; \tau)$ and $P^*_{max}(a'; \tau)$, that in turn depend on the value functions $V_s(a, P_s; \tau)$ and $V_b(a'; \tau)$ through the indifference conditions in equations (33) and (34). The dependence on the joint distribution function arises from the fact that the total rate of meeting a seller or a buyer on the OTC market depends not only on the meeting rate $\eta > 0$ at which a bilateral exchange opportunities arises, but also on the stationary density functions $g_s(a, P_s; \tau)$ and $g_b(a'; \tau)$ for seller and buyer types, respectively, as in the canonical model of over-the-counter markets à la Duffie et al. (2005).

Equation (49) represents the stacked system of KF equations. The left-hand side equals the zero vector $\mathbf{0}$ as one is considering the stationary equilibrium in which the joint distribution is

not time-varying. The right-hand side details the matrix product between the transpose of the composite transition matrix and the joint distribution vector. Intuitively, the joint distribution vector is given as the solution to an eigenvector problem given by equation (49).

Discussion. There are two challenges in solving the system of matrix equations (48) and (49), one of which is commonplace and the other of which is novel to this paper. The commonplace challenge is that the matrix system is non-linear in the endogenous vectors $(\boldsymbol{V}, \boldsymbol{g})$, and as such, one must employ an iterative scheme to solve this system. The novel challenge arises from the fact that the entire joint distribution vector $\boldsymbol{g}$ enters the HJB equation (48). By contrast, the standard heterogeneous agent model in continuous time à la Achdou et al. (2021) features a HJB equation that does not depend directly on the joint distribution. I propose a computational algorithm below that can accommodate the joint distribution function entering the HJB equation directly.

Numerical Algorithm. Table 1 presents the numerical algorithm with which I solve for the endogenous vectors $(\boldsymbol{V}, \boldsymbol{g})$. The essence of the numerical algorithm is that one must iterate on the value function and joint distribution function jointly. By contrast, in the standard heterogeneous agent model in continuous time à la Achdou et al. (2021), one need only iterate on the value function, and obtain the joint distribution function "for free" by solving an eigenvector problem.

The numerical algorithm contains three auxiliary parameters. The first is a standard threshold tolerance parameter $\epsilon > 0$ that regulates the norm threshold at which convergence is attained. The second and third parameters are the step size parameters, namely Δ_V and Δ_g, for updating the value function vector $\boldsymbol{V}$ and joint distribution vector $\boldsymbol{g}$, respectively.

Table 1: Numerical Algorithm for Stationary Equilibrium

ALGORITHM: SOLVE FOR VALUE FUNCTION AND JOINT DISTRIBUTION FUNCTION IN STATIONARY EQUILIBRIUM

1 *Propose an initial guess for the endogenous vectors* $(\boldsymbol{V^0}, \boldsymbol{g^0})$

while *joint distribution function has not converged* $\|\boldsymbol{g^{i+1}} - \boldsymbol{g^i}\| > \boldsymbol{\epsilon}$

  while *value function has not converged* $\|\boldsymbol{V^{j+1}} - \boldsymbol{V^j}\| > \boldsymbol{\epsilon}$

2 *Iterate on the value function vector* $\boldsymbol{V^j}$ *using implicit updating*

$$\frac{1}{\Delta_V} \cdot \left(\boldsymbol{V}^{j+1} - \boldsymbol{V}^j\right) + \rho \cdot \boldsymbol{V}^{j+1} = U\left[\boldsymbol{V}^j\right] + \left[A + B\left(\boldsymbol{V}^j\right) + C\left(\boldsymbol{V}^j, \boldsymbol{g}^i\right)\right] \cdot \boldsymbol{V}^{j+1}$$

  end

3 *Iterate on the joint distribution function vector* $\boldsymbol{g^i}$ *using implicit updating*

$$\frac{1}{\Delta_g} \cdot \left(\boldsymbol{g}^{i+1} - \boldsymbol{g}^i\right) = \left[A + B(\overline{\boldsymbol{V}}) + C\left(\overline{\boldsymbol{V}}, \boldsymbol{g}^i\right)\right]^T \cdot \boldsymbol{g}^{i+1}$$

  $\overline{\boldsymbol{V}}$ *denotes the converged value function from the inner loop when the joint distribution function is* $\boldsymbol{g^i}$

end

6.2 Calibration

This subsection describes my calibration strategy. I calibrate the model at an annual frequency so that one unit of time is one year. I focus on the year 2022 for which the SCF data are available.

Model Parameters. The quantitative model contains a set of parameters are commonplace in a heterogenous agent model of private business à la Cagetti & De Nardi (2006) and its modern rendition in Moll (2014). The remaining parameters are unique to the presence of a frictional OTC market for the bilateral exchange of business productivity levels. I discuss each set of parameters in turn.

Standard Parameters. I group the standard parameters into three subcategories: (i) preferences; (ii) technology; (iii) financial frictions; and (iv) input markets. The two preference parameters are the discount rate $\rho > 0$ and the coefficient of relative risk aversion $\sigma > 0$. The technology parameters include the two parameters of the production function, namely the capital share $\alpha \in (0,1)$ and the returns to scale $\nu \in (0,1)$, the two productivity levels $z_\ell > 0$ and $z_h > z_\ell > 0$, and the two Poisson rate parameters $\gamma_{\ell h}$ and $\gamma_{h\ell}$. The single parameter regulating financial frictions is the leverage ratio $\lambda \in [1, \infty)$. The input market parameters are the wage rate $w > 0$ and interest rate $r > 0$.

Novel Parameters. There are two parameters that are novel to the frictional OTC market for bilateral exchanges. The first is the meeting rate $\eta > 0$ at which bilateral exchange opportunities on the frictional OTC market arise. The second is the bargaining parameter $\kappa \in (0,1)$ that regulates the bargaining power of the seller relative to that of the buyer in determining the exchange asset price function (36).

Fixed Parameters. I set the coefficient of relative risk aversion to $\sigma = 1.5$, the capital share to $\alpha = 0.33$, and the returns to scale to $\nu = 0.88$, all of which are standard calibration values in the related literature on heterogeneous agent models in private business (Cagetti & De Nardi, 2006). I set the interest rate to $r = 0.03$, a standard value. I set the leverage ratio parameter $\lambda = 1.5$, mirroring the value used to evaluate the effects of the capital income tax versus the wealth tax (Guvenen et al., 2023). Lastly, I set the meeting rate $\eta = 0.1$ for bilateral exchange opportunities to the maximum value that is computationally feasible.

Calibrated Parameters. I calibrate the discount rate ρ, the productivity Poisson transition rates $(\gamma_{\ell h}, \gamma_{h\ell})$ to match moments of the income and wealth distribution for private business owners using the 2022 SCF data.

Table 2: Model Parameterization

	Parameter Description	Value	Source
	A. Fixed Parameters		
σ	Coefficient of Relative Risk Aversion	1.5	Cagetti & De Nardi (2006)
α	Capital Share	0.33	Cagetti & De Nardi (2006)
ν	Returns to Scale	0.88	Cagetti & De Nardi (2006)
w	Wage Rate	1.0	Normalization
r	User Cost of Capital	0.03	Itskhoki & Moll (2019)
λ	Leverage Ratio	1.5	Guvenen et al. (2023)
η	Meeting Rate on OTC Market	0.1	Computational maximum
κ	Seller's Relative Bargaining Power	0.5	Equal bargaining power
	B. Calibrated Parameters		
ρ	Discount Rate	0.05	-
$\gamma_{\ell h}$	Transition Rate from Low to High	0.01	-
$\gamma_{h\ell}$	Transition Rate from High to Low	0.06	-
z_ℓ	Low Productivity Level	1.5	-
z_h	High Productivity Level	3.0	-

Table 2 summarizes the values for the fixed and calibrated model parameters.

6.3 Quantitative Results

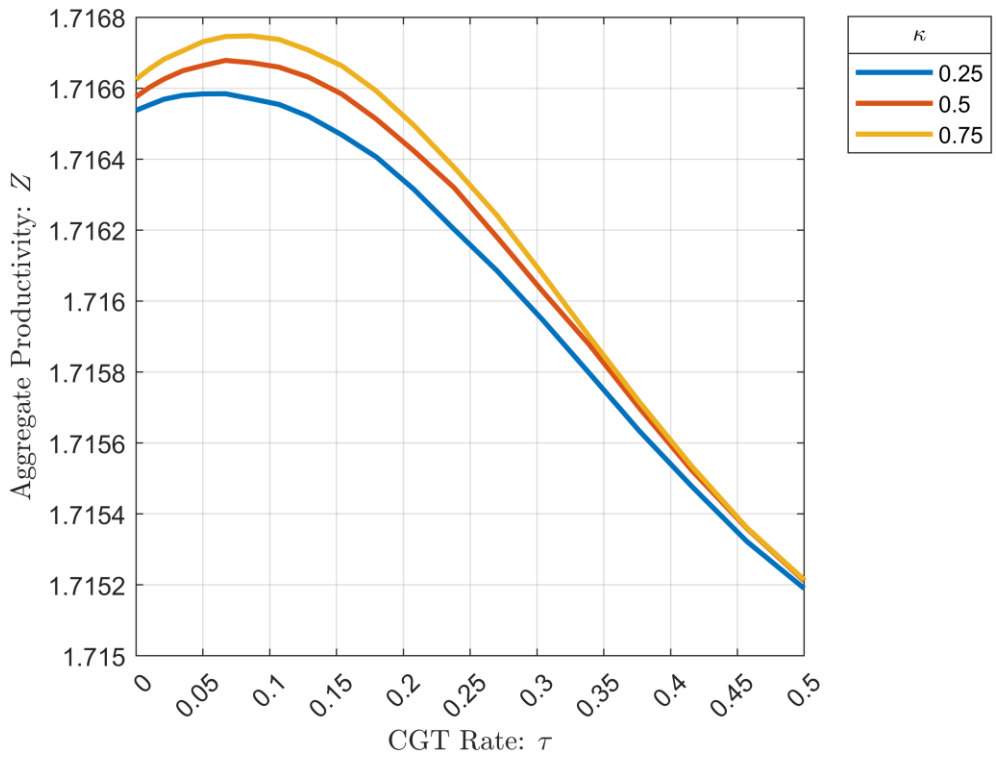
The quantitative analysis aims to quantify the effects of an increase in the CGT rate on aggregate productivity $Z(\tau)$ and aggregate tax revenues $T(\tau)$. Three model parameters play an important

role in determining the quantitative effects. The first is the leverage ratio parameter λ that governs the severity of financial frictions. The second is the bargaining power parameter κ that governs the balance of bargaining power between sellers and buyers. I discuss the importance of each parameter where appropriate.

6.3.1 The Effect of Capital Gains Taxation on Aggregate Productivity

I characterize the effects of the CGT rate τ on aggregate productivity (42) by solving for the stationary equilibrium of the quantitative model under a set of different CGT rates τ, and for different values of the leverage ratio parameter λ and the bargaining parameter κ. Figure 9 presents plots of aggregate productivity $Z(\tau)$ by the CGT rate τ for a scenario in which financial frictions are severe (i.e., $\lambda = 1.5$, implying a net leverage ratio of 1.5 times net worth), and for two different bargaining power parameter values.

Figure 9: Aggregate Productivity by CGT Rate ($\lambda = 1.5$)



The striking result is that a strictly positive CGT rate maximizes aggregate productivity, albeit the productivity gains in moving from a zero CGT rate $\tau = 0$ to the optimal CGT rate $\tau > 0$ are small. Figure 10 and Figure 11 and decompose the change in aggregate productivity into its constituent parts using the decomposing formula in equation (47). For a fixed leverage ratio parameter λ, the source of aggregate productivity gains varies depending on the value of the bargaining parameter κ. If bargaining power is distributed equally between sellers and buyers (i.e., $\kappa = 0.5$), then the positive contribution to aggregate productivity at the optimal CGT rate stems from using a smaller aggregate input more efficiently. The aggregate output of both sellers and buyers is lower at the optimal CGT rate that maximizes aggregate productivity. This indicates that the productivity gains do not stem from producing more output, holding inputs fixed.

Figure 10: Decomposition of Change in Aggregate Productivity ($\lambda = 1.5, \kappa = 0.5$)

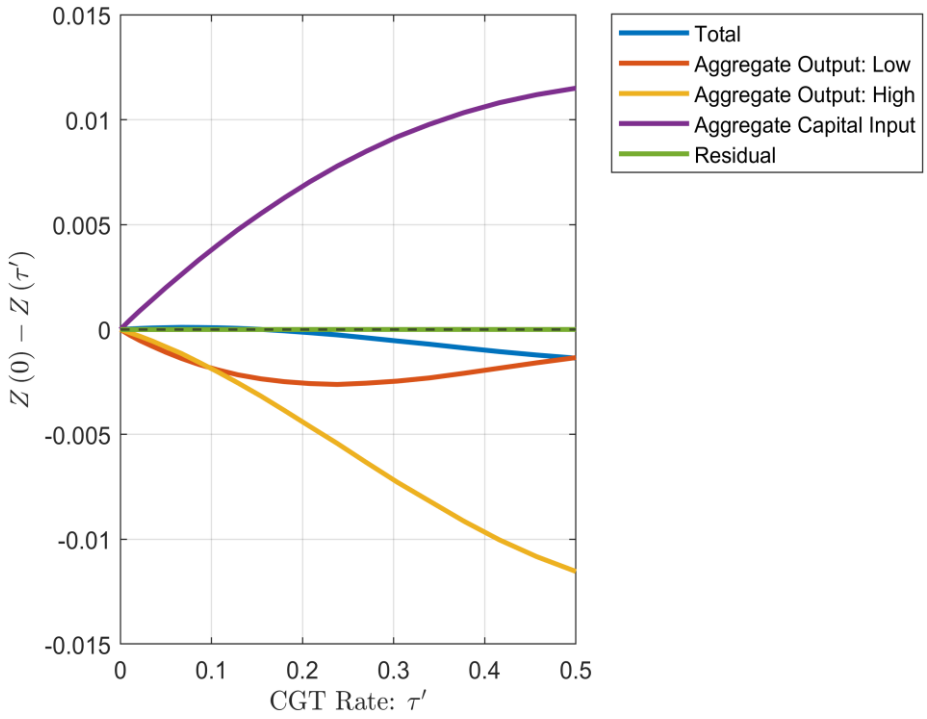
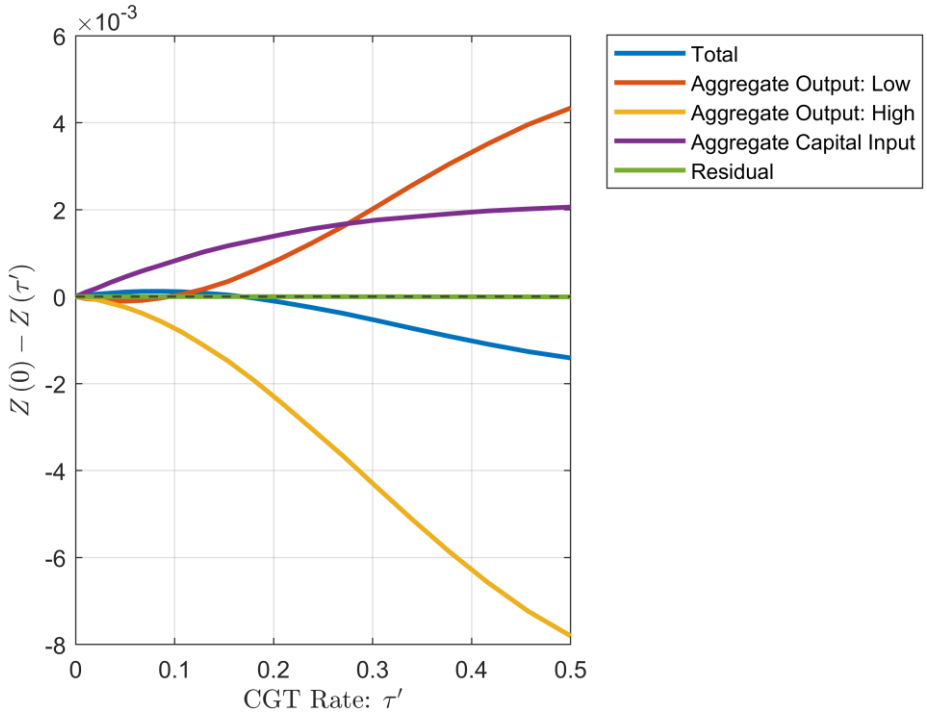


Figure 11: Decomposition of Change in Aggregate Productivity ($\lambda = 1.5, \kappa = 0.75$)

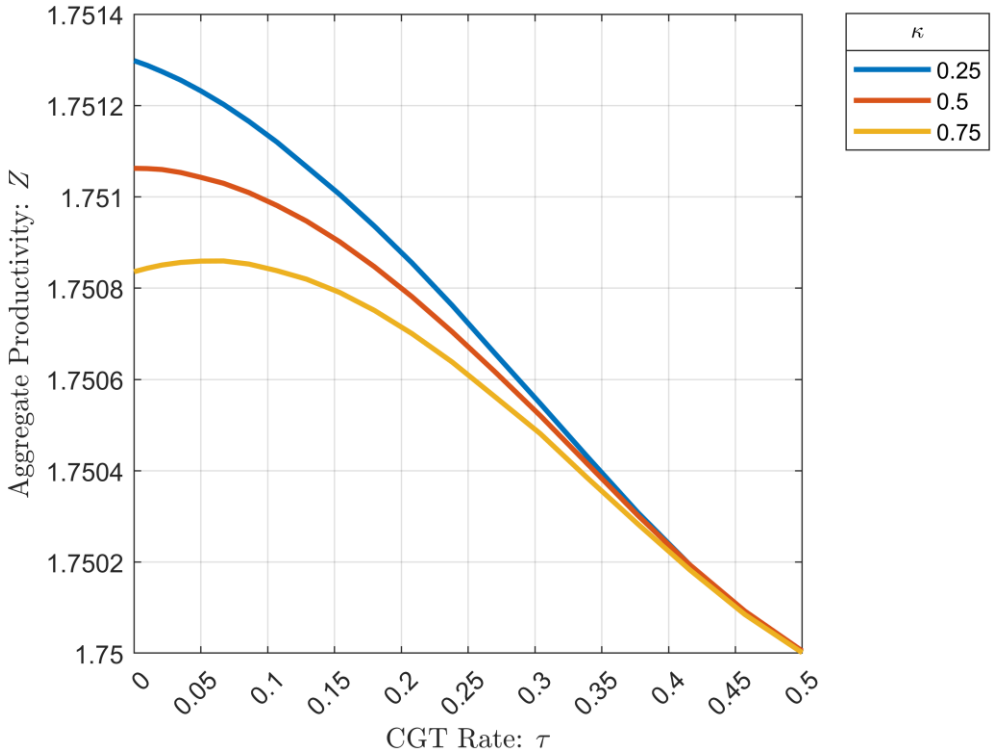


Meanwhile, if bargaining power favors sellers (i.e., $\kappa = 0.75$), then the positive contribution to the change in aggregate productivity at the optimal CGT rate τ' stems from an increase in the aggregate output of buyers (indexed by low productivity). Intuitively, if bargaining power favors sellers, then all else equal, bilateral exchanges occur at higher asset prices, since the exchange asset price attaches more weight to the buyer's maximum asset price and less weight to the seller's minimum asset price. This means that, as sellers transition to the buyer state, their post-exchange wealth is greater because the higher exchange asset price results in a greater lump-sum payment from sellers to buyers. The additional wealth post-exchange enables the buyers to overcome the collateral constrain more effectively, and hence produce more output.

Now, consider an alternative scenario in which financial frictions are mild. Figure 12 presents plots of aggregate productivity $Z(\tau)$ by the CGT rate τ for a scenario in which financial frictions are mild (i.e., $\lambda = 3.0$, implying a leverage ratio of three times net worth). In this instance, whether zero CGT rate or a strictly positive CGT rate maximizes aggregate productivity depends on the severity of the bargaining power imbalance. If bargaining power is equal between sellers and buyers (i.e., $\kappa = 0.5$), then a zero CGT rate maximizes aggregate productivity. However, if bargaining power favors sellers, then a strictly positive CGT rate maximizes aggregate productivity.

Figure 13 and Figure 14 decompose the change in aggregate productivity, from a zero CGT rate to a strictly positive CGT rate, into its constituent parts. Unlike the case with severe financial frictions $(\lambda = 1.5)$ and bargaining power favoring sellers $(\kappa = 0.75)$, the case with a mild financial friction $(\lambda = 3.0)$ and bargaining power favoring sellers $(\kappa = 0.75)$ showcases that productivity gains from stemming from using the aggregate input more efficiently.

Figure 12: Aggregate Productivity by CGT Rate ($\lambda = 3.0$)



Taken together, the results from the quantitative analysis reveal that the effects of capital gains taxation on aggregate productivity depend crucially on the interaction between financial frictions and bargaining power. If financial frictions are severe, then a strictly positive CGT rate maximizes aggregate productivity, albeit the source of the productivity gains differs depending on the balance of bargaining power across sellers and buyers. If financial frictions are mind, then a strictly positive CGT rate maximizes aggregate productivity only if bargaining power favors sellers. In this instance, the source of aggregate productivity gains lies from using the aggregate input more efficiently, contrasting with the source of aggregate productivity gains when financial frictions are severe.

Figure 13: Decomposition of Change in Aggregate Productivity ($\lambda = 3.0, \kappa = 0.5$)

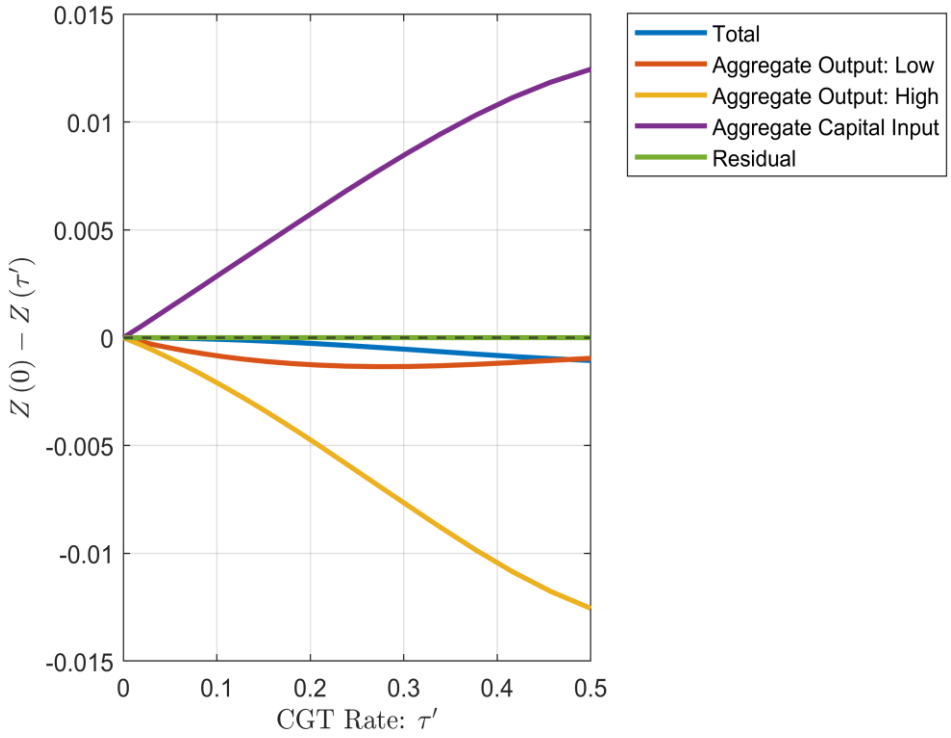
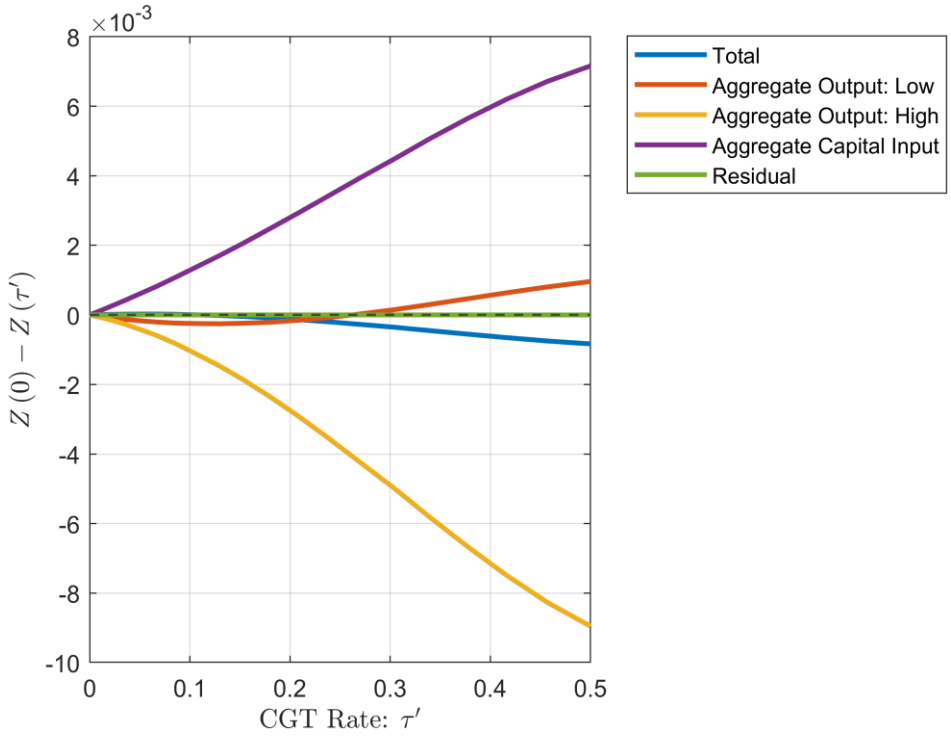


Figure 14: Decomposition of Change in Aggregate Productivity ($\lambda = 3.0, \kappa = 0.75$)



6.3.2 The Effect of Capital Gains Taxation on Aggregate Tax Revenues

The classic trade-off in designing tax policy is between efficiency and equity. Having established the efficiency effects of capital gains taxation on aggregate productivity, I quantify the equity effects of capital gains taxation on aggregate tax revenues. The main result is that aggregate tax revenues exhibit a Laffer curve, i.e., an inverted-U shape.

Figure 15 and Figure 16 plot aggregate tax revenues (37) by the CGT rate τ for a scenario with a severe financial friction $(\lambda = 1.5)$ and a scenario with mild financial friction $(\lambda = 3.0)$, respectively. The peak of the Laffer curve lies at the CGT rate $\tau = 0.2$, suggesting that a capital gains tax rate of 20% would maximize aggregate tax revenues. This optimal CGT rate under the equity criterion of maximizing aggregate tax revenues is not sensitive to the severity of financial frictions, albeit the amount of aggregate tax revenues raised does depend on the severity of financial frictions.

Lastly, the optimal CGT rate at the peak of the Laffer curve is strictly greater than that which maximizes aggregate productivity. This gives rise to the classic efficiency-equity trade-off. That is, maximizing tax revenues in order to accomplish redistribution objectives comes at the expense of efficiency losses manifesting as a lower aggregate productivity level.

Figure 15: Aggregate Tax Revenues by CGT Rate ($\lambda = 1.5$)

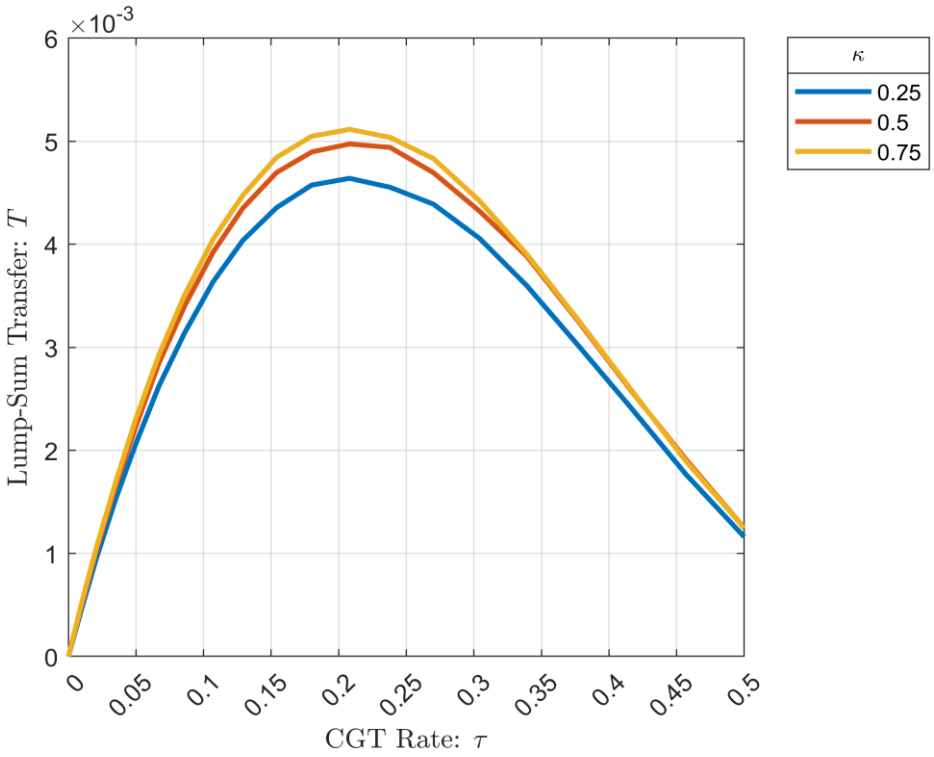


Figure 16: Aggregate Tax Revenues by CGT Rate ($\lambda = 3.0$)

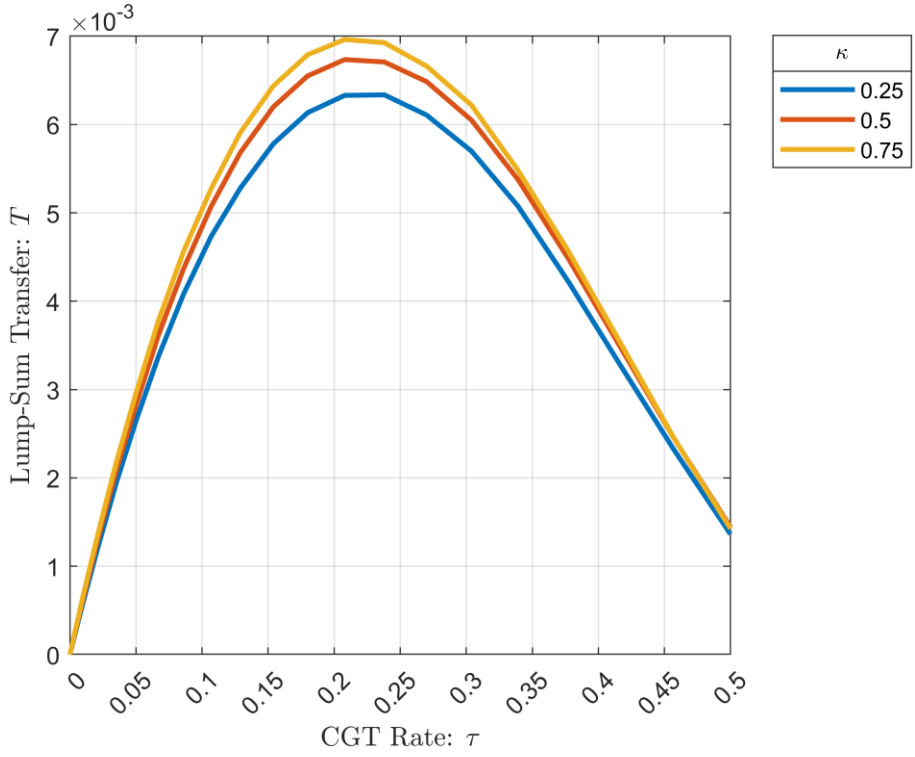


Figure 15 and Figure 16 highlight the revenue-maximizing CGT rate lies within the range 20-22%. This revenue maximizing rate is greater than those in the existing empirical literature that estimates the tax elasticity of capital gains and infers the resulting revenue maximizing tax rates (Agersnap & Zidar, 2021).

6.3.3 Efficiency-Equity Frontier

To take stock of the effects of capital gains taxation on aggregate productivity and aggregate tax revenues, I trace out the efficiency-equity frontier as a function of the CGT rate τ. This frontier characterizes the trade-off between maximizing efficiency, as measured by aggregate productivity $Z(\tau)$, and maximizing equity, as measured by aggregate tax revenues $T(\tau)$.

One complication in characterizing this trade-off is that aggregate productivity $Z(\tau)$ and aggregate tax revenue $T(\tau)$ differ in their units. Aggregate productivity in equation (42) is measured as the ratio of aggregate output to the aggregate input, and hence inherits the units of the output to input ratio. Meanwhile, the units of aggregate tax revenues in equation (37) are flow dollars per unit of time. To construct unit-free measures of aggregate productivity and aggregate tax revenues, I compute the standardized values of these variables.

$$\tilde{X}(\tau) := \frac{X(\tau) - \frac{1}{N_\tau}\sum_{\tau' \in \mathcal{T}} X(\tau')}{\sqrt{\frac{1}{N_\tau}\sum_{\tau' \in \mathcal{T}} \left[X(\tau') - \frac{1}{N_\tau}\sum_{\tau' \in \mathcal{T}} X(\tau')\right]^2}}, X \in \{Z, T\} \tag{50}$$

Equation (50) defines the standardized variable $\tilde{X}(\tau)$, which is given by subtracting the mean and dividing by the standard deviation. The variation stems from evaluating $X \in \{Z, T\}$ for different CGT rates in the set $\mathcal{T}$.

Figure 17: Efficiency-Equity Frontier (Aggregate Productivity: $\lambda = 1.5$)

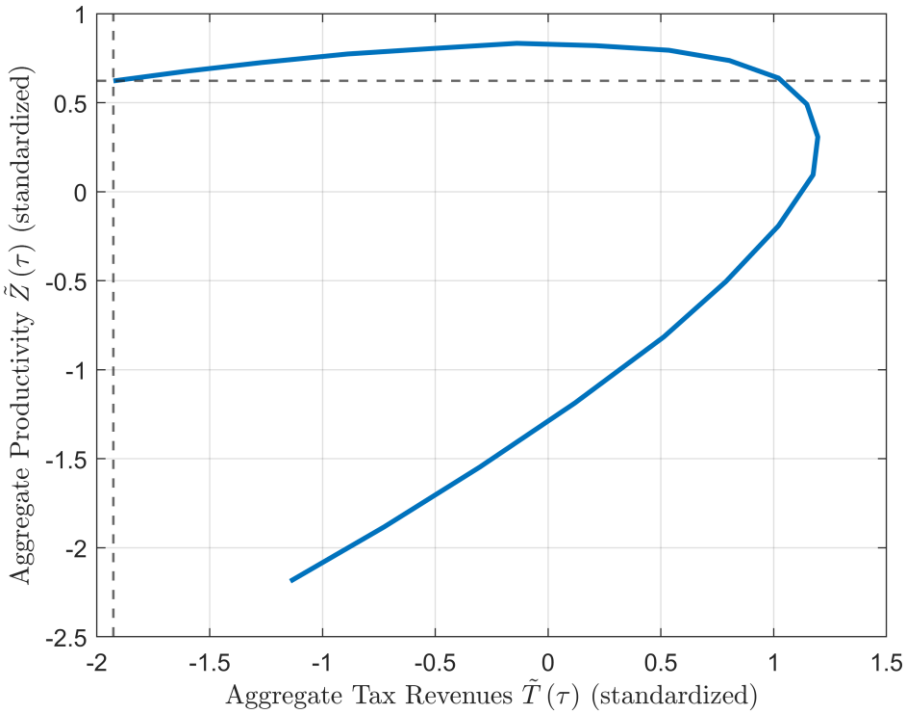
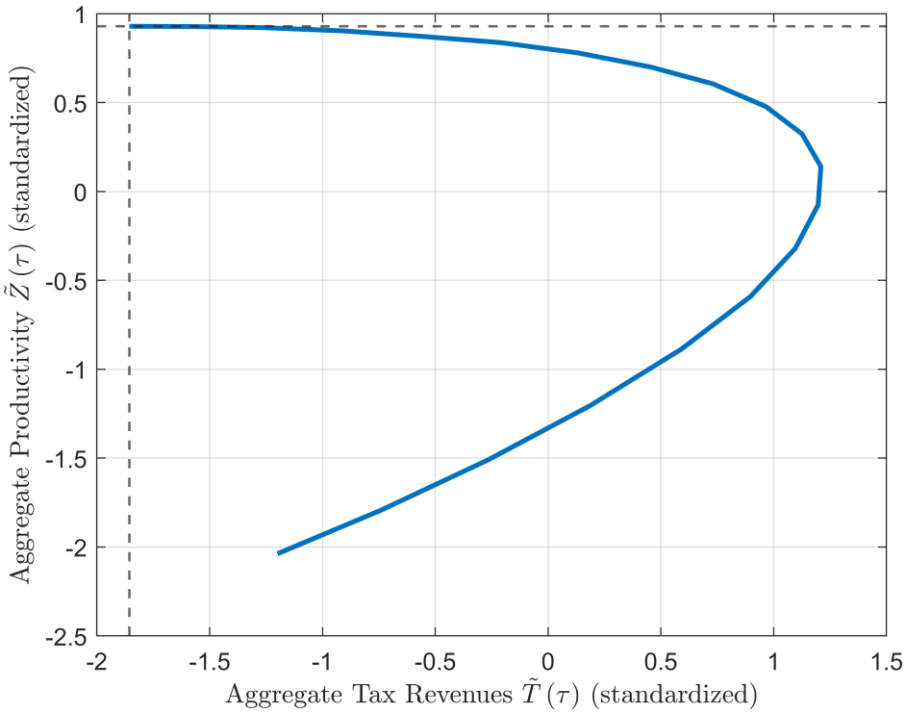


Figure 18: Efficiency-Equity Frontier (Aggregate Productivity: $\lambda = 3.0$)



Aggregate Productivity. Figure 17 and Figure 18 plot the efficiency-equity frontier for the case in which the measure of efficiency is aggregate productivity, for two different values for the leverage ratio parameter λ. Each point in this figure pertains to a stationary equilibrium with a particular CGT rate τ, and associated values for aggregate productivity $\tilde{Z}(\tau)$ and aggregate tax revenues $\tilde{T}(\tau)$. The dashed vertical line and dashed horizontal pertain to the values for aggregate tax revenues and aggregate productivity, respectively, in the stationary equilibrium with a zero CGT rate (i.e., $\tau = 0$).

The dashed vertical line and dashed horizontal line segment the space into four quadrants. However, only the two quadrants to the right of the vertical dashed line are relevant because it is not possible to generate tax revenues that are lower than those generated under a zero CGT rate. The upper right quadrant details the set of strictly positive CGT rates such that both aggregate productivity and aggregate tax revenues are greater, relative to the stationary equilibrium with a zero CGT rate. This is the region with a "free lunch" in which a strictly positive CGT rate is optimal for maximizing both aggregate productivity and aggregate tax revenues. No such region exists in Figure 18 as a strictly positive CGT rate necessarily reduces aggregate productivity if the financial friction is less severe (i.e., $\lambda = 3.0$).

The lower right quadrant details the set of strictly positive CGT rates such that aggregate productivity is lower than that in the stationary equilibrium with a zero CGT rate. Initially, there frontier line in this quadrant has a negative slope, indicating that increasing tax revenues requires incurring productivity losses. That is, there is a trade-off between maximizing aggregate productivity and aggregate tax revenues. However, at one point, the slope of the frontier line becomes positive, indicating that further increases in the CGT rate reduce both aggregate productivity and aggregate tax revenues. This case arises once the economy reaches the "wrong"

side of the Laffer curve, namely points appearing to the right of the peak of the Laffer curve in Figure 15 and Figure 16.

Figure 19: Efficiency-Equity Frontier (Aggregate Output: $\lambda = 1.5$)

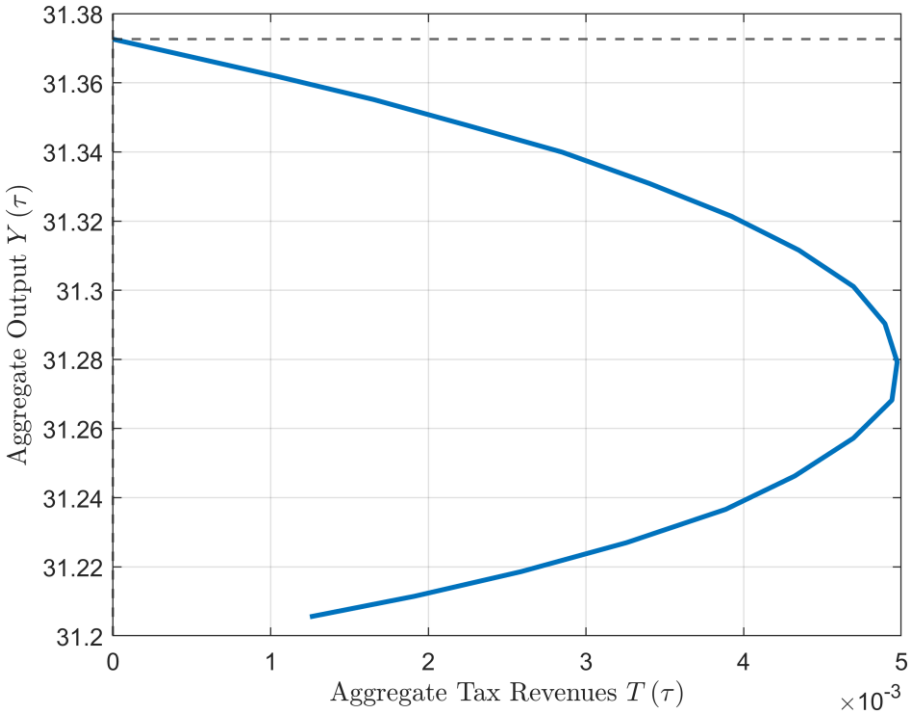
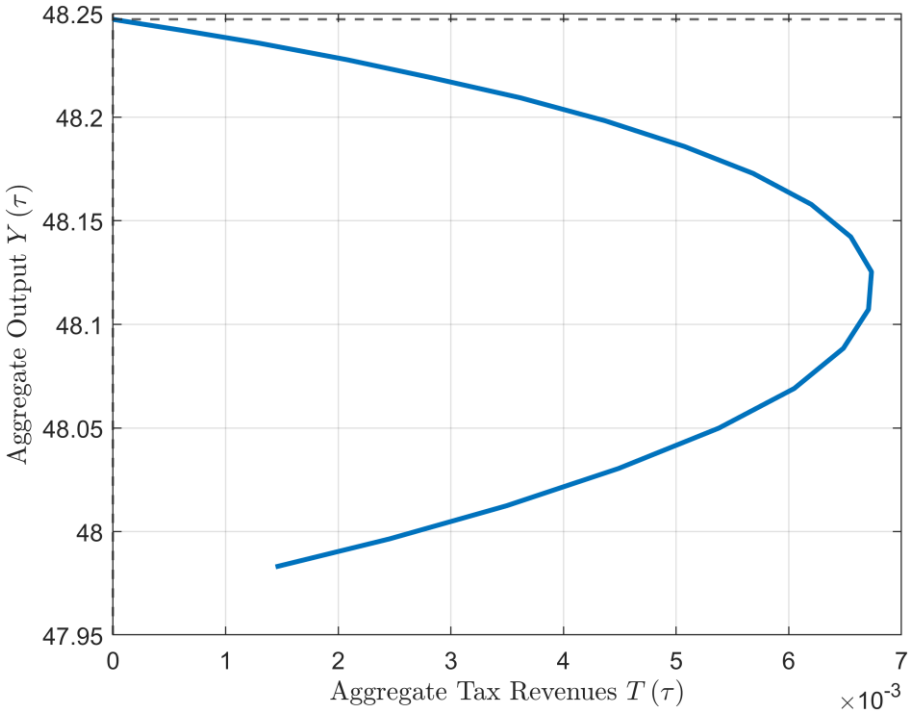


Figure 20: Efficiency-Equity Frontier (Aggregate Output: $\lambda = 3.0$)



Aggregate Output. Figure 19 and Figure 20 plot the efficiency-equity frontier for the case in which the measure of efficiency is aggregate productivity $Y(\tau)$, for two different values for the leverage ratio parameter λ. Once more, the dashed vertical line and the dashed horizontal line pertain to the values for aggregate tax revenues and aggregate output, respectively, in the stationary equilibrium with a zero CGT rate. Unlike the efficiency-equity frontier in Figure 17, the efficiency-equity frontier Figure 19 and Figure 20 does not appear in the upper right quadrant. The intuitive interpretation is that there is no "free lunch," since a strictly positive CGT rate always results in lower aggregate output relative to the equilibrium with a zero CGT rate.

The segment of the frontier in Figure 19 and Figure 20 with a negative slope characterizes the trade-off between maximizing aggregate tax revenues and maximizing aggregate productivity. Specifically, the slope of the line measures the loss in aggregate output, at the margin, to obtain an additional unit of tax revenues. Meanwhile, the segment of the frontier with a positive slope pertains to all CGT rates on the "wrong" side of the Laffer curve. At this point, any further increases in the CGT reduce both aggregate output and aggregate tax revenues. The role of financial frictions in influencing the efficiency-equity frontier is primarily in a level shift of the frontier schedule. Specifically, an economy with less severe financial frictions, as detailed in Figure 20, exhibits a higher level of aggregate output for a given CGT rate compared to an economy with more severe financial frictions, as detailed in Figure 19.

The Role of Bargaining Power. In addition to financial frictions, (as measured by the leverage ratio parameter λ), bargaining power (as measured by the parameter κ) also plays a role in shaping the efficiency-equity frontier. Figure 21 and Figure 22 detail the efficiency-equity frontier, for different values of the leverage ratio parameter λ and the bargaining parameter κ.

Figure 21: Efficiency-Equity Frontier and Bargaining Power ($\lambda = 1.5$)

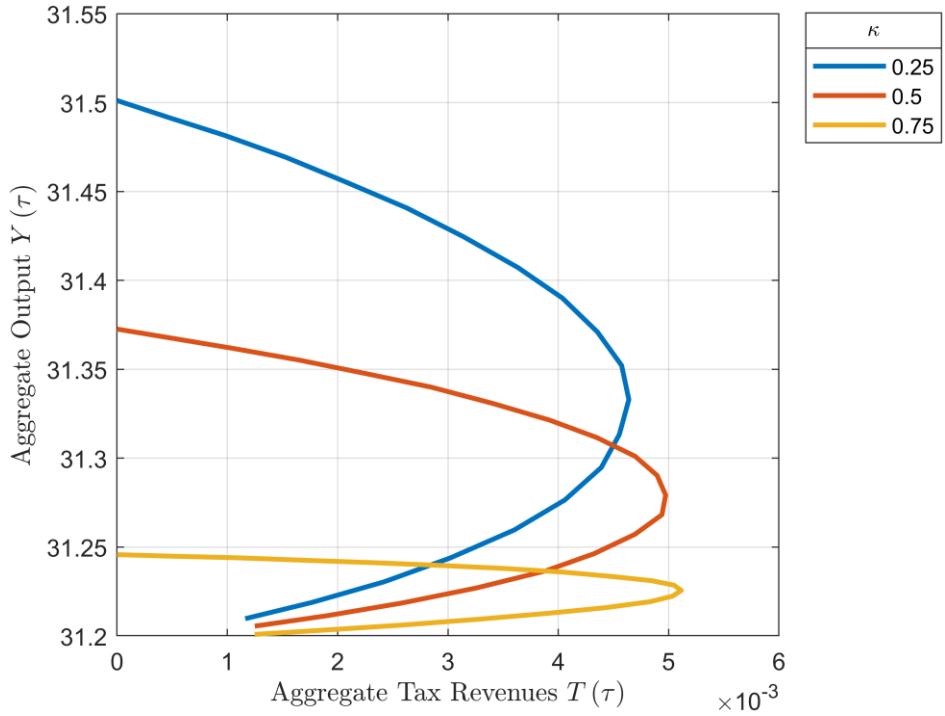
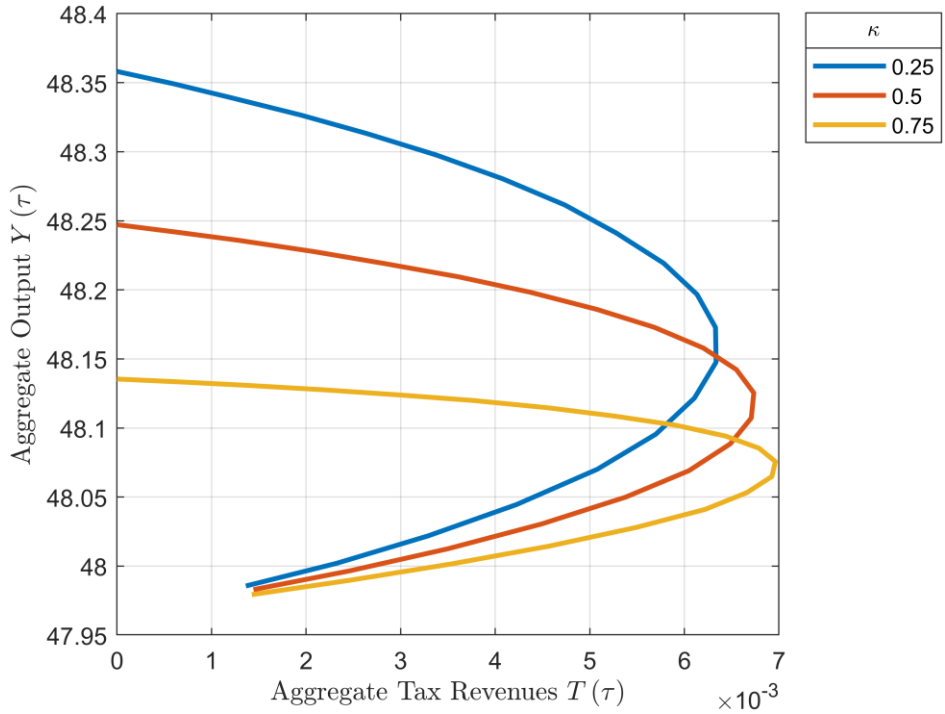


Figure 22: Efficiency-Equity Frontier and Bargaining Power ($\lambda = 3.0$)



Holding financial frictions fixed, an increase in the seller's bargaining (i.e., an increase in the parameter κ) has two effects on the efficiency-equity frontier. The first effect is a downward level shift in the frontier. For a given CGT rate, aggregate output is smaller when the seller has greater bargaining power in determining the exchange asset price (36). This negative effect on aggregate output stems from the fact that prospective buyers are "paying too much" to acquire the high productivity business, thereby leaving the buyers with less wealth post-exchange. In turn, lower wealth, all else equal, implies that the collateral constraint (2) is more binding. The second effect is that the slope of the frontier becomes shallower. Since the slope of the frontier characterizes the trade-off between aggregate output and aggregate tax revenues, a shallower slope implies increasing tax revenues involves forgoing smaller decreases in output. The reason stems from the fact that buyers are paying higher exchange asset prices. Consequently, the realized capital gains are larger, resulting in greater tax revenues from capital gains taxation.

Both the level effect and the slope effect of an increase in the seller's bargaining power depend quantitatively on the severity of financial frictions, as measured by the leverage ratio parameter λ. In particular, the slope effect is more pronounced if financial frictions are severe, as in Figure 21, compared to the case with less severe financial frictions, as in Figure 22.

Taking Stock. In summary, the efficiency-equity frontier for the capital gains tax depends on the two underlying economic frictions: (i) the collateral constraint (2); and (ii) bargaining power. An implication is that the design of capital gains tax policy should incorporate such characteristics of the market in which the asset (i.e., private business) is exchanged, as these characteristics shape the trade-off between aggregate output and aggregate tax revenues in a quantitatively meaningful sense.

7. Conclusion

This paper studies the effects of capital gains taxation in the decentralized market for private businesses, an asset class with empirically large unrealized capital gains. On the methodological front, I extend a heterogeneous agent model of private business to incorporate a frictional over-the-counter market for exchanging business productivities, and in which sellers pay taxes on realized capital gains. Armed with this quantitative model, I make two main contributions.

The first contribution is characterizing the theoretical predictions of a change in the CGT rate on endogenous outcomes in the model. All else equal, an increase in the CGT rate induces a lock-in effect whereby sellers require a higher minimum asset price in order to accept a bilateral exchange. This lock-in effect distorts both exchange allocations and exchange asset prices. The elasticity summarizing the effect on exchange asset prices depends on the balance of bargaining power between sellers and buyers in determining the exchange asset price.

The second contribution is quantifying the effects of capital gains taxation on aggregate outcomes. If financial frictions are severe, then a strictly positive CGT rate maximizes aggregate productivity, albeit the productivity gains stem from a decrease in input use, rather than an increase in aggregate output. Moreover, aggregate tax revenues exhibit a Laffer curve. The CGT rate that maximizes aggregate tax revenues is strictly greater than that which maximizes aggregate output, generating the classic efficiency versus equity trade-off in tax policy design.

A fruitful direction for future research is studying the nexus between capital gains taxation and traditional capital taxes, namely the capital income tax and wealth tax. The quantitative model that I develop in this paper provides a useful economic framework for doing so.

BIBLIOGRAPHY

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., & Moll, B. (2021). Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach. *The Review of Economic Studies*, *89*(1), 45–86. https://doi.org/10.1093/restud/rdab002

Agersnap, O., & Zidar, O. (2021). The Tax Elasticity of Capital Gains and Revenue-Maximizing Rates. *American Economic Review: Insights*, *3*(4), 399–416. https://doi.org/10.1257/aeri.20200535

Bhandari, A., Martellini, P., & McGrattan, E. (2021). A Theory of Business Transfers. *Working Paper*.

Boar, C., & Midrigan, V. (2023). Should We Tax Capital Income or Wealth? *American Economic Review: Insights*, *5*(2), 259–274. https://doi.org/10.1257/aeri.20220192

Buera, F. J., Kaboski, J. P., & Shin, Y. (2011). Finance and Development: A Tale of Two Sectors. *American Economic Review*, *101*(5), 1964–2002. https://doi.org/10.1257/aer.101.5.1964

Buera, F. J., & Shin, Y. (2013). Financial Frictions and the Persistence of History: A Quantitative Exploration. *Journal of Political Economy*, *121*(2), 221–272. https://doi.org/10.1086/670271

Cagetti, M., & De Nardi, M. (2006). Entrepreneurship, Frictions, and Wealth. *Journal of Political Economy*, *114*(5), 835–870. https://doi.org/10.1086/508032

Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica*, *54*, 607–622. https://api.semanticscholar.org/CorpusID:154421900

Chari, V. V, Christiano, L. J., & Kehoe, P. J. (1994). Optimal Fiscal Policy in a Business Cycle Model. *Journal of Political Economy*, *102*(4), 617–652. https://doi.org/10.1086/261949

Chari, V. V, Golosov, M., & Tsyvinski, A. (2005). Business Start-Ups, the Lock-In Effect, and Capital Gains Taxation. *Working Paper*.

Dai, Z., Maydew, E., Shackelford, D. A., & Zhang, H. H. (2008). Capital Gains Taxes and Asset Prices: Capitalization or Lock-in? *The Journal of Finance*, *63*(2), 709–742. https://doi.org/https://doi.org/10.1111/j.1540-6261.2008.01329.x

Duffie, D., Gârleanu, N., & Pedersen, L. H. (2005). Over-the-Counter Markets. *Econometrica*, *73*(6), 1815–1847. https://doi.org/https://doi.org/10.1111/j.1468-0262.2005.00639.x

Fagereng, A., Gomez, M., Gouin-Bonenfant, E., Holm, M., Moll, B., & Natvik, G. (2022). Asset-Price Redistribution. *Working Paper*.

Farhi, E. (2010). Capital Taxation and Ownership When Markets Are Incomplete. *Journal of Political Economy*, *118*(5), 908–948. https://doi.org/10.1086/657996

Guntin, R., & Kochen, F. (2023). Financial Frictions and the Market for Firms. *Working Paper*.

Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo, S., & Chen, D. (2023). Use It or Lose It: Efficiency and Redistributional Effects of Wealth Taxation. *The Quarterly Journal of Economics*, *138*(2), 835–894. https://doi.org/10.1093/qje/qjac047

Holmes, T. J., & Schmitz, J. A. (1990). A Theory of Entrepreneurship and Its Application to the Study of Business Transfers. *Journal of Political Economy*, *98*(2), 265–294. https://doi.org/10.1086/261678

Itskhoki, O., & Moll, B. (2019). Optimal Development Policies With Financial Frictions. *Econometrica*, *87*(1), 139–173. https://doi.org/https://doi.org/10.3982/ECTA13761

Judd, K. L. (1985). Redistributive Taxation in a Simple Perfect Foresight Model. *Journal of Public Economics*, *28*(1), 59–83. https://doi.org/https://doi.org/10.1016/0047-2727(85)90020-9

Moll, B. (2014). Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *American Economic Review*, *104*(10), 3186–3221. https://doi.org/10.1257/aer.104.10.3186

Myerson, R. B., & Satterthwaite, M. A. (1983). Efficient Mechanisms for Bilateral Trading. *Journal of Economic Theory*, *29*(2), 265–281. https://doi.org/https://doi.org/10.1016/0022-0531(83)90048-0

Saez, E., & Stantcheva, S. (2018). A Simpler Theory of Optimal Capital Taxation. *Journal of Public Economics*, *162*, 120–142. https://doi.org/https://doi.org/10.1016/j.jpubeco.2017.10.004

Straub, L., & Werning, I. (2020). Positive Long-Run Capital Taxation: Chamley-Judd Revisited. *American Economic Review*, *110*(1), 86–119. https://doi.org/10.1257/aer.20150210

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, *49*(2), 326–365. https://doi.org/10.1257/jel.49.2.326

APPENDIX A: PROOFS

A.1. Proof of Lemma 1

Recall the private business owner's static profit maximization problem.

$$\pi^*(a,z) := \max_{n,k}\{(z\cdot k)^{\alpha}\cdot n^{1-\alpha} - w\cdot n - r\cdot k\}$$

$$\text{s.t. } k \in [0, \lambda\cdot a]$$

This proof proceeds in three steps. First, I optimize over the choice of labor demand n.

$$n^*(a,z) = \left[\frac{1-\alpha}{w}\right]^{\frac{1}{\alpha}}\cdot z\cdot k^*(a,z)$$

Substituting the labor demand policy function $n^*(a,z)$ back into the objective function yields the following simplified problem over the choice of capital demand k.

$$\pi^*(a,z) := \max_{k}\{[MPK(z) - r]\cdot k\}$$

$$\text{s.t. } k \in [0, \lambda\cdot a]$$

The marginal product of capital, denoted by $MPK(z)$, evaluated at the optimal labor demand $n^*(a,z)$, does not depend on the level of capital demand k.

$$MPK(z) := \left.\frac{\partial \mathcal{F}(n,k;z)}{\partial k}\right|_{n=n^*(a,z)} = \alpha\cdot\left[\frac{1-\alpha}{w}\right]^{\frac{1-\alpha}{\alpha}}\cdot z$$

Since the simplified objective function is linear in the choice of capital demand k, it follows that the capital demand policy function exhibits a bang-bang property.

$$k^*(a,z) = \begin{cases}\lambda\cdot a & MPK(z) > r\\ 0 & MPK(z)\leq r\end{cases}$$

Substituting the capital demand policy function $k^*(a,z)$ into the simplified objective function yields the ex-post profit income function.

$$\pi^*(a,z) = \begin{cases} [MPK(z) - r] \cdot \lambda \cdot a & MPK(z) > r \\ 0 & MPK(z) \leq r \end{cases}$$

A.2. Proof of Lemma 2

Recall the HJB equation (9).

$$\rho \cdot V(a,z) = \max_{c>0} \left\{ \ln c + [r \cdot a + R(z) \cdot a - c] \cdot \frac{\partial V(a,z)}{\partial a} + \gamma_{z'z} \cdot [V(a,z') - V(a,z)] \right\}$$

This proof follows a conjecture and verify method. I conjecture that the value function admits a logarithmic solution.

$$V(a,z) = B(z) + A \cdot \ln a$$

The constant term A and the function $B(z)$ are to be determined. Given the conjecture, the flow consumption policy function satisfies the following first order condition.

$$\frac{1}{c^*(a,z)} = \frac{A}{a}$$

Solving the first order condition yields the flow consumption policy function.

$$c^*(a,z) = \frac{1}{A} \cdot a$$

Substituting the conjecture for the value function and the flow consumption policy function into the HJB equation yields the following equation for a private business owner with high productivity z_h.

$$\rho \cdot B(z_h) + \rho \cdot A \cdot \ln a = -\ln A + [r \cdot A + R(z_h) \cdot A - 1] + \gamma_{\ell h} \cdot [B(z_\ell) - B(z_h)] + \ln a$$

There is a similar equation for a private business owner with low productivity z_ℓ.

$$\rho \cdot B(z_\ell) + \rho \cdot A \cdot \ln a = -\ln A + [r \cdot A + R(z_\ell) \cdot A - 1] + \gamma_{h\ell} \cdot [B(z_h) - B(z_\ell)] + \ln a$$

Equating coefficients on both sides for each equation and solving the system of two linear equations for the two unknown variables $\{B(z_\ell), B(z_h)\}$ yields the following solutions for the unknown variables $\{A, B(z_\ell), B(z_h)\}$.

$$A = \frac{1}{\rho}$$

$$B(z_\ell) = \frac{1}{\rho} \cdot \left[\ln\rho + \left(\frac{r-\rho}{\rho}\right) + \frac{1}{\rho} \cdot \left[\frac{\rho + \gamma_{\ell h}}{\rho + \gamma_{\ell h} + \gamma_{h\ell}} \cdot R(z_\ell) + \frac{\gamma_{h\ell}}{\rho + \gamma_{\ell h} + \gamma_{h\ell}} \cdot R(z_h) \right] \right]$$

$$B(z_h) = \frac{1}{\rho} \cdot \left[\ln\rho + \left(\frac{r-\rho}{\rho}\right) + \frac{1}{\rho} \cdot \left[\frac{\rho + \gamma_{h\ell}}{\rho + \gamma_{\ell h} + \gamma_{h\ell}} \cdot R(z_h) + \frac{\gamma_{\ell h}}{\rho + \gamma_{\ell h} + \gamma_{h\ell}} \cdot R(z_\ell) \right] \right]$$

Hence, the conjecture for the functional form of the value function is verified because there exists a constant term A and a function $B(z)$ satisfying the conjecture.

A.3. Proof of Lemma 3

Recall the static profit maximization problem.

$$\pi^*(a,z) := \max_{n,k} \{[(z \cdot k)^\alpha \cdot n^{1-\alpha}]^\nu - w \cdot n - r \cdot k\}$$

$$\text{s.t. } k \in [0, \lambda \cdot a]$$

The first order condition with respect to labor n characterizes the labor demand policy function.

$$n^*(a,z) = \left[\frac{\nu(1-\alpha)}{w}\right]^{\frac{1}{1-\nu(1-\alpha)}} \cdot (z \cdot k)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}$$

A.4. Proof of Proposition 2

Consider the seller's indifference condition (33) for the case in which a capital gain is realized when the bilateral exchange occurs at the seller's minimum asset price, i.e., $P^*_{min} > P_s$.

$$V_b(a + P^*_{min} - \tau \cdot (P^*_{min} - P_s); \tau) - V_s(a, P_s; \tau) = 0$$

Applying the implicit function theorem to the seller's indifference condition yields the partial derivative of the seller's minimum asset price P^*_{min} with respect to the capital gains tax rate τ.

$$\frac{\partial P^*_{min}}{\partial \tau} = \frac{P^*_{min} - P_s}{1 - \tau}$$

Rearranging yields the desired elasticity.

$$\frac{\tau}{P^*_{min}} \frac{\partial P^*_{min}}{\partial \tau} = \frac{\tau}{1 - \tau} \cdot \frac{P^*_{min} - P_s}{P^*_{min}}$$

A.5. Proof of Proposition 3

By the chain rule, the partial derivative of the exchange asset price P^* with respect to the CGT rate τ can be decomposed into the product of two partial derivatives.

$$\frac{\partial P^*}{\partial \tau} = \frac{\partial P^*}{\partial P^*_{min}} \cdot \frac{\partial P^*_{min}}{\partial \tau}$$

I rewrite the partial derivative above in elasticity form.

$$\frac{\tau}{P^*} \frac{\partial P^*}{\partial \tau} = \frac{P^*_{min}}{P^*} \frac{\partial P^*}{\partial P^*_{min}} \cdot \frac{\tau}{P^*_{min}} \frac{\partial P^*_{min}}{\partial \tau}$$

Using equation (36) and Proposition 2, one can simplify the elasticity of the exchange asset price with respect to the CGT rate to obtain the desired result.

$$\frac{\tau}{P^*}\frac{\partial P^*}{\partial \tau} = (1-\kappa)\cdot\frac{\tau}{1-\tau}\cdot\frac{P^*_{min}-P_s}{P^*}$$

A.6. Proof of Proposition 4

This proof proceeds in three steps. First, I characterize aggregate output $Y(\tau)$.

$$\begin{aligned} Y(\tau) &:= \sum_{z\in\{z_\ell,z_h\}}\int_0^\infty y^*(a,z)\cdot g(a,z;\tau)\cdot da \\ &= \sum_{z\in\{z_\ell,z_h\}}\int_0^\infty [[z\cdot k^*(a,z)]^\alpha\cdot n^*(a,z)^{1-\alpha}]^\nu\cdot g(a,z;\tau)\cdot da \\ &= \left[\frac{\nu(1-\alpha)}{w}\right]^{\frac{\nu(1-\alpha)}{1-\nu(1-\alpha)}}\cdot\left[\sum_{z\in\{z_\ell,z_h\}}\int_0^\infty [z\cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}\cdot g(a,z;\tau)\cdot da\right] \end{aligned}$$

The first line follows by definition of aggregate output. The second line follows by the production function (19). The third line follows by the labor demand policy function (21).

Second, I characterize aggregate labor demand $N(\tau)$.

$$\begin{aligned} N(\tau) &:= \sum_{z\in\{z_\ell,z_h\}}\int_0^\infty n^*(a,z)\cdot g(a,z;\tau)\cdot da \\ &= \left[\frac{\nu(1-\alpha)}{w}\right]^{\frac{1}{1-\nu(1-\alpha)}}\cdot\left[\sum_{z\in\{z_\ell,z_h\}}\int_0^\infty [z\cdot k^*(a,z)]^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}\cdot g(a,z;\tau)\cdot da\right] \end{aligned}$$

The first line follows by definition of aggregate labor demand. The second line follows by the labor demand policy function (21). Similarly, I characterize aggregate capital demand $K(\tau)$.

$$K(\tau) := \sum_{z\in\{z_\ell,z_h\}}\int_0^\infty k^*(a,z)\cdot g(a,z;\tau)\cdot da$$

Third, I conjecture that the aggregate production function inherits the functional form of the microeconomic production function.

$$Y(\tau) = [[Z(\tau) \cdot K(\tau)]^{\alpha} \cdot N(\tau)^{1-\alpha}]^{\nu}$$

Fourth, I verify that this conjecture is correcting by proving that there exists a unique aggregate productivity level $Z(\tau)$ that satisfies such an aggregate production function.

$$Z(\tau) = \left[\sum_{z \in \{z_\ell, z_h\}} \left[\frac{\int_0^\infty k^*(a,z)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \cdot g(a,z;\tau) \cdot da}{\left(\sum_{z \in \{z_\ell, z_h\}} \int_0^\infty k^*(a,z) \cdot g(a,z;\tau) \cdot da\right)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}} \right] \cdot z^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \right]^{\frac{1-\nu(1-\alpha)}{\nu\alpha}}$$

It is useful to rewrite the endogenous weights in the CES aggregator for aggregate productivity. Let $\bar{K}(z;\tau)$ denote the numerator of the endogenous weight.

$$\bar{K}(z;\tau) := \int_0^\infty k^*(a,z)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \cdot g(a,z;\tau) \cdot da$$

The denominator of the endogenous weigh is simply the aggregator capital demand $K(\tau)$ to an exponent that depends on the production parameters (ν, α). Given the definitions of $\bar{K}(z;\tau)$ and $K(\tau)$, one can rewrite aggregate productivity as follows:

$$Z(\tau) = \left[\left[\frac{\bar{K}(z_\ell;\tau)}{K(\tau)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}} \right] \cdot z_\ell^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} + \left[\frac{\bar{K}(z_h;\tau)}{K(\tau)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}} \right] \cdot z_h^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \right]^{\frac{1-\nu(1-\alpha)}{\nu\alpha}}$$

For the purposes of decomposing the change in aggregate productivity following a change in the CGT rate τ, I define the following aggregate productivity function.

$$\bar{Z}(\tau_\ell, \tau_h, \tau_K) = \left[\left[\frac{\bar{K}(z_\ell; \tau_\ell)}{K(\tau_K)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}} \right] \cdot z_\ell^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} + \left[\frac{\bar{K}(z_h; \tau_h)}{K(\tau_K)^{\frac{\nu\alpha}{1-\nu(1-\alpha)}}} \right] \cdot z_h^{\frac{\nu\alpha}{1-\nu(1-\alpha)}} \right]^{\frac{1-\nu(1-\alpha)}{\nu\alpha}}$$

Importantly, the following relationship holds.

$$Z(\tau) = \bar{Z}(\tau, \tau, \tau), \forall \tau \in [0, 1)$$

The change in aggregate productivity from the old CGT rate τ to a new CGT rate $\tau' \neq \tau$ can be decomposed as follows, where $\tau_\ell = \tau_h = \tau_K = \tau$ and $\tau'_\ell = \tau'_h = \tau'_K = \tau'$.

$$Z(\tau') - Z(\tau) = \underbrace{[\bar{Z}(\tau'_\ell, \tau_h, \tau_K) - \bar{Z}(\tau_\ell, \tau_h, \tau_K)]}_{\textit{Reallocation: Low Productivity}} + \underbrace{[\bar{Z}(\tau_\ell, \tau'_h, \tau_K) - \bar{Z}(\tau_\ell, \tau_h, \tau_K)]}_{\textit{Reallocation: High Productivity}}$$

$$+ \underbrace{[\bar{Z}(\tau'_\ell, \tau'_h, \tau_K) + \bar{Z}(\tau_\ell, \tau_h, \tau_K) - \bar{Z}(\tau'_\ell, \tau_h, \tau_K) - \bar{Z}(\tau_\ell, \tau'_h, \tau_K)]}_{\textit{Reallocation: Interaction}}$$

$$+ \underbrace{[\bar{Z}(\tau'_\ell, \tau'_h, \tau'_K) - \bar{Z}(\tau'_\ell, \tau'_h, \tau_K)]}_{\textit{Change in Capital Stock}}$$

A.7. Proof of Lemma 5

This proof proceeds in two steps. First, integrate the seller's KF equation (31) over the seller's wealth a and cost-basis P_s to obtain the following equation.

$$\gamma_{\ell h} \cdot g_z(z_\ell) - \gamma_{h\ell} \cdot g_z(z_h) = 0$$

In addition, there is an adding up constraint requiring that the sum of the marginal densities sum to one.

$$g_z(z_\ell) + g_z(z_h) = 1$$

Solving this system of two simultaneous linear equations in the two unknown variables $g_z(z_\ell)$ and $g_z(z_h)$ gives the desired result.

APPENDIX B: COMPUTATION

B.1. Proof of Lemma 6

The burden in this proof lies in deriving the stacked and discretized HJB equation (48). To do so, I employ the finite difference method and upwinding scheme of Achdou et al. (2021). This procedure approximates the flow consumption policy function using the dynamic first order conditions (27) and (28), and hence approximates the flow saving policy functions in equations (29) and (30). The primary complication in my quantitative model, relative to the standard framework of Achdou et al. (2021), arises from the presence of the frictional OTC market for bilateral exchanges.

Endogenous Consumption-Saving. The first step involves approximating the derivative of the value function with respect to wealth a using either the forward derivative or the backward derivative. Let $i \in \{1, \cdots, I\}$ index points on the wealth grid and let $j \in \{1, \cdots, J\}$ index points on the cost-basis grid (pertinent for the seller only). Then, the value function derivatives with respect to wealth can be characterized as follows.

$$V_{s,F,ij} := \frac{V_s(a_{i+1}, P_{s,j}; \tau) - V_s(a_i, P_{s,j}; \tau)}{a_{i+1} - a_i}$$

$$V_{s,B,ij} := \frac{V_s(a_i, P_{s,j}; \tau) - V_s(a_{i-1}, P_{s,j}; \tau)}{a_i - a_{i-1}}$$

$$V_{b,F,i} := \frac{V_b(a_{i+1}; \tau) - V_b(a_i; \tau)}{a_{i+1} - a_i}$$

$$V_{b,B,i} := \frac{V_b(a_i; \tau) - V_s(a_{i-1}, P_{s,j}; \tau)}{a_i - a_{i-1}}$$

Given the approximations of the value function derivatives, one can approximate the flow consumption policy function using either the forward derivative or the backward derivative by imposing the dynamic first order conditions (27) and (28).

$$c^*_{s,F,ij} = V^{-\frac{1}{\sigma}}_{s,F,ij}$$

$$c^*_{s,B,ij} = V^{-\frac{1}{\sigma}}_{s,B,ij}$$

$$c^*_{b,F,i} = V^{-\frac{1}{\sigma}}_{b,F,i}$$

$$c^*_{b,B,i} = V^{-\frac{1}{\sigma}}_{b,B,i}$$

Similarly, one can approximate the flow saving policy function using either the forward derivative or the backward derivative from equations (29) and (30).

$$\dot{a}^*_{s,F,ij} = r \cdot a_i + \pi^*(a_i, z_h) + T(\tau) - c^*_{s,F,ij}$$

$$\dot{a}^*_{s,B,ij} = r \cdot a_i + \pi^*(a_i, z_h) + T(\tau) - c^*_{s,B,ij}$$

$$\dot{a}^*_{b,F,i} = r \cdot a_i + \pi^*(a_i, z_\ell) + T(\tau) - c^*_{b,F,i}$$

$$\dot{a}^*_{b,B,i} = r \cdot a_i + \pi^*(a_i, z_\ell) + T(\tau) - c^*_{b,B,i}$$

The upwind scheme posits the forward derivative to approximate the flow consumption policy function of the seller if $\dot{a}^*_{s,F,ij} > 0$ (similarly, $\dot{a}^*_{b,F,i} > 0$ for the buyer), the backward derivative to approximate the flow consumption policy function of the seller if $\dot{a}^*_{s,B,ij} < 0$ (similarly, $\dot{a}^*_{b,B,i} < 0$ for the buyer), and total flow income if otherwise (i.e., the steady state scenario).

$$c^*_{s,ij} = c^*_{s,F,ij} \cdot 1\{\dot{a}^*_{s,F,ij} > 0\} + c^*_{s,B,ij} \cdot 1\{\dot{a}^*_{b,B,i} < 0\}$$
$$+[r \cdot a_i + \pi^*(a_i, z_h) + T(\tau)] \cdot 1\{\dot{a}^*_{s,F,ij} \leq 0 \leq \dot{a}^*_{s,B,ij}\}$$

$$c^*_{b,i} = c^*_{b,F,i} \cdot 1\{\dot{a}^*_{b,F,i} > 0\} + c^*_{b,F,i} \cdot 1\{\dot{a}^*_{b,B,i} < 0\}$$
$$+[r \cdot a_i + \pi^*(a_i, z_\ell) + T(\tau)] \cdot 1\{\dot{a}^*_{b,F,i} \leq 0 \leq \dot{a}^*_{b,B,i}\}$$

Given the approximation of the flow consumption policy function, the approximation of the flow saving policy function follows naturally from equations (29) and (30).

$$\dot{a}^*_{s,ij} = r \cdot a_i + \pi^*(a_i, z_h) + T(\tau) - c^*_{s,ij}$$
$$\dot{a}^*_{b,i} = r \cdot a_i + \pi^*(a_i, z_\ell) + T(\tau) - c^*_{b,i}$$

This completes the discretization of the flow utility component and discretization of the flow value from endogenous changes in wealth a due to consumption-saving.

Bilateral Exchanges on OTC Market. Next, I turn to the discretization of the flow value from endogenous changes in wealth a and productivity z due to bilateral exchanges on the frictional OTC market. First, I approximate the reservation asset price functions $P^*_{min}(a, P_s; \tau)$ and $P^*_{max}(a'; \tau)$, defined implicitly by the indifference conditions (33) and (34) respectively, using a linear interpolation procedure. This results in the approximated reservation asset prices $P^*_{min,ij}$ and $P^*_{max,i}$, for a given CGT rate τ.

Given the approximated reservation asset prices $\left(P^*_{min,ij}, P^*_{max,i}\right)$, I approximate the bilateral exchange policy function $D^*(a, P_s, a'; \tau)$ using equation (35).

$$D^*_{iji'} = \begin{cases} 1 & P^*_{max,i'} > P^*_{min,ij} \\ 0 & P^*_{max,i'} \leq P^*_{min,ij} \end{cases}$$

If a bilateral exchange occurs $D^*_{iji'} = 1$, then I approximate the exchange asset price function $P^*(a, P_s, a'; \tau)$ using equation (36).

$$P^*_{iji'} = \kappa \cdot P^*_{max,i'} + (1 - \kappa) \cdot P^*_{min,ij}$$

Given the approximated exchange asset price function $P^*_{iji'}$, I approximate the post-exchange wealth levels of the seller and the buyer by identifying the closest grid point on the wealth grid to the post-exchange wealth level.

$$\tilde{a}_{s,iji'} := \underset{a\in\mathbb{A}}{\operatorname{argmin}} \left| a - \left[a_i + P^*_{iji'} - \tau \cdot \max\left\{ P^*_{iji'} - P_{s,j}, 0 \right\} \right] \right|$$

$$\tilde{a}_{b,iji'} := \underset{a\in\mathbb{A}}{\operatorname{argmin}} \left| a - \left[a_{i'} - P^*_{iji'} \right] \right|$$

Similarly, I approximate the new cost-basis in the seller state by identifying the closest grid point on the cost-basis grid to the approximated exchange asset price $P^*_{iji'}$.

$$\tilde{P}_{s,iji'} := \underset{P_s\in\mathbb{P}}{\operatorname{argmin}} \left| P_s - P^*_{iji'} \right|$$

Discretized HJB Equation. Combining all approximations together, I characterize the discretized HJB equation of a seller with wealth a_i and cost-basis $P_{s,j}$ below.

$$\begin{aligned}
&\rho \cdot V_s(a_i, P_{s,j}; \tau) \\
&= \frac{\left(c^*_{s,ij}\right)^{1-\sigma} - 1}{1-\sigma} \\
&+ \left[-\frac{\min\{\dot{a}^*_{s,B,ij}, 0\}}{a_i - a_{i-1}} \quad \frac{\min\{\dot{a}^*_{s,B,ij}, 0\}}{a_i - a_{i-1}} - \frac{\max\{\dot{a}^*_{s,F,ij}, 0\}}{a_{i+1} - a_i} \quad \frac{\max\{\dot{a}^*_{s,F,ij}, 0\}}{a_{i+1} - a_i} \right] \cdot \begin{bmatrix} V_s(a_{i-1}, P_{s,j}; \tau) \\ V_s(a_i, P_{s,j}; \tau) \\ V_s(a_{i+1}, P_{s,j}; \tau) \end{bmatrix} \\
&+ [-\gamma_{h\ell} \quad \gamma_{h\ell}] \cdot \begin{bmatrix} V_s(a_i, P_{s,j}; \tau) \\ V_b(a_i; \tau) \end{bmatrix} \\
&+ \eta \cdot \sum_{i'} \left[-D^*_{iji'} \cdot g_b(a_{i'}, ; \tau) \cdot \Delta a_{i'} \quad D^*_{iji'} \cdot g_b(a_{i'}, ; \tau) \cdot \Delta a_{i'} \right] \cdot \begin{bmatrix} V_s(a_i, P_{s,j}; \tau) \\ V_b(\tilde{a}_{s,iji'}; \tau) \end{bmatrix}
\end{aligned}$$

Similarly, I characterize the discretized HJB equation of a buyer with wealth a_i.

$$\cdot V_b(a_i;\tau)$$

$$= \frac{\left(c^*_{b,i}\right)^{1-\sigma} - 1}{1-\sigma}$$

$$+\left[-\frac{\min\{\dot{a}^*_{b,B,i},0\}}{a_i - a_{i-1}} \quad \frac{\min\{\dot{a}^*_{b,B,i},0\}}{a_i - a_{i-1}} - \frac{\max\{\dot{a}^*_{b,F,i},0\}}{a_{i+1} - a_i} \quad \frac{\max\{\dot{a}^*_{b,F,i},0\}}{a_{i+1} - a_i}\right] \cdot \begin{bmatrix} V_b(a_{i-1};\tau) \\ V_b(a_i;\tau) \\ V_b(a_{i+1};\tau) \end{bmatrix}$$

$$+[-\gamma_{\ell h} \quad \gamma_{\ell h}] \cdot \begin{bmatrix} V_b(a_i;\tau) \\ V_s(a_i,0;\tau) \end{bmatrix}$$

$$+\eta \cdot \sum_{i',j} \left[-D^*_{i'ji} \cdot g_s\left(a_{i'}, P_{s,j},;\tau\right) \cdot \Delta a_{i'} \quad D^*_{i'ji} \cdot g_s\left(a_{i'}, P_{s,j},;\tau\right) \cdot \Delta a_{i'}\right] \cdot \begin{bmatrix} V_b(a_i;\tau) \\ V_s\left(\tilde{a}_{b,i'ji}, \tilde{P}_{s,i'ji};\tau\right) \end{bmatrix}$$

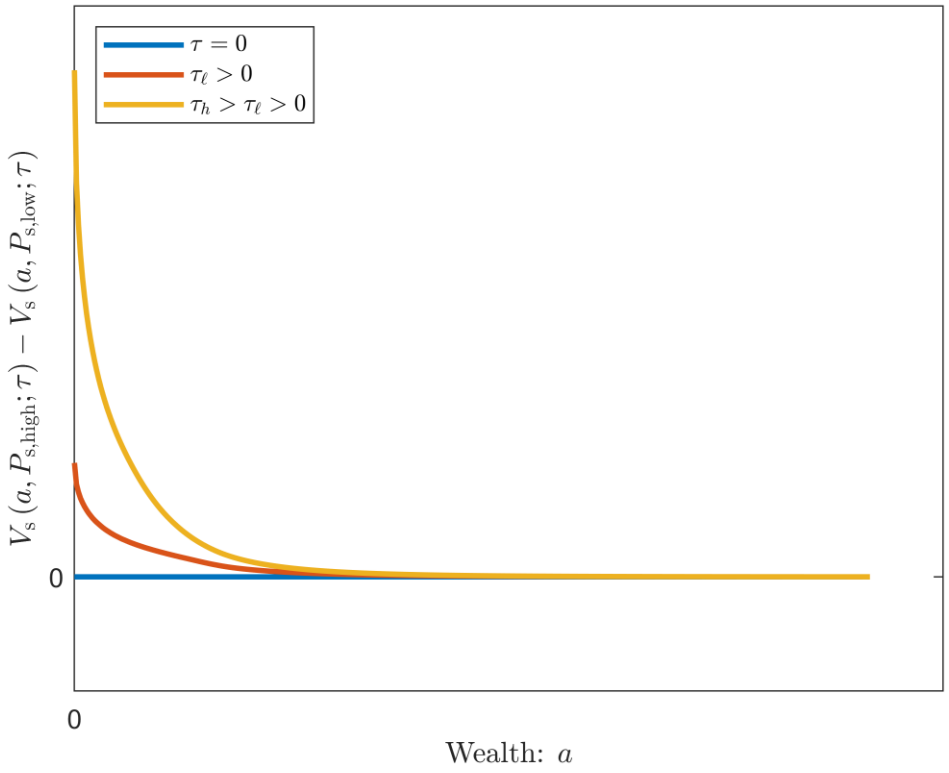
Stacking the discretized HJB equations above across all grid points for wealth a and cost-basis yields the desired nonlinear HJB matrix equation (48).

Lastly, following the arguments in Achdou et al. (2021), the nonlinear KF equation (49) can be derived using the transpose of the composite transition matrix from the nonlinear HJB equation (48). This is the sense in which, after solving the HJB equation for the value function vector $\boldsymbol{V}$, one obtains the joint distribution vector $\boldsymbol{g}$ "for free" by solving a straightforward eigenvector problem using standard linear algebra techniques.

APPENDIX C: SUPPLEMENTARY FIGURES

C.1. Value Function and Cost-Basis

Figure 23: Seller's Value Function Difference between Low and High Cost-Basis



C.2. Wealth Distribution

Figure 24: Wealth Distribution (Aggregate)

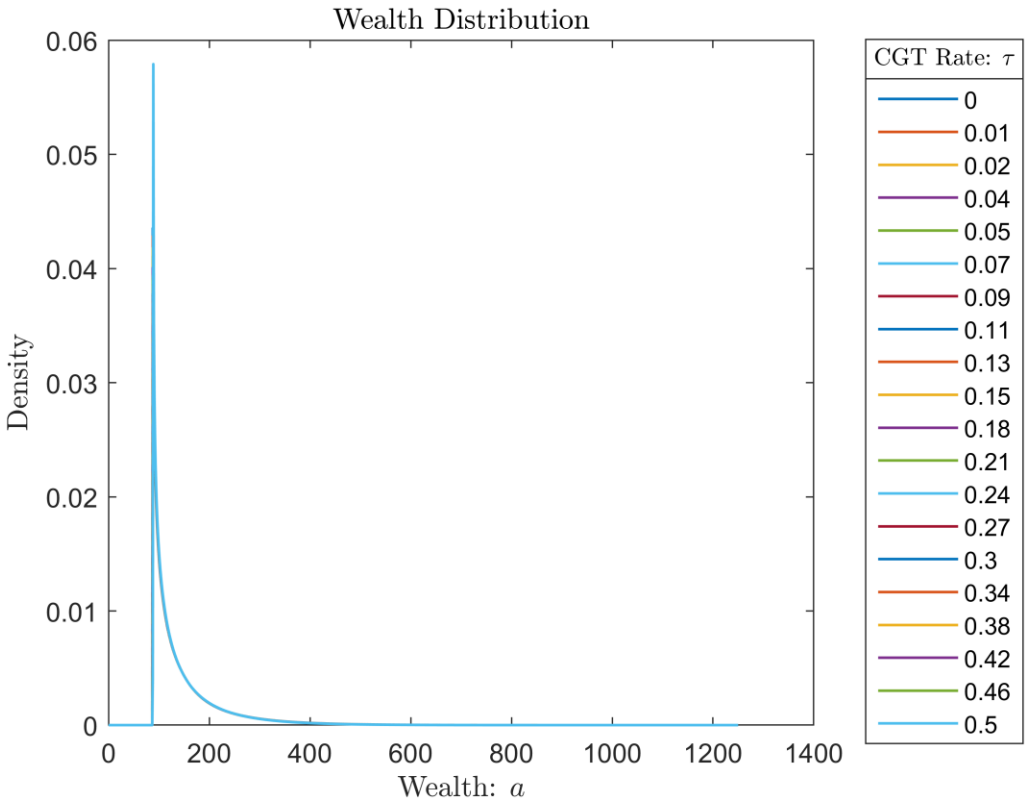


Figure 25: Wealth Distribution (Sellers)

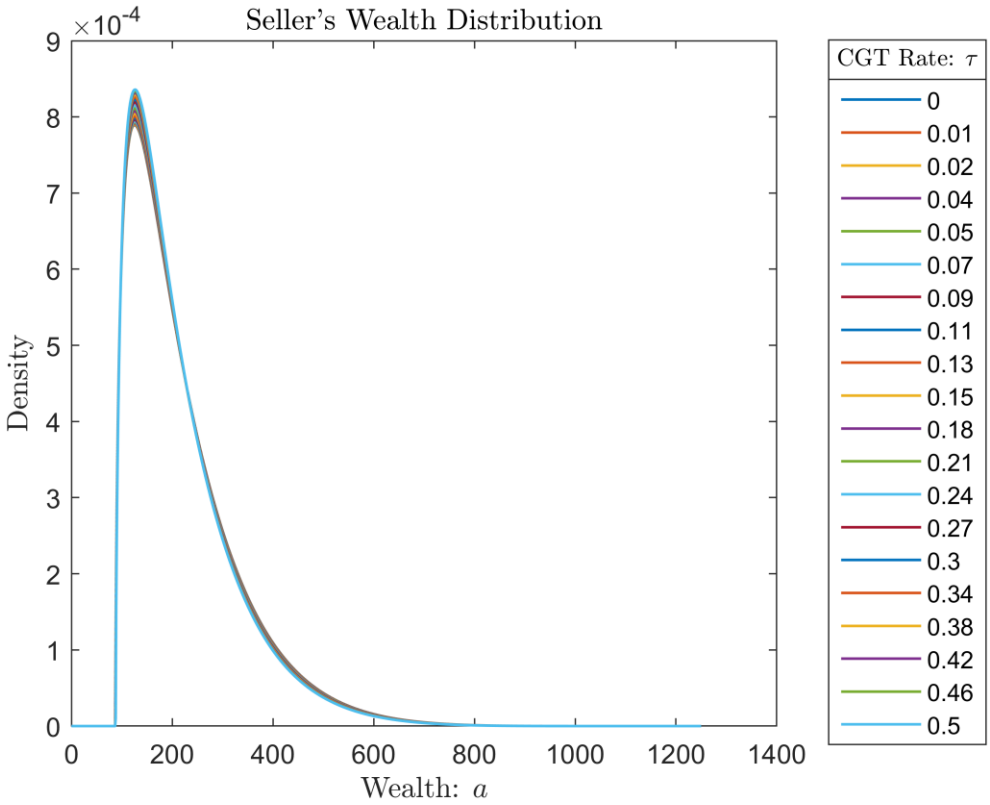


Figure 26: Wealth Distribution (Buyers)